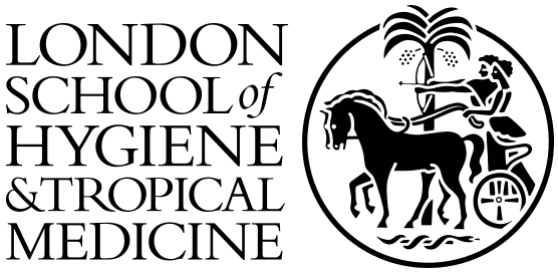


Mendelian Randomisation and Cardiovascular Disease

Juan-Pablo Casas

Thesis submitted to the University of London for the degree of

Doctor of Philosophy

London School of Hygiene & Tropical Medicine

Submitted by February 2009

Declaration of work by the candidate

I wrote the entire thesis presented hereafter. Some of the work conducted and publications written involved collaborative efforts, as mentioned below.

For the work in chapter 2, I conducted the search strategies, data extraction and analysis. I interpreted the findings of statistical analyses and wrote the first draft of the published manuscript. I received assistance in statistical analysis by Dr Leonelo Bautista and in the search strategy by Dr Pankaj Sharma.

For the work in Chapter 3, I conducted the search strategies, data extraction, and analyses of the information. I led the interpretation and drafting of the report to be submitted for publication (work in progress). I received assistance from Mr Paul Newcombe, under my supervision, in the search strategy and data extraction. Together, we contacted more than 70 investigators from new genetic studies, and invited them to participate in the collaboration that originated the results for Chapter 3.

For work on Chapter 4, I developed and specified the analytical strategy that was provided to statisticians from individual studies participating in the project. I established communication with each statistician to resolve any arising queries or inconsistencies. After this quality control process I utilised meta-analytic techniques to combine information across studies. I led the interpretation of findings and drafted the published manuscript. I utilised a Stata do-file generated by Dr Leonelo Bautista to conduct the simulation analysis.

For the work in Chapter 5, I also developed the analytical strategy that was applied by Ms Jackie Cooper for NPHS-2 study and Dr. Ionna Tsoulaki for the Edinburgh Artery Study. After a quality control process, I then meta-analysed the information and led the interpretation of findings and drafting of the report.

For the work on Chapter 6, I elaborated the analyses plan for the Lp-PLA2 genetics collaboration, which was then applied by statisticians from individual studies, whilst again I established close contact to resolve any doubts or inconsistencies from the results generated. I synthesized the Information from 11 studies, led the data interpretation and drafting of the report to be submitted for publication (work in progress).

For work on Chapter 7, Prof Hingorani and I jointly developed the analytical strategy which was conducted by Dr Fotio Drenos (Centre for Cardiovascular Genetics at UCL) utilising information available in the NPHS-2 study. I actively participated in the data interpretation and in the writing of the report. The work of this Chapter forms part of a more comprehensive manuscript that has been re-submitted to Human Molecular Genetics.

Juan-Pablo Casas

23rd February 2009

Date

Abstract

Background and aims: Homocysteine (Hcy), C-reactive protein (CRP), and Lipoprotein-associated phospholipase A_2 (Lp-PLA2) have been associated with a high risk of cardiovascular disease. If casual, they are expected to provide additional tools for prevention. Utilising the unique properties of genetic variants (randomly allocated and unmodifiable), they could be used as unconfounded proxies of environmental exposures to investigate disease aetiology, known as Mendelian randomisation. Herein I conduct a series of Mendelian randomisation experiments to: (i) investigate the role of Hyc in stroke; (ii) judge causality of CRP in cardiovascular disease; (iii) investigate the validity of Lp-PLA2 as a therapeutic target in coronary heart disease (CHD) and; (iv) describe how the integration of *cis*-acting variants and their cognate proteins can be used to dissect causal pathways.

Methods: For aim (i), I conducted synthesis research of published and unpublished studies investigating the *MTHFR*/C677T variant, Hcy and stroke. For aims (ii) to (iv), a series of prospective collaborations conducting *de novo* genotyping for *CRP* and *PLA2G7* genes, were established using European-based cardiovascular genetic studies of adults.

Results: The meta-analyses of studies on *MTHFR*/C677T-Hcy-Stroke to 2003, showed that subjects with the TT genotype had on average 1.93 μmol/L higher levels of Hcy compared with subjects with CC genotype, and an odds ratio (OR) of stroke of 1.26 (95%CI: 1.14, 1.40). An update analyses to 2008 showed that in both *MTHFR*-Hcy and *MTHFR*-stroke associations, studies in Asia had the largest effect, followed by Europe with intermediate effect, and lower or negligible effect in the Americas and Australasia. Analysis on CRP, indicated that subjects homozygous for the T-allele of *CRP*/+1444C>T variant despite having 0.68 mg/L higher levels of CRP, had no increase in risk of myocardial infarction (OR of 1.01 [95CI: 0.74, 1.38]). A tagging-haplotype approach showed a gradual increase on CRP levels by haplotype, but no effect on CHD, diabetes or stroke. Analysis of the seven *PLA2G7* tagging-SNPs showed that the best variant (rs1051931) had small to moderate effects on the Lp-PLA2 activity (up to 7% relative differences). No genetic signal with CHD was observed for any *PLA2G7* variant, despite some comparisons including up to 8412 CHD cases. A description of the proof of principle, illustrating how to utilise *cis*-acting variants and their cognate proteins to distinguish causal from non-causal effects among correlated blood proteins was presented, using as an example 6 proteins that have been associated with CHD risk.

Conclusions: Genetic evidence reported in this document suggested that the MTHFR effect is more pronounced in geographic regions associated with low folate intake. Genetic studies on CRP presented in this document indicated that CRP is unlikely to be causal in cardiovascular disease. A genetic approach using common variants in PLA2G7 had a reduced ability to confirm or reject Lp-PLA2 as a valid drug target. Finally, integration of *cis*-acting variants with their cognate proteins seems to be an effective way of dissecting biological pathways.

Acknowledgements

I am indebted to the many people and organizations for their support in conducting this work. I would especially like to thank Professor Aroon D Hingorani, with whom I have had the pleasure to work alongside for the last 7 years.

I am particularly grateful to both my supervisor and co-supervisor at the London School of Hygiene and Tropical Medicine, Professor Liam Smeeth for all his patience, guidance and support throughout this thesis, and Professor John Whittaker for his advice and statistical assistance.

I would also like to express my thanks and gratitude to all those with whom I have collaborated in this thesis, especially Dr Tina Shah, Ms Jackie Cooper, Dr Fotio Drenos, Dr Ionna Tsoulaki, Dr Leonelo Bautista, Dr Pankaj Sharma, and Mr Paul Newcombe.

Thanks also to all those people who have provided me with the data sets used in this thesis, in particular Professor Philippa Talmud, Professor Steve Humphries, Professor Gerry Fowkes, and Dr Manjinder Sandhu. Also thanks to my friends at LSHTM for all their help and support, especially thanks to Pablo Perel, Valerie McCormack, and Beatriz Diaz-Apodaca.

To my parents, Luis-Felipe and Nora,

especially to my mother who passed away a year ago, during work on this thesis

Table of Contents

List of Tables

List of Figures

Abbreviations

UK:	United Kingdom
CVD:	Cardiovascular disease
CHD:	Coronary heart disease
MI:	Myocardial infarction
CRP:	C-reactive protein
Hcy:	Homocysteine
Lp-PLA2:	Lipoprotein-associated phospolipase A_2
SNP:	Single nucleotide polymorphism
t:	Tagging
LD:	Linkage disequilibrium
OR:	Odds ratio
HR:	Hazard ratio
UCL:	University College London
pQTG:	Protein quantitative trait gene
pQTSNP:	Protein quantitative trait SNP

Chapter 1 Background and Aims

1.1 Risk factors for cardiovascular disease, with focus on coronary heart disease and stroke

Cardiovascular diseases (CVD) account for 20.3% of disability-adjusted life years lost in more developed countries and already 8.1% in less-developed countries.[1] Cardiovascular diseases as a group are the leading cause of death in the European Union, accounting for over 1.5 million deaths each year. Nearly half (42%) of all deaths in the European Union are from CVD.[2] In the United Kingdom (UK) although the age-adjusted CVD mortality has decreased in the last three decades,[3] it remained the primary cause of mortality responsible for 39% of all deaths in 2002.[3] Coronary heart disease (CHD), the main form of CVD, causes over 170,000 deaths a year in the UK: approximately one in five deaths in men and one in six deaths in women. In 2003 the prevalence estimated in UK adult population of CHD was 2.65 million, and it is expected that every year about 85,000 British individuals under 65 living in UK will have a new heart attack.[3]

Coronary heart disease and stroke, the most common disorders within the CVD group, have both been associated with similar environmental and behavioural risk factors.[4] Smoking, hypertension, elevated blood cholesterol, diabetes, are considered well-accepted risk factors for CHD.[5] For stroke, hypertension, smoking, diabetes and atrial fibrillation are considered well-established risk factors, while in contrast to CHD, the role of cholesterol remains controversial.[6]

Despite a great advance in the understanding of the causes of CVD, mainly CHD, it is known that 30% of cardiovascular events occurring in middle age subjects take place in individuals within the lowest portion (in quintile terms) of the continuous distribution of the established cardiovascular risk factors.[7] In addition, a considerable proportion of subjects with *high* levels of those established risk factors do not develop a cardiovascular event. Between 58% to 85% of subjects with at least one established risk factor within the clinically elevated range fail to develop a cardiovascular event after long follow-up (range from 14.4 to 24.4 years).[7,8] This has motivated the search for additional risk factors for CVD that might enhance disease prediction, and offer the prospect of new therapies. As a consequence, in recent years increasing attention has focused on the so-called emerging risk factors such as infection,[9] homocysteine (Hcy),[10] and markers of the inflammatory response (e.g. C-reactive protein [CRP], and Lipoprotein-associated phospholipase A_2 [Lp-PLA2]).[11,12] However the precise causal role of these putative factors on the natural history of different types of cardiovascular disorders remains to be established.

To date, the main disciplines contributing to the identification of possible new risk factors for CVD are the laboratory research and the traditional observational epidemiology. Despite its inherent limitations, observational epidemiology has played an important role in our understanding of the established risk factors for CVD, and has had a considerable impact in public health and clinical practice. However, in recent years well-known discrepancies between the findings from

observational studies and randomised clinical trials in cardiovascular disease, (and also cancer), has brought to focus the need for careful design to control for the main threats of validity in observational research.[13,14] Spurious associations in observational epidemiology may be the result of uncontrolled confounders and reverse causation, but such studies are also prone to other types of bias. Assessing the likelihood and magnitude of such biases or systematic errors is complicated because some confounding factors may be unknown or difficult to measure and ascertaining the starting time for exposure and disease may not be possible raising the potential for reverse causation.[15]

In contrast, randomised trials, although aimed to at evaluating the efficacy of interventions and the reversibility of an exposure, arise as a better design in the hierarchy of research to help in determine causal role of environmental exposures. The basic concept behind that superiority of clinical trials over observational studies is the randomisation process, which makes the groups being evaluated comparable for any known and unknown characteristics, which abolishes the possibility of spurious results due to confounding.[16] If high levels of a certain exposure are associated with an increase in disease risk, an intervention that specifically reduces the levels of that exposure in subjects randomised to that intervention should lead to a lower event rate if that exposure is causal. That has been the case for blood pressure and cholesterol, in which randomised clinical trials have provide strong evidence of a causal role on coronary heart disease, which was previously suggested by prospective cohort studies.[17] However, because of the lack of specific interventions, or because the conduct of a clinical trial may be unethical or logistically unachievable, randomised trials to evaluate causality may not be practical. Inference on causality then mainly resides on non-randomised observations (on humans: classic observational epidemiology) or experimentation (in animal models, systems, or cells/tissues). An emerging genetic approach offers an opportunity to substantially reduce some of the limitations from observational studies, when inference on causality from environmental risk factors is evaluated, that might help to prioritise the development of new therapies, which is a costly endeavour.

1.2 Limitation of observational epidemiology for aetiological studies, with a focus on biomarkers in cardiovascular disease

In an ideal setting, to evaluate causality a study should compare individuals with themselves in both an exposed and unexposed state over the same time. Such as approach is clearly not realistic and is referred to as counterfactual.[18] Since we cannot achieve the counterfactual ideal, we strive to come as close to it as possible. The aim would be to find an unexposed population that would give a result close to, if not identical, with that, of the counterfactual comparison.[18] An strategy that attempts to generate comparable groups is the "randomisation" process involved in a clinical trial, in which two generated groups are comparable for any characteristic, with the exception of the intervention to which they are randomised. Because of the impossibility of

generating comparable groups fulfilling the counterfactual ideal, results from any epidemiologic study deviate from the truth, due to the presence of systematic and random errors.[18] The systematic error is what is left after an study became infinitely large in which the random error disappears. Systematic error affects the validity of the results from an epidemiologic study. There are three main groups of systematic error or biases that affect the results of an observational epidemiology: selection bias, information bias and confounding.[18]

Case-control and prospective cohort studies are the main designs used to identify associations in terms of relative risks between certain biomarkers (a term used to reflect uncertainty regarding their aetiological role) and later cardiovascular events such as myocardial infarction (MI) or stroke. These biomarkers include representatives from lipid, inflammatory, metabolic and pro-oxidant pathways e.g. lipoprotein (a), high-density lipoprotein, Hcy, CRP, and interleukin-6, among many others. However, the validity of these associations is often called into question due to the presence of systematic error which will compromise the causal inference on the role of a certain biomarkers as a relevant therapeutic targets for which a specific intervention and then a randomised trials need to be developed. In the cardiovascular arena, systematic error is mainly due to the possibility of residual confounding and reverse causation, among the many possible biases. Since several novel biomarkers are often associated with one or more exposures that causally affect risk of CVD such as blood pressure, smoking, obesity, diabetes and physical activity, confounding becomes an important issue.[19] Despite, several existing strategies to reduce confounding by matching, restriction or statistical adjustment, not all the relevant confounders are measured in every single study as some of them are unknown at the time of conducting the study. Also a problem is the presence of unavoidable errors in measurement of potential confounders.[14,15,18] An additional limitation, that is becoming more relevant with the ample use of biomarkers is that for a variable to be classified as a confounder of the exposure (biomarker) of interest, it should not lie in the causal pathway of linking the biomarker of interest with the outcome. If it does inappropriate adjustment will arise and it will tend to bias the results. For example, if the association between a biomarker and CVD is mediated through blood pressure, or a second biomarker, rather than confounded by them, adjustment for these factors will lead to an inappropriate loss of association.

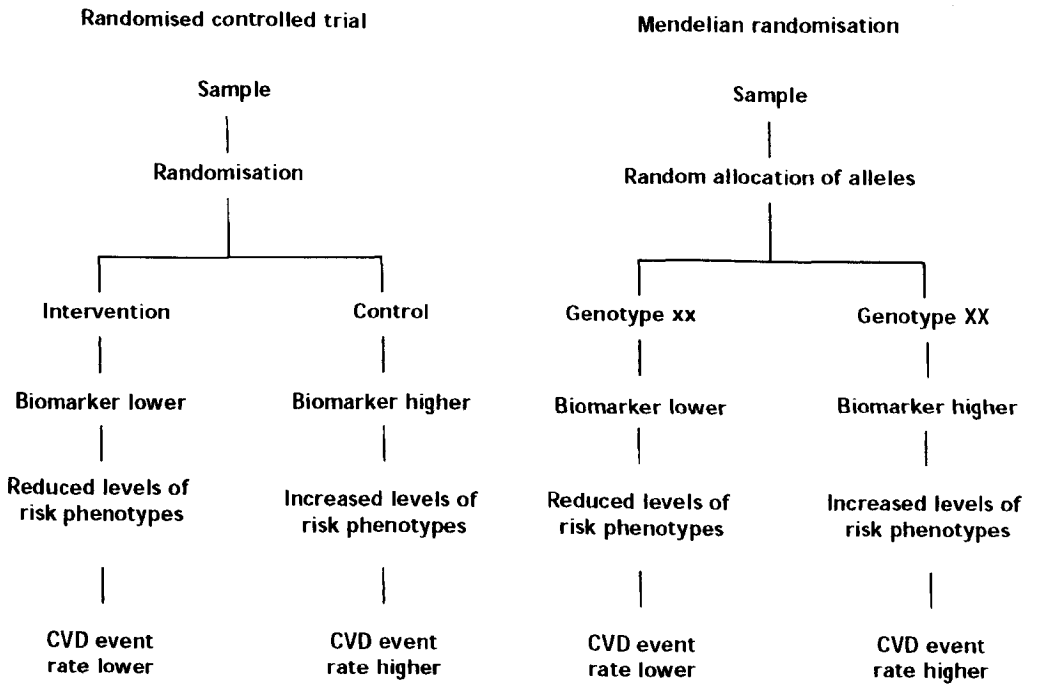
Many biomarkers studied (e.g. CRP, Hcy and LpPLA2) are also higher in subjects with established CVD which makes reverse causation a possible explanation. This is because, the natural history of the atherosclerotic process, the key factor in CHD and ischaemic stroke, begins in the early adulthood with the presence of fatty streaks that, with time, accumulates leading to the development of the atherosclerotic plaque. Therefore, even in prospective studies of usually middle-age populations there is a potential, though reduced compared to case-control studies, for reverse causation is not abolished and could still lead to an observed association of biomarker and CHD risk.[20]

On the other hand, regression dilution bias, which arises due to biological variability in the level of the biomarker and measurement error, can lead to an underestimation of the association between a biomarker and disease.[21] Dealing with these potential sources of error to distinguish causal from non-causal associations is of critical importance, because, if CRP, Hcy, or other biomarkers were causal in CVD, they would represent legitimate therapeutic targets, and provide additional opportunities for disease prevention.

1.3 Advantages and limitations of the genetic approach, Mendelian randomisation to establish unbiased biomarker-outcome associations

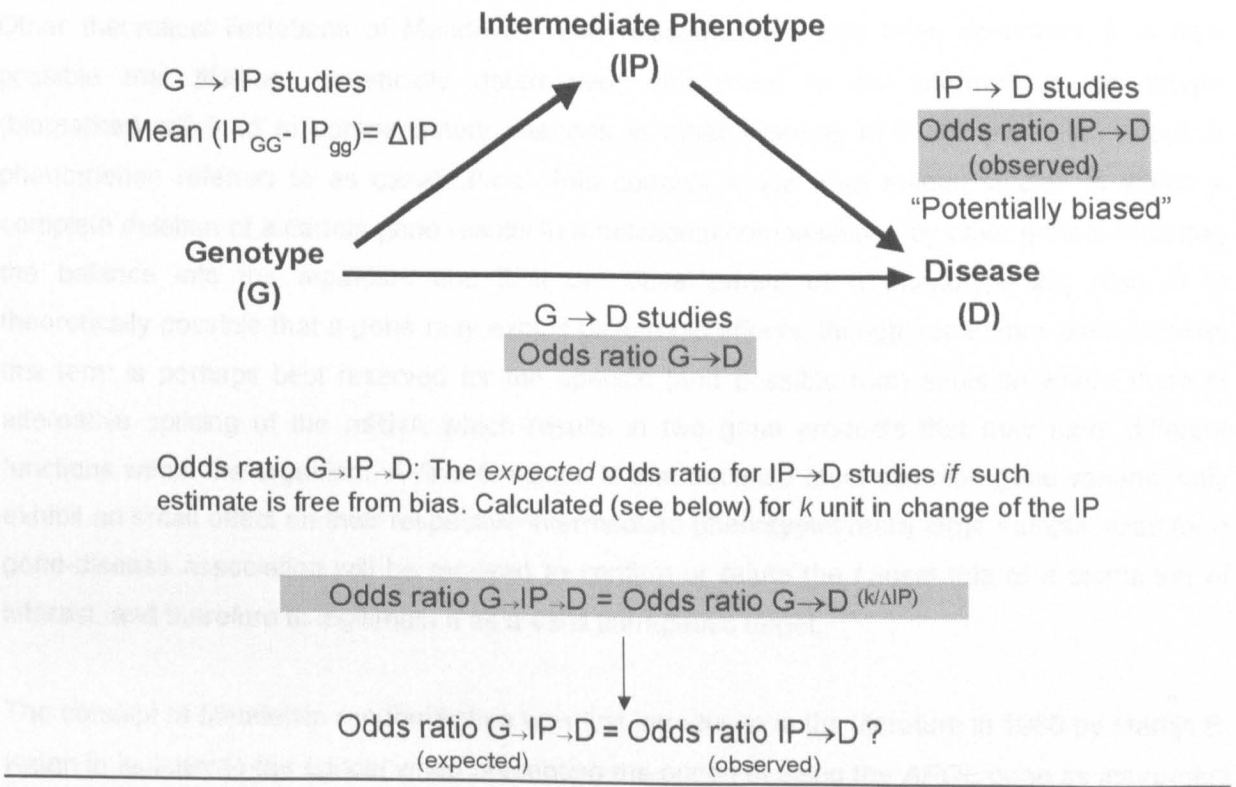
Because of the random assortment of paternal and maternal alleles at conception, according to Mendel's 2nd law, the carriage by an individual of a certain allele, genotype or haplotype that directly governs the circulating level of a biomarker of interest will be a random process independent of any other characteristic that may confound the plasma biomarker-disease association. This mimics the random allocation operating in a clinical trial (Figure 1.1).[22,23] Therefore, a genetic approach might allow unbiased insight into the link between biomarkers and high-risk phenotypes (e.g. high blood pressure, diabetes) and cardiovascular outcomes.

Figure 1.1. Parallel of the Mendelian randomisation approach with a randomised clinical trial. Figure adapted from the Figure of the manuscript: Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005;366:1906-8.



If we consider the case of a specific circulating biomarker associated with CVD in observational studies, where the concentration of the biomarker is influenced by common polymorphisms in a gene **X**. Let alleles **X** and **x** be associated with high and low levels of biomarker, if the circulating biomarker were a true cause of disease, individuals with genotype **XX** should have an increased risk of CVD commensurate with the effect of this genotype on its intermediate phenotype free from confounding (Figure 1.2).[24] Because genotype is a non-modifiable characteristic, and determined far in advance of disease, any gene-disease association should also be free of reverse causation and regression dilution bias.[22,23]

Figure 1.2. Components of the Mendelian randomisation strategy.



This concept, is highly attractive since it is well known that many (if not all) circulating biomarkers are highly heritable. Genes contributing to their regulation are therefore plausible candidates for CVD susceptibility.

Although, in contrast to the biomarker-cardiovascular outcome association, a genetic approach that uses a gene variant that reliably affects the concentration of such biomarker in a direct manner should not be affected by confounding and reverse causation, for reasons mentioned above, it may suffer from other limitations. Although confounding is reduced in genetic studies, it may not be abolished because specific type of confounding can affect even this design. The first one is confounding by linkage disequilibrium (LD). If the gene variant of interest (gene-1) is in LD

with a second gene variant (gene-2), in an adjacent gene, the association of gene-1 with disease might not be mediated through its direct biomarker (biomarker-1), but rather it may be the result of confounding by LD with the adjacent gene, that modifies disease risk through a second biomarker. In this way, spurious evidence on causation could be obtained for biomarker-1.[22] The second source of confounding, could arise when the genotype (or allele) frequency differs in groups of differing ancestral origin, when ancestry is associated with disease risk, and when the case and control groups in an association study differ in their ethnic composition. This is referred to as population stratification.[25] However, population stratification, to an extent large enough to distort results, has been considered unlikely to occur in well-designed studies,[25,26] and moreover, nowadays it is possible to adjust for this type of confounding.[27,28]

Other theoretical limitations of Mendelian randomisation have also been described. It is also possible that lifelong, genetically determined, differences in an intermediate phenotype (biomarker) will lead to compensatory changes in other systems to inhibit the initial effect, a phenomenon referred to as canalization. This concept arose from animal studies in which a complete deletion of a certain gene results in a functional compensation by other genes, restoring the balance into the organism and limit the development of a phenotype.[22] Also, it is theoretically possible that a gene may exhibit pleiotropic effects, though sometimes used loosely, this term is perhaps best reserved for the specific (and possible rare) situation where there is alternative splicing of the mRNA which results in two gene products that may have different functions within the organism. A final limitation is that because most common gene variants only exhibit an small effect on their respective intermediate phenotypes really large sample sizes for a gene-disease association will be required to confirm or refute the causal role of a biomarker of interest, and therefore to legitimate it as a valid therapeutic target.

The concept of Mendelian randomisation was first introduced in the literature in 1986 by Martijn B. Katan in its letter to the *Lancet* when presenting the option of using the *APOE* gene as instrument for cholesterol to help to clarify the controversial role (at the time) of low levels of cholesterol on cancer.[29] Later, some of the principles of Mendelian randomisation were sporadically mentioned by other investigators,[24] but it was not until recent years when Professors G. Davey-Smith and S. Ebrahim, formally introduce the theoretical grounds of this novel approach as way to understand the environmental determinants of disease.[22,23] Part of the results presented in this thesis actually constitutes some of the pioneers examples of Mendelian randomisation experiments to help to clarify a possible causal role of several biomarkers in CVD. As the development of this thesis occurred, there has been immense interest in this approach, that not only lies in the generation of empirical evidence for this approach in different diseases,[30,31,32] but also in the development of more adequate statistical tools to overcome some of the challenges that this strategy brought.[33,34]

1.4 Thesis aims and objectives

1.4.1 Aim 1: Homocysteine and stroke

Aim: To conduct a series a systematic reviews and meta-analyses of genetic studies investigating the *MTHFR*/C677T variant, Hcy and stroke, and contextualize the findings with evidence derived from observational studies in Hcy and stroke and clinical trials using Hcy-lowering therapies.

Objectives:

- To synthesised genetic evidence from published and unpublished studies evaluating the *MTHFR* effect on Hcy.
- To synthesised genetic evidence from published and unpublished studies investigating the *MTHFR* effect on Stroke.
- To synthesised the results from clinical trials using Hcy-lowering therapies in stroke.
- To investigate, whether the *MTHFR*-Hcy-Stroke association is modified by geographical location, used as proxy of the different policies for folic acid fortification.

1.4.2 Aim 2: C-reactive protein and cardiovascular disease

Aim: To investigate the potential causal role of CRP in CHD, stroke, and diabetes adopting the *CRP* gene (using *de novo* genotyping on a single-variant or a tagging-haplotype) as unconfounded proxy of plasma CRP.

Objectives:

- To evaluate the effect of the *CRP*/+1444C>T variant on CRP levels, and to generate empirical evidence of the random distribution by genotype of variables considered as confounders of the plasma CRP associations.
- To evaluate the effect of *CRP*/+1444C>T variant on the risk of MI.
- To evaluate the effect of the *CRP* tagging-haplotypes on CRP levels, and to obtain evidence of a random distribution by haplotype of variables that act as confounders of the plasma-CRP associations.
- To evaluate the effect of the *CRP* tagging-haplotypes on risk of CHD, stroke and diabetes.

1.4.3 Aim 3: Lp-PLA2 and coronary heart disease

Aim: To investigate the validity of Lp-PLA2 enzyme as a therapeutic target in CHD by using *de novo* data on the Lp-PLA2 gene (known as *PLA2G7*) as an unbiased tool for Lp-PLA2 activity.

Objectives:

- To evaluate the shape of the association between Lp-PLA2 activity and CHD risk, and to investigate the potential for confounding on that association.
- To evaluate the effect of the *PLA2G7* tagging-SNPs on Lp-PLA2 activity, and on cardiovascular traits that act as confounders of Lp-PLA2 activity-CHD association.
- To evaluate the effect of the *PLA2G7* tagging-SNPs on CHD risk.

1.4.4 Aim 4: Mendelian randomisation and biological pathways

Aim: To describe how the integration of protein quantitative trait SNPs (pQTSNPs) and their cognate proteins can be used to dissect biological pathways among blood proteins.

Objectives:

- To evaluate the degree of correlations present among 6 blood proteins (CRP, fibrinogen, Lp-PLA2, Factor VII, apo-A1, and apo-B) that have been associated with high risk of cardiovascular disease.
- To select the most potent *cis*-acting variant to be used as unconfounded tool of its respective cognate protein (e.g. *CRP* variants on CRP levels)
- To quantify the effect of each one of the selected genetic-tool on its respective cognate protein (e.g. *CRP*-variant on CRP levels), as well as on the other 5 blood proteins (e.g. *CRP*-variant on fibrinogen, Lp-PLA2, Factor VII, apo-A1, and apo-B)

Chapter 2 Homocysteine and Stroke

2.1 Homocysteine and stroke: use of Mendelian randomisation to support a causal role

Stroke is the third most common cause of death in developed countries,[1] and approximately 80 percent of strokes are ischaemic in origin and the remainder haemorrhagic.[35,36] In the United Kingdom, stroke is the largest single cause of severe disability with more than 125,000 incident strokes annually, and approximately 60,000 deaths due to stroke each year.[35] The best approach to reduce mortality and morbidity from the disease is by primary prevention through modification of acquired risk factors.[36]

Data derived from non-genetic observational studies (cohort and case-control studies), suggest that a prolonged increase in total Hcy (sum of homocysteine, homocysteine-homocysteine, and homocysteine-cysteine disulfide and that now on will be referred in this document as Hcy) levels is associated with a higher risk of stroke.[37-39] However, since Hcy concentration is also related to smoking status, blood pressure, and social class, and is also higher in subjects with existing atherosclerosis, this relationship, may be subject to residual confounding, reverse causation or both.[40,41]

Homocysteine is a sulphur-containing aminoacid that is formed during the methionine metabolism and is regulated by the B vitamins: folate, B6 and B12. [42] The first evidence for a vascular effect of Hcy in humans was derived from anecdotal reports from individuals suffering from homocystinuria, a metabolic disorder associated with very high levels of Hcy consequence from a homozygous deficiency of the cystathionine beta-synthase.[43] In 1985, Mudd et al. [44] reported the natural history of 629 subjects with homocystinuria before treatment with Hcy-lowering therapies was established. A time-to-event analysis indicated that by age 29 the probability of suffering a thromboembolic event was 50%. During the follow up, a total of 253 thromboembolic events in 158 patients (1 event per 25 patient-years of follow up) were observed, distributed as follow: 130 (51%) peripheral venous events, of which 32 were pulmonary embolism, 81 (32%) cerebrovascular accidents, 28 (11%) peripheral arterial events, 10 (4%) MI and 4 (2%) events did not fall into none of those categories. Of interest, there were 8 times more cerebrovascular events than coronary events. The evidence derived from this rare genetic disorder associated with very high levels of Hcy, is in a manner analogous to the inference that in past was derived for LDL cholesterol in CHD, using to individuals suffering from familial hypercholesterolemia.[22]

In 1995, Frost et al, [45] reported the existence of a C-to-T substitution at nucleotide 677 (rs1801133) in the gene encoding methylenetetrahydrofolate reductase (*MTHFR*), an enzyme involved in Hcy metabolism. This substitution that converts an alanine to a valine residue at position 222 (A222V) has been associated with a decrease enzyme activity (35% to 40% in subjects with the TT genotype compared with subjects with the CC genotype) and differences in Hcy concentration.[46,47] Since carriage of this variant is subject to the random assortment of maternal and paternal alleles at the time of gamete formation,[48] associations between *MTHFR*

genotype and Hcy or stroke should not be subject to reverse causality bias and should also be largely free from confounding by other determinants of Hcy or risk factors for stroke.[22,23] Therefore, if the Hcy-stroke association is unbiased, carriage of the *MTHFR* polymorphism, that exposes individuals to a long term elevation in Hcy, should confer an increased risk of stroke proportional to the difference in Hcy attributable to variant and the risk observed in non-genetic observational studies.

The evaluation of consistency between unbiased risk estimates obtained from genetic studies with those from non-genetic observational studies, to provide insight into the nature of the observed associations, as described in Chapter 1 has been referred to as Mendelian randomisation approach. This approach was used previously to test the nature of the association between Hcy and CVD.[37] However, in the prior genetic analyses of the *MTHFR* polymorphism and stroke, the number of available studies was relatively small. Also, information on the relative effect of the *MTHFR* polymorphism on Hcy level in different geographical locations, ethnic groups and age bands was limited. [37]

If Hcy is shown to be causally related to the development of stroke this would have important implications in primary prevention, since folic acid is known to reduce Hcy levels,[49] and a policy of fortification of cereal and flour with folic acid to reduce the incidence of neural tube defects has been initiated since 1996 in United States, Canada, Australia and New Zealand and more recently in Ireland, Brazil, Chile, Jordan and South-Africa. Indirect evidence derived from two non-randomised intervention studies in subjects with homocystinuria indicated that uses of Hcy-lowering therapies during 18 years was associated with a substantially lower than expected (compared with event-rate observed in the study by Mudd et al in 1985) number of vascular events, including stroke. [50,51] Using a quasi-experimental intervention Yang et al,[52] observed an accelerated decline in stroke mortality in post-fortification period (1998-2002) compared with the pre-fortification period (1990-1997). In contrast there was not a significant change in the decline of stroke mortality, for the same period in England and Wales where fortification is not required.[52]

I have now performed two up-dated and comprehensive meta-analyses. In the first, I evaluated to what extent Hcy levels are determined by the *MTHFR*/C677T polymorphism. I also evaluated the effect of pre-existing CVD, geographical location, ethnic background and age on this association. In the second, I estimated the odds of stroke conferred by this variant. Compared to a previous meta-analysis, my meta-analysis of *MTHFR*/C677T polymorphism and Hcy concentration includes data from 48 additional studies and extends the observations to subjects without CVD.[37] Compared to a previous meta-analysis by Wald et al,[37] my meta-analysis of the *MTHFR*/C677T polymorphism and stroke includes data from 23 additional studies and evaluates the robustness of the results by sensitivity analysis.

2.2 Systematic review and meta-analysis of the effect of the MTHFR/C677T polymorphism in homocysteine and stroke

2.3 Methods

Two electronic databases [MEDLINE and EMBASE] were searched up to June 2003 for all studies evaluating the association between the *MTHFR*/C677T polymorphism and homocysteine levels, and up to December 2003 for studies on the association between the *MTHFR*/C677T polymorphism and stroke. For the first search, I used the text words, which were also MeSH terms, "polymorphism", "mutation", "genes", and "cardiovascular disease" in combination with "homocysteine". For the second search, the terms used were "stroke", "brain infarction", "cerebral ischaemia", "hemorrhagic stroke" and "silent brain infarction" in combination with "genetic", "polymorphism", "mutation" or "genes". Both, literature searches were limited to "human", and "English Language". I searched for any additional studies in the references of all identified publications, including previous relevant meta-analyses.[37]

2.3.1 Selection Criteria

2.3.1.1 Homocysteine levels and MTHFR Polymorphism

For inclusion, studies had to have an analytical design and examine the association between Hcy levels and the *MTHFR*/C677T polymorphism. Studies were included only if they were published as full-length articles or letters in peer-reviewed journals. For duplicate publications the smaller dataset was excluded.

2.3.1.2 MTHFR and Stroke

For inclusion, studies had to involve unrelated subjects, and examine the associations between ischaemic or hemorrhagic stroke and the presence of the *MTHFR*/C677T polymorphism. For the main comparison, only studies published as full-length articles or letters in peer-reviewed journals in English language were included. A sensitivity analysis was conducted which included silent brain infarction as an outcome, non-full text papers, and non-English language publications. In all searches, when relevant information was not reported, or there was doubt regarding duplicate publications, the authors were contacted to obtain the required information.

2.3.2 Data Extraction

I extracted and entered the data for analysis into two databases. The results were compared and disagreements resolved by consensus with Prof Aroon D Hingorani and Dr Pankaj Sharma. The data extracted and entered for the analysis was country of origin, study design, mean age of participants, frequency of genotypes, and alleles, Hcy and serum folate levels, ethnic background, and frequency of cardiovascular risk factors.

2.3.3 Statistical Analysis

I obtained a summary estimate of the effect of raised plasma Hcy on risk of stroke from a recently published meta-analysis.[37] In this analysis, an increase of 5 μmol/L in plasma Hcy was associated with an OR for stroke of 1.59 (95%CI: 1.29-1.96).[37] I conducted a meta-analysis to obtain the weighted mean difference (WMD) in plasma Hcy levels between individuals homozygous for the T allele and subjects homozygous for the C allele. The WMD was obtained separately for subjects with and without known cardiovascular disease and for both groups combined. For these analyses a random effect model was used.[53] Then, I conducted a second meta-analysis of all published studies to obtain a OR for all strokes for individuals homozygous for the T allele compared with homozygotes for the C allele. Fixed effect ORs and 95% confidence intervals (CI) were calculated using the Mantel-Haenszel method,[54] while the DerSimonian and Laird method[53] was used to calculate random effects ORs and their 95%CI. Also, to evaluate the robustness of our findings I calculated different ORs according to outcome sub-type, ethnic background, publication language and type of publication.

The DerSimonian and Laird *Q* test[55] was used to evaluate the degree of heterogeneity between studies, and funnel plots and the Egger regression asymmetry test to evaluate small-study bias.[56] Meta-regression was also used to evaluate the extent to which different variables explained heterogeneity in WMD and in the summary ORs.[57] Finally, I used the WMD in Hcy concentration by *MTHFR*/C677T polymorphism to estimate an expected increase in the risk of stroke (non-genetic OR) assuming that an increase of 5 μmol/L in plasma Hcy would be associated with an OR for stroke of 1.59 (95%CI: 1.26-1.96), and that this association follows a log-linear relationship.[37] The uncertainty surrounding the estimate of the non-genetic OR is a function of the variability of the WMD in Hcy by genotype and the variability of the summary OR from Wald's meta-analysis,[37] and cannot be directly calculated. Therefore, to obtain a 95%CI for this OR, I generated one million values from a normal distribution with mean and standard deviations equal to the WMD and its standard error, and one million values from a normal distribution with mean and standard deviation equal to the natural logarithm of the summary OR from Wald's meta-analysis [37] and its standard error (calculated from its 95%CI). I used the simulated values to calculate one million estimates of the non-genetic OR and took the 2.5% and 97.5% percentile values of the so created empirical distribution as 95% confidence limits. The original do-file program in Stata for the simulations was generated by Dr. Leonelo Bautista from Department of Population Health Sciences at University of Wisconsin School of Medicine and Public Health.[58]

Then, the non-genetic OR was compared with the summary OR obtained from the meta-analysis of genetic studies by means of an interaction test.[59] Consistency between the two ORs would indicate that the association between plasma Hcy and stroke seen in non-genetic epidemiological studies is unlikely to be the result of confounding or reverse causality bias. Data were analysed

using the Review Manager software (version 4.2) from the Cochrane Collaboration 2003 and Stata 8.0.

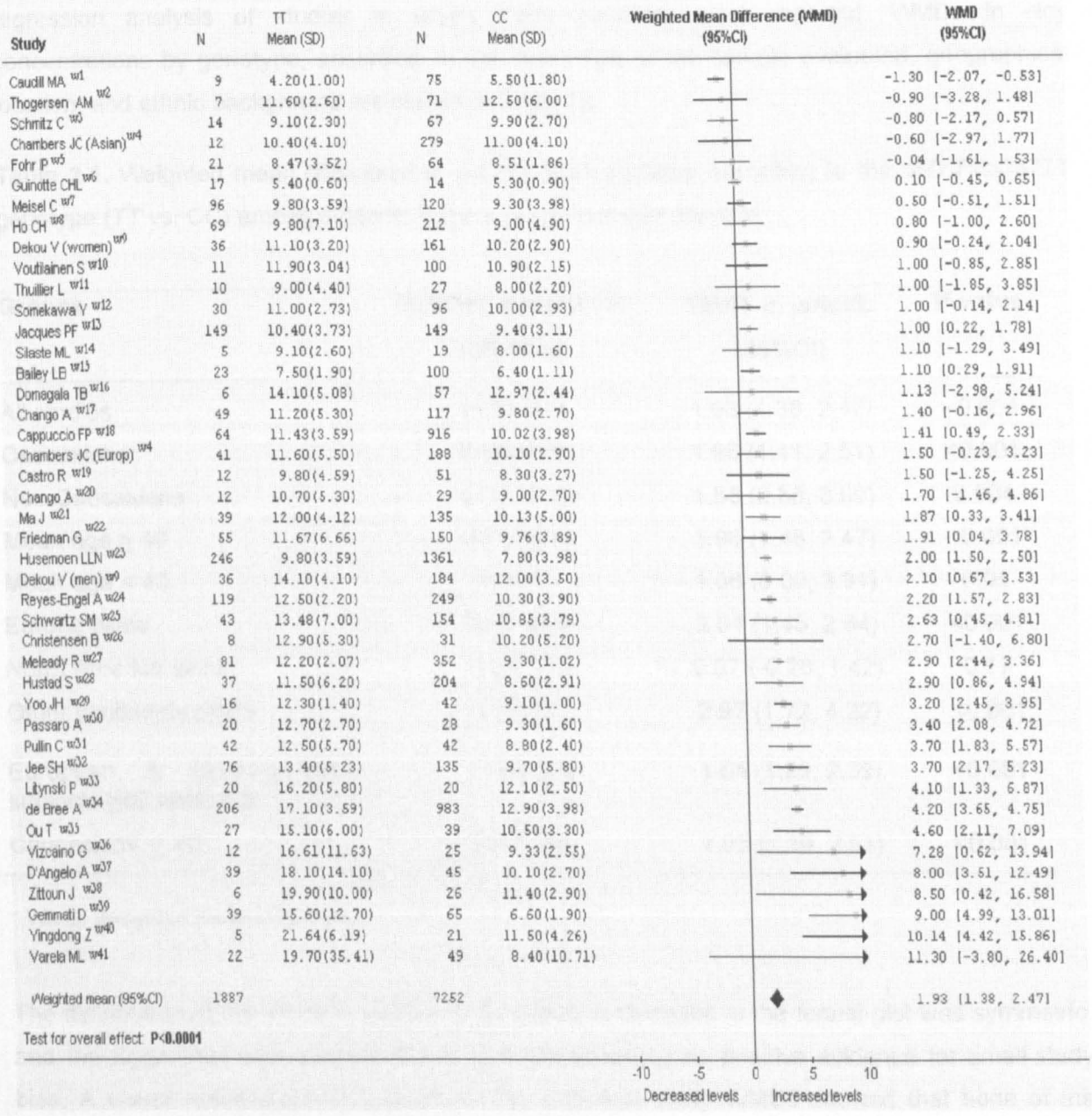
2.4 Results

2.4.1 Homocysteine and MTHFR/C677T polymorphism

The primary search generated 104 potentially relevant studies of which 81 met the selection criteria.[w1-w81] From the 23 articles excluded, 16 did not provide sufficient data to calculate the WMD between the genotypes,[w82-w97] and the relevant information could not be obtained from the authors. Two were discarded as likely duplication[w98,w99] and five only reported the Hcy levels for TT subjects and C-allele carriers, but not for the CC genotype alone.[w100-w104] Of the 81 studies included (31,355 subjects), information for the main comparison (TT vs. CC genotype) was available for 15,635 subjects without known CVD,[w1-w41] for 6312 subjects with CVD,[w2,w4,w7,w11,w16,w21,w26,w27,w29,w32,w37,w39-w55] and for 9408 subjects reported only as the combination of individuals with and without CVD.[w56-w81]

Among subjects without CVD, the WMD in plasma Hcy between subjects homozygous for the T allele compared with homozygotes for the C allele was 1.93 μmol/L (95%CI: 1.38, 2.47; p<0.001, Figure 2.1).[w1-w41] Significant inter-study heterogeneity was observed (P-value for heterogeneity; PHet <0.001).

Figure 2.1. Weighted mean differences in plasma Hcy (μmol/L) according to the *MTHFR*/C677T genotype (TT vs. CC) among subjects without known cardiovascular disease.



*; The total number of subjects without known cardiovascular disease included in the meta-analysis was 15,635 (TT: 1887, CT: 6496 and CC: 7252).

From the variables evaluated in a meta-regression analysis, the only major source of heterogeneity detected was the mean level of serum folate. The WMD in plasma Hcy levels comparing the TT and CC genotypes was 0.048 μmol/L less for each increase in mean serum folate of 1 nmol/L, p=0.035. In studies conducted in North America the WMD in Hcy concentration (0.57 μmol/L) was considerably lower than in studies conducted in Europe (2.04 μmol/L) or other

continents (2.97 μmol/L; Table 2.1). Conversely, the weighted mean serum folate concentration was higher in North American (25.3 nmol/L; 95%CI: 7.3, 43.3) than European studies (13.6 nmol/L; 95%CI: 11.4, 15.6) and studies conducted in other continents (15.3 nmol/L; 95%CI: 9.0, 21.6). No other sources of heterogeneity by age (p=0.94), ethnic background (Caucasians vs Non-Caucasians) (p=0.62), smoking (p=0.41) or gender (p=0.57) were observed in a meta-regression analysis of studies in which these variables were reported. WMDs in Hcy concentrations by genotype, according to the mean age of the sample evaluated, geographical location, and ethnic background are shown in Table 2.1.

Table 2.1. Weighted mean difference in plasma homocysteine according to the *MTHFR*/C677T genotype (TT vs. CC) among subjects without cardiovascular disease.

Groups	Number of subjects (studies)	WMD* in μmol/L (95%CI)	P-value
All studies	9139 (43)	1.93 (1.38, 2.47)	<0.001
Caucasians	6948 (32)	1.96 (1.41, 2.51)	<0.001
Non-Caucasians	2191 (11)	1.83 (0.58, 3.09)	0.004
Mean age ≥ 40	8575 (36)	1.96 (1.46, 2.47)	<0.001
Mean age < 40	564 (7)	1.66 (0.00, 3.31)	0.05
Europe alone	7031 (26)	2.04 (1.45, 2.64)	<0.001
North America alone	1027 (8)	0.57 (-0.28, 1.42)	0.19
Other continents alone	1081 (9)	2.97 (1.72, 4.22)	<0.001
European & North-American, subjects ≥40 years old	7742 (29)	1.84 (1.29, 2.39)	<0.001
Caucasians, ≥ 40	6747 (29)	1.95 (1.39, 2.51)	<0.001

*;WMD: Weighted mean difference.

The distribution of the WMD in relation to its standard deviation in the funnel plot was symmetric, and the Egger test was non-significant (p=0.72) providing no positive evidence for small-study bias. A visual assessment of a graph of the individual study WMDs showed that none of the studies had an undue influence on the overall WMD.

Among subjects with CVD, homozygous for the T allele, Hcy levels were 4.35 μmol/L (95%CI: 3.22, 5.49; p<0.001) [w2,w4,w7,w11,w16,w21,w26,w27,w29,w32,w37,w39-w55] higher than subjects homozygous for the C allele. A meta-regression analysis indicated that the greater difference in Hcy by genotype among people with CVD was partially explained by lower serum folate (the crude β-coefficient for disease status was 3.04, falling to 2.73 after adjustment for serum folate). When the data from the previous groups were combined with data from studies that evaluated individuals with and without CVD without distinction, the WMD for homozygotes for T allele against homozygotes for the C allele was 3.10 μmol/L (95%CI: 2.54, 3.65; p<0.001).[w1-w81]

2.4.2 Odds ratio estimated from non-genetic observational studies and the mean difference in homocysteine by the *MTHFR*/C677T polymorphism

According to a previous meta-analysis of prospective studies,[37] a plasma Hcy concentration higher by 5 μmol/L corresponds to an OR for stroke of 1.59 (95%CI: 1.29, 1.96). An increase in Hcy concentration of 1.93 μmol/L (95%CI: 1.38, 2.47) in healthy TT-genotype subjects, would therefore result in an expected OR for stroke of 1.20 (95%CI: 1.10, 1.31) compared with individual homozygous for the C allele, if the association between Hcy and risk of stroke follows a log-linear relationship (Table 2.2), and is free from confounding and reverse causality bias. Expected ORs for stroke derived from differences in Hcy by genotype in all subjects without CVD, and also separately by ethnic background, geographical location and mean age are summarised in Table 2.2.

Table 2.2. Evaluation of consistency between odds ratios (ORs) derived by extrapolation from non-genetic observational studies with those derived from meta-analysis of genetic studies. All ORs are based on comparisons of TT vs. CC genotypes.

	Expected OR for stroke ‡ (95%CI)	Genetic OR† (95%CI)	P-value for comparison¶
Main comparison			
All studies	1.20 (1.10, 1.31)	1.26 (1.14, 1.40)	0.29
Secondary comparisons			
Studies with mean age ≥40 years	1.20 (1.10, 1.30)	1.26⊥ (1.14, 1.40)	0.30
Europe & North America alone	1.17 (1.09, 1.27)	1.21 (1.02, 1.43)	0.37
Caucasians alone	1.20 (1.10, 1.31)	1.19 (1.02, 1.39)	0.39

‡ ORs were calculated using data from subjects without cardiovascular disease. Mean expected odds ratios were calculated using the formula: expected OR= 1.59 raised to the power of d/5, where d=WMD in homocysteine by genotype, using the assumption that a 5 μmol/L increase in homocysteine is associated with an OR for stroke of 1.59 [37]. †; ORs obtained from the genetic meta-analysis of association studies of MTHFR/C677T and stroke. ¶; P-value for the comparison using an interaction test (see methods) ⊥; The weighted mean age for cases and controls were 58 and 53 years old respectively.

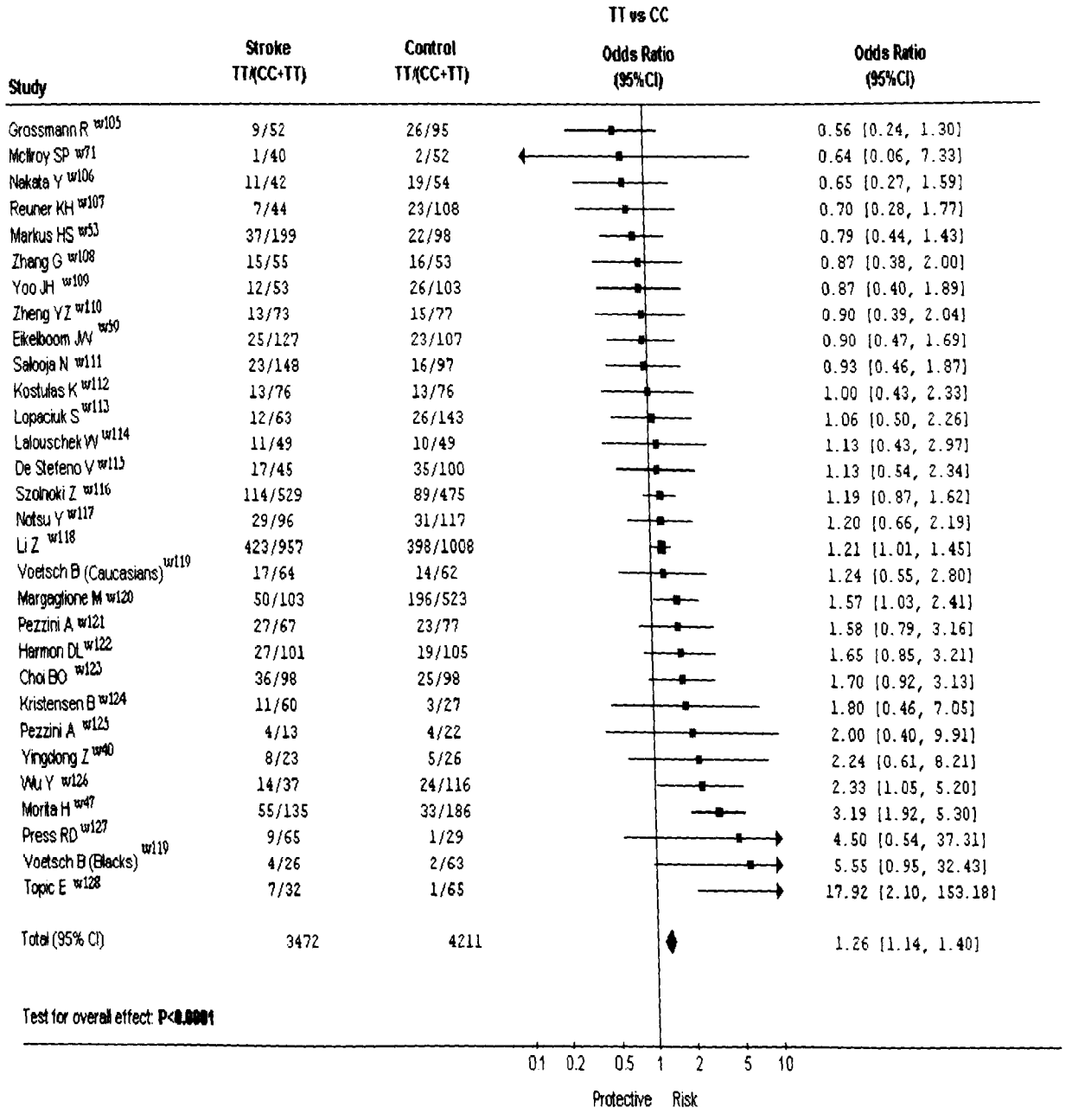
2.4.3 *MTHFR*/C677T polymorphism and stroke

Of the 49 potentially relevant studies identified in the primary search, 30 met the selection criteria.[w40,w47,w53,w59,w71,w105-w128] From the 19 articles excluded for the purpose of the main comparison, three were published in non-English journals,[w129-w131] two reported the outcome silent brain infarction,[w132,w133] and five were published only as abstracts.[w134-w138] Though excluded from the main comparison, these studies were used in a sensitivity analysis. Of the remaining nine publications excluded, five studies encompassed some duplication,[w62,w139-w142] and four did not report the genotype frequency and the relevant information could not be obtained from the authors.[w103,w143-w145] Of the 30 studies (6324 cases and 7604 controls) included, 19 were in Caucasians subjects,[w53,w59,w71,w105,w107,w111-w116,w119-w122,w124,w125,w127,w128] 10 involved Asian subjects[w40,w47,w106,w108-w110,w117,w118,w123,w126] and one study included both Caucasian and subjects of Afro-Caribbean origin.[w119]

The summary OR, under a fixed effect model, indicated that individuals with the TT-genotype compared to subjects homozygous for the C allele had an OR for stroke of 1.26 (95%CI: 1.14, 1.40; p<0.001)(Figure 2.2).[w40,w47,w53,w59,w71,w105-w128] Significant heterogeneity among the results of individual studies was observed (PHet=0.034). A sensitivity analysis revealed that the study by Morita H. et al,[w47] was mainly responsible for the heterogeneity. After excluding this study, the heterogeneity was no longer significant (p=0.32), but the estimate of the overall effect changed very little and remained significant (OR: 1.20 [95%CI: 1.08, 1.34]; p=0.0006). Likewise, a random effect model that takes into account the intra and inter-study variability resulted in a similar overall estimate (OR: 1.26 [95%CI: 1.07, 1.47]; p =0.004). A meta-regression analysis showed that ethnic background (Caucasians vs. non-Caucasians; p=0.23), the presence of risk factors such as age (p=0.38), gender (p=0.46), hypertension (p=0.15), smoking (p=0.43) and diabetes (p=0.10) were not significant sources of heterogeneity in a group of 20 studies with information on all these variables.[w47,w53,w59,w105,w108,w109,w111,w116,w117,w118,w119,w120,w121,w122,w123,w124,w125,w126,w127]

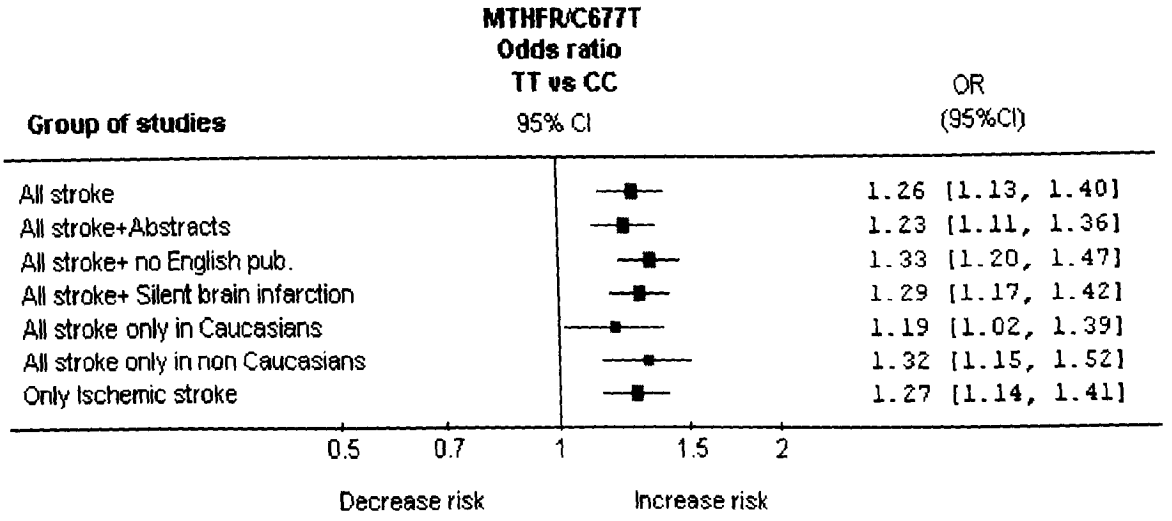
The distribution of the OR in relation to its standard deviation in the funnel plot was symmetric, and the Egger test was non-significant (p=0.32) providing no positive evidence of small-study bias. No individual study had an undue influence on the summary OR. A sensitivity analysis revealed a robust association between the *MTHFR*/C677T polymorphism and stroke (Figure 2.3). No significant changes in the summary OR were detected after adding studies published in non-English language,[w129-w131] in abstract form,[w134-w138] and those using the outcome silent brain infarction.[w132,w133] Similarly, no differences in the summary OR were observed when the analysis was restricted according to ethnic background (Caucasians or non-Caucasians) or to studies in which the outcome was solely ischaemic stroke (Figure 2.3). Data from five studies (611 cases and 2405 controls)[w106,w110,w111,w118,w129] where information was available on the association of the *MTHFR*/C677T polymorphism and haemorrhagic stroke (TT vs. CC), yielded a summary OR of 1.16 (95%CI: 0.90, 1.50; p=0.25), under a fixed effect model.

Figure 2.2. Odds ratio of stroke in subjects with the TT-genotype in comparison with subjects homozygous for the C allele of the *MTHFR*/C677T polymorphism.



*; The total number of stroke cases was 6324 (TT: 1041, CT: 2852 and CC: 2431) and the total number of controls was 7604 (TT: 1140, CT: 3393 and CC: 3071).

Figure 2.3. Sensitivity analysis of the *MTHFR*/C677T polymorphism and stroke association. Summary odds ratio of stroke for TT subjects compared to CC subjects by publication type, stroke classification or type, and ethnic group



2.4.4 Consistency between the odds ratio estimated from non-genetic and genetic studies

OR estimates for TT homozygous subjects in comparison with CC genotype subjects, based on the WMD in Hcy by genotype together with the risk observed in non-genetic observational studies were, mathematically, very similar to those observed in the genetic meta-analysis (Table 2.2). The P value, for interaction, for the main comparison of the study, the non-genetic and genetic OR for stroke in TT vs. CC subjects, derived from all the studies was 0.29. Therefore, there was no evidence of a significant difference between the OR estimated from the increment in the Hcy levels by genotype (*MTHFR*/C677T) and that observed from the genetic meta-analysis. Moreover, no significant differences between estimated ORs were observed when similar comparisons were conducted among more homogenous groups defined by geographical location, ethnic background and mean age (Table 2.2). In all the comparisons, the expected ORs were within the 95%CI of the corresponding observed OR obtained from the genetic association studies.

2.5 Discussion

The main finding of these meta-analyses was that the OR for stroke conferred by the *MTHFR*-TT genotype (1.26 [95%CI: 1.14, 1.40]) was similar to that estimated by extrapolation from non-genetic observational studies and from the difference in the Hcy levels generated by the polymorphism (1.20 [95%CI: 1.10, 1.31]). Indeed after exclusion of the study of Morita et al.,[w47]

which was responsible for much of the heterogeneity in the genotype-stroke analysis, the summary OR for the *MTHFR*-TT genotype was 1.20 (95%CI: 1.08, 1.34), identical to the predicted effect of the polymorphism estimated from the studies of Hcy level and stroke risk. Since the genotype-Hcy and genotype-disease associations should both be unbiased and non-confounded, these results imply that the relationship between Hcy and stroke seen in prospective non-genetic observational studies is not subject to substantial residual confounding or reverse causality bias, and provides evidence for a causal role for Hcy in stroke pathogenesis.

My first meta-analysis examining the association between Hcy and MTHFR genotype, involving more than 31,000 subjects allowed us to refine the size estimate of the effect of the *MTHFR*/C677T polymorphism on plasma Hcy concentration, and to explore sources of heterogeneity. The absolute difference in Hcy conferred by the genotype though consistent in direction was greater in subjects with established atherosclerosis (4.35 μmol/L) than in individuals who were healthy at the time of measurement (1.93 μmol/L). Similar findings were obtained in a recent individual-patient data meta-analysis.[46] Therefore, my data suggests the presence of an additional effect of disease status on the MTHFR-Hcy association, partially explained by lower serum folate levels in subjects with disease. I therefore used estimates from healthy subjects when calculating the predicted OR for stroke among TT subjects. When making estimates of the genotype-intermediate phenotype association, for the Mendelian randomisation approach the potential modifier effect of disease status should be evaluated.[60,61]

The second genetic meta-analysis (*MTHFR*/C677T and stroke), which included ~14,000 individuals, allowed us to obtain a precise estimate of the effect of *MTHFR* genotype on stroke risk. Subjects homozygous for the T allele had an OR for stroke of 1.26 (95%CI: 1.14, 1.40) compared with subjects homozygous for the C allele. This difference in risk is similar to that estimated from the difference in Hcy levels by genotype in healthy individuals, (OR: 1.20 [95%CI: 1.10, 1.31]). Moreover, when the comparison was restricted to Caucasian individuals or to studies conducted in Europe or North America, within a comparable age range as the population included in the meta-analysis of prospective studies of Hcy and stroke risk,[37] similar results were obtained.

Clinical studies have shown that supplementation with B-vitamins (folic acid, B-6 and B-12), decreases Hcy levels by an average of 3 μmol/L.[49] If Hcy is causally associated with an increased risk of stroke, as this and other studies suggest,[46] then nutritional interventions to lower Hcy might be expected to produce a relative risk reduction in the incidence of stroke of about 23%.[38] A number of randomised clinical trials of the effects of lowering Hcy with B vitamins (folic acid, B-6 and B-12) on cardiovascular outcomes have been conducted and other are still ongoing and their results are placed in the context of an updated analyses of genetic studies in the Chapter 3. Despite evidence for causality, from this and other similar analyses, such randomized intervention trials of Hcy lowering are important firstly because of several

potential limitations of the type of study I have conducted, secondly to establish precisely the magnitude of the treatment effect, and lastly to ascertain whether there are any unexpected harms (e.g. colorectal cancer) from such therapies.

Despite the consistent association demonstrated between the *MTHFR*/C677T polymorphism, Hcy and stroke risk, the size of the effect is modest in comparison to orthodox cardiovascular risk factors and does not necessarily provide a rational basis for screening for the *MTHFR*/C677T polymorphism or for the measurement of Hcy in isolation, in the prediction of stroke. Whether either of these measurements would add useful predictive information to more established risk prediction tools (e.g. Framingham risk equation) will require further evaluation.

The interpretation of the analysis must be made in the context of the limitations of the available data. In the meta-analysis of the *MTHFR* variant and Hcy levels, significant heterogeneity was observed (PHet<0.001). This is perhaps not surprising since the effect of the mutation could be modified by environmental factors such as folate levels, as detected in my meta-regression analysis. This is supported by the smaller effect of the *MTHFR*/C677T polymorphism on the Hcy in studies conducted in North America, in which the mean levels of serum folate were higher, presumably as a result of folic acid fortification of cereals, and flour initiated in 1996 and fully established by 1998. However, larger data-sets using individual-patient data will be required for a more precise quantification of this potential gene-environment interaction. For this reason, only the WMDs derived from a random effect model were used in the current analysis.

Publication bias is an unlikely explanation of the observed association between the *MTHFR* polymorphism and Hcy, and between this variant and stroke risk, as indicated by the results obtained from the funnel plots and the Egger's tests. Although confounding is less likely in analyses of an association of a genotype with disease, some imbalance in the distribution of cardiovascular risk factors by *MTHFR* genotype cannot be totally excluded. However, previous studies that evaluated the effect of the *MTHFR* variant on CHD suggested that systematic confounding from other cardiovascular risk factors (e.g. age, gender, hypertension, diabetes, obesity, or alcohol intake) is of little relevance.[46,62]

The Mendelian randomisation approach is a potentially useful tool to evaluate the nature of the observed associations between putative risk factors and disease. This approach overcomes some limitations of observational studies.[22-24] However, this approach has its own theoretical limitations, which were mentioned previously in Chapter 1, section 1.3 in this document. Despite evidence for causality, from this and other similar analyses, randomised intervention trials of Hcy lowering are important firstly because of several potential limitations of the type of study I have conducted, secondly to establish precisely the magnitude of the treatment effect, and lastly to ascertain whether there are any unexpected harm from such therapies.[63] However, the data

obtained from this study and previous analyses of the same type, provide a source of optimism that such therapies will be beneficial.

In conclusion, synthesis of the genetic evidence to December2003 on *MTHFR*-Hcy levels-Stroke indicated a considerable consistency in the risk of stroke observed for subjects carrying the *MTHFR* variant with that suggested from observational studies. Nonetheless, there were two important limitations, publication bias and the presence of an effect modification of the *MTHFR*-Hcy-Stroke association by folate levels that unfortunately could not be analysed in a detailed manner due to limited number of evidence for such analyses. In order to address these limitations in detail the work obtained in Chapter 3 was initiated.

Chapter 3 Homocysteine and stroke: Gene-environment interaction

In Chapter 2, I synthesised the published genetic evidence to December 2003 on the *MTHFR*/C677 variant-Hcy and stroke. Although the results were promising, there were important limitations that were not adequately addressed. Firstly, a possible overestimation of a genetic effect due to small-study bias, and secondly, the presence of effect modification on the *MTHFR*-Hcy-Stroke association by folic acid intake suggested by the results from Chapter 2. These are now addressed in detail in the present Chapter. In addition, I also contextualized the genetic evidence with that from that of clinical trials in stroke using Hcy-lowering therapies.

3.1 Homocysteine and stroke: gene-environment interaction and implications for clinical trials of folic acid

Evidence from prospective cohort studies have estimated that a reduction of 3 μmol/L in serum homocysteine, after a mean follow-up of 7.3 years, is associated with decreased risk of CHD by 18% and stroke by 24%.[37,38] Nonetheless, residual confounding and reverse causation, as discussed in Chapter 2, are still possible explanations for these estimations.[40,41]

A genetic approach that overcomes reverse causality and minimises residual confounding has been used previously to judge the possible causal role of Hcy in stroke and CHD.[37,46] In Chapter 2, I described the results of the Mendelian randomisation experiment of Hcy in stroke. In those analyses, I utilised published genetic association studies on the *MTHFR*/C677T variant (rs1801133) that serves as an unconfounded proxy for Hcy level. Although the results from those literature-based meta-analyses provided some support for a causal role of Hcy in stroke, there are several limitations that currently limit definitive conclusions to be drawn. First, since my Mendelian randomisation experiment was based on published studies, there is scope for small-study bias. This has been suggested as an explanation for the association of *MTHFR*/C677T in CHD.[64] Indeed, in my previous meta-analysis of genetic studies on stroke only two studies included more than 400 cases.

Second, my analysis on stroke (Chapter 2, section 2.4) as well as data from other studies,[65] suggests that the effect of *MTHFR*/C677T variant on the Hcy concentrations could be modified by the prevailing intake of folic acid. In individuals with high levels of folic acid the effect of *MTHFR*/C677T variant on Hcy levels will be lower compared to individuals with low levels of folic acid. If this is the case, and also taking into consideration that mandatory policies of folic acid fortification of cereals and flour has been established in several countries,[66] it would affect not only the interpretation of genetic studies on *MTHFR*/C677T and stroke, but also the results from emerging and ongoing randomised clinical trials of Hcy-lowering interventions.[67]

In order to further evaluate the potential limitations of findings on stroke and Hcy presented in Chapter 2 in a reliable manner, I have updated the previous literature-based meta-analyses to

January 2008, and also assembled a network of genetic studies on stroke that now includes both large studies and unpublished studies to provide a much larger aggregate sample size and to help overcome small-study bias. Moreover in this update analysis I have also been able to investigate, in more detail, the potential modifying effect of folic acid consumption. The results from this updated meta-analysis has further been contextualised with a pooled analyses of published randomised clinical trials evaluating the effect of Hcy-lowering interventions (B-vitamins) on cardiovascular end-points, specifically stroke, to assess the concordance between the findings of genetic and interventional studies.

3.2 Establishment of a large scale collaboration to evaluate the effect of MTHFR/C677T polymorphism in homocysteine and stroke

Relying exclusively on literature-based meta-analyses for the investigation of a gene-environment interaction in stroke may have problems. Meta-analyses using "only published data" are prone not only to small-study bias, but also to missing data of relevance to the investigation of gene-environment interaction (e.g. levels of Hcy and folate). In addition, there is also the scope for reporting bias, mainly from prospective cohort studies that obtain information on multiple incident disease outcomes. Such studies may have obtained information on incident stroke and have typed the *MTHFR*/C677T for other reasons may not have reported the association due to the relative small number of strokes per-study that poses an obstacle to publication. For these reasons, I supplemented information from published studies with unpublished genetic data obtained by assembling a large collaborative network of investigators allowed me to obtain information on the *MTHFR*/C677T effect on stroke, as well as on Hcy and folic acid concentrations for a more precise estimation of genetic effect sizes, while minimising the scope for reporting and publication bias. The collaborative group for the genetic analysis was assembled by direct contact with principal investigators of any study that had previously reported at least one genetic finding in stroke (for case-control studies) or in cardiovascular disease (for cohorts) in a peer reviewed journal, and using a database I have developed and maintained for this and other studies [68,69,70]. This approach allowed me to amalgamate a total of 109 genetic studies evaluating the effect of the *MTHFR* gene on stroke including a total of 92,944 individuals (18,533 stroke events and 74,411 controls) covering the main geographic regions around the world. This database is valuable not only because it contains three times more information than my previous analyses reported in Chapter 2, but also because it provides information on the *MTHFR* effect on Hcy and stroke risk in regions with very different prevailing levels of folic acid consumption, and with different policies on the folic acid fortification of cereals and flour.

3.3 Methods

3.3.1 Search strategy genetic studies

Two electronic databases (MEDLINE and EMBASE) were searched up to January 2008 by Mr Paul Newcombe and I, for all studies on the association between the *MTHFR*/C677T polymorphism and stroke. I used the text words, which were also MeSH terms, "polymorphism", "mutation", "genotype", "genetic", "gene(s)", "allele(s)" in combination with "stroke", "cerebrovascular disorder/disease", "cerebral ischemia", "hemorrhagic stroke", "(silent) brain infarction". Literature searches were limited to "human". All languages were included. Mr Paul Newcombe and I searched for any additional studies in the references of all identified publications, including previous relevant meta-analyses. This sensitivity search strategy allowed me to identify comprehensively published genetic studies in stroke. These were then contacted an invited to participate in the collaboration.

3.3.2 Search strategy for randomised trials

I conducted a literature search in MEDLINE up to August 2008 to search for randomised clinical trials using homocysteine-lowering interventions that evaluated cardiovascular end-points. The MesH terms utilised were "cardiovascular disease", "coronary heart disease", "coronary stenosis", "myocardial infarction", "cerebrovascular accident", "stroke", and "randomised controlled trial", "clinical trial", and "folic acid". I also searched in previous published meta-analyses addressing this question.[67,71]

3.3.3 Selection criteria

3.3.3.1 MTHFR polymorphism and homocysteine level

For inclusion, studies had to have an analytical design (case-control, cohort or cross-sectional) and examine the association between Hcy levels and the *MTHFR*/C677T polymorphism. Studies were included only if they were published as full-length articles or letters in peer-reviewed journals. For duplicate publications the smaller dataset was excluded.

3.3.3.2 MTHFR polymorphism and stroke risk

For inclusion, studies had to have an analytical design (case-control, cohort, or cross-sectional), and had to examine the associations between ischaemic or haemorrhagic stroke, or silent brain infarction and the presence of the *MTHFR* polymorphism. All languages were included. For duplicate publications the smaller dataset was excluded. In all searches, when relevant information was not reported or there was doubt about duplicate publications, Mr Paul Newcombe and I contacted the authors to obtain the required information. In addition, principal investigators from any genetic study that has published in a peer-reviewed journal was also invited to participate in this collaboration by providing genotype counts by case-control status, according to different stroke sub-types as well as Hcy and folic acid levels by genotype.

3.3.3.3 Randomised clinical trials of homocysteine lowering interventions in stroke

Studies had to be randomised, and parallel in design in adults, and examine the effect folic acid supplementation (with or without additional vitamin B supplementation) on stroke. Studies had to have a minimum follow up of 12 months. Only studies published as full-length papers or letters in peer-reviewed journals in English Language were included.

3.3.4 Data Extraction

For the genetic studies, the following variables were extracted: country and year in which the study was conducted, study design, proportion of males, mean age of participants, frequency of genotypes and alleles by case control status, Hcy and folate concentrations, stroke sub-type, language of publication, and ethnic background were extracted and entered into a database by myself and Mr Paul Newcombe. Results were compared and disagreements resolved by consensus. Authors from a total of 72 new identified studies were contacted (on at least three occasions) to obtain data on *MTHFR*/C677T genotype, Hcy and folate concentrations (to avoid reporting bias) by case-control status, including stroke sub-type, as well as to obtain any relevant missing data or when other clarifications of the published information were deemed necessary. A positive reply was obtained and information provided by authors of 28 studies.

For the randomised trials the following variables were extracted: year of conduction of the trial, geographic region in which the trial was conducted, type of blinding, number of participants in each arm, number of stroke events in each arm, the diagnostic criteria used for the stroke outcome, dose of folic acid in the intervention group, type of intervention used in the control arm, mean duration of the trial, and baseline and post-intervention levels of Hcy in each arm.

3.3.5 Statistical Analysis

Statistical analysis was conducted following the guidelines from the HuGENet published in the HuGE Review Handbook for meta-analysis of genetic association studies.[72]

3.3.5.1 Homocysteine levels and MTHFR polymorphism

I used meta-analytic techniques, using random effect models, to obtain the weighted mean difference (WMD) in plasma/serum level of Hcy according to the *MTHFR*/C677T genotype (CC: homozygous common allele; CT: heterozygous and TT: homozygous rare allele). Individuals homozygous for the C allele were used as the reference group. The estimate of WMD was restricted to analyses made in individuals without clinical evidence of CVD (ischaemic heart disease, stroke, or venous thrombosis). In order to explore the modifying effect of folic acid levels on the *MTHFR*/C677T-Hcy association two strategies were utilized both incorporating meta-regression. First, I used the information on the geographical location and date on which the genetic study was conducted to generate geographical regions to serve as proxies of the folate

supplementation status. For this classification, I considered whether policies for folic acid fortification had been initiated and the date of implementation. The categorization was developed by Dr. Robert Clarke from University of Oxford, a collaborator in the analyses described in this Chapter. Categories were sorted according to the level of folic supplementation from the lowest to the highest as follows. (1) Asia: in addition to Asians countries (e.g. China, India, Korea, Thailand, Mongolia) this category also included countries from North and sub-Saharan Africa. (2) Europe-low: including studies from Scandinavia or Netherlands conducted at any time, and European countries (including Turkey and Russia) conducted prior to folic acid fortification of cereal and flour initiated 1995. (3) Europe moderate: including European countries (including Turkey and Russia) in 1995 or thereafter (except Scandinavia and Netherlands). (4) America & Australia & New Zealand low-moderate (pre-fortification): which included studies from North-America Australia and New Zealand prior to 1995, and studies from Central and South America conducted at any time. (5) America & Australia & New Zealand high (post-fortification): which included studies from North-America Australia and New Zealand conducted in 1995 and thereafter. For a total of 24 Studies (of 96) on which information was unavailable from the publication, or previous publications from the same study, or after contacting study authors on the timing of the study, the submission/publication date was used as the year the study was conducted. The second broad approach was possible in a sub-set of 34 studies with information on folic acid level. For these, a meta-regression analysis of the differences in Hcy by *MTHFR* genotype against the folic acid concentrations was conducted. Previous analyses were then repeated only in studies with sample size equal or greater than 500 individuals in order to explore the influence of small-study bias.

3.3.5.2 MTHFR and Stroke

I conducted a second meta-analysis of published and unpublished genetic studies on stroke to obtain a per-genotype summary OR and their 95%CI (using random effect models) for all strokes types combined, using individuals homozygous for the C-allele as the reference group. This outcome referred to as "main stroke comparison" was selected in order to increase comparability with the results from prospective cohort studies and randomised clinical trials. This main stroke comparison included studies reporting: (i) only ischaemic stroke, (ii) only haemorrhagic stroke, (iii) both (ischaemic and haemorrhagic by separate) stroke types, or (iv) only unclassified stroke from studies in which neuroimaging tools were not available to classify the event as haemorrhagic or ischaemic. For studies reporting ischaemic and/or haemorrhagic stroke, the diagnosis was done by neuroimaging (MRI or CT). For this meta-analysis, I assumed equivalence between risk ratio and OR. As described in section 3.3.5.1, I utilised the geographical categories (or region of folate supplementation) to explore whether the *MTHFR*/C677T effect on stroke is modified according to the concentrations of folic acid. Previous analyses were then restricted only to studies with ≥ 400 stroke events, to evaluate the potential influence of small-study bias.

3.3.5.3 Randomised clinical trials of homocysteine lowering medications in Stroke

Finally, I conducted a third meta-analysis of randomised controlled trials that compared the effect of folic acid supplementation (with or without additional B vitamins) on stroke outcomes in trials that used placebo, or low dose-folic acid, or usual care in the comparator arm. The trial arm containing folic acid (with or without additional B vitamins) was assigned as the experimental group. For the stroke outcome (incidence of stroke) a pooled relative risk and 95%CI was calculated using the method of DerSimonian & Laird. When data was available from the publication, I also calculate the mean difference in the change of Hcy between randomised groups (Δ-experimental arm minus Δ-comparator).

In all meta-analyses, the DerSimonian and Laird Q test was used to evaluate the degree of heterogeneity between studies, and the I^2 measure was used to describe the proportion of total variation in study estimates that is due to heterogeneity.[72] Funnel plot and the Egger regression test were conducted to evaluate the presence of small-study bias.[72] For all previous analyses a random effect model was used, in order to allow for any heterogeneity across studies.

3.4 Results

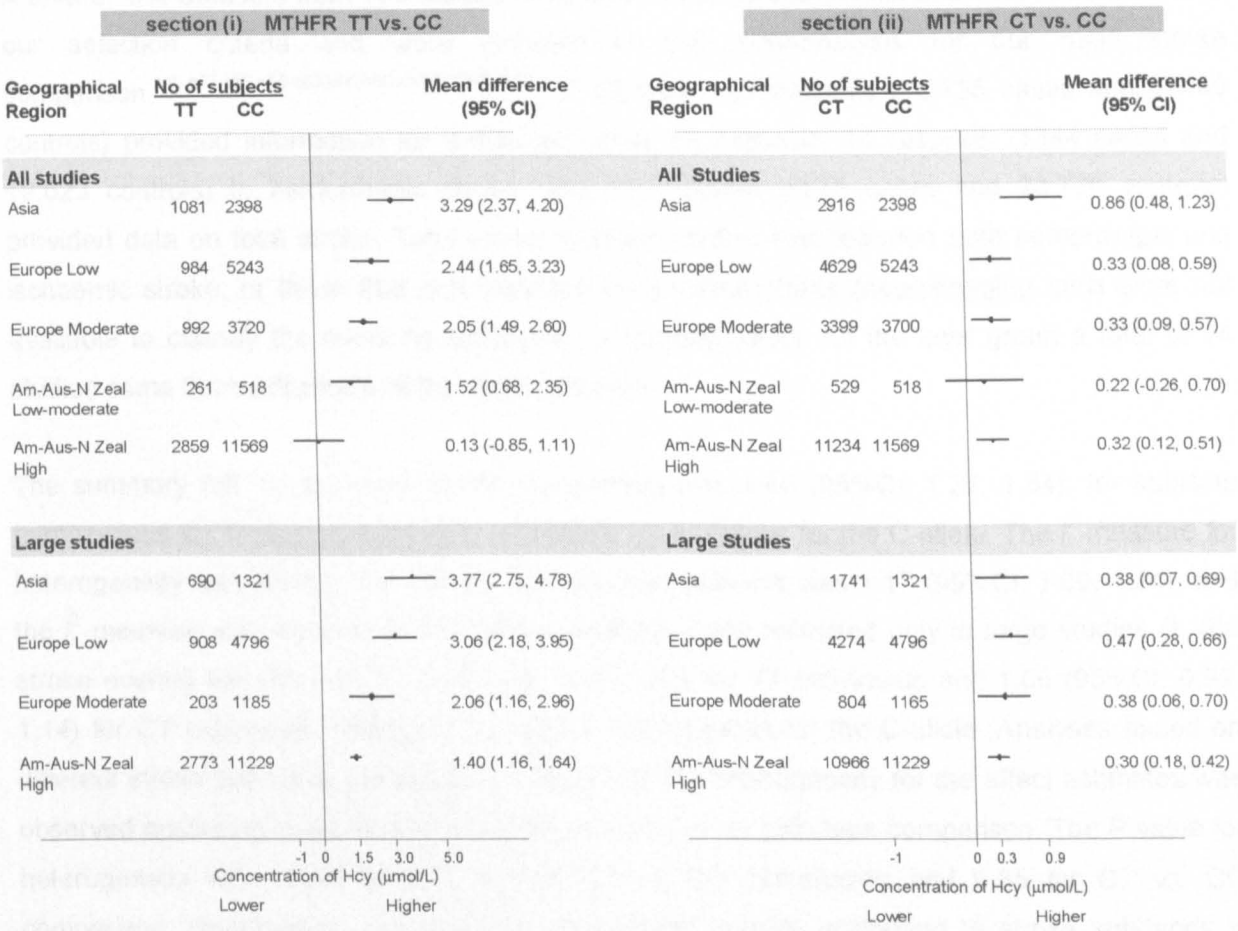
3.4.1 Homocysteine and MTHFR/C677T polymorphism

A total of 96 datasets (from 79 studies) including a total of 58,644 individuals met our selection criteria.[s1-s79] From the 96 datasets, a total of 66 datasets including 52,332 subjects without evidence of CVD were utilised for the analyses, which is more than three times the information included in the analyses in Chapter 2. The additional 30 datasets (of 96) evaluating individuals with CVD were excluded. Among subjects without CVD, the WMD in plasma Hcy between subjects homozygous for the T allele compared with individuals homozygotes for the C allele was 2.15 μmol/L (95%CI: 1.74, 2.55). The I^2 measure was 84.1%. For subjects heterozygous the mean difference in Hcy levels was of 0.42 μmol/L (95%CI: 0.29, 0.55) compared with subjects with the CC-genotype. The I^2 was equal to 50.6%.

The *MTHFR*/C677T effect on Hcy was highly dependent of the geographical region in which the study was conducted. There was a clear trend for subjects homozygous for the T-allele when compared with individuals homozygous for the C-allele, to have higher levels of Hcy in regions with no supplementation of folic acid (Asia). The effect was smaller in the studies conducted in geographical regions with policies for folic acid fortification (Americas-Australia-New Zealand), for details see section (i) Figure 3.1. The X^2 value for heterogeneity between sub-groups of geographical locations was equal to 157.06 ($p< 0.0001$) for the TT vs. CC comparison. When the analyses were limited to large studies (≥ 500 individuals), similar results were obtained, see section (i) Figure 3.1. The patterns were similar when heterozygous subjects were compared with subjects homozygous for the C-allele, though the magnitude of the genetic effect was smaller

(see section (ii) of Figure 3.1). The X^2 for heterogeneity between geographical regions was 9.19 (p= 0.057) for the CT vs. CC comparison.

Figure 3.1. Homocysteine differences in subjects without cardiovascular disease by *MTHFR*/C677T genotype according to regions of supplementation. Upper panel includes all studies and bottom panel includes only studies with more than 500 individuals.



A meta-regression analysis of the differences in Hcy by genotype (TT vs. CC) against the geographical location, coded in ascending order of the levels of folic acid supplementation (Asia= 1; Europe-low= 2; Europe-moderate= 3; America-Australia-New Zealand low-mod= 4 and America-Australia-New Zealand high= 5) showed a reduction in the effect of genotype on Hcy levels. For a 1 category-increase in geographical location, the genetic effect on Hcy levels was reduced by 0.66 µmol/L (p< 0.0001). In order to explore the possible effect of small-study bias, the previous mete-regression was then restricted only to large-studies (≥ 500 individuals) and the pattern was similar (beta-coefficient for a 1 category-increase= -0.59 µmol/L; p= 0.02). In a subset of 34 studies with information on folic acid, a meta-regression analysis indicated that for every increase of 1 ng/mL in folic acid concentration, the mean difference in Hcy levels for TT-subjects vs. CC-subjects decreased by 0.18 µmol/L, p= 0.03.

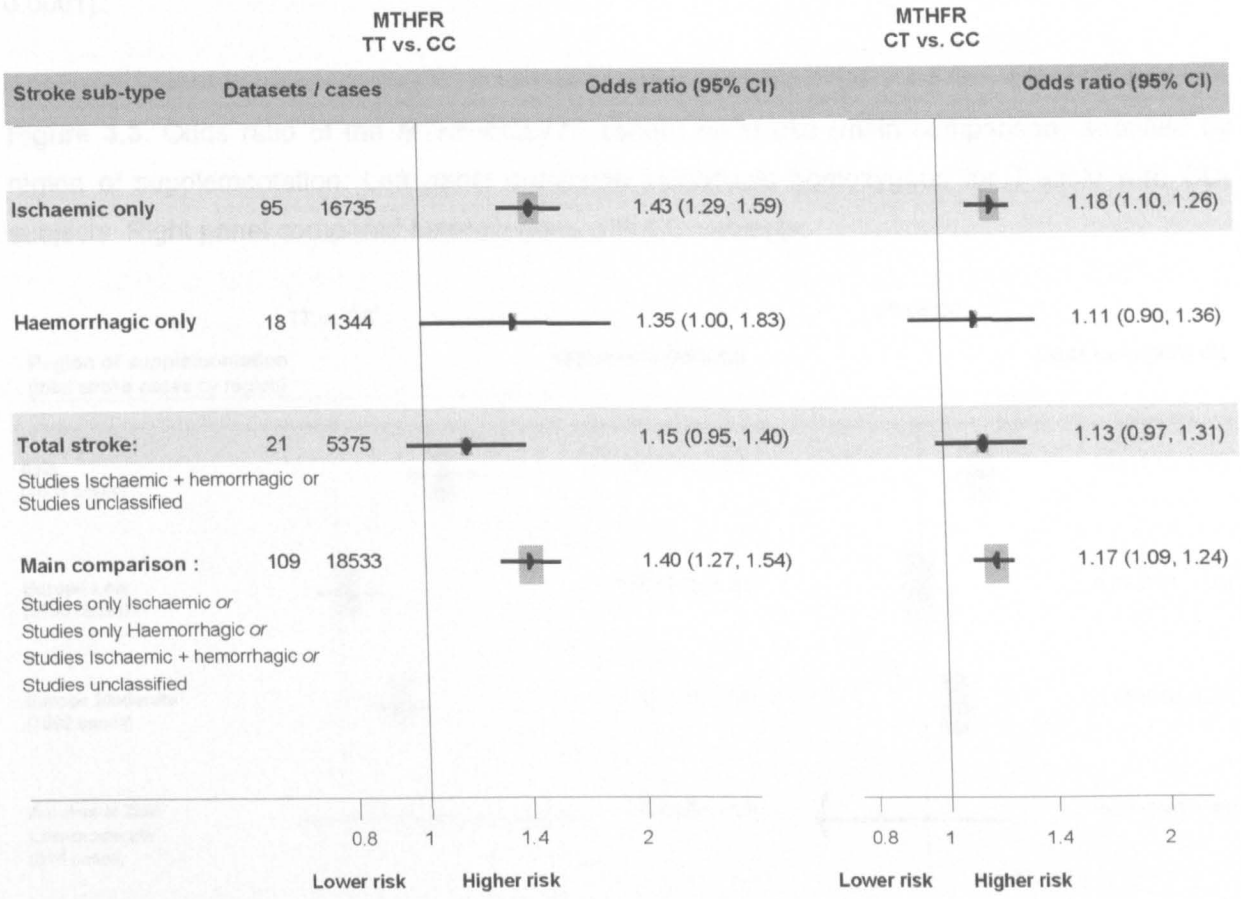
The P value for the Egger regression asymmetry test (including all 66 studies regardless of the geographical location) was 0.06 for TT vs. CC comparison, and 0.13 for the CT vs. CC comparison.

3.4.2 MTHFR/C677T Polymorphism and Stroke

A total of 134 datasets from 109 studies including 18,533 stroke events and 74,411 controls met our selection criteria and were included in the meta-analysis for our main stroke comparison.[s40,s47,s53,s56-s63,s65,s66,s68-s75,s77,s79-s167] Of these, 95 datasets (16,735 cases and 69440 controls) provided information for ischaemic stroke as outcome, 18 datasets (1344 cases and 19,623 controls) to hemorrhagic stroke, and 21 datasets (5375 cases and 23,776 controls) provided data on total stroke. Total stroke included studies that reported both hemorrhagic and ischaemic stroke, or those that only reported unclassified stroke (neuroimaging tools were not available to classify the event as ischaemic or hemmorhagic). Of the later group a total of 14 studies came from individuals of European ancestry.

The summary OR for the main stroke comparison was 1.40 (95%CI: 1.27, 1.54), for subjects homozygous for T-allele compared to individuals homozygous for the C-allele. The I^2 measure for heterogeneity was 46.9%. The OR for heterozygous subjects was 1.17 (95%CI: 1.09, 1.24), and the I^2 measure was equal to 45.8%. When analyses were restricted only to large studies (≥ 400 stroke events) the OR was 1.12 (95%CI: 0.99, 1.27) for TT-individuals and 1.06 (95%CI: 0.99, 1.14) for CT individuals compared to subjects homozygous for the C-allele. Analyses based on different stroke sub-types are shown in Figure 3.2. No heterogeneity for the effect estimates was observed according to the stroke sub-types in either of the genotype comparison. The P value for heterogeneity was equal to 0.17 for the TT vs. CC comparison and 0.85 for CT vs. CC comparison. Nonetheless, the power to detect heterogeneity according to stroke sub-types is limited due to reduce number of certain stroke types, especially hemorrhagic.

Figure 3.2. Odds ratio of the *MTHFR*/C677T variant on different stroke sub-types. Left panel compares individuals homozygous for T-allele with CC-subjects. Right panel compared heterozygous with CC-subjects.

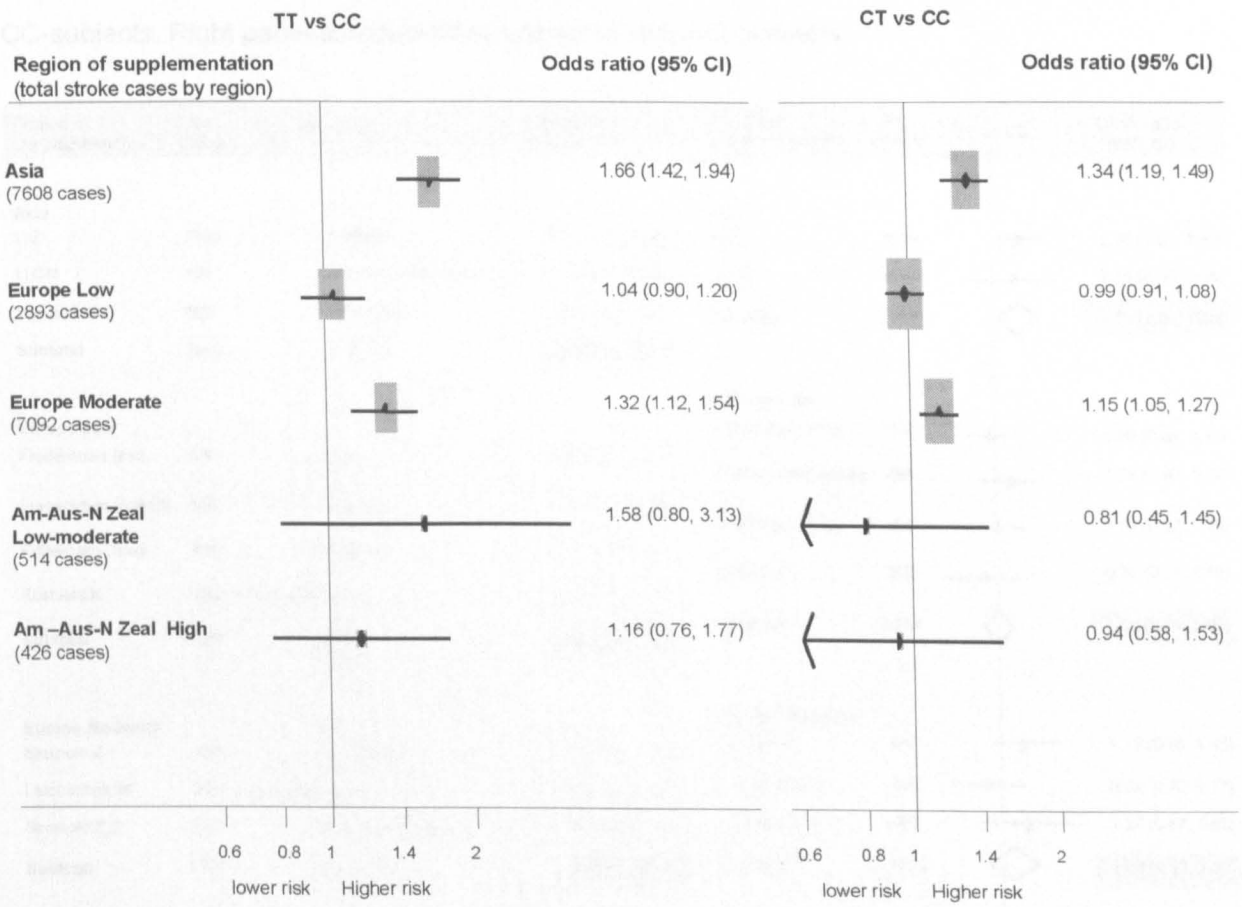


When studies were categorized by geographical region according to the folate supplementation, important differences were observed in the magnitude of the genotypic risks. In Asia, region associated with lower population levels of folic acid,[46,73] and also a region in which the *MTHFR*/C677T variant exhibited the biggest effect on Hcy levels (Figure 3.1), showed the highest increase in risk of stroke, summary OR of 1.66 (95%CI: 1.42, 1.94) for the TT vs. CC genotype comparison(Figure 3.3). By comparison, in regions such as Europe low and moderate with intermediate levels of folic acid and smaller differences in Hcy levels by *MTHFR* genotype (Figure 3.1), the magnitude of the association between stroke and *MTHFR* genotype were null or lower; summary OR for Europe low of 1.04 (95%CI: 0.90, 1.20) and Europe moderate of 1.32 (95%CI: 1.12, 1.54). Regions in which there are policies for folate fortification (America, Australia, and New Zealand post-fortification) the point estimate is smaller (OR of 1.16), however the confidence intervals are still wide, ranging from 0.76 to 1.77 (Figure 3.3). The X^2 value for heterogeneity among geographical regions was equal to 16.77 (p= 0.002) for the TT vs. CC comparison.

Similar differences in the genotypic risks, although of lower magnitude were also observed for heterozygous individuals compared with subjects homozygous for the C-allele (Figure 3.3). The OR for Asian studies was 1.34 (95%: 1.19, 1.49), for Europe-moderate was 1.15 (95%: 1.05,

1.27), while non-significant associations were observed for studies conducted in America, Australia or New Zealand before or after the implementation of folic acid fortification policies (Figure 3.3). The X^2 value for heterogeneity among geographical regions was equal to 24.98 ($p< 0.0001$).

Figure 3.3. Odds ratio of the *MTHFR*/C677T variant on stroke (main comparison) stratified by region of supplementation. Left panel compares individuals homozygous for T-allele with CC-subjects. Right panel compared heterozygous with CC-subjects.

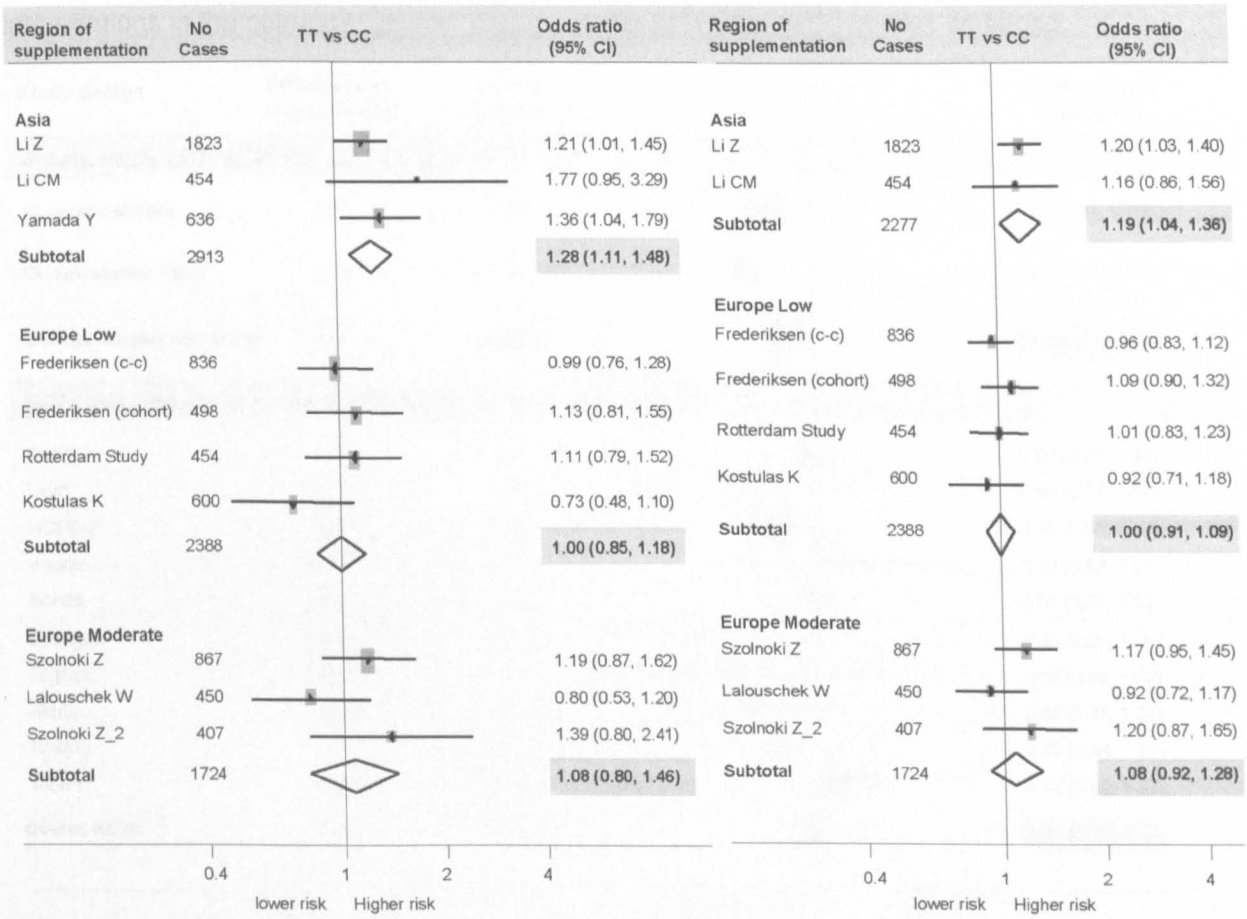


When the *MTHFR*/C677T effect on stroke by geographical region was restricted to large studies (≥ 400 stroke events) a similar pattern of associations was also observed in both genotype comparisons. The summary OR of stroke for studies in Asia was 1.28 (95%CI: 1.11, 1.48) for the TT vs. CC comparison and, 1.19 (95%CI: 1.04, 1.36) for the CT vs. CC comparison. In contrast, studies in Europe low or moderate no increase in the risk of stroke was observed for the *MTHFR*/C677T variant. Unfortunately no large studies have been conducted (or those were not identified by our extensive search strategy) in some geographical regions, such as America, Australia or New Zealand (Figure 3.4).

After including all studies regardless of geographical region by folate supplementation policy, the P value for the Egger regression asymmetry test was < 0.001. However, this asymmetry could be

due to differences in the genotypic effect by geographical location (and therefore by folate levels) as suggested in Figure 3.1 & Figure 3.3. Therefore, I obtained the Egger test for each geographical region, the P values were as follow: Asia= 0.002, Europe-low= 0.75, Europe-moderate= 0.04, Am-Aus-NZ low/mod= 0.05 and Am-Aus-NZ high= 0.6. These analyses suggest that the genetic risks may have been overestimated in smaller studies, and that a less biased genetic estimate was obtained from large studies (bottom panel of Figure 3.1 and Figure 3.4).

Figure 3.4. Odds ratio of stroke (main comparison) by region of supplementation restricted to large studies (≥ 400 stroke events). Left panel compares individuals homozygous for T-allele with CC-subjects. Right panel compared heterozygous with CC-subjects.

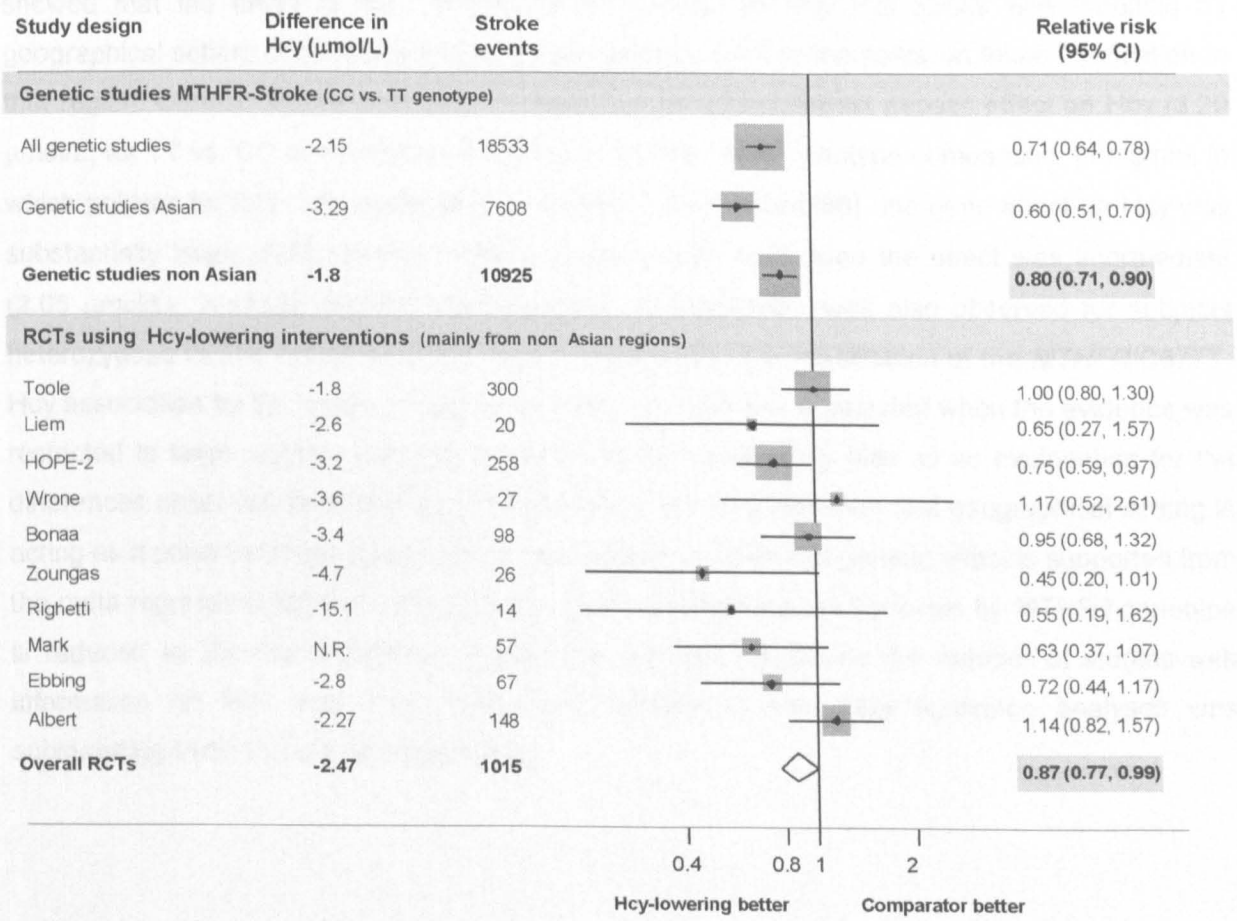


3.4.3 Genetic studies of MTHFR/C677T-stroke in the context of randomised trials of Hcy lowering interventions.

A total of 10 randomised trials of Hcy-lowering interventions including 25,379 individuals, (1015 stroke events) met our selection criteria.[74,75,76,77,78,79,80,81,82,83] The main study characteristics are described in Table 3.1. None of the RCTs was stopped prematurely and, with the exception of two small trials including only 681 individuals (Liem et al, and Righetti et al),[74,75] all trials were double blind. Nine out of ten trials, providing 95 percent of the stroke outcomes, have been conducted outside Asia, i.e. in Europe, America, and Australia.

All individuals included in the trials had a previous history of CHD, stroke or renal disease. In all trials, except the VISP trial [76] that evaluated only ischaemic stroke, the stroke outcome evaluated was composed of both fatal and non-fatal ischaemic, or hemorrhagic events. The weighted mean follow was 5.3 years. After pooling data from ten trials the summary relative risk for stroke in subjects randomised to active treatment was 0.87 (95%CI: 0.77, 0.99) and was associated with a mean reduction in Hcy levels of 2.47 μmol/L in favour of the folic acid arm (with or without B-vitamins), see Figure 3.5. A meta-regression analysis of the log of the relative risk for stroke against the mean difference in Hcy achieved by the intervention was not significant (b-coefficient= 0.06, p= 0.16).

Figure 3.5. Pooled relative risk of stroke from randomised clinical trials of Hcy-lowering interventions in the context of genetic studies of the *MTFHR*/C677T variant on stroke.



When the results of randomised trials were compared with the effect of allocation to the naturally randomised *MTHFR*/C677T variant as proxy for an Hcy-lowering intervention, a similar degree of reduction in the risk of stroke was observed especially when trials were compared with genetic studies conducted in the same geographical region in which trials have been conducted, i.e. non-Asians. The relative risk reduction observed in trials of 13% (95%CI: 1% to 23%) was within the range predicted by genetic studies of 20% (95%CI: 10% to 29%) (Figure 3.5). The random error in the effect estimate from genetic studies was considerably smaller than that observed for

randomised trials. Indeed, genetic studies from a comparable geographical region (i.e. in Europe America and Australia) included up to ten times more stroke events than randomised trials (Figure 3.5).

3.5 Discussion

In order to investigate the possible modifying effect of folate consumption on the Hcy-stroke association, that may affect the interpretation of a possible causal role for Hcy in stroke, I conducted a systematic evaluation of two randomised sources of evidence, genetic studies and clinical trials.

A systematic evaluation of the evidence from (published and unpublished) genetic studies showed that the effect of the *MTHFR*/C677T variant on Hcy and stroke was modified by geographical setting of the study which, in turn corresponded to the policy on folate fortification in that region. Genetic studies conducted in Asia uncovered the largest genetic effect on Hcy (3.29 μmol/L; for TT vs. CC comparison). In contrast, for the same genotype comparison, in regions in which policies for folic acid fortification have been implemented,[66] the gene effect on Hcy was substantially lower (0.13 μmol/L), while in regions such as Europe the effect was intermediate (2.05 μmol/L). A similar pattern, although lower in magnitude, was also observed for subjects heterozygous for the *MTHFR*/C677T variant. Importantly, the modification of the *MTHFR*/C677T-Hcy association by the region of supplementation of folate was unaffected when the evidence was restricted to large studies, reducing the possibility of small-study bias as an explanation for the differences observed (Bottom panel of Figure 3.1). The interpretation that geographical setting is acting as a proxy for folate consumption and that this modifies the genetic effect is supported from the meta-regression analyses that indicated that the difference in Hcy levels by *MTHFR* genotype is reduced as the concentrations of folic acid increase. However, the number of studies with information on folic acid levels that were included in the meta-regression analyses was substantially lower, 34 out of 66 datasets.

Table 3.1. Characteristics of Hcy-lowering trials with information on stroke identified in the systematic review.

Trials (year pub)	No randomised	Blinding	Homocysteine lowering intervention (mg/d)			Reference arm	Length of follow up (months)	Mean age (years)	Geographical location (year conducted)	Outcome
			Folic acid	B_{12}	B_6					
Toole (2004)	3680	Double	2.5	0.4	25	20 mg/d folic acid	24	66	USA, Canada, Scotland (1996, 2003)	Only recurrent Ischaemic stroke, fatal and non-fatal
Liem (2005)	593	Open	0.5	—	—	Usual care	42	65	Netherlands (1998, 2002)	Cerebrovascular accident or transient ischemic attack
HOPE-2 (2006)	5522	Double	2.5	1.0	50	Placebo	60	69	Canada, USA (72.1 %) Brazil, Western Europe & Slovakia (27.9 %) (2000, 2005)	Fatal & non-fatal stroke, including hemorrhagic
Bonaa (2006)	2815	Double	0.8	0.4	40	Placebo	36	63	Norway (1998, 2002)	Fatal & non-fatal stroke, including hemorrhagic
Zoungas (2006)	315	Double	15	—	—	Placebo	43	56	Australia, New Zealand (1998, 2003)	Fatal & non-fatal stroke
Wrone (2004)	510	Double	5 or 15	0.006	12.5	1 mg/d folic acid	24	60	USA (1998, 2000)	Cerebrovascular accident or transient ischemic attack
Righetti (2006)	88	Open	5	—	—	Usual care	29	64	Italy (2003, 2005)	Fatal and non-fatal stroke
Mark (1996)	3318	Double	0.8	0.018	6	Placebo	72	54	China (1985, 1991)	Fatal stroke, including hemorrhagic
Albert (2008)	5442	Double	2.5	1.0	50	Placebo	87	63	USA (1998, 2005)	Ischaemic and hemorrhagic stroke.
Ebbing (2008)	3096	Double	0.8	0.4	40	Placebo	38	61	Norway (1999, 2006)	Fatal & non-fatal stroke, including hemorrhagic

When evaluating the effect of the *MTHFR*/C677T variant on risk of stroke, a similar modification of the genotypic risk by the region of supplementation of folate was also observed, which mirrored the *MTHFR*-Hcy association. Studies in Asia had the largest genotypic risk on stroke, studies in Europe an intermediate effect, and studies in America, Australia or New Zealand showed the lowest or null effect stroke. However, the amount of evidence from the America, Australia or New Zealand is considerably smaller, up to 17 times fewer cases when compared to studies in Asia and Europe (Figure 3.3). This pattern was not only evident for TT homozygous when compared with subjects homozygous for the C-allele, but also for subjects heterozygous for the *MTHFR*/C677T variant. As for the *MTHFR*-Hcy association, the findings of an effect modification on the *MTHFR*-stroke associations by region of folate supplementation remained largely unaltered when the analyses were restricted to large studies, diminishing the possibility of small-study bias as an explanation (Figure 3.4). Although the modifying effect on *MTHFR*-Stroke association by geographical location could be due to differences in the levels of folate consumption, a detailed investigation of this possibility was hampered due to lack of studies with all the necessary information.

Despite synthesis of the genetic evidence suggested an increase risk of stroke associated with the *MTHFR*/C677T variant (used as unconfounded proxy of Hcy), it does not necessarily follow that reducing Hcy levels will reverse the risk of stroke in adult life. Such evidence, if feasible, should be derived from randomised clinical trials using an intervention that specifically reduces the levels of Hcy, such as folic acid.

A pooled analysis of 10 trials published up to August 2008 using folic acid interventions with information on stroke (including 1015 events) indicated that after a mean follow up of 5.3 years and an average reduction in Hcy levels of 2.47 μmol/L, the risk of stroke was reduced by 13% (95%CI: 1% to 23%). However, I was unable to investigate whether the reduction in the risk of stroke associated with Hcy-lowering interventions in trials is modified by the region of folate supplementation (and ideally by levels of folic acid) in a similar manner to that conducted for genetic studies. The reasons for this are that the large majority of trials were based in Europe, North-America and Australia and also, because only aggregate data were obtained for the meta-analyses of clinical trials.

However, it is important to highlight a few important differences between the evidence obtained from randomised clinical trials and genetic studies. First, the majority of trial evidence comes from settings where policies for folic acid fortification had already been implemented.[66,84] By contrast, the large majority of the genetic evidence judged by the number of stroke events is derived from unfortified regions (17,593 out of 18,533 stroke events). Second, despite the level of reduction in Hcy concentrations being largely comparable across the two levels of evidence (2.47 μmol/L for clinical trials and 1.8 μmol/L for genetic studies), the length of follow up is considerable different. Only three trials including 463 stroke events had five or more years of follow

up,[77,81,82] and for all trials randomisation to the intervention occurred in adult life, mean-age range from 54 to 69 years old. In contrast genetic studies can be conceived as life-time natural randomised trials since randomisation to the allele variants (used as instrument for Hcy levels) occurred at time of gamete formation. Third, the number of stroke events in published trials is substantially lower than that included in my genetic analyses (1015 events in trials vs. 10925 events in genetic studies of similar geographical location). All these reasons explain why evidence on trials is still not-definitive, but also why the expected reductions derived from genetic studies are perhaps greater than those observed in trials conducted to date, for detailed discussion see section 8.5 in Chapter 8.

Despite, all these differences a considerable concordance between randomised trials and genetic studies was observed, utilising the evidence available to August 2008. Genetic studies conducted outside Asia (as the trials) showed that individuals homozygous for the C-allele that carried lower levels of Hcy by 1.8 μmol/L compared with TT individuals, also had a reduced risk of stroke by 20% (95%CI: 10% to 29%). Reduction that is similar to that observed in clinical trials, 13% (95%CI: 1% to 23%; see Figure 3.5]). However, as observed from the analyses restricted to large genetic studies, it is entirely possible that the expected reduction for trials conducted outside Asia to be null or a much lower benefit of around 8%, compared with the current point estimate of 13%.

If adequately powered genetic studies of the *MTHFR*/C677T variant on stroke can be envisaged as a type of randomised evidence of the expected benefit that could be achieved by long term reductions in Hcy levels, the genetic analyses presented in this Chapter suggest that the relative risk reduction should be greater in Asian populations compared to populations of European descent (Asian: 40% [95%CI: 30% to 49%] vs. Non Asian: 20% [95%CI: 10% to 29%]). In addition, since the event rates of stroke are higher in Asian countries compared to those from European descent populations,[85] and also considering that Asian countries account for a large proportion of the worldwide population (and therefore of stroke events) the public health impact benefit could be of great importance.[85]

Although the analyses I present here suggest that reducing Hcy may reverse the risk of stroke, there is still considerable controversy whether these findings also apply to CHD. There is the possibility that Hcy is indeed causally associated to both stroke and CHD, but that the magnitude of the association, as reported from observational studies, is greater for stroke than for CHD. A similar situation is seen for systolic blood pressure which has a stronger association with stroke than CHD. If this is case, the genetic signal to be detected by studies evaluating the *MTHFR*/C677T variant on CHD would also be smaller. Along with the fact that the majority of large CHD-genetic studies are being conducted in individuals of European descent, this could partially explain why a clear genetic signal for *MTHFR* has not yet been detected for CHD.[64] Also, if this is correct the expected benefit from randomised trials in CHD would be smaller, suggesting existing trials may be underpowered for CHD.

The analyses I presented in this Chapter have to be considered within the context of their strengths and limitations. The establishment of a large genetic collaboration including a considerable amount of previously unpublished information of relevance to the gene environment interaction (folic acid and Hcy in genetic association studies of stroke), have diminished the possibility of publication and reporting bias. The large sample size included in both genetic meta-anlaysis (58,644 for *MTHFR*-Hcy and 92,944 for *MTHFR*-Stroke) increased the ability to detect differences according to regional policies on folate supplementation as well as by the level of folic acid from individual studies. Finally, I also contextualized the genetic findings with a systematic evaluation of findings from published trial using Hcy-lowering interventions (i.e. I compared the observed effect on stroke from genetic studies with the observed effect size from trials) rather than focusing solely on statistical tests in the genetic associations.

In conclusion, the concordance of findings on the genetic effects of *MTHFR* on Hcy levels and the genetic association with stroke risk argues in favour of the existence of a gene-environment interaction. This has important repercussions for the interpretation of emerging results of clinical trials using Hcy-lowering interventions, which to date seem to suggest that the high risk of stroke conferred by high Hcy levels is reversible.

Chapter 4 C-reactive protein and myocardial infarction

In Chapters 2 and 3, I utilised a Mendelian randomisation approach to evaluate the potential causal role of Hcy in stroke. For this, I took advantage of the large amount of available evidence on a well known gene variant (*MTHFR*/C677T) that was used as a proxy for an indirect biomarker, Hcy. The Mendelian randomisation approach described in this Chapter has two differences compared with the previous Chapters. First, this is the result of a prospective collaboration of genetic studies in which *de novo* genotyping of a single (the most promising at the time) *CRP*-variant was utilised. Second, in contrast the *MTHFR*/C677T-Hcy-Stroke experiment, due to the particularity that the genetic-variant used as proxy lies in the gene (*CRP*) that codifies for the biomarker of interest (plasma CRP), it allows me to obtain empirical evidence on a different type of Mendelelian randomisation experiment. Further details about the different uses of Mendelian randomisation are discussed in Chapter 8.

4.1 C-reactive protein and coronary events in observational studies

Prospective studies indicate a robust association, in healthy individuals, between levels of the acute phase reactant CRP and later coronary events,[86-88] and it has been proposed that measurement of CRP may be a useful adjunct to coronary risk assessment and, perhaps, causally involved in atherosclerosis.[11] However, CRP concentrations are associated with smoking status, blood pressure, abdominal obesity, diabetes, physical activity, and other products of the inflammatory response, and are also higher in individuals with clinical CVD.[89-92] Despite statistical adjustment, this association may therefore be subject to residual confounding or bias due to reverse causation.

A randomised controlled clinical trial of a selective intervention to reduce CRP would provide an unbiased insight into the nature of the association. Unfortunately, no such selective intervention currently exists for use in humans. Although inhibitors of HMG-CoA reductase (statins) and glitazones reduce CRP,[93,94] they also have major effects on lipid profile and glucose metabolism. An alternative approach to control for confounding and reverse causality is to identify a common polymorphism in the *CRP* gene reliably associated with differences in circulating CRP concentration. Due to second Mendel's law, if CRP actually increases the risk of coronary events, then carriage of an allele that exposes individuals to a long term elevation in CRP should confer an increased risk of coronary events proportional to the difference in CRP attributable to the allele. This relation should be largely unconfounded and free of reverse causality bias.[22,23] Therefore, if non-genetic observational studies were unbiased, the increase in risk estimated from these studies should be consistent with the increase in risk conferred by carriage of the allele.

CRP concentration is a heritable trait,[95,96] and in three small studies I and others have found that homozygosity for the rarer T allele of a +1444C>T polymorphism (rs1130864) in the 3' untranslated region of the *CRP* gene was associated with higher basal and/or stimulated CRP

concentrations.[97,98] In order to test whether the increase in risk of coronary events in individuals homozygous for this polymorphism (rs1130864) is close to that predicted under the assumption that the CRP-coronary event association from previous non-genetic studies is unbiased, a large number of European descent men were genotyped de novo.

This approach, of conducting new genotyping within the framework of a prospective collaboration of studies, contrasts with that approach I conducted for Chapters 2 and 3, for which a genetic variant that modify the concentration of the biomarker of interest had been widely studied for several years and allowed me to conduct, in the first instance, a synthesis of the available evidence.

4.2 Methods

4.2.1 Study populations

After obtaining ethical approvals, genotyping for the *CRP* polymorphism and measurement of plasma CRP concentration were conducted in samples obtained from a number of cross-sectional or prospective studies, or randomised controlled trials summarized in Table 4.1.[99-104]

From the Second Northwick Park Heart Study (**NPHS-2**)[99] genotyping was conducted in 2676 Caucasian men of whom 103 developed a non-fatal MI after a median follow-up of 10.6 years.

In a nested case-control genetic study from the West of Scotland Coronary Prevention trial (**WOSCOPS**),[100] which randomised 6595 male participants, initially free from clinically evident cardiovascular disease to either pravastatin or a placebo, 348 incident non-fatal MIs were identified and defined as cases and were matched by age and smoking status with 1103 controls drawn from the remainder of the cohort.

In the Lower Extremity Arterial Disease Event Reduction (**LEADER**) trial [101] a randomised controlled trial of bezafibrate in men with lower extremity arterial disease, a total of 647 Caucasian men in the active treatment and 419 in the placebo arm were included. There were 43 incident non-fatal MIs during follow-up and these subjects were defined as cases.

The Hypercoagulability and Impaired Fibrinolytic function MECHanisms predisposing to MI (**HIFMECH**)[102] study was a European multicentre case-control study of MI. For the present study a total of 491 cases and 517 controls were available for genotyping.

The cross-sectional "**Army-study**" recruited Caucasian men from the Army Training Regiment, (Bassingbourn, UK) to evaluate the relationship between genetic factors and inflammatory response to exercise. DNA was extracted from 219 individuals for *CRP* genotyping and the measurement of CRP plasma concentration.[98]

The UCL DiAbetes and Cardiovascular disease Study (**UDACS**) study[103] is a cross sectional case-control study to evaluate risk factors for coronary heart disease in subjects with diabetes mellitus. For the current study only Caucasian males (n=348) without coronary heart disease were included and contributed to the first analysis.

The Electron Beam Computerised Tomography (**EBCT**)[104] study was designed to compare coronary artery calcification and coronary risk factors in Caucasian Type-1 diabetic patients and non-diabetic participants. Controls were a random sample of the general population (94 men and 107 women), stratified to have a similar age and gender distribution to the patients with diabetes. For the present study the male controls were included (n=72) and contributed to the first analysis.

To test the association of +1444C>T genotype with CRP concentration, male subjects without clinically evident cardiovascular disease from six studies with available DNA and plasma samples were studied; Table 4.1. A second analysis was conducted to evaluate the association between genotype and risk of MI in male subjects from four studies; Table 4.1.

Table 4.1 Design characteristics of the studies included in the present analysis. All subjects included in the studies were of European ancestry.

Name of the Study	Study Design & Median follow-up (years)	Sample size Original/Genotyped	Country(ies)	Study population	Main exclusion criteria	Primary analysis Genotype & CRP level (# subjects)	Primary analysis Genotype & non-fatal MI (# subjects)
NPHS-2	Prospective Cohort (10.6)	3012/2676	UK	Healthy middle-aged men.	▪ Pre-existing cardiovascular disease. ▪ Coronary surgery. ▪ Aspirin or anticoagulant therapy. ▪ Malignant disease.	Yes (2221)	Yes (2676)
LEADER	Nested case-control study from a clinical trial of bezafibrate treatment. (4.6)	1568/1066	UK	Men with lower extremity arterial disease.	▪ Unstable angina. ▪ Total cholesterol <3.5 or >8.0 mmol/l. ▪ Significant renal or hepatic disease or malignant disease.	No	Yes (1066)
WOSCOPS	Nested case-control in a clinical trial of pravastatin treatment (4.9)	6595/1451	UK	Moderately hypercholesterolemic men.	▪ MI, or angina pectoris requiring hospitalisation. ▪ Life-threatening non-cardiac illness.	Yes (1334)	Yes (1451)
HIFMECH	Case-control (N/A)	Cases: 533/491 Controls: 575/517	UK Sweden France Italy	Cases: male MI survivors. Controls: men matched by age and regional areas.	▪ Familial hypercholesterolaemia. ▪ Insulin-dependent diabetes mellitus.	Yes (465 controls)	Yes (1008)
Army	Cross-sectional (N/A)	250/219	UK	Healthy UK-Army recruits.	N/A	Yes (219)	No

‡; DM: diabetes mellitus.

Table 4.1 Continuation.

Name of the Study	Study Design & Median follow-up (years)	Sample size Original/Genotyped	Country(ies)	Study population	Main exclusion criteria	Primary analysis Genotype & CRP level	Primary analysis Genotype & non-fatal MI
UDACS	Case-control (N/A)	Controls: 449/348	UK	Cases: men and women with DM‡ and cardiovascular disease. Controls: men and women with DM but without cardiovascular disease.	N/A	Yes (348 Male controls)	No
EBCT	Case-control (N/A)	Controls: 94/72	UK	Cases: men and women with type-1 DM. Controls: healthy men and women matched by age and gender.	N/A	Yes (72 Male controls)	No

4.2.2 Data collection

Data on demographical variables, cardiovascular risk factors and plasma CRP were obtained from the original studies. Subjects were classified using unified definitions of hypertension, hypercholesterolemia, type-2 diabetes mellitus and obesity from the guidelines on primary prevention of the American Heart Association.[5] For the studies relating genotype and coronary events,[99,100,101,102] non-fatal MI according to WHO criteria,[105] was considered the primary outcome, as this end-point had been uniformly used across all studies. Analysis of genotype and CRP were limited to male subjects to preserve consistency with available studies of genotype and coronary events, which were all conducted in men.

4.2.3 Laboratory analysis

All studies included in the present report used high sensitivity assays to measure plasma CRP concentrations. Plasma CRP concentrations for the NPHS-2 (717 subjects) and LEADER studies were measured by use of commercial assays (R&D Systems). Inter-assay and intra-assay coefficients of variations were 6.2% and 1.9% respectively, with a detection limit of 0.1 mg/L. For an additional 2221 subjects from the NPHS-2 study an Enzyme Immunoassay (Kordia Life Sciences) was used to measure CRP concentrations. For the EBCT and HIFMECH studies CRP was measured with a highly sensitive in-house enzyme-linked immunosorbent assay with rabbit anti-human CRP (Dako, Copenhagen, Denmark) as a catching and tagging antibody with inter and intra-assay coefficients of variations of 4.7 and 3.8% and a limit of detection of 0.15mg/L. For the WOSCOPS study a validated in-house assay was used with a lower limit of detection of 0.1 mg/L. The intra-assay and inter-assay coefficients of variation were 1.9% and 6.2%, respectively. In the Army-study, plasma CRP was measured on a BN Prospec (Dade Behring, Milton Keynes, UK). Inter-assay and intra-assay coefficients of variation were <4% and <2% respectively with a detection limit of 0.2 mg/L. In the UDACS study, CRP was measured using a highly sensitive ELISA assay (Dako A/S, Glostrup, Denmark). Inter-assay and intra-assay coefficients of variation were 8% and 10% respectively with a detection limit of 0.26 mg/L.

4.2.4 C-reactive protein genotyping

The *CRP*/+1444C>T single nucleotide polymorphism (SNP; rs1130864) was genotyped by PCR and RFLP analysis using primer pairs described previously,[98] and the restriction enzymes *Sdu*I or *Bsp*1286I which cleave the 181bp PCR product into 23bp and 158bp fragments only in the presence of the common C allele. All DNA analysis was performed by staff unaware of the clinical status of the

subjects. Except the WOSCOPS study (genotyping conducted by members of the laboratory of Dr Naveed Sattar), all genotyping was conducted by Dr Tina Shah at the Centre for Cardiovascular Genetics and Centre for Clinical Pharmacology, Department of Medicine, UCL.

4.2.5 Statistical Analysis

4.2.5.1 Genotype and CRP concentration

To quantify the effect of the *CRP* gene polymorphism on CRP concentration, a total of 4659 men from six studies were genotyped; Table 4.1. I calculated the within-study mean difference in CRP concentration between individuals homozygous for the T allele and carriers of the C allele and then weighted each mean by the inverse of its variance to obtain an overall WMD. In the calculation of the WMD, I limited the analysis to subjects without known coronary or peripheral artery disease at the time of blood sampling in order to avoid the potential for established disease to modify the size of the genotype-CRP association. However, an analysis including all subjects was also conducted. Therefore, for prospective studies, baseline CRP data were used from all available subjects without clinically evident atherosclerosis. For case-control studies, genotype-CRP associations were analysed solely using control subjects. Because of its skewed distribution, CRP values were log-transformed, using the natural log-scale before the analysis. I calculated the absolute geometric-WMD in CRP concentration by *CRP* genotype, using the following formula: geometric WMD= [(relative difference in CRP concentration between TT homozygotes and C-allele carriers × mean CRP concentration in C-allele carriers) – mean CRP concentration in C-allele carriers]. In addition, to test whether the association between the *CRP*/+1444C>T polymorphism and CRP concentration could be confounded by other risk factors, I compared the distribution of cardiovascular risk factors in groups defined by *CRP* genotype. Random and fixed effect models were used for these analyses.

4.2.5.2 Genotype and non-fatal myocardial infarction

To examine the effect of *CRP* genotype on risk of non-fatal MI, a total of 6201 men from four studies were genotyped; Table 4.1. The adjusted ORs and 95%CI for subjects homozygous for the T allele compared with carriers of the C allele were calculated from each study. In addition, for the two intervention trials [100,101] the potential interaction of the genotype-MI association with the active therapy was also evaluated. I then pooled the within-study ORs to obtain a summary adjusted-OR and 95%CI for non-fatal MI, under both fixed and random effect models. Fixed effect summary-ORs were calculated using the inverse variance-weighted method,[106] and the DerSimonian and Laird Q test was used to evaluate the degree of heterogeneity between studies.[55]

4.2.5.3 Consistency between the odds ratios from genetic and from non-genetic observational studies

I estimated the expected OR for non-fatal MI, corresponding to the WMD in CRP between TT subjects and C allele carriers, based on data from prior non-genetic observational studies that examined the CRP-coronary event association. To do this I used information from the most recent meta-analysis of observational studies of CRP and coronary events, as well as data from the Reykjavik Heart Study. The first estimate, based on an OR of 1.58 (95%CI: 1.48, 1.68), for the top vs. bottom tertiles of the CRP distribution, came from the recent meta-analysis of 22 prospective studies including 7068 cases with different degrees of adjustment for traditional cardiovascular risk factors across the studies included and no adjustment for regression dilution bias.[107] The second estimate was based on an maximum-adjusted OR of 1.45 (95%CI: 1.26, 1.68) between top and bottom tertiles of the CRP distribution reported in the Reykjavik Heart Study,[107] the largest prospective observational study of CRP which contributed 2459 cases to the meta-analysis and also undertook a comprehensive adjustment for potential confounders (age, sex, period of recruitment, smoking, systolic blood pressure, total cholesterol level, triglyceride levels, body mass index, forced expiratory volume in one second, diabetes, and socio-economic status). The third estimate from Reykjavik Heart Study,[107] was based on an OR of 1.92 (95%CI: 1.68, 2.18) for top vs. bottom tertile of CRP, based on a more limited degree of adjustment (age, sex and period of recruitment). The final estimate also from Reykjavik study was based on the adjusted standard 10-year follow-up risk of 1.84 (95%CI: 1.49, 2.28) for top vs. bottom tertile of CRP.[107]

For these estimates I assumed that the usual mean difference in CRP between the individuals in the top and bottom tertiles was 1.4mg/L,[90] and that the CRP-coronary event relationship was log-linear. The expected OR for TT homozygous subjects with reference to C allele carriers was calculated using the formula: expected OR= OR-non genetic studies$^{(WMD/1.4)}$. The 95%CI for the expected OR for TT homozygous subjects with reference to C allele carriers was obtained by simulation, as described in Chapter 2. Briefly, one million replications of the expected OR were obtained using the WMD in CRP by genotype and the OR from non-genetic observational studies with their corresponding standard errors, therefore the uncertainties surrounding the two associations, genotype-intermediate phenotype and intermediate phenotype-disease risk were take into account.[58] The values for the 2.5 and 97.5 percentiles of the simulated distribution were used as limits of the 95% confidence interval for the expected OR.[58] This expected OR was compared with the observed summary-OR obtained from the genetic studies, by means of an interaction test, as described in Chapter 2.[59] Data were analysed using the Review Manager software (version 4.2) from the Cochrane Collaboration 2003 and Stata 8.2.

4.3 Results

4.3.1 Allele and genotype frequencies

The allele and genotype frequencies of the +1444C>T polymorphism were in Hardy-Weinberg equilibrium for all studies included in the present report, apart from a marginal distortion in the NPHS-2 study (Table 4.2). The frequencies of the rare allele in disease-free subjects from all studies were very similar (range: 26% to 33%).

Table 4.2 Allele and genotype frequencies of *CRP*/+1444C>T polymorphism in the studies evaluated.

Study	Controls (number)	P value for H-W equilibrium	Cases (number)	P value for H-W equilibrium
NPHS-2				
+1444CC	1261	0.03	57	0.48
+1444CT	1047		41	
+1444TT	265		5	
T-allele frequency	0.30		0.24	
WOSCOPS				
+1444CC	511	0.88	175	0.17
+1444CT	481		136	
+1444TT	111		37	
T-allele frequency	0.32		0.30	
LEADER				
+1444CC	514	0.27	21	0.31
+1444CT	412		16	
+1444TT	97		6	
T-allele frequency	0.29		0.32	
HIFMECH				
+1444CC	254	0.27	246	0.28
+1444CT	224		196	
+1444TT	39		49	
T-allele frequency	0.29		0.30	
Army				
+1444CC	122	0.42	-	-
+1444CT	92		-	
+1444TT	13		-	
T-allele frequency	0.26		-	
UDACS				
+1444CC	167	0.12	-	-
+1444CT	139		-	
+1444TT	42		-	
T-allele frequency	0.32		-	
EBCT				
+1444CC	39	0.85	-	-
+1444CT	29		-	
+1444TT	6		-	
T-allele frequency	0.28		-	

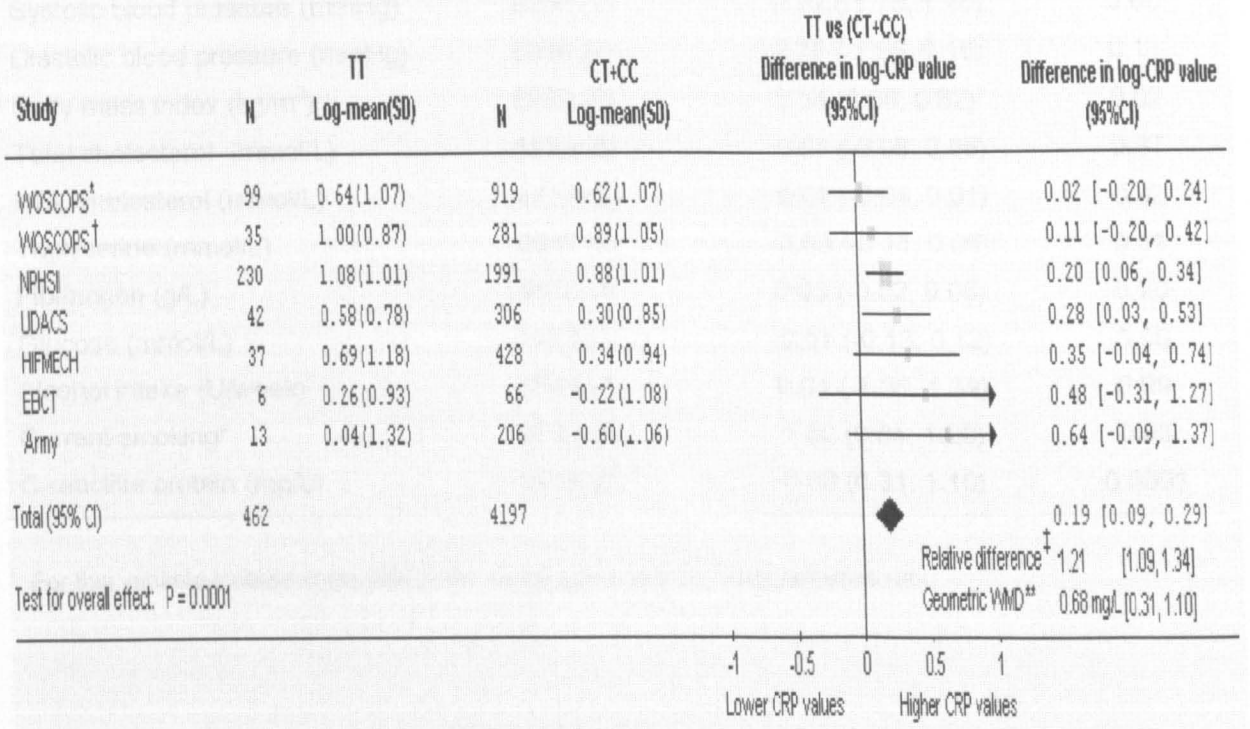
- : Not applicable.

4.3.2 *CRP*/+1444C>T polymorphism and CRP concentrations

Using data from 4659 men from six studies, the weighted mean CRP concentration in C-allele carriers without known cardiovascular disease was 2.01 mg/L (95%CI: 1.94, 2.07). Under a fixed effect model, the geometric-weighted mean difference in CRP concentration between individuals homozygous for the T allele compared with carriers of C allele was 0.68 mg/L (95%CI: 0.31, 1.10). There was no significant between-study heterogeneity (P value for heterogeneity [P_{Het}]= 0.47; Figure 4.1. When alternative models of the effect of genotype on CRP were evaluated, only the (TT) vs. (CC) comparison was significant (WMD= 0.78 mg/L [0.41, 1.20]; p<0.0001), while heterozygosity (CT vs.

CC) was not (WMD= 0.06 [-0.24, 0.44]; p=0.66). When subjects with coronary or peripheral atherosclerosis at the time of blood sampling were included in this analysis (n= 5658), the WMD between individuals homozygous for the T allele compared with carriers of C was very similar (0.74 mg/L [95%CI: 0.37, 1.10]; p<0.0001).

Figure 4.1 Mean log-CRP and mean difference in log-CRP (mg/L) by study among subjects without known cardiovascular disease.



*; Data for control subjects. †; Data for subjects with non-fatal MI. ‡; Relative difference was obtained by antilog of the difference in log-CRP values. **; Geometric WMD= [(relative difference in CRP concentration between TT homozygotes and C-allele carriers × mean-CRP concentrarion in C-allele carriers) – mean-CRP concentrarion in C-allele carriers].

To evaluate potential confounding of the genotype-CRP association, I quantified the association between TT genotype and age, systolic and diastolic blood pressure, body mass index, current smoking, glucose, alcohol intake, fibrinogen, triglyceride, total and HDL cholesterol. No significant difference was observed for any of these risk factors, with the exception of a slightly higher body mass index among TT subjects (mean difference=0.34 kg/m^2; p=0.02; Table 4.3). In view of the multiple comparisons made, however, this significance level is not particularly extreme. Moreover, this apparent association was not significant when subjects with and without CHD were considered together, and it is likely, but not yet certain, to reflect the play of chance.

Table 4.3 Cardiovascular risk factor distribution according to the *CRP*/+1444C>T polymorphism. Comparisons are made between TT homozygotes and C-allele carriers.

Variable	Subjects (studies)	Weighted mean difference [TT minus C-carriers] (95%CI)	P value
Age (years)	6358 (7)	0.17 (-0.19, 0.52)	0.35
Systolic blood pressure (mmHg)	6356 (7)	-0.31 (-1.78, 1.16)	0.68
Diastolic blood pressure (mmHg)	6356 (7)	-0.70 (-1.56, 0.16)	0.11
Body mass index (kg/m^2)	6359 (7)	0.34 (0.06, 0.62)	0.02
Total cholesterol (mmol/L)	6115 (6)	-0.01 (-0.08, 0.06)	0.87
HDL-cholesterol (mmol/L)	4714 (5)	-0.01 (-0.04, 0.01)	0.32
Triglyceride (mmol/L)	6041 (5)	-0.03 (-0.11, 0.06)	0.51
Fibrinogen (g/L)	5656 (4)	0.03 (-0.02, 0.08)	0.20
Glucose (mmol/L)	1913 (3)	-0.01 (-0.13, 0.12)	0.90
Alcohol intake (U/week)	4549 (4)	0.01 (-1.38, 1.39)	0.99
Current smoking*	6132 (6)	1.00 (0.84, 1.20)	0.98
C-reactive protein (mg/L)	4659 (6)	0.68 (0.31, 1.10)	0.0001

*; For this variable instead of the WMD, the value reported is the weighted odds ratio.

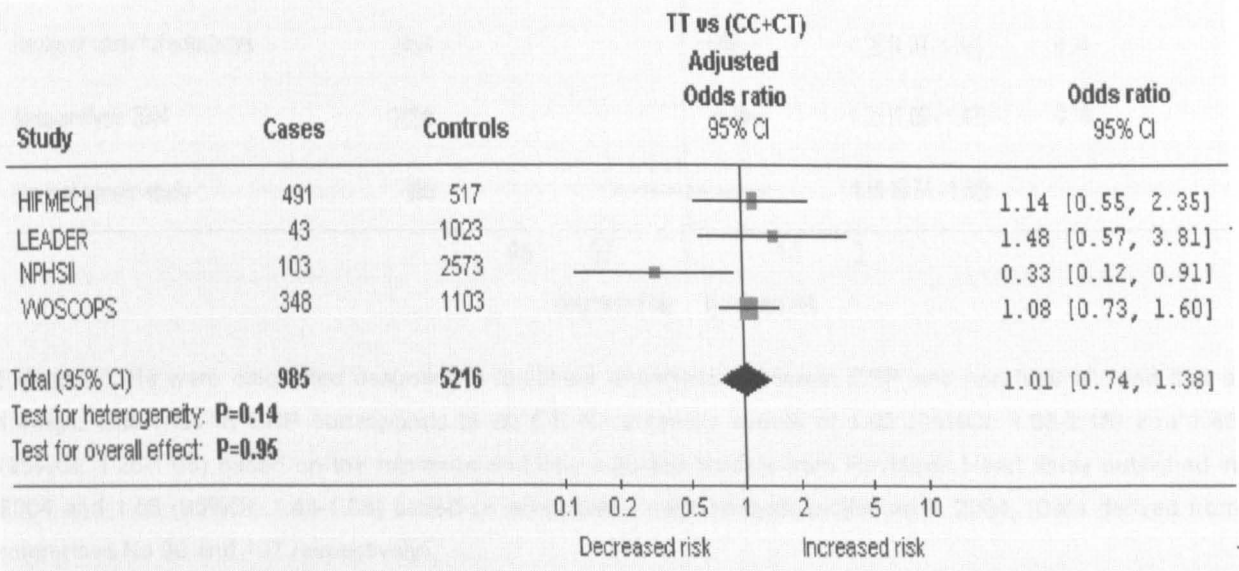
4.3.3 Odds ratio estimated from observational studies and the expected mean difference in plasma CRP

Expected ORs for non-fatal MI among TT homozygotes compared to C allele carriers based on a between-genotype WMD difference in CRP of 0.68 mg/L (95%CI: 0.31, 1.10) were 1.37 (95% CI: 1.14, 1.68), 1.20 (95% CI: 1.07, 1.38), 1.34 (95%CI: 1.12, 1.67) and 1.25 (95% CI: 1.09, 1.43) based respectively on the Reykjavik Heart Study minimally and maximally adjusted models, Reykjavik study 10-year estimate, and on the meta-analysis of studies to 2004.[107]

4.3.4 Odds ratio of non-fatal myocardial infarction for the *CRP*/+1444C>T polymorphism

Data from four new studies involving 985 subjects with non-fatal MI and 5216 control subjects were pooled to obtain a summary adjusted OR. After combining the studies under a fixed effect model, subjects homozygous for the T allele compared with C allele carriers, had no significant increase in the risk of non-fatal MI (summary adjusted-OR= 1.01 [95%CI: 0.74, 1.38]; p=0.95). No significant inter-study heterogeneity was observed (P_{Het}=0.14); Figure 4.2. When the risk of non-fatal MI conferred by homozygosity for the T allele was restricted to prospective genetic studies, there was still no significant association with non-fatal MI (OR=0.98 [95%CI: 0.70, 1.38]; p=0.93). No significant increase in risk was observed in the individual studies, and there was no significant interaction between treatment and genotype-outcome association in the two clinical trials involved in this analysis: for LEADER, p=0.53 and WOSCOPS, p=0.44.

Figure 4.2 Adjusted odds ratio* for non-fatal MI among subjects with the +1444TT-genotype in comparison with carriers of the +1444C allele.



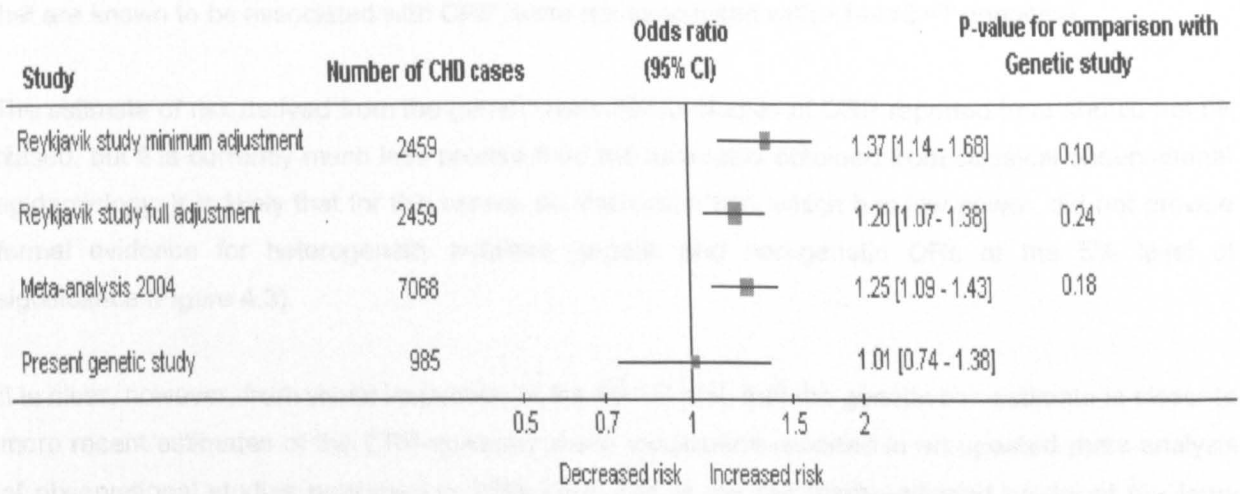
*; A multivariate logistic regression model was used to adjust the OR from each study by age, total-cholesterol, alcohol intake, body mass index, hypertension, diabetes, and smoking.

4.3.5 Consistency between the odds ratios estimated from the mean difference in CRP level and the odds ratio from the genetic study

The estimate of the observed OR from the genetic study (1.01 [95%CI: 0.74, 1.38]) was mathematically smaller than the expected OR (1.37 [95%CI: 1.14, 1.68] or 1.34 [95%CI: 1.12, 1.67]),

calculated using data from the minimally adjusted estimate or the 10-year estimate from Reykjavik Heart Study. It was, however, closer to estimates based on data from the maximally adjusted model of the large Reykjavik Heart Study (1.20 [95%CI: 1.07, 1.38]) and to an estimate based on an updated meta-analysis of studies to 2004 (1.25 [95%CI: 1.09, 1.43]; Figure 4.3). The *P* value for interaction for the comparisons of the minimum and maximally adjusted estimates from Reykjavik study, and from an updated meta-analysis of studies to 2004 with the observed genetic estimate is outlined in Figure 4.3.

Figure 4.3 Comparison of expected and observed odds ratios for non-fatal MI for individuals homozygous for the *CRP* +1444C>T polymorphism.



Expected ORs were calculated assuming a log-linear association between CRP and non-fatal MI, and that a 1.4mg/L difference in CRP corresponds to an OR for coronary events of 1.92 (95%CI: 1.68-2.18) and 1.45 (95%CI: 1.25-1.68) based on the minimum and fully adjusted models from Reykjavik Heart study published in 2004 and 1.58 (95%CI: 1.48-1.68) based on an updated meta-analysis published in 2004, (Data derived from references No 90 and 107 respectively).

4.4 Discussion

Using pooled data from 4659 individuals, I found that men homozygous for the T allele of the +1444C>T polymorphism of the human *CRP* gene had a circulating CRP concentration 0.68 mg/L (95%CI: 0.31, 1.10) higher than carriers of the C-allele, confirming the results of earlier smaller studies.[97,98] *CRP* genotype therefore contributes to the variation in CRP concentration observed in the population.

It was estimated that this difference in CRP of 0.68 mg/L (95%CI: 0.31, 1.10) would confer an OR for non-fatal MI of 1.25 (95% CI: 1.09, 1.43) or 1.20 (95% CI: 1.07, 1.38) using data from a meta-analysis of prospective cohort studies published in 2004,[107] and from the maximally-adjusted model of the Reykjavik Heart Study[107] respectively, assuming the observational associations have been unbiased. However, despite possessing a genotype that would have exposed them to a long-term elevation in CRP concentration, for men with the TT genotype, the point estimate of the OR for non-fatal MI was 1.01 (95% CI 0.74, 1.38). This finding is important since, in contrast to the reported association between CRP and coronary events in previous studies, the association between *CRP* genotype and events should not be subject to reverse causality bias, regression dilution bias or to confounding by other cardiovascular risk factors related to CRP, though some doubt about a marginal effect of the genotype on body mass index remains to be evaluated. This is supported by the data in Table 4.3, which indicate that traditional cardiovascular risk factors and some inflammatory markers, that are known to be associated with CRP, were not associated with +1444C>T genotype.

The estimate of risk derived from the genetic association studies of CRP reported here should not be biased, but it is currently much less precise than the estimates obtained from classical observational epidemiology. It is likely that for this reason an interaction test, which has low power, did not provide formal evidence for heterogeneity between genetic and non-genetic ORs at the 5% level of significance (Figure 4.3).

It is clear, however, from visual inspection of the Forest plot, that the genetic risk estimate is closer to more recent estimates of the CRP-coronary event association reported in an updated meta-analysis of observational studies published in 2004,[107] and in the maximally-adjusted model of the large Reykjavik Heart Study cohort.[107] Thus, recent results from prospective observational studies in which the random error has been reduced substantially by considering more studies in a meta-analysis, and in which more extensive statistical adjustment has been conducted, as exemplified in the Reykjavik Heart Study, indicate that the association of the CRP with coronary heart disease risk has decreased substantially in comparison to earlier estimates. The strong potential for confounding in observational studies of the measured phenotype (CRP), and the utility of genotype as a proxy for CRP was also illustrated by a recent study [108] in which a range of covariates that exhibited strong correlations with CRP were distributed evenly among 3500 British women separated according to genotypes for a *CRP*/1059G>C (rs1800947) polymorphism that is in LD with the variant reported here.[109]

Although a null association was identified between *CRP* genotype and coronary events in this study the width of the confidence limits is such that our observation could be compatible with a modest but

potentially important causal link. It is estimated that about 20,000 CHD cases and a similar number of controls will be required to confirm or refute an small but important causal effect, although even larger numbers will be needed for more detailed analyses.[110] However, evidence from different sources indicate that, if anything, the discrepancy observed between the genetic and non-genetic-ORs is likely to be greater. First, the observational estimate of the CRP-coronary event association [107] on which I based the non-genetic OR, did not adjust for regression dilution bias and so may have been underestimated. On the contrary this bias is unlikely to affect the genetic estimate.[22,23] Second, emerging evidence supporting the current null genetic estimate comes from three genetic studies in European descent individuals, with an additional 1858 cases and 1347 controls,[111,112,113] where carriers of the C allele of the 1059G>C variant (in LD with the +1444/C>T variant) did not have an increased risk of cardiovascular events despite having a CRP concentration that was 0.3-0.6 mg/L higher than GG homozygous subjects. Moreover, recent data from two other genetic studies in European descent individuals found no association of the +1444/C>T variant with cardiovascular events despite a similar effect of this variant on CRP concentration.[111,114]

An additional approach in future Mendelian randomisation studies is to genotype several SNPs at the *CRP* locus to generate all common haplotypes, and to ensure inclusion of SNPs with demonstrated functionality from *in vitro* studies in such analyses.

In conclusion, European descent individuals men with a genotype that would have exposed them to a long-term elevation in CRP concentration (0.68 mg/L [95%CI: 0.31, 1.10]) were not at increased risk of non-fatal MI (OR: 1.01; 95% CI: 0.74, 1.38), though the confidence limits currently encompass substantial uncertainty. Taking this into account, it is at the very least likely that unbiased and non-confounded estimates of the effect of CRP on coronary events are smaller than earlier studies estimated, a finding that is in agreement with the latest evidence derived from prospective observational studies and also with more refined and carefully conducted mechanistic studies.[115] If our null point estimate of the genetic effect is stable with the addition of further studies, it might lead to the re-evaluation of CRP both as a risk marker and as a potential therapeutic target.

Chapter 5 C-reactive protein and cardiovascular disease: A tagging-haplotype approach for Mendelian randomisation

In Chapter 4, I described the proof of principles of a Mendelian randomisation experiment in CRP and coronary events, using a single-variant approach. In this Chapter, I expanded the results from my previous Chapter in two ways. Firstly, I used a tagging-haplotype that captures most of the genetic variation in individuals of European ancestry, and secondly I increased the number of cardiovascular outcomes to be evaluated by a genetic approach.

5.1 CRP as a therapeutic target in cardiovascular disease

Multiple observational studies have reported consistent associations of CRP with presence or progression of asymptomatic atherosclerosis in coronary, carotid, peripheral vascular beds, as well as with the development of incident diabetes and stroke.[115] Although these observations support the view that inflammation plays a critical role in atherogenesis and in the development of some cardiovascular disorders (e.g. diabetes), it is uncertain whether CRP (an inflammatory biomarker) itself is causally linked with the progression of atherosclerosis or with cardiovascular events.[115] If CRP is causally involved in cardiovascular disease, it would legitimise to CRP as a valid therapeutic target. Indeed, interventions that specifically modify CRP are being developed for clinical testing.[116]

However associations of CRP with indices of atherosclerosis, diabetes, and stroke could equally be explained by confounding or reverse causation as described in Chapter 4. In addition, controlling of confounding through statistical models requires a judgement about the causal pathway. For example, if high blood pressure or diabetes were to mediate rather than confound the CRP-cardiovascular event association, as some studies suggest [117,118], adjustment for these factors would underestimate the true causal association. Likewise, many emerging biomarkers (e.g. fibrinogen, interleukin-6, vonWillebrand Factor, E-selectin, ICAM-1 and VCAM) exhibit associations both with CRP and with risk of cardiovascular disease, but it remains unclear whether they lie on the CRP causal pathway.[115] Although some experimental studies and animal models have suggested CRP as a pro-inflammatory and pro-thrombotic molecule (that may explain the association with those biomarkers), these studies are also affected by important sources of bias that limit their findings.[115] All these make it difficult to assess the independent effect of CRP in observational studies.

An altemative approach to confirm (or refute) the causal effect of CRP on (i) other emerging risk factors, (ii) surrogate end-points and (iii) cardiovascular events is to identify common genetic variants in the *CRP* gene that are reliably associated with differences in concentrations of CRP, its cognate protein.[68,119] If the purported effects of plasma CRP on emerging risk factors, end-points or clinical events are real, a tagging-SNP approach within the *CRP* gene that captures most of genetic variation with in the *CRP* gene should exhibit an effect on those different

outcomes in a manner proportional to its effect on the levels CRP. This genetic strategy (or Mendelian randomisation), minimises confounding, and abolishes reverse causation due to the particular properties of genetic variants described in Chapter 1.[120]

In Chapter 4, I described the proof of principle of how to conduct a Mendelian randomisation experiment for CRP by using a single *CRP* gene-variant (rs1130864). Since then multiple genetic studies have shown that a range of different *CRP*-genetic variants are all associated with CRP concentrations,[68] but their effects on surrogate end-point and cardiovascular disorders have been conflicting.[108.119,121,122,123,124] Recently, the Cardiovascular Health Study reported on consistent associations of tagging (t) single nucleotide polymorphisms (SNPs) in the *CRP* gene with CRP level in both European descent and African-American individuals. Associations of some but not all *CRP* genotypes were seen with an increase in risk of cardiovascular mortality, but not with index of sub-clinical atherosclerosis (carotid intima media thickness). Interestingly, no *CRP* genotypes was associated with non-fatal myocardial infarction or with stroke.[125] The results were interpreted as indicating a possible causal role of CRP in fatal but not non-fatal cardiovascular events through a process that does not involve enhancement of atherosclerosis. Although the associations of genotype with CRP concentration in the Cardiovascular Health study appear robustly consistent with other studies, there is the possibility that with the relatively small number of clinical events, and the multiple comparisons made, by ethnic group, by genotype and by clinical end-point that associations with fatal disease outcome arose by chance. These associations therefore require replication, particularly because several prior studies did not observe associations of *CRP* genotypes with cardiovascular risk factors such as hypertension (see Table 4.3, Chapter 4), diabetes or cardiovascular events despite their consistent association with CRP level. [108,119,121,122,123,124]

Using data from two population-based studies of European descent individuals in the UK, the Edinburgh Artery study and Northwick Park Heart Study-2, I evaluated the effect of the *CRP* tagging SNPs on established and emerging cardiovascular risk factors, index of the progression of atherosclerosis in the lower extremity (ankle brachial pressure index) and incident cardiovascular events. Conducting the analyses in two independent data sets allowed me to check the consistency of the associations evaluated.

5.2 Methods

5.2.1 Study design

5.2.1.1 NPHS-2

This study was previously described in Chapter 4. NPHS-2 is a prospective study of 3,012 healthy men of European descent (age range at recruitment 50 to 64 years), with enrolment commencing

in 1989.[99] Nine general practices participated in the study. The study was approved by the local institutional review committee and all subjects provided written informed consent. None of the participants had a clinical history of unstable angina, myocardial infarction (MI; including silent infarction), coronary surgery, other cardiovascular diseases, aspirin or anticoagulant use, or malignant disease (except skin cancer other than melanoma) at the time of recruitment. Demographic information was obtained at baseline by means of a structured questionnaire. Patients still alive were recalled annually for measurement of blood pressure and certain other circulating factors as reported previously. The cardiovascular endpoints examined were fatal CHD events, non-fatal MI (based on World Health Organisation criteria), coronary artery surgery, and silent MI on follow-up ECG (Minnesota criteria), and these are referred to collectively as CHD events. Cerebrovascular events (fatal and non-fatal) were defined using the ICD9 codes 430.0, 431.0, 434.9, and 436.0. Incident diabetes was defined by practice note search for physician diagnosed and treated Type 2 diabetes mellitus according to current national guidelines.

5.2.1.2 Edinburgh Artery Study

The Edinburgh Artery Study (EAS) is a prospective study of 1592 European descent individuals (809 men and 783 women) aged 55 to 74 years enrolled in 1988. Individuals recruited were randomly selected from 11 general practices serving a range of socioeconomic and geographic areas throughout the city of Edinburgh. Details of the study recruitment, and examination process have been described elsewhere.[126] Ethics committee approval was given for this study, and informed consent was obtained from each subject. Participants completed a self-administered questionnaire at baseline and at 5 and 12 years of follow-up that included demographical and lifestyle variables, the WHO angina and intermittent claudication questionnaires. Individuals were also invited for a clinical examination at baseline and at 5 and 12 years after enrolment. Ankle systolic pressures were measured first in the right leg and then in the left leg at the posterior tibial artery using standard methods previously described.[126] Ankle-brachial index (ABI) was calculated by dividing the ankle systolic pressure by the brachial systolic pressure. The lower of the two indices was used in the analysis as indicative of worse disease. Reproducibility of the ABI and its quality control measures have been described previously.[126] Throughout the follow-up, fatal and non-fatal cardiovascular events were recorded and defined with criteria adapted from the American Heart Association. Subjects who developed major CHD during the follow up years were those who had experienced an event of fatal/ non-fatal MI or who had a coronary artery bypass or coronary angioplasty. Cerebrovascular events (includes fatal and non-fatal events) were defined according to that used by the American Heart Association. To identify all deaths occurring in the cohort, each participant's record was flagged at the UK National Health Service Central Registry. If a post-mortem examination was not performed, causes of death were verified by consulting hospital or general practitioner records. Diabetes at baseline was defined using recall of doctors' diagnosis and WHO 1999 criteria (fasting glucose $\geq$7 mmol/L or 2-h glucose $\geq$11.1 mmol/L) applied to data collected during the oral glucose tolerance test. At baseline, 277 subjects had a history of MI, stroke, angina or intermittent claudication and were excluded from further analysis.

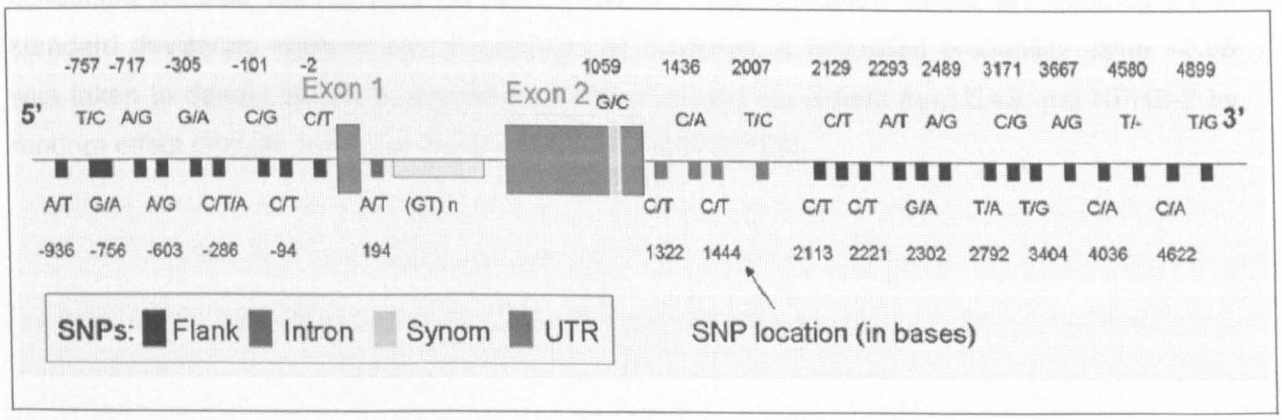
5.2.2 Measurement of CRP concentrations

In NPHS-2, CRP was measured using commercial high-sensitivity assays (R&D Systems or Kordia Life Sciences). Inter-assay and intra-assay coefficients of variation were <6.2% and <1.9% respectively, with a detection limit of 0.1 mg/L. Of 2,738 men with CRP measures, 1,541 were from samples obtained at baseline, 724 from samples obtained at year 1, 341 at year 2, 100 at year 3, 27 at year 4, and 5 at year 5. Descriptions of the measurement of other biomarkers have been reported previously.[99] Baseline CRP in Edinburgh Artery Study was measured immunologically using a high-sensitive assay (BN ProSpec nephelometer [Dade Behring]). Intra-assay and inter-assay variability coefficient were 4.7% and 8.3% respectively. Technical details and quality control for other biomarkers included have been described elsewhere.[126]

5.2.3 Genotyping of single nucleotide polymorphisms in the *CRP* gene

Polymorphisms in the human *CRP* gene (OMIM: C-reactive protein, pentraxin related; 1q21-q23) were identified by reference to public-domain databases of human sequence variation. This information was combined with previous polymorphism screen of the *CRP* gene conducted at the Centre for Clinical Pharmacology in UCL by Dr Tina Shah, to generate a consensus map of polymorphic sites (Figure 5.1).

Figure 5.1 Summary map of all reported SNPs in the *CRP* gene. Figure adapted from Figure 2 manuscript of CRP CHD Genetics collaboration. Eur J Epidemiol 2008;23:531-40. Authors/Writing committee: J Danesh, A Hingorani, F Wensley, JP Casas, L Smeeth.



Using validated genotype data (minor allele frequency > 5%) from individuals of European descent from the NHLBI PGA database (http://pga.mbt.washington.edu/), and the human HapMap database (http://www.hapmap.org/), and based on the pattern of linkage disequilibrium across the *CRP* gene, Dr Tina Shah (Centre for Clinical Pharmacology, Department of Medicine, UCL) selected a set of tSNPs capable of capturing maximum haplotype diversity among individuals of European descent using the programme TagIT (http://popgen.biol.ucl.ac.uk/software.html) Figure 5.2.

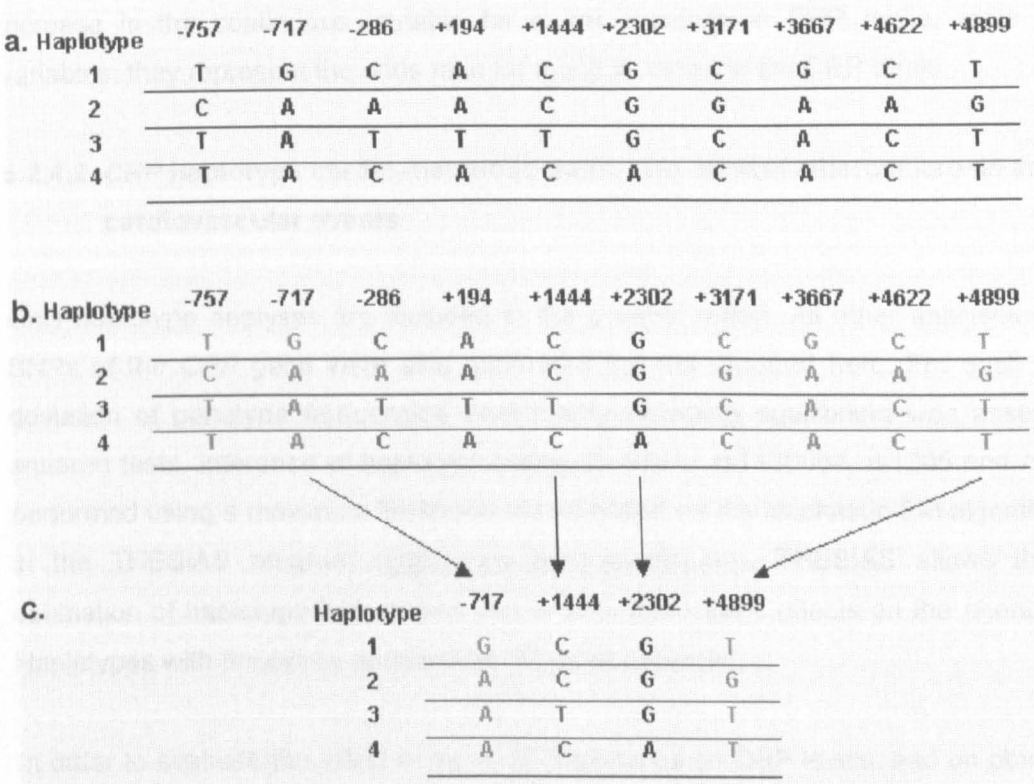
In NPHS-2, DNA was extracted for 2,775 men, and genotyping data was available from 2,676. Genotyping for db SNP rs1130864 (+1444C>T) was by PCR and RFLP as described in Chapter 4, but genotyping was repeated in 200 subjects using a 5'nuclease assay (Taqman) with an inter-assay agreement rate of ≥99%. The dbSNP rs1205 (+2302G>A) and dbSNP rs3093077 (+4899T>G) were also determined by a 5'-nuclease assay (TaqMan). Primers and probe sets were designed and manufactured using Applied Biosystems 'Assay-by-Design' custom service. In Edinburgh Artery Study, DNA was extracted for 913 individuals, and genotyping data was available from 908 subjects for three *CRP* tSNPs (three tagging).

Genotyping for db SNP rs1130864 (+1444C>T), dbSNP rs1205 (+2302G>A) and dbSNP rs3093077 (+4899T>G) was determined by a 5'-nuclease assay (TaqMan). Genotyping for both studies was conducted at the Dept. Medicine at UCL by Dr Tina Shah (for NPHS-2) and members of Dr Helen Ireland's group (for EAS). In both studies, positive and negative control samples were included on each assay plate for quality control, and cases and controls were distributed over all study plates, minimising the risk of differential mistyping by case control status. Genotyping was conducted in a blind fashion to the case-control status.

5.2.4 Statistical analysis

Data were analyzed with the use of the SPSS version 12.0 software package for EAS study (by Dr Ionna Tsoulaki) and Stata 9.0 in NPHS-2 study (by Ms Jackie Cooper) following a pre-specified analysis plan I developed. Log-transformations using the natural log-scale were conducted for data not normally distributed and for these variables means are geometric and standard deviations approximate. Throughout all analyses, a two-sided probability value <0.05 was taken to denote statistical significance. I then pooled the results from EAS and NPHS-2 by random effect models, using the inverse variance method.[106]

Figure 5.2 Identification of (t)SNPs that capture haplotype diversity at the *CRP* locus among individuals of European ancestry. Panel (a): Ten common SNPs with minor allele frequency greater than 5% lie on 4 major haplotypes. Panel (b): Strong LD between markers of the same colour allows typing of a single marker from each group without loss of information on variation. Panel (c): typing of 3 markers (+1444, +2302 and +4899) allows reconstruction of the four common haplotypes. Figure adapted from Figure 3 manuscript of CRP CHD Genetics collaboration. Eur J Epidemiol 2008;23:531-40. Authors/Writing committee: J Danesh, A Hingorani, F Wensley, JP Casas, L Smeeth.



5.2.4.1 CRP, cardio-metabolic traits, sub-clinical atherosclerosis and cardiovascular events

Associations between plasma CRP concentrations and incident cardiovascular outcomes (hypertension diabetes, CHD and stroke) were assessed using Cox proportional hazards regression. Hazard ratios and 95%CI were obtained for the top versus the bottom tertile of the CRP distribution. A progressive adjustment for potential confounders were conducted as follows: for CHD and stroke the model-1 includes age, gender and practice (for NPHS-2 only); model-2: previous model plus smoking, diabetes, body mass index, systolic blood pressure, LDL-cholesterol, HDL-cholesterol and physical activity (for EAS only). Model-3 includes previous variables plus fibrinogen.

Statistical models conducted to explore the CRP effect on diabetes were: model-1 that includes age and gender and practice (only for NPHS-2); model-2: previous model-1 plus body mass index, triglycerides and physical activity (only for EAS). Model-3 includes previous model-2 plus fibrinogen.

When evaluating associations of plasma CRP levels with cardiovascular risk factors, the biomarker/complex trait (in a standardised scale) was regressed (using normal error regression) on the CRP tertile, while for categorical variables (which were converted into binary outcomes) a logistic regression model was used. In both regression models the effect of CRP tertile was estimated as a linear trend. Beta-coefficients represent the number of standard deviations increase in the continuous variable for a unit increase in CRP tertile, while for categorical variables; they represent the odds ratio for a unit increase in the CRP tertile.

5.2.4.2 *CRP* haplotype cardio-metabolic traits, sub-clinical atherosclerosis and cardiovascular events

Only haplotype analyses are included in the present report. All other analyses using individual SNPs of the *CRP* gene were also generated but not reported here. For each genetic variant, deviation of genotype frequencies from Hardy-Weinberg equilibrium was assessed using chi-squared tests. Inference of haplotype (using db SNPs: rs1130864, rs1205 and rs3093077) were performed using a maximum likelihood model based on the stochastic-EM algorithm implemented in the THESIAS program (http://www.genecanvass.org). THESIAS allows the simultaneous estimation of haplotype frequencies and of their associated effects on the phenotype of interest. Haplotypes with frequency of less than 5% were excluded.

In order to evaluate the effect of the *CRP* haplotypes on CRP levels, and on other cardiovascular risk factors, for continuous variables, the standardised biomarker/complex trait was regressed on the *CRP*-haplotype, while for categorical variables a logistic regression model was used. As for the plasma CRP analyses, for the *CRP* gene a linear effect was assumed. Beta-coefficients represent the number of standard deviations increase in the outcome/biomarker for a unit increase in haplotype, while for categorical variables they represent the odds ratio for a unit increase in haplotype.

Hazard ratios and 95%CI, adjusted for age, gender and practice (for NPHS-2 only) were obtained to measure the strength of the association between the *CRP*-haplotypes with incident cardiovascular outcomes: CHD, stroke and diabetes. For the haplotype trend the effect represents the change from two copies of one haplotype to two copies of the next. For all genetic analyses, the *CRP*-haplotype associated with lowest CRP value (direct intermediate phenotype) was considered the reference group.

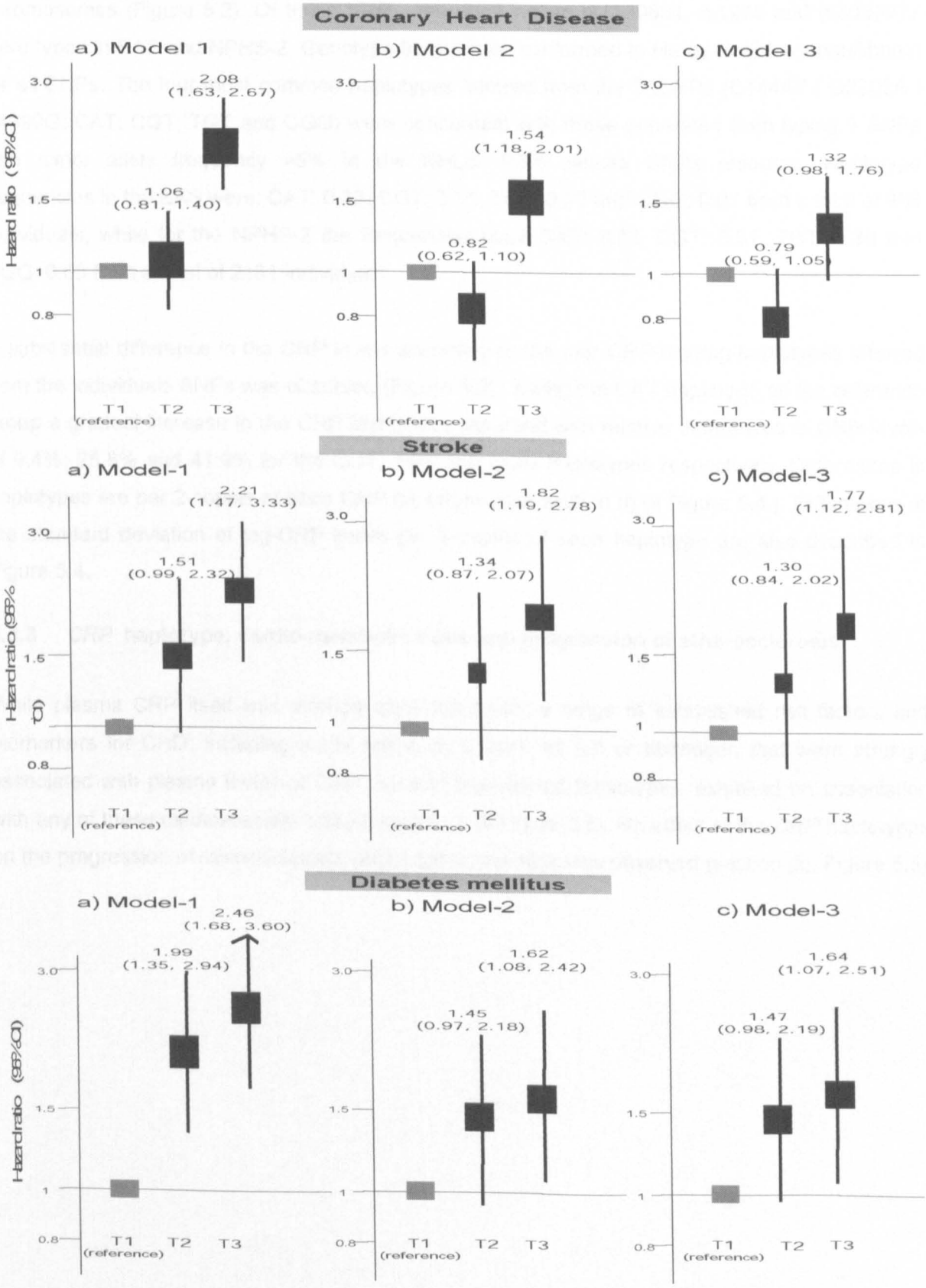
5.3 Results

5.3.1 Plasma CRP, cardiovascular events and cardiovascular traits

As reported in many previous prospective studies, higher CRP concentrations were associated with an increased risk of CHD events in both studies (403 CHD events: 178 in EAS and 225 in NPHS-2). A pooled HR of 2.08 (95%CI: 1.63, 2.67) for individuals in the top tertile of CRP distribution compared to the bottom tertile was observed for the statistical model with less adjustment (model-1). Similarly, high baseline CRP levels were associated with a hazard ratio of 2.21 (95%CI: 1.47, 3.33) for stroke (172 stroke events: 124 in EAS and 48 in NPHS-II) and 2.46 (95%CI: 1.68, 3.60) for type-2 diabetes mellitus (model-1 in Figure 5.3). However, for all three cardiovascular events, an important reduction in the magnitude of the effect was observed as the statistical models included additional potential confounders. For CHD the hazard ratio for the top tertile was reduced by a 70% from 2.08 (95%CI: 1.63, 2.67) to 1.32 (95%CI: 0.98, 1.76); for stroke by 36 percent from 2.21 (95%CI: 1.47, 3.33) to 1.77 (95%CI: 1.12, 2.81) and for diabetes by 35 percent from 1.99 (95%CI: 1.68, 3.60) to 1.64 (95%CI: 1.07, 2.51) (model-3 in Figure 5.3). A positive association of CRP was also observed for ABI. Individuals in the top tertile of the CRP distribution had on average lower values of ABI at baseline (-0.06 [95%CI: -0.08, -0.03]) and were also associated with a decrease in ABI after 12 years of follow up of -0.05 (95%CI: -0.07, -0.02). This indicates a faster progression of sub-clinical atherosclerosis in subjects with high values of CRP.

However, high levels of CRP were also associated with several established risk factors (Figure 5.5). Individuals with higher CRP concentrations had higher than average values of blood pressure, proatherogenic lipid fractions, body mass index, fasting glucose, and higher prevalence of smoking. CRP was also inversely associated with HDL-cholesterol, apolipoprotein-A1, and lower levels of physical activity. In addition a substantial number of emerging biomarkers which have also been associated with high risk of CHD, were also associated with CRP levels such as: fibrinogen, interleukin-6, Lp-PLA2, triglycerides, D-dimer, E-selectin, ICAM-1, vonWilebrand factor, tissue plasminogen activator, and Factor VII (Figure 5.5). However, these emerging biomarkers (with the exception of fibrinogen) were not included in any of the statistical models that evaluate an independent effect of CRP on cardiovascular events, due to the uncertainty of whether these biomarkers are real confounders or by contrast actually lie on the causal pathway of CRP and cardiovascular events. If they lay on the causal pathway, adjustment for these risk factors will partially or totally remove the causal effect of CRP. This illustrates the great potential for residual confounding, and difficulties in statistical adjustment to evaluate the independent effect of CRP in the multiple cardiovascular outcomes associated.

Figure 5.3 Hazard ratios of CHD (upper panel), stroke (middle panel) and diabetes mellitus (lower panel) for concentrations of CRP according to different statistical models.



5.3.2 *CRP* haplotype and CRP levels

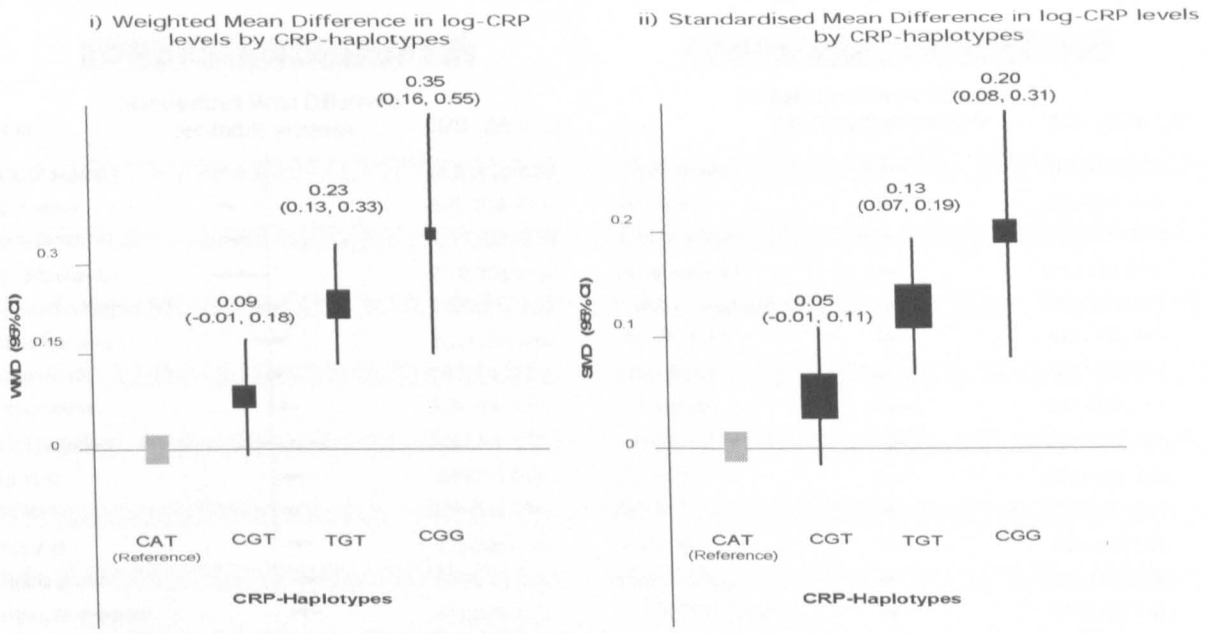
In subjects of European descent, several combinations of three tagging SNPs were equally capable of capturing the 4 most common *CRP* haplotypes, accounting for 94 percent of all chromosomes (Figure 5.2). Of these SNPs, the combination rs1130864, rs1205 and rs3093077 were typed in EAS and NPHS-2. Genotype frequencies conformed to Hardy-Weinberg equilibrium for all SNPs. The four most common haplotypes inferred from the 3 tSNPs (C1444T / G2302A / T4899G: CAT; CGT; TGT and CGG) were concordant with those generated from typing 9 SNPs with minor allele frequency >5% in the NHLBI PGA Seattle SNPs resource. Haplotype frequencies in the EAS were: CAT: 0.33, CGT: 0.31, TGT: 0.29 and CGG: 0.07 from a total of 908 individuals, while for the NPHS-2 the frequencies were CAT: 0.33, CGT: 0.31, TGT: 0.30 and CGG: 0.05 from a total of 2161 individuals.

A substantial difference in the CRP levels according to the four *CRP*-tagging-haplotypes inferred from the individuals SNPs was observed (Figure 5.4). Using the CAT haplotype as the reference group a gradual increase in the CRP levels was observed with relative differences in CRP levels of 9.4%, 25.8% and 41.9% for the CGT, TGT and CGG haplotypes respectively. Differences in haplotypes are per 2-copies of each *CRP* haplotype (see section (i) of Figure 5.4.). Differences in the standard deviation of log-CRP levels per 2-copies of each haplotype are also described in Figure 5.4.

5.3.3 *CRP* haplotype, cardio-metabolic traits and progression of atherosclerosis

While plasma CRP itself was strongly associated with a range of established risk factors and biomarkers for CHD, including some biomarkers, such as IL6 or fibrinogen that were strongly associated with plasma levels of CRP, none of the inferred haplotypes, exhibited an association with any of these cardiovascular traits (section (ii) of Figure 5.5). No effect of the *CRP* haplotypes on the progression of atherosclerosis (indicated by the ABI) was observed (section (ii), Figure 5.5)

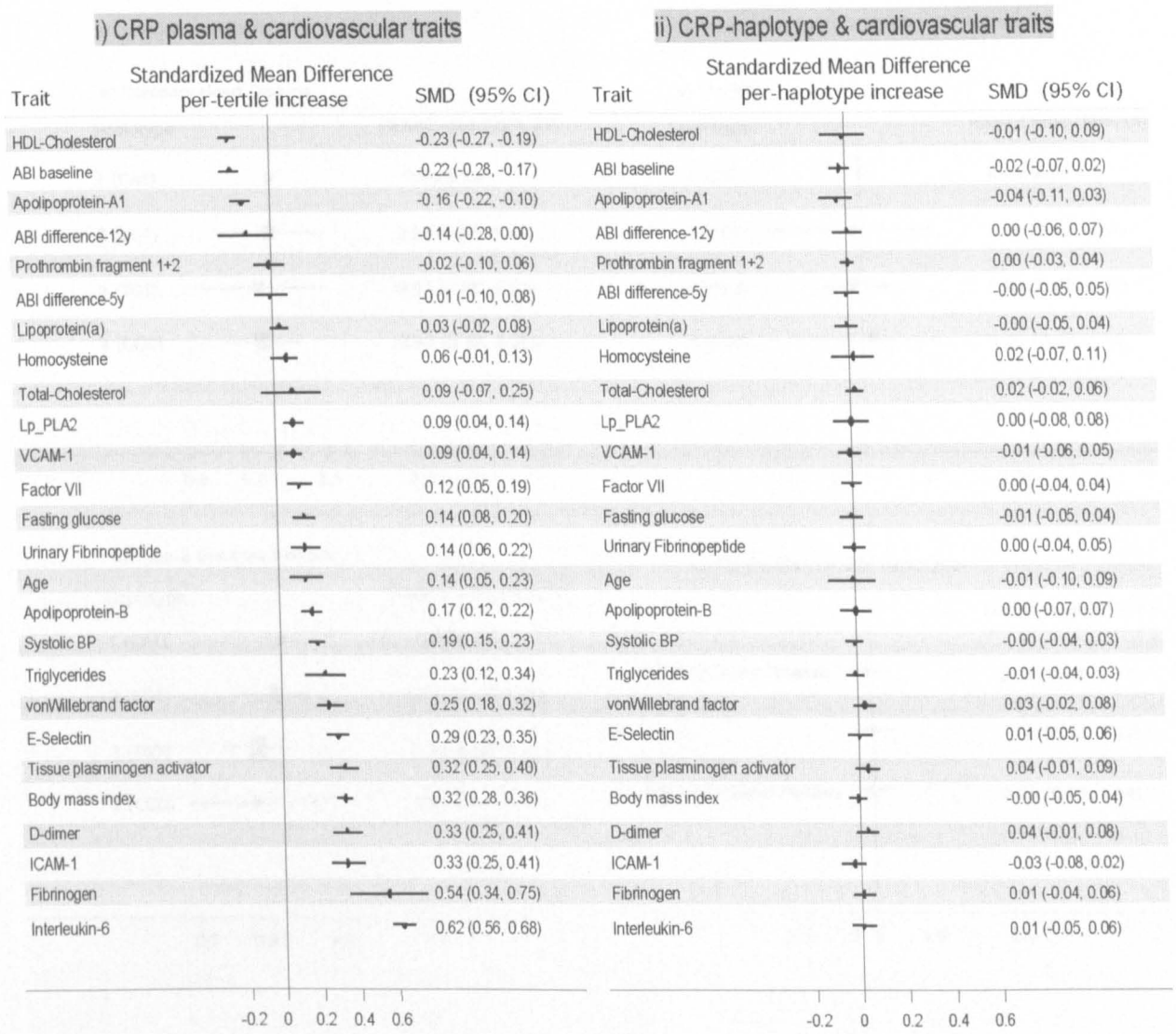
Figure 5.4. Effect of the *CRP*-haplotypes on CRP levels. i) Weighted mean difference, and ii) Standardised mean difference. Haplotype CAT is used as reference group.



5.3.4 *CRP* haplotype and cardiovascular events

Though *CRP* tagging-haplotypes exhibited a clear association with CRP concentration of up to a 41.9% relative difference for the extreme haplotype (CGG vs. CAT-reference), no increase in risk of CHD was observed for each haplotype and neither for a trend-effect for *CRP*-haplotype, HR of 0.95 (95%CI: 0.84, 1.07) (Figure 5.6). A similar scenario was observed for stroke (HR of 1.00 [95%CI: 0.85, 1.17]) and type-2 diabetes mellitus (HR of 0.99 [95%CI: 0.86, 1.14]), see Figure 5.6.

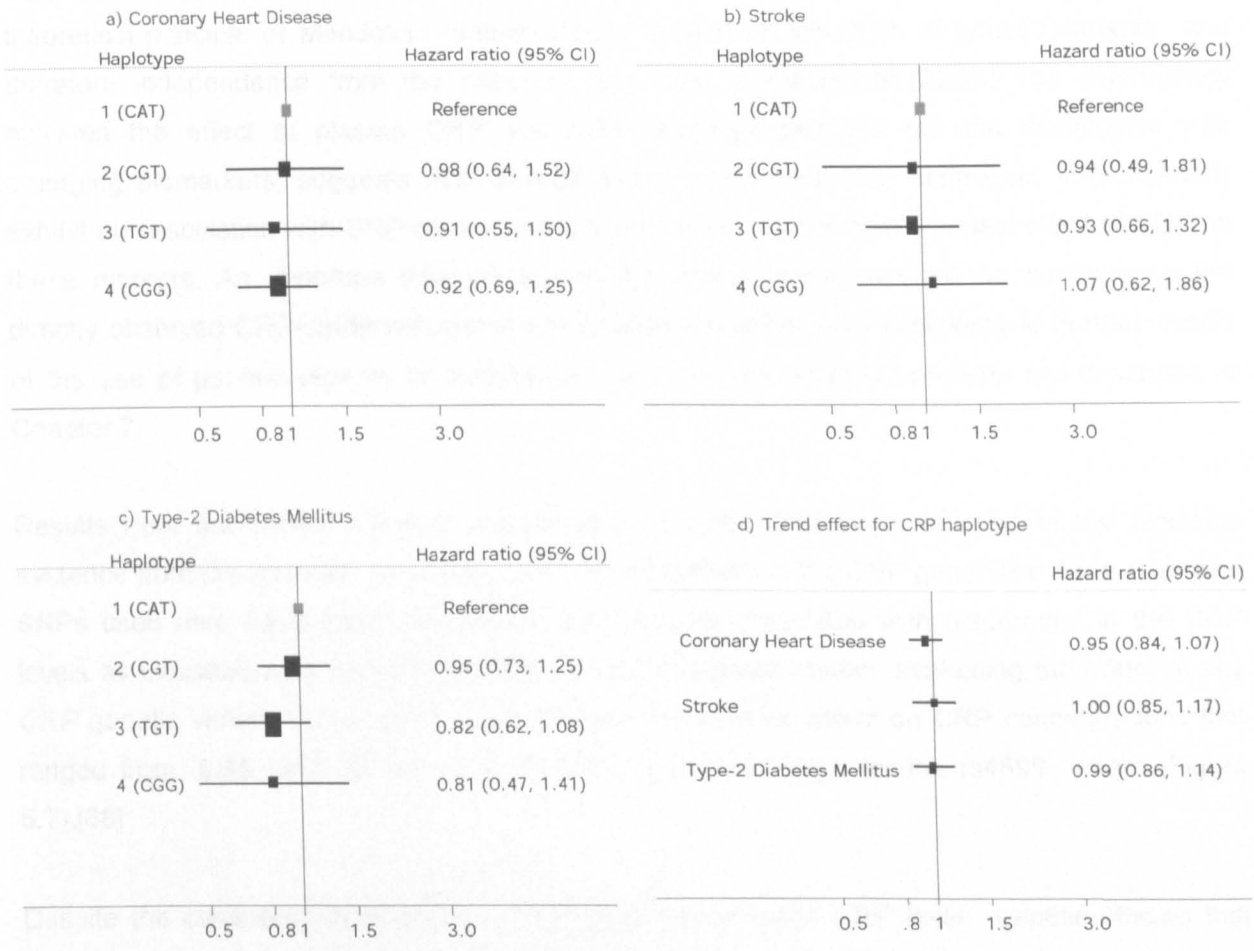
Figure 5.5 Standardised mean difference on cardiovascular traits according to concentrations of CRP (per-tertile increase) and by *CRP* haplotype (per-haplotype increase).



5.4 Discussion

Using data from two population-based cohorts, a *CRP* tagging-haplotype approach was not able to detect an increase in the risk of cardiovascular events for which consistent associations for plasma CRP have been reported. However, a substantial and clear effect of the *CRP* tagging haplotypes on CRP concentration was observed. No associations were observed between *CRP* tagging-haplotypes and progression of the ABI (surrogate end-point of peripheral and generalized atherosclerosis) or with any of the multiple emerging biomarkers that were associated with plasma CRP.

Figure 5.6 Hazard ratio of cardiovascular event according to the C-reactive protein tag-haplotypes. a) CHD; b) Stroke; c) Type-2 Diabetes Mellitus; and d) Trend-effect for *CRP*-haploytpe for the CHD, stroke and diabetes.



Since a genetic strategy, using tagging SNP approach that covers most of the genetic variation of the *CRP* locus, and the usual observational approach are prone to different sources of bias, the overall lack of consistency of these two strategies, across all the cardiovascular outcomes, between the genetic and the observational approach might at this stage indicate: i) that the plasma CRP effects observed in observational studies is perhaps an overestimation due to the presence of multiple biases (residual confounding and reverse causation), or ii) that association of plasma CRP reflect a true causal association, but that a gene-based approach, which overcomes some biases, is still largely underpowered to detect a small but important causal effect of CRP in cardiovascular disorders, if considering that such an effect will be concordant with the effect of the *CRP* gene on the plasma CRP.

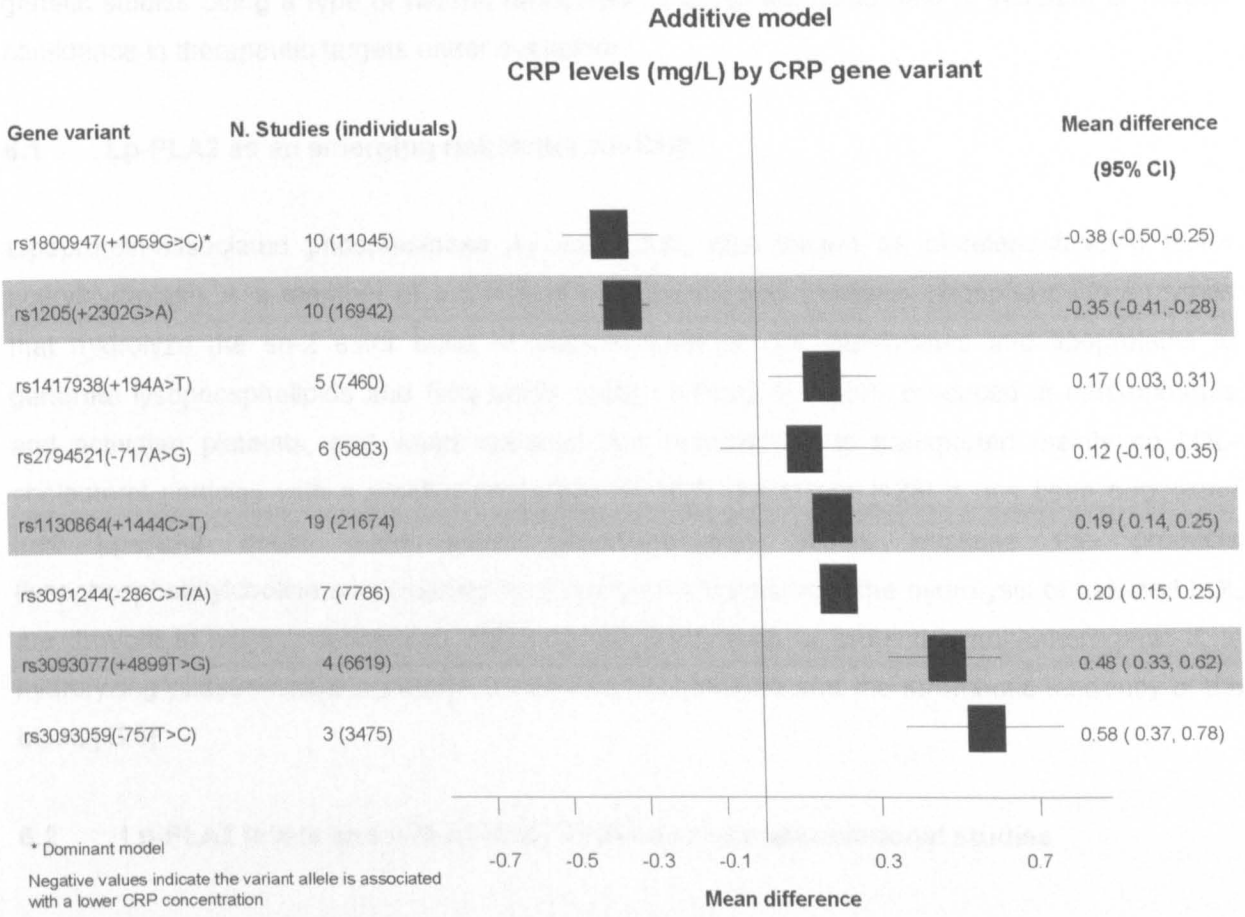
The lack of consistency between a genetic approach and the usual observational strategy was also observed for progression of the ankle-brachial index, as well as for all the emerging biomarkers, cardiovascular traits or behaviours with which plasma CRP has been consistently associated (Figure 5.5). The only exception to this was the presence of a clear effect of the *CRP* tagging-haplotypes on its cognate protein (Figure 5.4). This provides empirical evidence for the theoretical principle of Mendelian randomisation, of random allocation of genetic variants and therefore independence from the effect of any other characteristic.[22,23] The discrepancy between the effect of plasma CRP and *CRP* tagging-haplotypes (on the association with emerging biomarkers, suggests that perhaps those biomarkers (e.g. fibrinogen, IL-6, ICAM-1) exhibit an association with CRP because of a common link, rather than a causal effect of CRP on these markers. An important interpretation is that these biomarkers will be confounding the directly observed CRP-cardiovascular event association rather than explaining it. Further details of the use of genetic-variants to disentangle networks among blood proteins are described in Chapter 7.

Results from the current Chapter, should also be interpreted in the context of the available evidence from the literature on genetic association studies on the *CRP* gene. The three *CRP* tag-SNPs used here have been consistently and robustly associated with differences in the CRP levels as indicated by a recent systematic review of genetic studies evaluating the effect of any *CRP* genetic variant on the CRP concentrations. A per-allele effect on CRP concentrations that ranged from -0.35 mg/L for the rs1205 variant up to 0.48 mg/L for the rs4899 variant (Figure 5.7).[68]

Despite the consistent association of *CRP*-gene variants with CRP level, genetic studies that have investigated the effect on cardiovascular outcomes have often observed a null effect on indices of atherosclerosis (carotid intima media thickness),[121] CHD [122,123] and cardiovascular disorders such as type-2 diabetes mellitus,[124] high blood pressure,[108] or clustering of cardiovascular risk factors,[119] known as metabolic syndrome for which a positive association of plasma CRP have been reported.[127]

In conclusion, men of European ancestry with CGG haplotype of the *CRP* gene that exposed them to a 42% higher level of CRP compared with individuals with CAT haplotype, did not exhibit (i) an increased risk of any cardiovascular event (CHD, stroke or diabetes), or (ii) a difference in the progression of atherosclerosis (index by changes in the ABI) or (iii) a difference in the levels of other biomarkers. These are in clear contrast to what is observed for CRP itself.

Figure 5.7 Summary effects of eight *CRP* genetic-variants on CRP concentrations derived from published studies in European descent individuals. Highlighted in grey the 3 tag-SNPs used in NPHS-2 and EAS. Figure adapted from Figure 4 manuscript of Verzilli C, Shah T, Casas JP et al. Am J Hum Genet 2008;82:859-72.



Results from Chapter and Chapter 4, together with the available evidence indicates a clear effect of *CRP* variants on CRP levels but not on cardiovascular events,[122,123] suggesting that the plasma CRP-cardiovascular event association is due to residual confounding and or reverse causation, which are unavoidable limitations of observational studies. Nonetheless, before refuting a small but important causal effect of CRP on cardiovascular diseases, larger Mendelian randomisation experiments are needed before CRP is removed from the list of potential therapeutic targets for CHD prevention. The undergoing effort of the CCGC collaboration, which will include over 30,000 CHD events and 110,000 controls, using individual patient data will provide a unique opportunity to judge a possible causal role of CRP in CHD. [110]

Chapter 6 Lp-PLA2 and CHD

In Chapters 2 to 5, I utilised the Mendelian randomisation approach to investigate a potential causal role of biomarkers (Hcy and CRP) in cardiovascular disease (Stroke and CHD). In this Chapter, I describe an extension of the Mendelian randomisation paradigm, that capitalise on genetic studies being a type of natural randomised trial, which could help to increase or reduce confidence in therapeutic targets under evaluation.

6.1 Lp-PLA2 as an emerging risk factor for CHD

Lipoprotein-associated phospholipase A_2 (Lp-PLA2), also known as platelet-activating factor acetylhydrolase, is a member of a family of intracellular and secretory phospholipase enzymes that hydrolyze the sn-2 ester bond of phospholipids of cell membranes and lipoproteins to generate lysophospholipids and fatty acids. [128] Lp-PLA2 is mainly produced in macrophages and activated platelets, and when released into circulation it is transported mainly on LDL-cholesterol particles with a smaller proportion on HDL-cholesterol.[128] It has been suggested that Lp-PLA2 could exert either pro-atherogenic effects because the products (lysophosphatidylcholine and oxidized fatty acids) it releases from the hydrolysis of oxidized LDL are thought to have a deleterious effect on the artery wall, or protective anti-atherogenic if, in hydrolyzing platelet-activating factor, it reduces inflammation and the thrombotic tendency of the blood.[128]

6.2 Lp-PLA2 levels and risk of CHD: evidence from observational studies

Evidence from several prospective cohort studies (including studies from the general population and from individuals with established CHD) have reported a consistent association of Lp-PLA2 levels (either mass or activity) with an increase in the risk of CHD. Results from 14 prospective studies published by September 2006 were synthesised in a literature-based meta-analysis that showed an increase in risk of 1.21 (95%CI: 1.11, 1.32) for a 1 standard deviation increase in the Lp-PLA2 mass or activity, with a similar magnitude of effect observed among studies including mainly healthy individuals compared with studies of subjects with evidence of established CHD.[129] In addition, the strength of the Lp-PLA2-CHD association was of a similar size for studies measuring mass, compared with those measuring activity, of the Lp-PLA2 enzyme.[129]

Nonetheless, and as described previously for the CRP and Hcy, (see Chapters 2 and 4) the increase in risk of CHD associated with high levels of Lp-PLA2 reported in prospective cohorts could be entirely or largely explained by the multiple associations of Lp-PLA2 levels with other established as well as emerging risk factors for CHD, a phenomenon known as confounding (described in section 1.2 Chapter 1). Concentrations of Lp-PLA2 activity or mass have been associated with several established as well as emerging risk factors for CHD. Individuals with higher levels of Lp-PLA2 tend to be older, to have higher levels of total-cholesterol, LDL

cholesterol, triglycerides, apolipoprotein-B, blood pressure, body mass index, interleukin-6, CRP and fibrinogen. Likewise, individuals with higher than average levels of Lp-PLA2 exhibit lower levels of protective risk factors such as HDL-cholesterol and alcohol consumption.[130] These multiple associations of Lp-PLA2 means that despite statistical adjustments conducted by individual studies, residual confounding could still explain some or all of the observed association with CHD. In consequence, despite the encouraging evidence from observational studies, it is not sufficient to establish an unequivocally causal role of Lp-PLA2 in CHD.

6.3 Randomised evidence to elucidate the causal role of Lp-PLA2 on CHD

An approach that eliminates bias present in observational studies and therefore provides the strongest evidence on a causal role of Lp-PLA2, is to randomly allocate individuals to interventions that modify Lp-PLA2 activity in a selective manner. One possibility is to conduct a randomised clinical trial of a selective intervention such as the new drug "darapladib", a selective Lp-PLA2 inhibitor (formerly SB-480848).[131,132]

Recently two small, mechanistic, placebo-controlled randomised trials of darapladib were conducted providing the first indication of the effects of inhibiting the Lp-PLA2 enzyme.[131,132] The first trial, a dose-finding trial including 964 individuals treated for 81 days, provided suggestive evidence of a dose-dependent reduction of Lp-PLA2 activity of 43%, 55%, and 66% for darapladib doses of 40, 80 and 160 mg daily respectively. No clear effect on lipid particles (LDL, HDL or triglycerides) was observed, indicating the lipid-independent or specificity of the intervention.[131] In the same trial, inhibition of Lp-PLA2 with darapladib had no effect on several biomarkers (P-selectin, CD40 ligand, urinary 11-dehydrothromboxane B_2) measured as proxies of platelet activation. The second trial (330 individuals with 12 months follow up) evaluated the effect of 160 mg daily of darapladib on human atherosclerotic plaque measured using intravascular coronary ultrasound. There was no effect on the plaque deformability, the primary outcome, or on the total atheroma volume, even though the drug decreased the Lp-PLA2 activity by 59%. However, the intervention decreased the necrotic core volume of the plaque, a secondary outcome, by -5.2 mm ($p=0.012$) compared to the placebo group.[132]

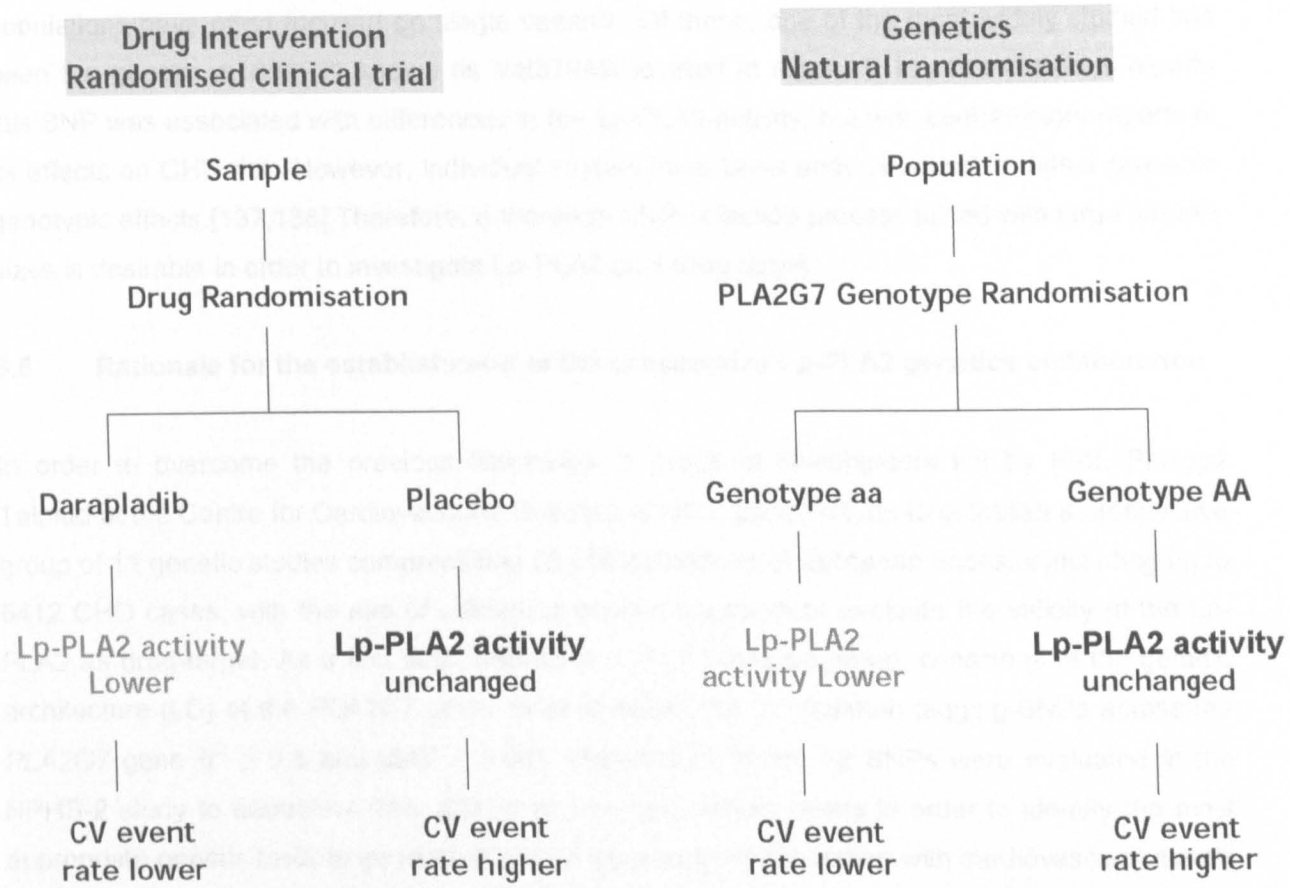
The efficacy profile of the Lp-PLA2 inhibitor has been sufficient to encourage further large and long term randomised trials evaluating hard clinical end-points. However, it seems reasonable to increase the confidence in the validity (efficacy and safety) of the drug target before proceeding into such costly trials and therefore minimize the risk of exposing individuals to an intervention with harmful effects. This is of great importance, considering the recent series of unexpected harmful effects of several new agents in recent phase-III trials in CVD, such as the case of the CETP-inhibitor torcetrabip and the dual- proliferator-activated receptors agonist muraglitazar.[133,134]

An approach that could help to increase or reduce the confidence in Lp-PLA2 as a valid drug target sufficient to inform on the decision to pursue or abandon LpPLA2 as a target that does not require the exposure of individuals to a drug, would be to investigate the effect on cardiovascular traits and CHD risk of randomly allocated variants within the Lp-PLA2 gene (known as *PLA2G7*) that modify levels of its cognate protein (Lp-PLA2) measured by activity or mass. This would generate a "like-with-like" comparison between the drug (darapladib: LP-PLA2 inhibitor) and the gene (*PLA2G7*: gene product Lp-PLA2). In both trials and genetic studies there is a random allocation procedure involved, in the case of the drug by trial randomisation procedures and in the case of the allelic-variant the natural randomisation resulting from Mendel's 2nd law, see Figure 6.1.

This extension of the Mendelian randomisation paradigm is complementary to that previously described in Chapters 2 to 5, in which Mendelian randomisation was used to provide unbiased evidence of causation of emerging biomarkers on cardiovascular disease. This new use of Mendelian randomisation capitalizes on genetic studies being a type of natural randomised trial (Figure 6.1).[120]

As genetic effects are smaller than the effects of potent drugs, the establishment of a robust natural randomised trial requires the investigation of several thousands of cases as well as individuals with a wide range of phenotypic data.

Figure 6.1. Parallel between randomised clinical trials and genetic studies addressing the same molecule target (e.g. Lp-PLA2 enzyme).



6.4 Genetic evidence on the regulation of Lp-PLA2 activity

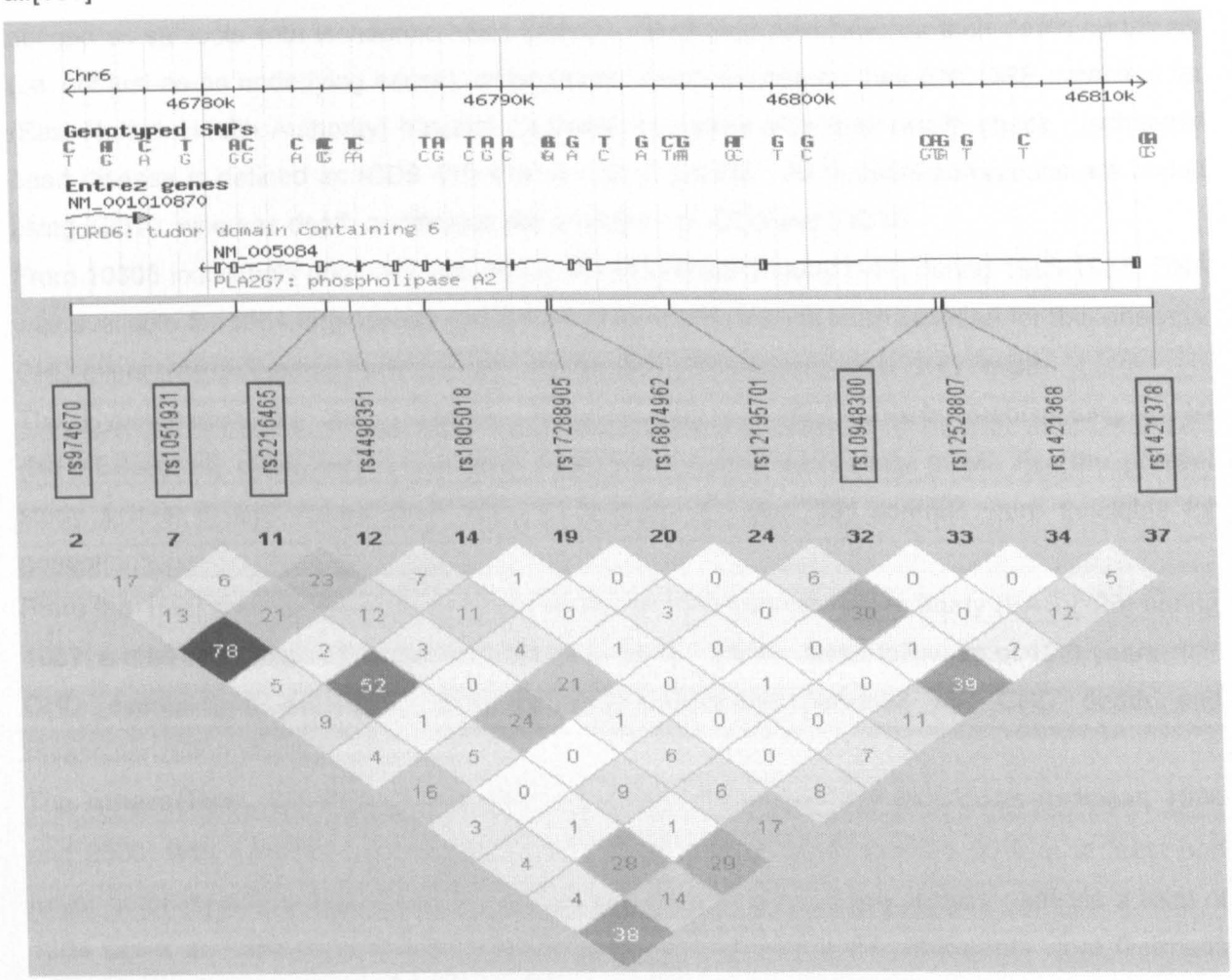
Lp-PLA2 activity is a highly heritable trait, with up to 62% of the plasma variation attributed to heritable factors.[135] It is almost certainly likely that this degree of heritability is accounted for by several genes acting together to regulate the Lp-PLA2 activity, as well as by shared, non-genetic familial factors (e.g. shared environment). However, of all the genes that influence LpPLA2 activity, it is *cis*-variants in the *PLA2G7* gene itself that encodes Lp-PLA2 (from now on referred to as a pQTG: protein quantitative trait gene) that provides the most specific instrument with which to model the pharmacological effect of LP-PLA2 inhibition by darapladib. The use of SNPs from other genes would lack specificity and therefore their findings would not be of help in assessing Lp-PLA2 as a drug-target (For more details see Chapter 7). The *PLA2G7* gene is organized into 12 exons spanning at least 45 kb of DNA sequence, and is located on chromosome 6p21.2-p12. The search for variants within the *PLA2G7* gene that determine differences in the Lp-PLA2 mass or activity has been complex and slow. In 1996, Stafforini D, et al. showed that a point mutation in exon 9, a G → T substitution, resulting in a Val → Phe transversion at position 279 of the mature protein (V279F) was responsible for the inherited deficiency in Lp-PLA2 activity present

in around 4% of the Japanese population.[136] More recently this mutation has also been reported in other populations (e.g. Turkish, Azerbaijan) but is almost absent in European descent populations.[128] Earlier attempts to evaluate genetic variants within *PLA2G7* in European populations have often focused on single variants. Of these, one of the most widely studied has been the genetic rs1051931 known as Val379Ala located at exon 11. In some previous reports this SNP was associated with differences in the Lp-PLA2 activity, but with contradictory reports of its effects on CHD risk. However, individual studies have been underpowered to detect plausible genotypic effects.[137,138] Therefore, a thorough SNP selection process joined with large sample sizes is desirable in order to investigate Lp-PLA2 as a drug target.

6.5 Rationale for the establishment of the prospective Lp-PLA2 genetics collaboration

In order to overcome the previous limitations, a group of investigators led by Prof. Philippa Talmud at the Centre for Cardiovascular Genetics at UCL, joined efforts to establish a prospective group of 11 genetic studies compromising 20,898 individuals of European ancestry including up to 8412 CHD cases, with the aim of utilising a genetic approach to evaluate the validity of the Lp-PLA2 as drug-target. As a first step, members of Prof Talmud's group, conscious of the genetic architecture (LD) of the *PLA2G7* gene, have identified the 12 common tagging-SNPs across the *PLA2G7* gene ($r^2 \geq 0.8$ and MAF ≥ 0.04). (Figure 6.2) These 12 SNPs were evaluated in the NPHS-2 study to determine their effects on Lp-PLA2 activity levels in order to identify the most appropriate genetic tools to be evaluated in a large scale collaboration with cardiovascular events as an outcome and in addition to evaluate the effect of the gene variants on continuous traits relevant to CHD risk. A second source of SNPs selection was derived from a comprehensive screening on the *PLA2G7* gene, aimed to select tagging SNP ($r^2 \geq 0.7$ and MAF ≥ 0.05) in both European descent and African-American populations living in the United Stated. [139]

Figure 6.2. Genetic architecture of the *PLA2G7* gene in subjects of European ancestry participating in NPHS-2 study. Displayed the 12 tagging SNPs (placed according to their genetic location) from *PLA2G7* gene typed in NPHS-2. Depicted is the LD plot visualised using the program Haploview. R^2 values among the tSNPs are displayed in the cells. Highlighted by a rectangle are the five tSNPs selected from NPHS-2 and typed in the genetic collaboration. In addition, two tSNPs rs1805018 and rs9381475 were selected from the publication by Sutton et al.[139]



6.6 Methods

6.6.1 Study population

After obtaining ethical approval from respective studies, genotyping for the *PLA2G7* polymorphisms and measurement of the Lp-PLA2 activity were conducted in samples obtained from a number of cross-sectional, case-only, case-control, or prospective studies summarised in Table 6.1.

From a total of 3012 participants from the Second Northwick Park Heart Study (**NPHS-2;** described in Chapters 4 and 5) [99] recruited between 1989-1994, genotyping was conducted in 2706 men of whom 229 developed an incident coronary event after 12 years of follow up. CHD

events include non-fatal MI, CHD deaths, revascularization procedures or a silent MI after the 5th phase of follow up.

The European Prospective Investigation into Cancer and Nutrition study, Norfolk-UK component **(EPIC-Norfolk)**[140] recruited between 1993 and 1997 a total of 25,663 men and women of European descent. For the current analyses, a nested case-control design based on participants free of cardiovascular disease at baseline was utilised and includes a total of 2159 incident CHD cases and 2159 controls with DNA available after mean follow up of 12 years. CHD cases were defined as subjects with ischaemic heart disease mentioned anywhere on their death certificate (i.e. not just as an underlying cause), or ischaemic heart disease on their ENCORE database for (East Norfolk Health Authority) hospital discharge occurring after their health check. Ischaemic heart disease is defined as ICD9 410-414 or ICD10 I20-I25. All hospital admissions are coded using ICD10, whereas death certificates are a mixture of ICD9 and ICD10.

From 10308 individuals originally included in the **Whitehall-2** study [141] during 1985-1988, DNA was available for 5664 individuals, and a total of 457 CHD events were included for this analysis. CHD event was defined as having a non-fatal MI (MONICA criteria), or a CHD death.

The Hypercoagulability and Impaired Fibrinolytic function MECHanisms predisposing to MI (**HIFMECH**)[138] study was a European multicentre case-control study of MI. For the present study a total of 527 cases (fatal and non-fatal events) and 566 controls were available for genotyping.

From the 1592 men and women originally recruited in Edinburgh Artery Study **(EAS)**[126] during 1987, a total of 913 had DNA available for genotyping. After a mean follow up of 12.8 years, 117 CHD events had accrued. CHD outcome comprised non-fatal MI, CHD death and revascularization procedures.

The **AtheroGene**[137] case-control study recruited a total of 1318 individuals (between 1996 and 2000) with coronary artery disease defined as a diameter stenosis $\geq$ 30% in at least one major coronary artery and a total of 485 healthy individuals. For the present analysis a total of 1304 cases and 468 controls with DNA available were included. All participants were Germans and inhabitants from the geographical region in which the recruitment was conducted.

The Southampton Atherosclerosis Study (**SAS**)[142] compromised a total of 1300 subjects recruited during 1999 to 2001, with coronary artery disease defined as a stenosis $\geq$ 50% of the diameter in at least one major epicardial coronary artery. For these analyses a total of 1164 individuals with DNA available were included.

The Ludwigshafen Risk and Cardiovascular Health Study (**LURIC**)[143] included 3148 European descent individuals hospitalised for coronary angiography between 1997 and 2000. A case of coronary artery disease was defined as a subject with coronary stenosis $\geq$ 20% in at least 1 of 15 coronary segments according to the American Heart Association. A control was a subject without angiographic evidence of coronary stenosis.

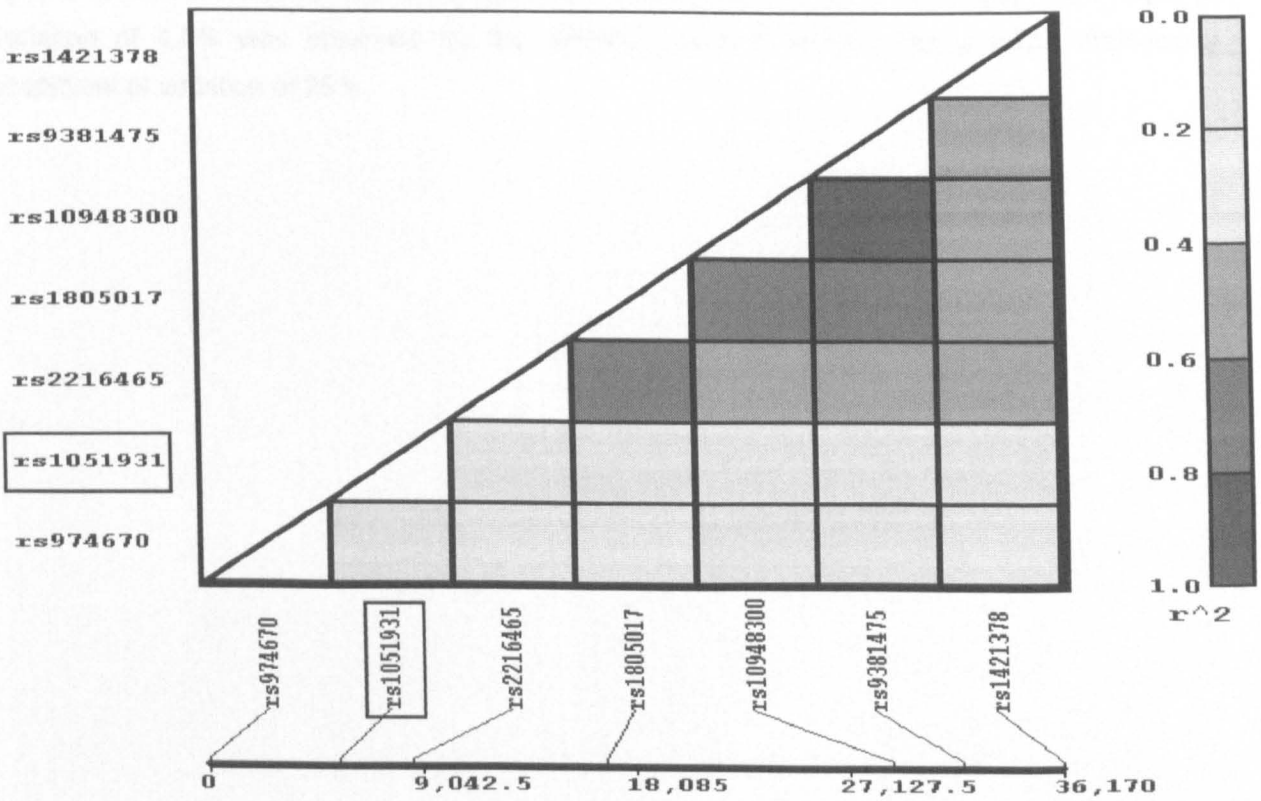
A total of 731 controls and 2581 cases with DNA available were included in this report. The **Cyprus** study [144] (still in recruitment phase) is a prospective cohort aiming to include 2000

individuals aged 40 or more. To date 767 subjects have been recruited and of which 739 individuals with DNA available have been included in this analysis.

The UCL DiAbetes and Cardiovascular disease Study (**UDACS**)[103] is a cross sectional case-control study to evaluate risk factors for coronary heart disease in subjects with diabetes mellitus. For the current study only European descent individuals (n=424) without coronary heart disease were included.

The THROMBOgenic Factors and Recurrent Coronary Events (**THROMBO**) study [145] involved a total of 766 patients who had survived a myocardial infarction, of whom 529 individuals with DNA available were included in this analysis. Blood markers (including Lp-PLA2 activity) included in the THROMBO study and presented here were determined 2 months after the index MI. For the current analyses only the baseline information was utilised, since coronary events were recurrent instead of incident events.

Figure 6.3. Linkage disequilibrium map of the seven PLA2G7 tagging-SNPs included in the Lp-PLA2 genetic collaboration. Highlighted with a black rectangle-box is the rs1051931 (A379V).



6.6.2 Data collection

Following a pre-determined analysis plan generated by myself, information was obtained on Lp-PLA2 activity levels, cardiovascular risk factors, *PLA2G7* variants and cardiovascular events from the 11 participating studies within the collaboration. The analysis plan included unified definitions

for coding of all relevant variables utilised in the analysis. For a sub-set of pre-defined variables (Lp-PLA2 activity, CRP, and triglycerides) natural-logarithmic transformations were used and these variables were always analysed in the log-scale. For the different multivariate models utilised to obtain risk of CHD, these included the same group of covariates across all the studies. I used the established definitions within each study for cardiovascular events. However, studies were then divided in two categories; those using CHD as an outcome (EPIC-Norfolk, NPHS-2, Whitehall-2, EAS, Cyprus, HIFMECH) and those using coronary artery disease as an outcome (Southampton Atherosclerosis, Athero*Gene* and LURIC). All the analyses were limited to individuals self-identified as being of European descent.

6.6.3 Laboratory analysis

In order to decrease any source of inter-study variability Lp-PLA2 activity in Athero*Gene*, EPIC-Norfolk, NPHS-2, UDACS, and Cyprus was determined in a single laboratory and investigator (Dr Ewa Ninio at INSERM U525, Paris), who assayed Lp-PLA2 activity using a radiometric assay as reported previously [146,147] For the THROMBO study, plasma Lp-PLA2 activity was determined by a commercial colorimetric assay (Cayman Chemical Co.) with 2-thio-PAF as substrate and following the manufacturer's directions.[145] For the LURIC study, LpPLA2 activity was measured by use of the Azwell Auto PAF-AH reagent set (Azwell).[148] A mean intra-assay coefficient of variation of 4.0% was observed for the NPHS-2 samples tested, and a mean inter-assay coefficient of variation of 25%.

Table 6.1 Design characteristics of the studies included in the present analysis.

Study Name	Study Design (location)	*PLA2G7* variants genotyped			Sampling frame	Main selection criteria	Primary analysis	
							Genotype & LpPLA2 level (# subjects)	Genotype & CHD (# subjects)
NPHS-2	Population-based Prospective Cohort (UK)	rs1051931 rs1805017 rs10948300	rs974670 rs2216465	rs1421378 rs9381475	All men aged between 50 to 63 years registered with 9 primary care practices.	Individuals were excluded if had: Pre-existing cardiovascular disease (CHD or stroke), Coronary surgery or malignant disease, or were taking Aspirin or anticoagulant.	Yes (2099)	Yes (2706)
EPIC-Norfolk	population-based Nested case-control study (UK)	rs1051931 rs2216465	rs1421378 rs9381475	rs1805017 rs10948300	All men and women aged between 40 and 79 years registered in primary care practices.	N/A	Yes (1394 controls)	Yes (4225)
CYPRUS	Prospective cohort (Cyprus)	rs1051931 rs1805017	rs974670 rs9381475	rs1421378	All men and women > 40 years identified through the population list at the Mayor's office	To be included subjects had to be inhabitants from Pedoulas or Nissou villages in Cyprus.	Yes (633)	No
HIFMECH	Case-control (UK, Sweden, France, Italy)	rs1051931			Cases: male MI survivors. Controls: men matched by age and regional areas.	Subjects with Familial hypercholesterolaemia, or Insulin-dependent diabetes mellitus were excluded.	No	Yes (1093)
THROMBO	Prospective cohort (United States of America)	rs1051931 rs1805017	rs974670 rs9381475	rs1421378	Consecutive series of post-infarction subjects	Individuals with diabetes were excluded.	Yes (524)	No

Table 6.1 (Continuation)

Name of the Study	Study Design (Location)	*PLA2G7* gene variants genotyped			Sampling frame	Main selection criteria	Primary analysis Genotype & LpPLA2 level (# subjects)	Primary analysis Genotype & CHD (# subjects)
Athero*Gene*	Case-control (Germany)	rs1051931 rs1805017	rs974670 rs9381475	rs1421378	Cases: patients with stable angina attending for diagnostic coronary angiography Controls : recruited either from GP's offices (routine check-up) or by newspaper announcement	Cases were excluded if there was evidence of significant co-morbidities: valvular heart disease, cardiomyopathy, malignant disease or febrile condition. Controls: individuals with evidence, from the interview, on atherosclerosis or with pathological ECG were excluded	Yes (468)	Yes (1304)
Whitehall-2	Prospective cohort (UK)	rs1051931			All civil servants in 20 departments at Whitehall	To be included individuals had to have a job-contract in a London-based civil service department	No	Yes (5611)
UDACS	Case-control (UK)	rs1051931 rs1805017 rs10948300	rs974670 rs2216465	rs1421378 rs9381475	Consecutive subjects recruited to diabetic clinic Cases: subjects with diabetes and CVD. Controls: individuals with diabetes but without CVD.	To be included individuals had to have T2DM according to WHO criteria	Yes (235 control subjects)	No
SAS	Collection of cases with coronary stenosis (UK)	rs1051931			Consecutive patients undertaking diagnostic and interventional coronary angiography	N/A	No	Yes (1091)
LURIC	Case-control (Germany)	rs1051931	rs1805017		Men and women hospitalized for coronary angiography.	N/A	Yes (728)	Yes (2581)
EAS	Population-based prospective cohort (UK)	rs1051931			A random sample stratified by age and selected from 10 general practises throughout the city of Edinburgh.	Subjects were excluded if they were unfit to participate (e.g. due to mental illness)	No	Yes (895)

6.6.4 Genotyping of single nucleotide polymorphisms in the *PLA2G7* gene

Following previous identification of tag-SNPs in the *PLA2G7* gene (i.e. a subset of all SNPs in the gene that captures information on untyped SNPs because of LD) by Prof Talmud's group (Figure 6.2), plus recent findings published by Sutton et al. [139] on *PLA2G7* variants, a total of seven *PLA2G7* variants were finally selected based on their effects on Lp-PLA2 activity or CHD risk, while considering their LD structure to avoid the inclusion of SNPs in strong LD (Figure 6.3). See Table 6.2 for the list of included variants.

Table 6.2 Description of coding of PLA2G7 variants included in the analysis.

rs_number	Common allele	Rare allele	Homozygous common allele (reference group) coded = 1	Heterozygous coded =2	Homozygous rare allele coded=3
rs974670	C	T	CC	CT	TT
rs1805017	G	A	GG	GA	AA
rs9381475	C	T	GG	GA	AA
rs10948300	A	T	AA	AT	TT
rs2216465	G	C	GG	GC	CC
rs1421378	A	G	AA	AG	GG
rs1051931 (A379V)	G	A	GG	GA	AA

For all the studies (except HIFMECH) the genotyping for the A379V (rs1051931) was done using Taqman assays by design. For HIFMECH study, genotyping was determined using an allele-specific polymerase chain reaction (PCR), using one sense primer (5′-AGG GAG ACA TAG ATT CAA CTG-3′) and one anti-sense primer (5′-CGT TTT GTA AGA ATG CTA ATG AA-3′). Use of these primers introduces a *PstI* site in the A379 allele in all subjects excluding rare homozygotes. The size of the undigested fragment is 69 base-pair (bp). After restriction, *PstI* generates fragment sizes of 49/20 bp in common homozygotes, 69/49/20 bp in heterozygotes and 69 bp in rare homozygotes. Restriction enzyme digestion was performed in a volume of 13 µl containing 5 µl of the PCR product and the buffer recommended by the manufacturer (Promega Inc.) for 240 min at 37 °C. Fragments were resolved using Microtitre Array Diagonal Gel Electrophoresis on ethidium-bromide stained 7.5% polyacrylamide gels. One heterozygote individual, as confirmed by sequencing and one negative control were included in each PCR run. All rare homozygote samples were re-amplified and re-digested to confirm genotype. All other additional SNPs (rs974670, rs1805017, rs9381475, rs10948300, rs2216465 and rs1421378) were genotyped using Taqman assays by design. With the exception of EPIC-Norfolk and LURIC studies, whose genotyping was done using local laboratory facilities available to the investigators, all genotyping was conducted by members of Philippa Talmud's group at the Centre for Cardiovascular Genetics

at UCL. All genotyping was performed by investigators unaware of subject status and confirmed by a second viewer.

6.6.5 Statistical analysis

Following the pre-specified analysis the following variables were log-transformed using the natural log-scale: triglycerides, and CRP. All other variables were analysed in their original scale. Throughout all analyses, a two-sided probability value <0.05 was taken to denote statistical significance. Results across studies were pooled by random model effects, using the inverse variance method.

6.6.5.1 Lp-PLA2 activity, cardiovascular traits and risk of CHD

For this section, only data from the NPSH-2 and EPIC-Norfolk prospective studies was used as they were the only prospective studies with coronary events and levels of Lp-PLA2 activity.

To evaluate the associations between levels of Lp-PLA2 activity with cardiovascular traits, quartiles of log-Lp-PLA2 activity were generated, and the unadjusted means with their respective standard deviations were obtained for the following cardiovascular traits: age (years), body mass index (kg/m^2), systolic blood pressure (mm Hg), total-cholesterol (mmol/L), LDL cholesterol (mmol/L), HDL-cholesterol (mmol/L), log-triglycerides (mmol/L), log-CRP (mg/dL) and alcohol consumption (units/week). Then, using the bottom quartile (Q1) as the reference category, the mean difference for each quartile was obtained and results were pooled across studies using random model effects. For these associations, only individuals without known CHD at the time of measuring the Lp-PLA2 activity were included, to reduce inflation of the effect size possible in individuals with established disease.

The association between Lp-PLA2 activity and incident CHD was assessed using Cox proportional hazards regression. Hazard ratios (HR) and 95%CI were obtained for each quartile, using the bottom quartile as the reference group. Hazard ratios and their 95%CI for each quartile were then pooled across the two studies to obtain a summary effect for each quartile. A progressive adjustment for potential confounders was conducted as follows: model-1 included age (continuous), gender and practice (for NPHS-2) and enrolment date (for EPIC); model-2: adjusted for variables in previous model plus body mass index (continuous), smoking (current or ex vs. never), diabetes (yes vs. no), systolic blood pressure (continuous), CRP (continuous), fibrinogen (continuous), alcohol (continuous as: units/week). Model 3: adjusted for variables in model-2 plus: total-cholesterol (continuous), triglycerides (continuous) apo-A1 (continuous), apo-B (continuous). If apo-A1 was not available, HDL-cholesterol (continuous) was used. Similarly, if apo-B was not available LDL-cholesterol (continuous) was used. For details of CHD outcome definitions see section 6.6.1 from this Chapter.

6.6.5.2 *PLA2G7* gene variants and Lp-PLA2 activity levels

For this section, data from 11 studies were included. Coding of the genotypes for each of the *PLA2G7* variants was conducted as described in Table 6.2. For all the analyses, individuals homozygous for the common allele served as the reference group. For associations of *PLA2G7* variants with Lp-PLA2 levels and cardiovascular traits only individuals without known CHD were included. Levels of LpPLA2 activity were log-transformed and all analyses were conducted using log-transformed values.

To evaluate the effect of the different *PLA2G7* variants on levels of Lp-PLA2 activity, 6081 individuals without clinically evident cardiovascular disease from 7 studies with available DNA and Lp-PLA2 activity levels were studied; see Table 6.1. The unadjusted mean value of the log-LP-PLA2 activity with its respective standard deviation was obtained for each genotype from the seven *PLA2G7* variants. Mean differences in log-Lp-PLA2 (log-ratio) for each genotype were obtained and results were pooled across studies using random model effects.

6.6.5.3 *PLA2G7* gene variants and cardiovascular traits

From the previous analyses, the gene-variant exhibiting the largest effect on Lp-PLA2 activity was then used as proxy of the likely effect on cardiovascular traits to be observed when using an Lp-PLA2 inhibitor in a clinical trial. The effect of the gene variant on cardiovascular traits is expected to be proportional to the magnitude of the effect that the gene-variant exhibits on the Lp-PLA2 activity levels. For this analysis, unadjusted mean values of cardiovascular traits with their respective standard deviations were obtained for each genotype. Mean differences in the cardiovascular traits for each genotype were calculated and then data were pooled across studies using random effect models. For cohort studies with repeated measures of some continuous traits, the baseline or first available measure (in the case of missing data) was utilised. Cardiovascular traits included here are those evaluated in the observational component, in order to allow a more like-with-like comparison.

6.6.5.4 *PLA2G7* gene variants and risk of CHD

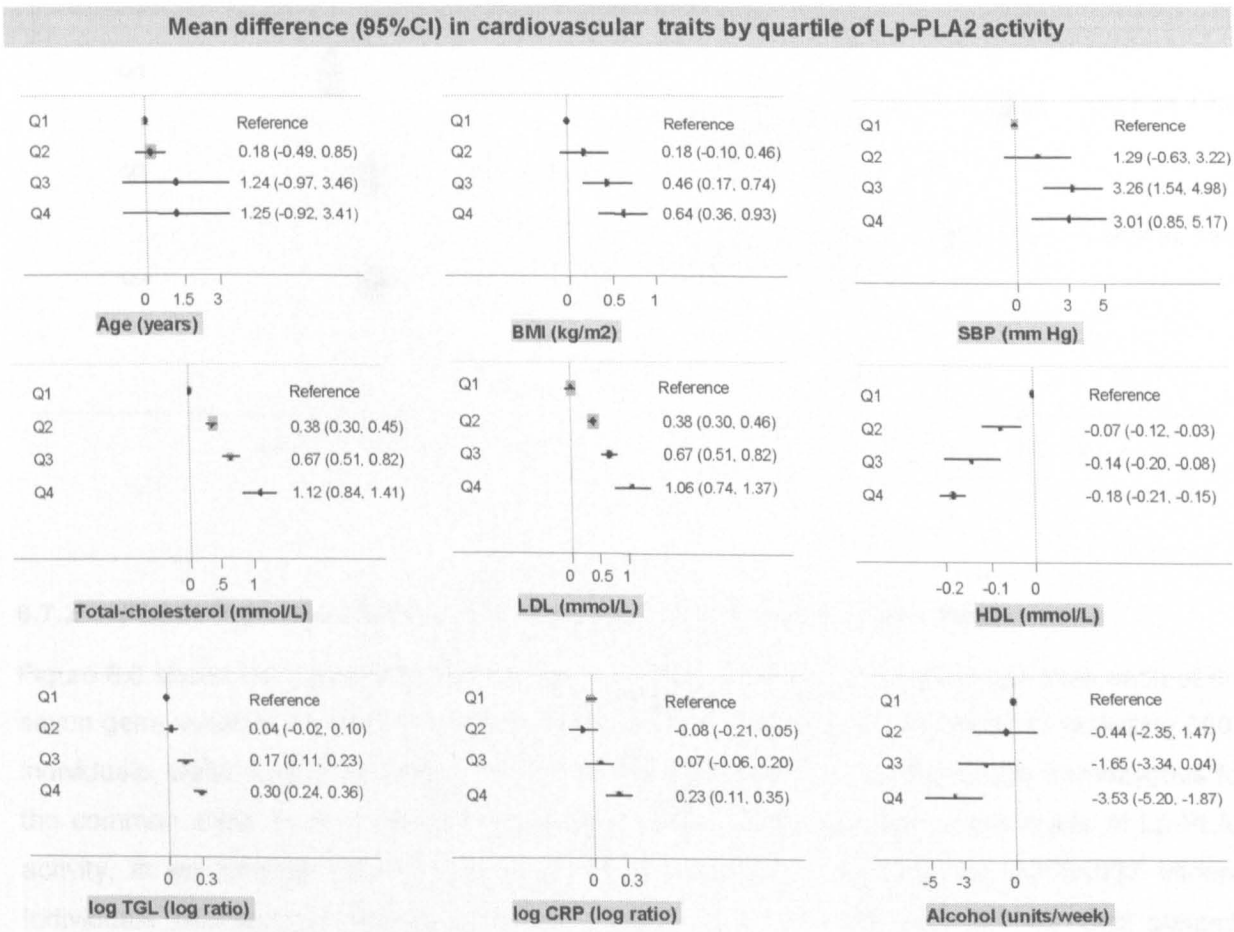
To evaluate the effect of the *PLA2G7* variants on the risk of CHD, a total of 8 studies including 19,506 individuals were included. For each gene-variant, using individual homozygous for the common allele as the reference group, a measure of association was obtained for each genotype comparison. Specifically, the log-odds ratios and their standard errors for case-control studies and log-hazard ratios with their standard errors for cohort studies were obtained for each genotype comparison from each gene-variant analysed. All measures of effect were adjusted by age (continuous), and gender. Enrolment date or practice centre was also included if relevant to each study. Assuming equivalence of log-odds ratio with log-hazard ratios, these measures of effect were then pooled across studies using random effect models and from now on these will be refer red to as the odds ratios. Initially, pooled OR were obtained separately according to the outcome reported; i.e. coronary artery disease (determined by coronary angiography) or CHD. If no evidence of heterogeneity was observed between the outcomes a summary OR was calculated.

6.7 Results

6.7.1 Lp-PLA2 activity, cardiovascular traits and risk of CHD

Using data from the NPHS-2 and EPIC-Norfolk cohort studies, it was observed that healthy individuals with higher than average levels of Lp-PLA2 activity tend to be older, to have higher levels of established risk factors such as body mass index, blood pressure, total-cholesterol, LDL-cholesterol and triglycerides. Similar positive correlations were also observed for the emerging risk factors fibrinogen and CRP. Lp-PLA2 activity was also inversely correlated with HDL-cholesterol and alcohol consumption (for details see Figure 6.4). These multiple associations of Lp-PLA2 activity illustrate the great potential for confounding of the Lp-PLA2 activity-CHD association.

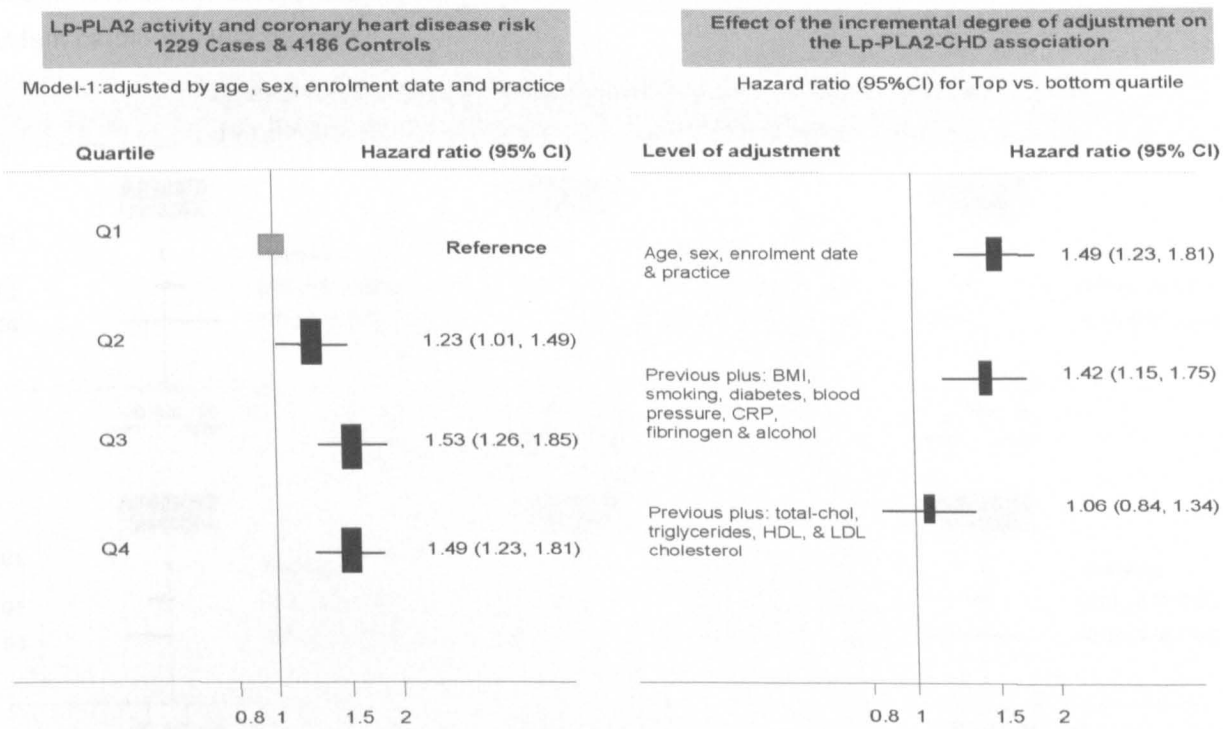
Figure 6.4. Mean differences in cardiovascular traits by quartile of Lp-PLA2 activity. Subjects from the bottom quartile (Q1) were used as the reference group.



As indicated by previous observational studies,[129] high levels of Lp-PLA2 activity were associated with a higher risk of CHD. The large number of cases accrued in the NPHS-2 and EPIC-Norfolk cohorts (1229 CHD events) allowed me to explore the shape of the Lp-PLA2 activity-CHD association, which appears to be continuous with no clear evidence of threshold effect across the quartiles of Lp-PLA2 activity (Figure 6.5). Individuals in the top quartile had a HR

of 1.49 (95%CI: 1.23, 1.81) compared with individuals from the bottom quartile in a minimally adjusted model (Model-1). However, as the level of adjustment increased the magnitude of the Lp-PLA2 activity-CHD association diminished to a HR of 1.06 (95%CI: 0.84. 1.34), see right panel Figure 6.5.

Figure 6.5. Hazard ratio (HR) of CHD by quartiles of Lp-PLA2 activity. Left panel describe the shape of the association using the minimally adjusted model (model-1). Right panel describe the effect on the HR (for the top vs. bottom quartile comparison) of progressive levels of adjustment.

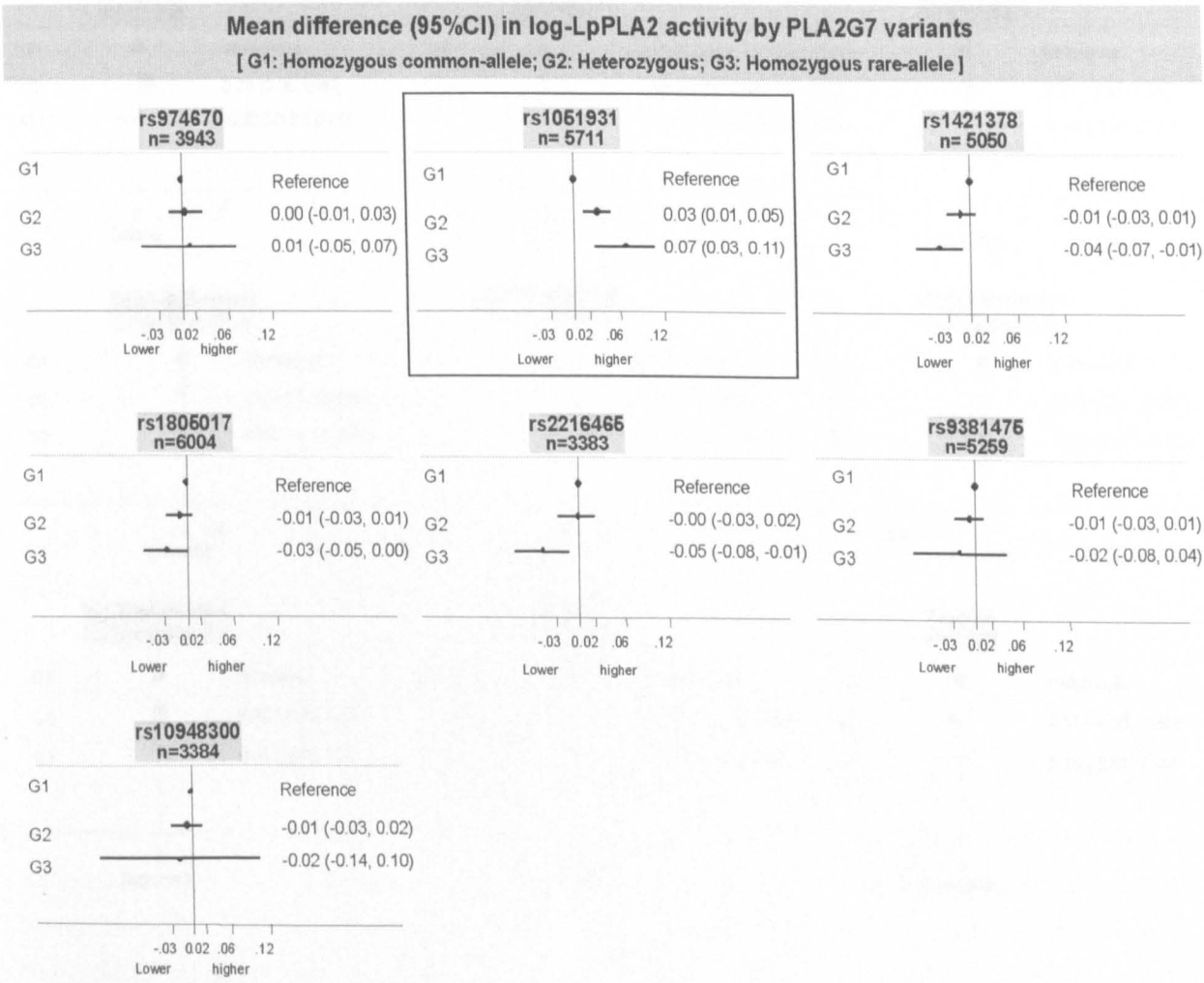


6.7.2 *PLA2G7* gene variants, Lp-PLA2 activity and cardiovascular traits

Figure 6.6 shows the mean difference in Lp-PLA2 activity (log-ratio) by genotype from each of the seven gene variants included. The most widely studied variant was the rs1805017 including 6004 individuals, while variant rs2216465 included 3383 individuals. Using individuals homozygous for the common allele as the reference group (G1), a significant increase in the levels of Lp-PLA2 activity, in an additive fashion was observed for individuals carrying the rs1051931 variant. Individuals homozygous for the rare allele had a relative difference of 7%, and subjects heterozygous had a 3% relative difference compared with those homozygous for the common allele. A trend towards an inverse association, lower levels, especially for subjects homozygous of the rare allele was observed for the following gene-variants: rs1421378, rs1805017 and rs2216465. None of these variants exhibited a strong LD (r^2 <0.2) with the variant rs1051931. A null association with levels of Lp-PLA2 activity was observed for the variants rs974670, rs9381475 and rs10948300. For details see Figure 6.6.

Plasma values of Lp-PLA2 activity itself were strongly associated with a range of established risk factors and emerging biomarkers for CHD (see Figure 6.4). However, the gene variant rs1051931 selected as an unconfounded tool of the Lp-PLA2 activity (as well as a proxy of Lp-PLA2 inhibitor, darapladib) did not exhibit any systematic association with any of the cardiovascular traits evaluated, including some lipid fractions that had strong association with Lp-PLA2 activity (Figure 6.7).

Figure 6.6. Mean differences by genotype in the Lp-PLA2 activity levels according the *PLA2G7* variants evaluated. Data pooled from 7 studies including up to 6081 individuals without clinically evident cardiovascular disease.

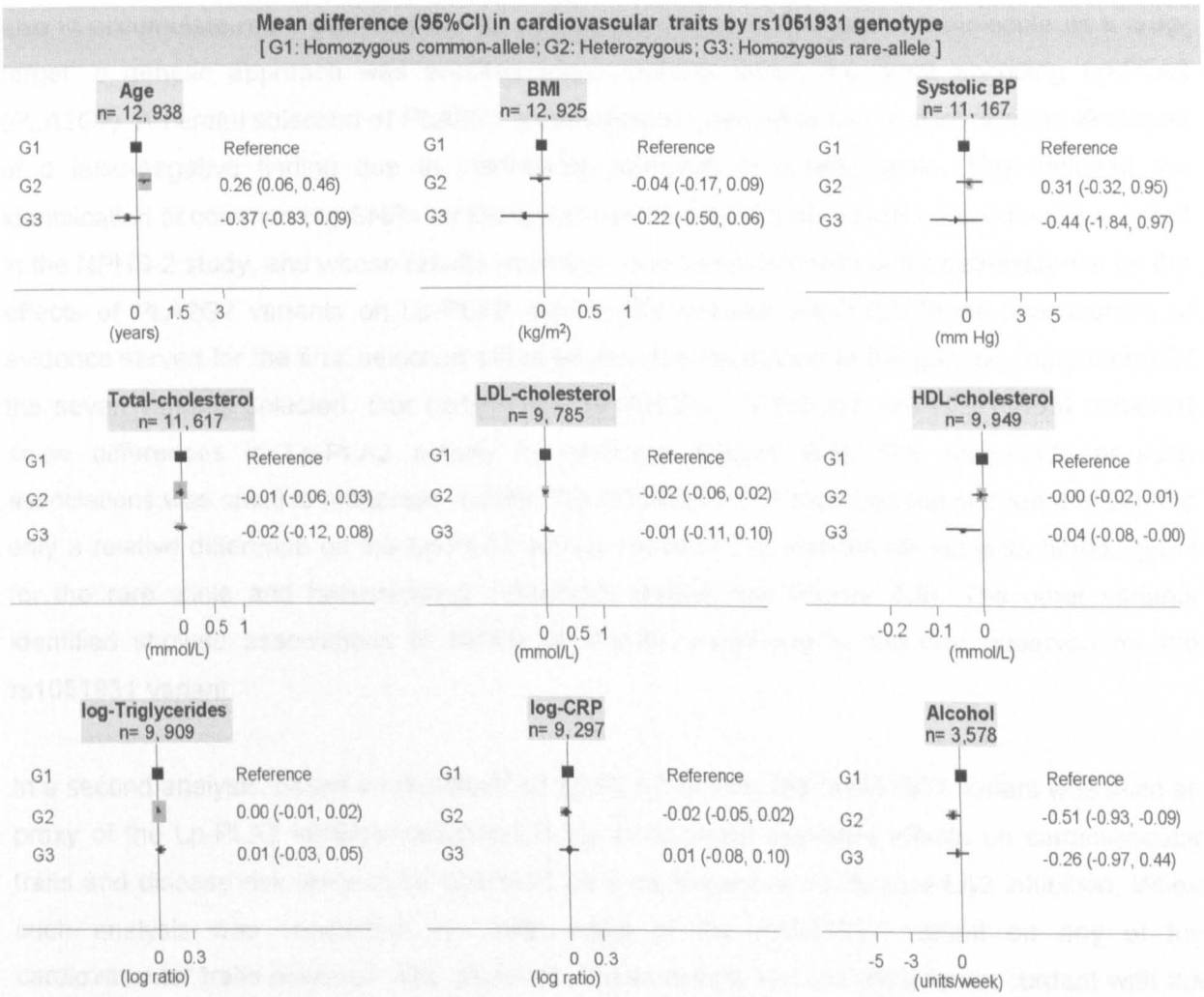


6.7.3 *PLA2G7* gene variants and risk of CHD

After pooling data from 8 studies with a variable number of CHD cases per-variant, ranging from 1528 events for the rs974670 up to 8412 events for the rs1051931, no evidence of a clear association with CHD was observed for any of the seven variants evaluated (Figure 6.8). Indeed, the variant rs1051931 that was consistently associated with a small-to-moderate difference in the levels Lp-PLA2 activity and included more than 8000 cases did not show an increase in the risk of CHD (see Figure 6.8). Nonetheless, the effect of the rs1051931 variant on Lp-PLA2 activity was

small-to-moderate in magnitude, therefore it is possible that the expected effect on CHD risk could lie within the confidence intervals of the effect observed for each of the genotype comparisons. No heterogeneity according to the outcome evaluated (CHD vs. Coronary artery disease) was observed (Figure 6.9)

Figure 6.7. Mean differences in cardiovascular traits by genotype according the *PLA2G7* variants evaluated. Data pooled from up to 10 studies and 12,938 individuals without cardiovascular disease.



6.8 Discussion

After including data from 1229 incident events an increase in risk of CHD was observed for individuals with high levels of Lp-PLA2 activity, as previously reported by several prospective studies.[129] The current analysis showed that the increase in risk seems to be present across all levels of Lp-PLA2 activity, without clear evidence of a threshold effect. However, the magnitude of the association diminished substantially as the number of cardiovascular risk factors adjusted for

increased, especially after the inclusion of lipid particles in the model (Figure 6.5). This finding, along with the multiple correlates observed for Lp-PLA2 activity (Figure 6.4) suggests that residual confounding is a potential explanation for the association observed. Alternatively, it is possible that some of the variables judged as confounders might mediate the effect of increased Lp-PLA2 activity on risk of CHD and if so, adjustment for such variables would be inappropriate, if the aim is to identify Lp-PLA2 activity as a causal factor in CHD.

In order to try to minimise previous uncertainties about a causal role of the Lp-PLA2 on CHD, and also to accumulate more evidence on the validity and safety of the Lp-PLA2 molecule as a drug-target, a genetic approach was adopted using variants within the gene encoding Lp-PLA2 (*PLA2G7*). A careful selection of *PLA2G7* gene-variants was conducted to minimise the likelihood of a false-negative finding due to inadequate selection of genetic tools. This included the identification of common tag-SNPs for European descent individuals which were initially evaluated in the NPHS-2 study, and whose results were then complemented with published-evidence on the effects of *PLA2G7* variants on Lp-PLA2 activity and disease risk.[139] These two sources of evidence served for the final selection of the seven variants utilised in the genetic component. Of the seven variants selected, four (rs1051931, rs1421378, rs1805017 and rs2216465) exhibited some differences in Lp-PLA2 activity by genotype (Figure 6.6). The magnitude of such associations was small to moderate. Indeed, the rs1051931 that exhibited the strongest effect had only a relative difference on the Lp-PLA2 activity levels of 7% and 3% for subjects homozygous for the rare allele and heterozygous individuals respectively (Figure 6.6). The other variants identified showed associations of similar or smaller magnitude to the one observed for the rs1051931 variant.

In a second analysis, based on its effects on Lp-PLA2 activity, the rs1051931 variant was used as proxy of the Lp-PLA2 inhibitor darapadib to try to elucidate expected effects on cardiovascular traits and disease risk likely to be observed as a consequence of the Lp-PLA2 inhibition. When such analysis was conducted, no clear effect of the rs1051931 variant on any of the cardiovascular traits analysed was observed. These results are directionally concordant with the preliminary data from the two recent short-term and randomised trials in humans using the Lp-PLA2 inhibitor darapladib conducted to date.[131,132] Despite reducing Lp-PLA2 activity, by 66% at 160 mg daily darapladib did not modify the concentrations of lipid particles (LDL, HDL or triglycerides). Similar findings have been recently observed in a pig model in which elevations in lipid concentrations and Lp-PLA2 activity were seen by rendering the pig diabetics. However, treatment with darapladib normalised Lp-PLA2 activity but had no effects on lipids.[149] Finally, I evaluated the effect of the seven *PLA2G7* variants on CHD risk, and despite that some comparisons included more than 8000 cases, no clear genetic signal could be detected for any of the genetic comparisons (Figure 6.8).

It is, however, important to take into consideration the limitations of the genetic approach adopted here. First, the use of a Mendelian randomisation strategy relies greatly on the ability (magnitude) of the gene-variants to serve as a proxy of the biomarker (Lp-PLA2 activity) or the drug (Lp-PLA2 inhibitor). A dose of 160 mg daily of darapladib, the dose suggested as the most appropriate to be used in trials evaluating clinical events, exhibited a relative reduction of Lp-PLA2 activity of 59% after 12 months and 66% after 4 weeks.[131,132] In contrast, the variant rs1051931 that exhibited the largest effect on Lp-PLA2 activity only showed a modest relative difference of 7%, albeit likely to be lifelong. Assuming the effect of Lp-PLA2 inhibitor on the Lp-PLA2 activity is linear and dose-dependent as suggested by the available evidence from trials,[131,132] this would suggest that the variant rs1051931 should be equivalent to a dose of around 17 mg of darapladib, almost ten times lower than the dose of darapladib intended to be used in trials evaluating clinical events. Although somewhat crude and simple, this calculation serves to illustrate that perhaps the lack of association observed with the gene-variants, especially with CHD risk, could be due to the limited ability of common variants within European descent individuals to behave as good proxies of both Lp-PLA2 levels and the Lp-PLA2 inhibitor. This restrained the power of the sample size included (8412 CHD cases) to detect a genetic signal on disease risk. Another alternative would be to use as a proxy the *PLA2G7* null variant found in individuals of Japanese or Chinese ancestry. This loss-of-function V279F variant occurs at polymorphic frequencies, with heterozygote frequencies of 27% and 4% homozygotes and is inherited in an autosomal recessive manner. Enzyme activity is significantly impaired in heterozygotes and is completely abrogated in homozygotes [136,150] This mutation has been associated with a significantly greater incidence of nonfamilial hypertrophic cardiomyopathy,[151] while the results from a study of almost 4,000 Japanese concluded that the variant occurred at a much higher frequency in those suffering from a myocardial or cerebral infarct.[152] Thus the consequence of natural Lp-PLA2 inhibition is suggestive of a detrimental effect on cardiovascular outcome.

In conclusion, I have shown a gradual increase in risk of CHD was observed across all the levels of Lp-PLA2 activity, though residual confounding may explain a considerable part of the association. A genetic approach using the Lp-PLA2 gene was limited by the inability of common variants in European descent to exhibit important differences on the Lp-PLA2 activity levels, limiting substantially the power of this nevertheless large collaboration. Future research on individuals of European ancestry should aim to investigate the existence of rare variants with greater effect on Lp-PLA2 activity and on cardiovascular risk, although this may require several tens of thousands of cases. This may be possible through the Lp-PLA2 Studies Consortium which is compiling information on 31 cohorts with nearly 75,000 participants.[153]

Figure 6.8 Relative odds of CHD associated with variants of the *PLA2G7* gene. Data pooled from up to 8 studies including up to 8412 CHD cases.

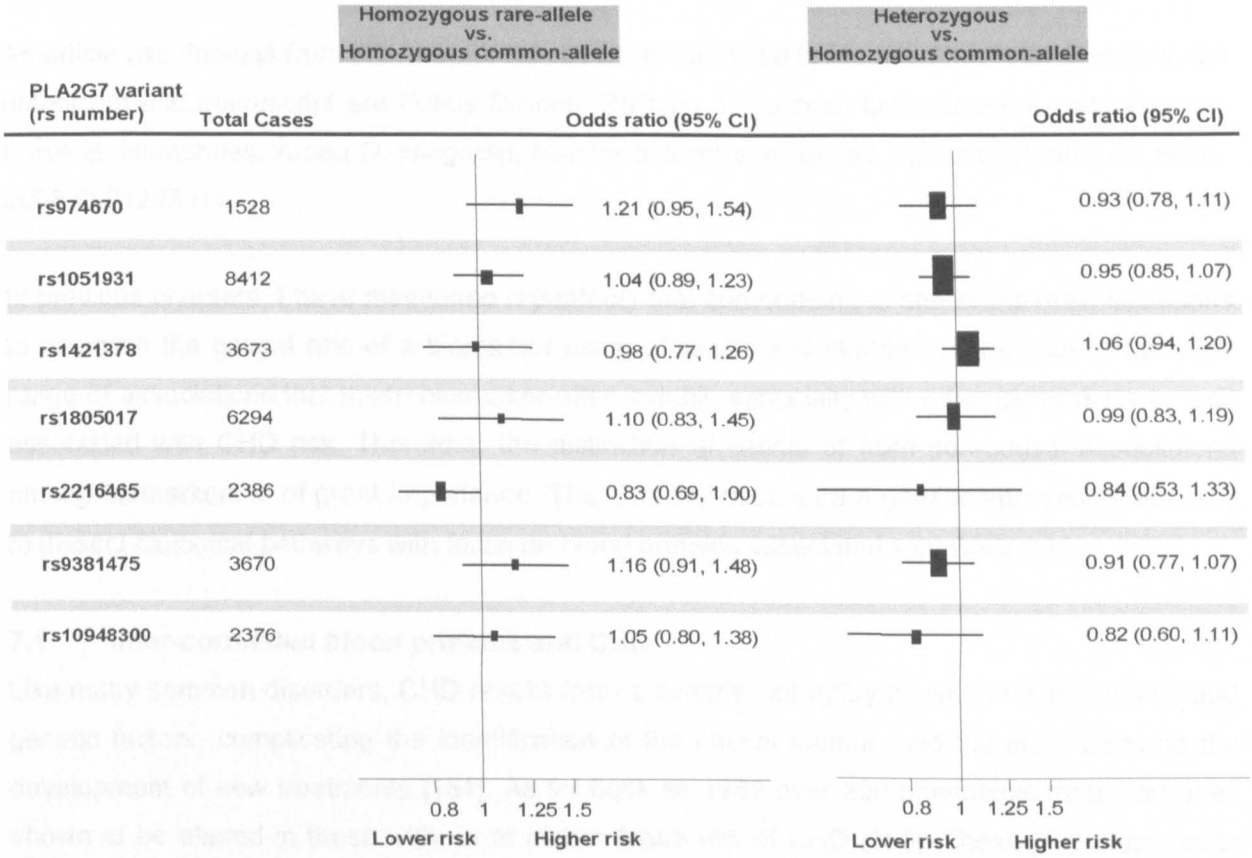
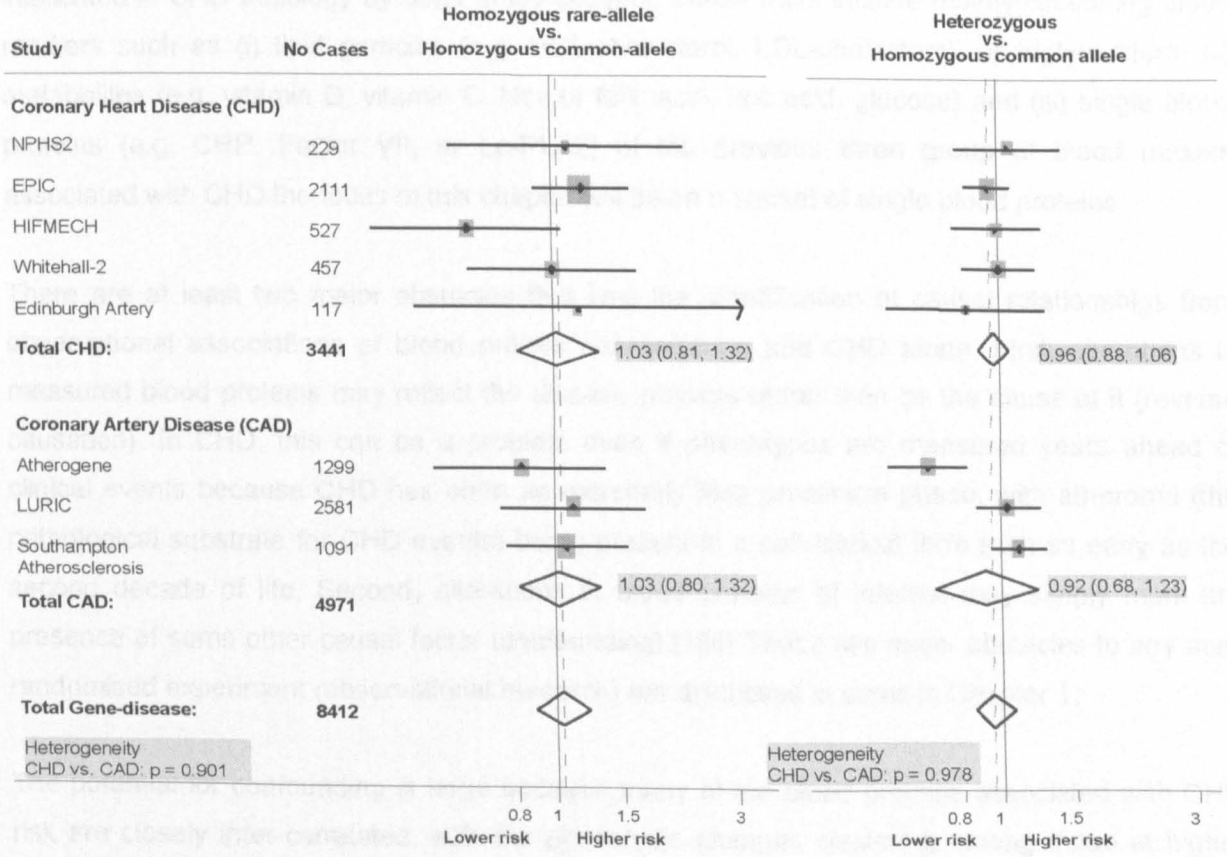


Figure 6.9 Odds ratio (95%CI) of the rs1051931 variant on the risk of CHD stratified by outcome.



Chapter 7 Integrating information on genotype and blood proteins to evaluate causal biological pathways in cardiovascular disease

An article use findings from this chapter has been re-submitted to Human Molecular Genetics, the authors of this manuscript are Fotios Drenos, Philippa J. Talmud, Liam Smeeth, Jutta Palmen, Steve E. Humphries, Aroon D. Hingorani (Senior author) and myself. Manuscript number: HMG-2008-D-01203.R1

In previous chapters, I have mentioned repeatedly that confounding is one of the main limitations to establish the causal role of a biomarker using observational studies. This is due to the wide range of associations that these biomarker often exhibit, especially with other biomarkers as well associated with CHD risk. Therefore, the distinction of causal of from non-causal associations among biomarkers is of great importance. This Chapter described a genetic approach that serve to dissect biological pathways with focus on blood proteins associated with CHD risk.

7.1 Inter-correlated blood proteins and CHD

Like many common disorders, CHD results from a complex interplay between environmental and genetic factors, complicating the identification of the causal factors, and therefore delaying the development of new treatments [154]. As far back as 1981 over 200 phenotypic traits had been shown to be altered in those with, or at higher future risk of CHD [155]. These phenotypic traits represent components of pathways or process (e.g. inflammation, coagulation) suggested to be implicated in CHD aetiology by other study designs. Those traits include mainly circulating blood markers such as (i) lipid particles (e.g. HDL-cholesterol, LDL-cholesterol, or triglycerides), (ii) metabolites (e.g. vitamin D, vitamin C, Hcy or folic acid, uric acid, glucose) and (iii) single blood proteins (e.g. CRP, Factor VII, or Lp-PLA2) of the previous three group of blood makers associated with CHD the focus of this chapter will be on a subset of single blood proteins.

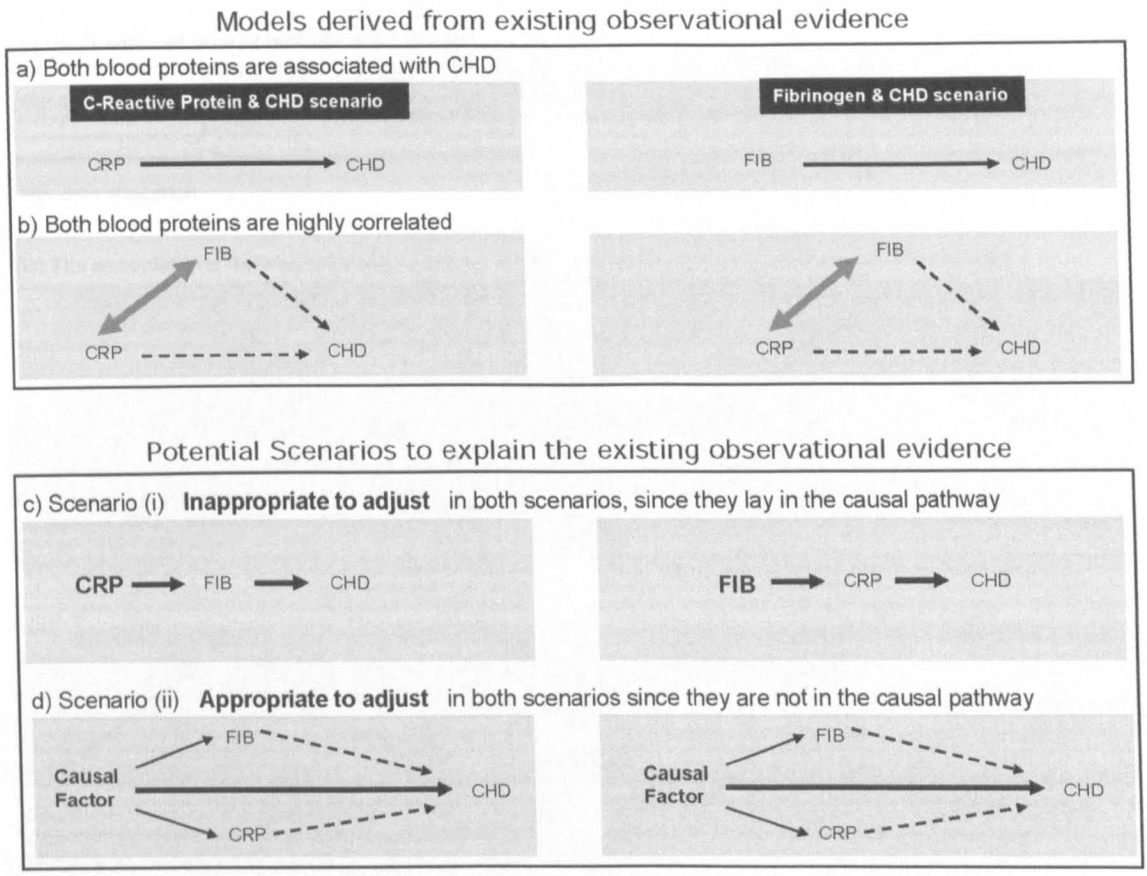
There are at least two major obstacles that limit the identification of causal relationships from observational associations of blood protein concentration and CHD alone. First, alterations in measured blood proteins may reflect the disease process rather than be the cause of it (reverse causation). In CHD, this can be a problem even if phenotypes are measured years ahead of clinical events because CHD has often an extremely long preclinical phase, with atheroma (the pathological substrate for CHD events) being present in a sub-clinical form from as early as the second decade of life. Second, alterations in blood proteins of interest may simply mark the presence of some other causal factor (confounding).[156] These two major obstacles to any non-randomised experiment (observational research) are discussed in detail in Chapter 1.

The potential for confounding is large because many of the blood proteins associated with CHD risk are closely inter-correlated, with the phenotypic changes clustering among those at higher risk of disease. This makes it challenging both to ascertain the nature and direction of the

biological inter-relationships among blood proteins and to assess the independent effect of any blood protein on disease risk. A relevant example that illustrates this comes from the blood proteins CRP and fibrinogen. Both proteins are associated with high risk of CHD and there is strong correlation between them.[12] Currently, the evidence to decide whether these markers confound the associations of each other with CHD, or on the other hand, whether they lie on the same causal pathway is derived from non-randomised observations, (mainly laboratory research in animals, cell and tissues), which are prone to considerable amount of bias. However the issue is critical because in the first scenario adjustment would be legitimate (indeed necessary) where as in the second it would be inappropriate.[157] This uncertainty introduces a large degree of subjectivity when deciding on the causal nature of an association between blood proteins and disease risk (Figure 7.1), and on the validity, or otherwise, of these markers as therapeutic targets.

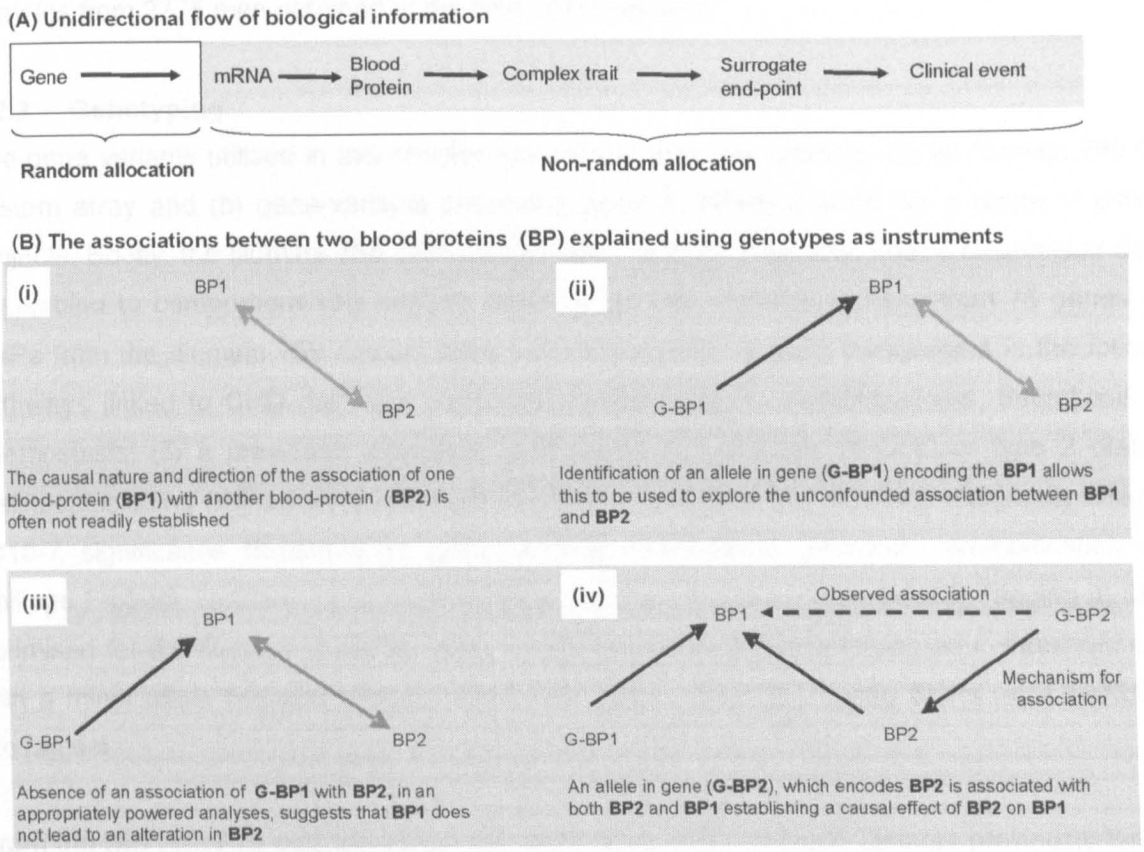
The ideal method to establish unequivocally a causal role in CHD for a particular blood protein of interest is through the use of an intervention that specifically modifies such blood protein in a randomised experiment. A similar approach has been conducted for the blood marker LDL-cholesterol that has proved to be causally involved on CHD, largely part by developing the HMG-CoA reductase inhibitor (statin) class of drugs that reduce blood levels of LDL-cholesterol, and testing their effects in randomised trials [158]. However, the cost and risk associated with the development of the many selective drugs needed to target the vast array of blood proteins implicated in CHD currently precludes a comprehensive and systematic approach to understand their causal relevance of all markers and limits translation of the basic science and epidemiological findings into new medications.

Figure 7.1. Difficulties in establishing causal pathways among correlated blood proteins and CHD risk using as an illustrative example the case CRP, and Fibrinogen (FIB) in CHD.



Many (if not all) of the blood proteins associated with CHD behave as heritable quantitative traits, and single nucleotide polymorphisms (SNPs) influencing their variance are now being identified both by candidate gene and genome wide studies [159,160]. Unlike the associations between blood proteins, or the association of blood proteins with CHD risk, genetic associations are protected from reverse causation because genotype is a fixed characteristic such that there is a unidirectional flow of information from common genome variation to mRNA to protein to complex phenotype and disease (see Figure 7.2). Moreover, since genotype is determined at random at conception, blood proteins residing off the causal pathway from SNP to disease should be balanced evenly among the different genotypic groups, as they are in a randomised clinical trial, thereby minimising confounding in genetic associations (see Figure 1.1). Therefore, SNPs that affect the variance in blood proteins linked to CHD could be used as unbiased tools with which to understand their aetiological relevance. This principle has been referred to as Mendelian randomisation and when applied to assess the potential causal role of blood proteins, genetic studies are utilised as a type of natural randomised trial.[22,23,120]

Figure 7.2. Exploiting unique properties of genetic associations to map causal relationships among blood proteins (BP).



In this chapter, I describe how to integrate and overlaying genetic information can be used to disentangle causal pathways among blood proteins by using SNPs (*cis*-acting) in the gene encoding the blood protein of interest (pQTSNPs: protein quantitative trait SNPs). The integration of SNPs and blood proteins information in a large prospective study can be used to help address the problem of identifying causal links between correlated blood proteins and between blood proteins and CHD risk. For this I utilised the available gene-centric data generated in the NPHS-2 study (This study has been described previously in methods section in Chapters 4 and 5).

7.2 Methods

7.2.1 Study design and phenotypic measures

As described in methods section in Chapters 4 and 5, the Northwick Park Heart Study 2 (NPHS-2) is a prospective study of 3012 healthy middle-aged men aged 50–64 years at recruitment, sampled from nine UK general practices between 1989 and 1994 [99]. Men were free from disease at the time of recruitment, and information on lifestyle habits, height, weight, blood pressure were recorded at baseline and on subsequent prospective follow-up. Measures were made of at least 15 circulating blood factors associated with CHD risk,[12,129,161,162] including six single blood proteins (CRP, fibrinogen, Lp-PLA2, apolipoprotein-A1, apolipoprotein-B, and

Factor-VII), seven complex biomarkers (Total-cholesterol, HDL, LDL, triglycerides, BMI, systolic and diastolic blood pressure) and two metabolites (homocysteine and folate). All the measures were obtained before the development of clinical events. A DNA repository was established using samples from 2775 men obtained at the time of recruitment.

7.2.2 Genotyping

The gene variants utilised in this chapter are derived from two sources; (a) an Illumina 768 SNP custom array and (b) gene-variants previously typed in NPHS-2 study for a range of different projects. Briefly, the Illumina 768 SNP custom array use gene-variants from a customised design assembled to comprehensively capture common genetic variation in more than 76 genes. The SNPs from the Illumina 768 custom array were chosen for: (i) their involvement in the following pathways linked to CHD risk; lipid metabolism, inflammation, oxidative stress, thrombosis and haemostasis; (ii) a previously described association with the risk of CHD or type 2 diabetes mellitus, including SNPs from a recent whole-genome analysis of T2D which marginally failed the $p<10\text{-}7$ significance threshold for genome wide significance (personal communication, Prof Philippe Froguel, London); or (iii) their ability to tag copy number variation.[163] Tagging (t) SNPs, optimised for the Illumina platform, were selected using HapMap, applying an r^2 threshold of 0.8 with a minor allele frequency threshold of 0.04. Where possible, coding SNPs were included in the tag set.

From the two previous sources of genetic variants (Illumina 768 and variants previously typed in NPHS-2), in this chapter I only utilised the gene-variants within genes that encode for each of the six blood proteins available to the NPHS-2 study (CRP, fibrinogen, Lp-PLA2, apoliporotein-A1, apolipoprotein-B and Factor-VII). Overall a total 124 gene variants were used as follows: 4 SNPs from *CRP* gene, 18 SNPs from Fibrinogen alpha (*FGA*), beta (*FGB*) and gamma (*FGG*) chains, 13 SNPs from *PLA2G7* gene (cognate protein: Lp-PLA2), 36 SNPs from the gene cluster *ApoA5-A4-C3-A1*, 42 SNPs from *APOB* gene and 11 SNPs from *F7* gene. See Table 7.1 for specific details of gene variants.

Table 7.1. List of gene-variants and their rs numbers from genes that encodes blood proteins (in grey cells) measured in NPHS-2.

rs number	Gene
Apolipoprotein B	
rs693	*APOB*
Ins/del[3]	*APOB*
rs10199768	*APOB*
rs1042031	*APOB*
rs11126598	*APOB*
rs11676704	*APOB*
rs12691202	*APOB*
rs12713450	*APOB*
rs12713540	*APOB*
rs12713554	*APOB*
rs12713559	*APOB*
rs12713675	*APOB*
rs12713843	*APOB*
rs12714097	*APOB*
rs12714214	*APOB*
rs12714225	*APOB*
rs12714264	*APOB*
rs12720791	*APOB*
rs12720801	*APOB*
rs12720843	*APOB*
rs12720848	*APOB*
rs12720854	*APOB*
rs13306187	*APOB*
rs1367117	*APOB*
rs1713223	*APOB*
rs17398765	*APOB*
rs1801700	*APOB*
rs1801701	*APOB*
rs1801702	*APOB*
rs1801703	*APOB*
rs2163204	*APOB*
rs2854725	*APOB*
rs520354	*APOB*
rs533617	*APOB*
rs550619	*APOB*
rs584542	*APOB*
rs585967	*APOB*
rs6752026	*APOB*
rs676210	*APOB*
rs679899	*APOB*
rs934197	*APOB*
rs952275	*APOB*
Apoliprotein AI	
rs670	*APOA1*
rs11216157	*ApoA5-A4-C3-A1*
rs11216159	*ApoA5-A4-C3-A1*
rs12284864	*ApoA5-A4-C3-A1*
rs12285095	*ApoA5-A4-C3-A1*
rs12365440	*ApoA5-A4-C3-A1*
rs12365462	*ApoA5-A4-C3-A1*
rs1263173	*ApoA5-A4-C3-A1*
rs12718464	*ApoA5-A4-C3-A1*
rs1729410	*ApoA5-A4-C3-A1*
rs2071521	*ApoA5-A4-C3-A1*
rs2071523	*ApoA5-A4-C3-A1*
rs2075291	*ApoA5-A4-C3-A1*
rs2075294	*ApoA5-A4-C3-A1*
rs2098453	*ApoA5-A4-C3-A1*
rs2216311	*ApoA5-A4-C3-A1*
rs2727789	*ApoA5-A4-C3-A1*
rs2849174	*ApoA5-A4-C3-A1*
rs33989105	*ApoA5-A4-C3-A1*
rs35625356	*ApoA5-A4-C3-A1*
rs35631472	*ApoA5-A4-C3-A1*
rs35833621	*ApoA5-A4-C3-A1*
rs4417316	*ApoA5-A4-C3-A1*
rs5072	*ApoA5-A4-C3-A1*
rs5101	*ApoA5-A4-C3-A1*
rs5142	*ApoA5-A4-C3-A1*
rs579163	*ApoA5-A4-C3-A1*
rs595049	*ApoA5-A4-C3-A1*
rs601634	*ApoA5-A4-C3-A1*
rs603446	*ApoA5-A4-C3-A1*
rs618354	*ApoA5-A4-C3-A1*
rs618923	*ApoA5-A4-C3-A1*
rs6589566	*ApoA5-A4-C3-A1*
rs6589567	*ApoA5-A4-C3-A1*
rs6589568	*ApoA5-A4-C3-A1*
rs9804646	*ApoA5-A4-C3-A1*
Factor VII	
-670A>C	*F7*
rs6046	*F7*
-402G>A	*F7*
In/del	*F7*
rs1475931	*F7*
rs3093248	*F7*
rs3211719	*F7*
rs510317	*F7*
rs555212	*F7*
rs574229	*F7*
rs6046	*F7*
Fibrinogen	
rs1800790	*FGB*
rs1800791	*FGB-G*
rs1118823	*FGA-B-G*
rs1800788	*FGA-B-G*
rs1800792	*FGA-B-G*
rs2066860	*FGA-B-G*
rs2066870	*FGA-B-G*
rs2070016	*FGA-B-G*
rs2070018	*FGA-B-G*
rs2070022	*FGA-B-G*
rs2070025	*FGA-B-G*
rs4220	*FGA-B-G*
rs4463047	*FGA-B-G*
rs4508864	*FGA-B-G*
rs6054	*FGA-B-G*
rs7659024	*FGA-B-G*
rs7659613	*FGA-B-G*
rs7673587	*FGA-B-G*
Lp-PLA2	
rs1051931	*PLA2G7*
rs10948300	*PLA2G7*
rs12195701	*PLA2G7*
rs12528807	*PLA2G7*
rs1421368	*PLA2G7*
rs1421378	*PLA2G7*
rs16874962	*PLA2G7*
rs17288905	*PLA2G7*
rs1805018	*PLA2G7*
rs2216465	*PLA2G7*
rs4498351	*PLA2G7*
rs6929105	*PLA2G7*
rs974670	*PLA2G7*
C-reactive protein	
rs1130864	*CRP*
rs1205	*CRP*
rs3093077	*CRP*
rs3091244	*CRP*

7.2.3 Statistical analysis

In order to illustrate the degree of clustering among blood proteins associated with CHD risk, pair-wise linear relationships between blood phenotypes were tested using Pearson correlation. Initially, all continuous phenotypes available in NPHS-2 and then restricted to blood proteins. In a second stage, the effect of all the *cis*-variants on their respective cognate proteins (e.g. *CRP* gene-variants on CRP levels) was evaluated in order to detect the most strongly associated SNP (judged as the variant with the lowest p-value) that could serve as unconfounded proxy for the cognate protein. Then, in order to investigate the nature of the associations between blood proteins the effect of the selected SNP on the blood proteins was obtained using a standardised scale in order to allow direct comparability across blood proteins. For all the genetic associations a model free approach was used. If the blood protein was measured once in the study, an ANOVA was used to test for its association with genotype. Where more than one measurement was available for the blood protein, a mixed model was fitted through the data using indicator variables. The regression coefficients obtained for the indicator variables compared to the reference category were assessed for significance using a Wald test.

7.3 Results

7.3.1 Between-phenotype correlations

To delineate the potential for confounding in the associations of biomarkers (including blood proteins) with CHD, the correlations between-phenotype was assessed (Figure 7.3). Seventy-two of the 105 possible pair-wise correlations between phenotypes were significant with a p value $\leq$ 0.05 (68 traits significant with a p value $\leq$ 0.01), which significantly exceeds the five expected by chance alone ($p < 1x10^{-8}$ for observed versus the expected). This illustrates the clustering of altered phenotypes among individuals at higher risk of CHD, which makes it challenging for observational studies alone to identify the independent causal effect of any one phenotype. When the analysis was restricted to blood proteins a similar scenario emerged with 12 out of the 15 pair-wise correlations being statistically significant (Figure 7.4). In some cases the magnitude of the correlation was large as that observed between CRP and fibrinogen ($r= 0.43$; $p<0.001$). Notably, some of the pair-wise correlations were between blood proteins from nominally different pathways, e.g. correlations were identified between markers of lipids, inflammation and coagulation, consistent with the view that some of these pathways may be causally linked.

Figure 7.3. Correlation matrix between multiple phenotypes linked to CHD in 2775 men from the NPHS-2 study. Values in cells indicate Pearson's correlation co-efficient r. * p<0.01, ** p<0.001. Baseline and five repeat measures were available for cholesterol, triglycerides, factor VII (FVII), fibrinogen (FIB), blood pressure and body mass index and single measures for the remaining traits.

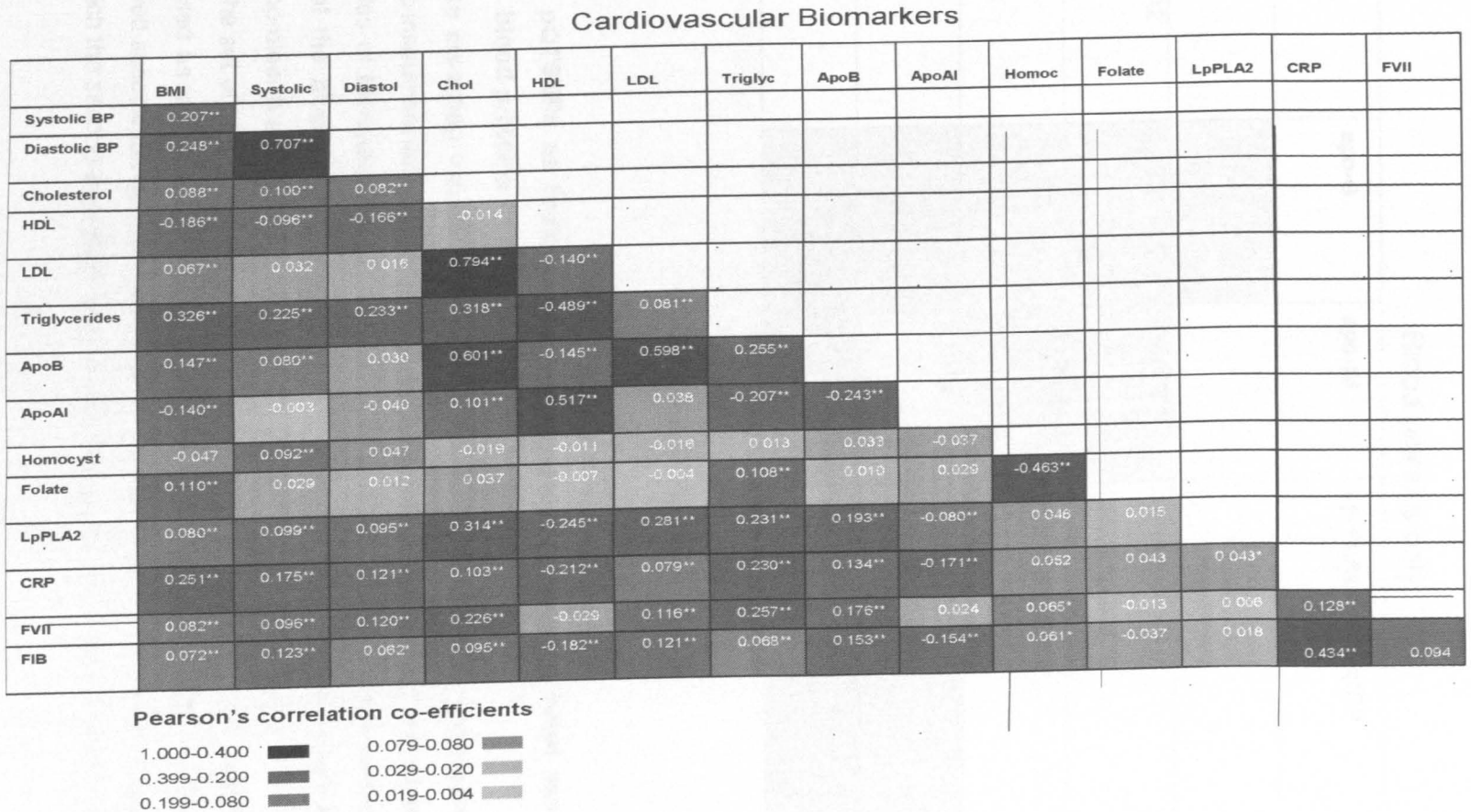


Figure 7.4. Correlations matrix between blood proteins linked to CHD in 2775 men from the NPHS-2 study. Values in cells indicate Pearson's correlation co-efficient r. * $p<0.01$, ** $p<0.001$. The intensity of the colour within the cells reflects the degree of correlation evaluated by using Pearson's correlation coefficient.

Blood proteins only

	apo-B	apo-AI	Lp-PLA2	CRP	FVII
apo-AI	-0.2433**				
Lp-PLA2	0.193**	-0.0802**			
CRP	0.1344**	-0.1707**	0.0433*		
FVII	0.1755**	0.0238	0.0058	0.1277**	
FIB	0.1528**	-0.1538**	0.0178	0.434**	0.0942

7.3.2 pQTSNPs as tools to distinguish causal from non-causal associations between blood proteins

Because *cis*-acting variants in a gene encoding a blood protein (pQTSNPs) provide *the* most specific instrument with which to investigate the causal effects of the encoded protein (utilising the principles of Mendelian randomisation), I identified the most strongly associating pQTSNPs for each of the blood protein available in NPHS-2 (CRP, fibrinogen, Lp-PLA2, apoliporotein-A1, apolipoprotein-B and Factor-VII), results are described in the Table 7.2.

From the set of pQTSNPs identified for each blood protein (Table 7.2), only the top-rank SNP was then used as unconfounded proxy to help distinguish causation from confounding in the directly observed association between these blood proteins, with the exception of apolipoprotein-B protein for which the second-best SNP was used due to low minor-allele frequency of the top-SNP.

Table 7.2. Associations of pQTSNPs with their respective cognate proteins. Only associations with a p value < 0.05 are presented. In *bold* the pQTSNP used as unconfounded instrument.

Cognate protein	Gene	rs number	P-value
Apo-B	APOB	rs585967	0.000853
	APOB	***rs952275***	0.000914
	APOB	rs533617	0.005622
	APOB	rs12714264	0.006086
	APOB	rs10199768	0.025601
Apo-AI	APOA5-A4-C3-A1	***rs1729410***	0.000758
	APOA5-A4-C3-A1	rs4417316	0.005714
	APOA5-A4-C3-A1	rs603446	0.009762
	APOA5-A4-C3-A1	rs12285095	0.015973
	APOA5-A4-C3-A1	rs9804646	0.023318
CRP	CRP	***rs1205***	0.00009559
	CRP	rs3091244	0.00021787
	CRP	rs3093077	0.00027444
	CRP	rs1130864	0.00158698
Lp-PLA2	PLA2G7	***rs2216465***	0.025633
	PLA2G7	rs1421378	0.03327
FVII	F7	***rs6046***	1.52E-180
	F7	rs555212	7.89E-36
	F7	rs3211719	1.04E-03
	F7	rs1475931	3.24E-03
Fibrinogen	FGA-B-G	***rs4508864***	1.00E-07
	FGA-B-G	rs1800790	1.68E-07
	FGA-B-G	rs4220	9.41E-07
	FGA-B-G	rs2070016	1.05E-05
	FGA-B-G	rs6054	9.64E-04

The Figure 7.5 a-f showed the effects (standardised mean difference) of each of the pQTSNP used as unconfounded instrument on the different blood protein phenotypes available in NPHS-2. As expected the association between each pQTSNP-and its cognate protein was strong, however none pQTSNPs was associated with any of the other blood proteins measured, despite 12 of the 15 pair-wise correlations between the blood proteins themselves being highly significant (Figure 7.4).

7.4 Discussion

Utilising the principles of Mendelian randomisation, the results from the present analysis demonstrated that integrating evidence from *cis*-acting SNPs in a gene with their respective cognate proteins is an effective approach to distinguish causal effects from non-causal associations among blood proteins, and therefore serve to reconstruct biological networks of blood proteins of interest for CHD aetiology.

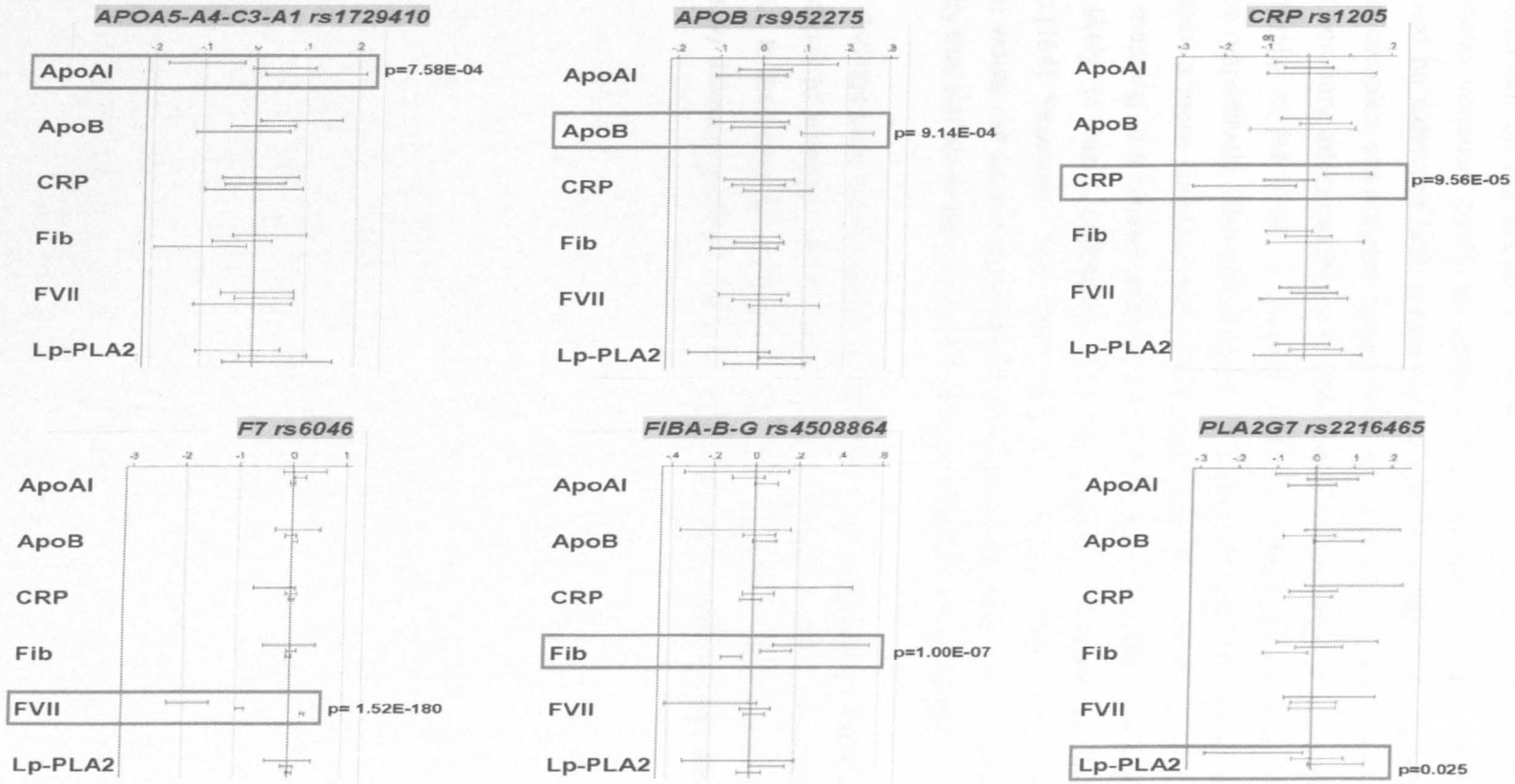
The efficacy of this genetic approach was exemplified when applied to a limited set of six blood proteins available in NPHS-2 for which the observational evidence indicates the existence of an association with CHD risk for each blood protein, but their causal role and validity as therapeutic target remains unproven due to problems of reverse causation and confounding. Twelve out 15 pair-wise comparison among the six blood proteins exhibited some degree of correlation (Figure 7.4), in some cases as strong as that observed between CRP and fibrinogen ($r=0.43$; $P<0.001$). However, when using *cis*-acting variants as highly specific and unconfounded instruments, it was not possible to replicate any of the 12 between blood protein associations (Figure 7.5). Nonetheless, all the genetic instruments proved to have a strong effect on their respective cognate protein for which they intended to be a proxy. This discordance provides evidence that perhaps most of the 12 between-protein associations identified were more likely to be explained by a shared association with an unmeasured factor than by a direct causal effect of one protein on the level of another (scenario (ii), Figure 7.1).

Some immediate translational application of this approach is in deciding which variables (blood proteins) adjust for when estimating the causal effect of a blood protein of interest. For example, CRP showed a strong correlation with fibrinogen ($r=0.43$). This association could arise because of a causal relationship (in either direction) or because of a common association of both these blood proteins with another factor (see Figure 7.1). This makes it difficult to assess whether CRP and fibrinogen lie on the same causal pathway to CHD, and makes it difficult to know whether it is legitimate to make statistical adjustment for fibrinogen in the many observational associations of CRP with CHD (and *vice versa*). For example, statistical adjustment, which is the orthodox approach to dealing with confounding in observational epidemiology, would be appropriate only if the association of CRP with CHD was not mediated (even in part) by an elevation in the level of fibrinogen. The absence of an association of *CRP* pQTSNPs with fibrinogen level (or *FIBA-B-G* with CRP level) (Figure 7.5), which is consistent with the findings from Chapters 4 (Table 4.3) and 5 (Figure 5.5) and other studies of these variants [108] indicates that CRP and fibrinogen do not exhibit a direct causal association with one another and that statistical adjustment is appropriate in observational studies of these biomarkers with CHD.

Some of the limitations and advantages of the approach described here deserved to be discussed. First, in theory, null associations observed using this pQTSNPs-blood protein integrative approach might be explained by insufficient statistical power to detect downstream effect. Although, that could be the case in the current study that includes a few thousand subjects, empirical evidence from this thesis as well as other studies confirms a null association for CRP (see Figure 5.5) and fibrinogen SNPs with multiple other phenotypic traits [108]. This limitation, however, could be overcome either by increasing the sample size or by using more potent genetic instrument. Another limitation to be consider is the presence of an effect modification in between blood protein association by a third factor (blood protein or complex trait) that may be in unnoticed to the investigators and therefore lead to exclude wrongly the absence of a causal effect. Such as the scenario described for the MTHFR/C677T-Hcy association by the folate intake (see Figure 3.1).

It is notable that most of the associations observed between the pQTSNPs and their cognate protein were highly significant despite the modest size of the data set. P-values were in the range of $x10^{-4}$ to $x10^{-180}$. There are several potential reasons for this. Variation in the level of circulating blood proteins is a much more proximal consequence of functional genetic variation (distal only to mRNA expression in the case of blood proteins) and so the signal to noise ratio in these associations is likely to be high. As a trait, the level of a blood protein is also likely to be influenced by a narrower range of genes and environmental factors than more distal biomarkers of CHD.

Figure 7.5. Mean values on a standardised scale and 95% confidence intervals for each genotypic class for the most strongly associating SNP in the pQTGs for: (a) apo-AI. (b) apo-B (b), (c) CRP , (d) factor VII, (e) fibrinogen, and (f) Lp-PLA2 each of which was used as unconfounded proxy for the cognate protein.



The approach described in this chapter can also be extended to include non-protein biomarkers. Whereas *cis*-acting SNPs in the encoding gene offer the most specific genetic instrument with which to assess the causal role of a protein, there may not be a single best instrument for a non-protein trait. Instead, SNPs from a range of genes (identified by gene centric or whole genome analysis) may be necessary. Although the approach described here took a gene-centric approach and focussed common alleles and a relatively narrow range of blood phenotypes, there is no theoretical constraint on the scope or breadth of the approach. For example, information on common genetic variation could be extended to incorporate rare SNPs. The phenotypes evaluated could be extended both proximally to include mRNA expression and more distally to include more complex phenotypes linked to CHD such as carotid-intima media thickness or measures of coronary artery calcification. The scope and breadth of the approach could also be extended laterally at the level of protein and non-protein phenotypes using proteomic and metabolomics respectively. The identification of SNP affecting the expression of an mRNA, its cognate protein, a more complex downstream phenotype and the risk of a clinical event would result in the mapping of a causal pathway, even though each element in the trail from protein to disease is likely to have additional, and increasingly complex, genetic and non-genetic determinants.[164] However, The SNP with its unique properties (random allocation and unmodifiable) would act as the dissecting tool to generate evidence on biological pathways with high credibility that will serve for a more efficient generation of drug-target.

In summary, findings from this Chapter demonstrate how integrating information on genotype and blood phenotypes in humans, can be used to construct association networks with a high-level of credibility, due to the particular properties of genetic variants, which are randomly allocated and unmodifiable by disease process, features which are not shared by any other natural biological exposure

Chapter 8. Discussion and conclusions

Herein the findings of this thesis are summarised. I conclude by considering this work and the broader perspective of Mendelian randomisation in general as an inferential instrument to distinguish causal from non-causal associations.

8.1 Homocysteine and stroke

Regarding the role of Hcy in stroke, the most comprehensive synthesis of genetic evidence generated to date and reported in Chapters 2 and 3 indicated that the *MTHFR* genetic signal on both Hcy and stroke is more pronounced in areas associated with a low folate intake at population level, such as Asia.

Nonetheless, since it has been reported, there has been a tendency for genetic studies conducted in Asia, and in particular in China, to systematically report stronger genetic signals for a wide variety of associations. [165] Therefore, there is the possibility that part of the findings on *MTHFR*-Hcy-Stroke could suffer from such overestimation, attributed to small-study bias.[165] Importantly, when the evidence was limited to large studies (defined as: MTHFR-Hcy ≥500 individuals and *MTHFR*-stroke ≥400 stroke events), the findings of a differential effect of the *MTHFR* gene on Hcy and stroke by region of folate supplementation remained largely unaltered. In addition, the only study with more than 1000 cases included in the analysis was conducted in Asia.[166] Although, far from definitive, this provides certain optimisms in the pronounced genetic effect observed in Asian studies.

Despite being a consistent finding, the presence of a modifying effect of *MTHFR*-Hcy, and *MTHFR*-stroke by geographical region acting as proxy of the folate consumption, the use of this proxy encompass considerable imprecision. In support of the use of geographical regions as a proxy of folate intake, were the meta-regression analyses which indicated that the difference in Hcy levels by *MTHFR* genotype is reduced as the concentrations of folic acid increase. The number of studies that participate in such analyses was however limited, it was not possible to reproduce a similar analysis for the *MTHFR*-stroke association due to the absence of necessary information. Nonetheless, the ideal approach to evaluate the presence of a gene-environment interaction will require the uses of individual participant data from a single study, large enough to provide sufficient cases, or from an international collaboration involving multiples studies.

Figure 8.1. Uses of Mendelian randomisation. Underlined are the three particular uses of Mendelian randomisation that were employed in this document.

Uses of Mendelian randomisation

1. **Proxy of external exposures**
   e.g. alcohol, heavy-metal Poisoning, life-time risk of infection.

2. **Proxy to evaluate trans-generational effects**
   e.g. early development of chronic diseases.

3. **Proxy of internal biomarkers to provide evidence on causality**

   3.a. Gene -> Direct intermediate phenotype: *CRP*->CRP plasma->CHD

   3.b. Gene -> Indirect intermediate phenotype: *MTHFR*->Hcy levels->Stroke

4. **Proxies of drug interventions:**

	Instrument	Molecule target	Outcome	Potency
Gene:	Lp-PLA2 gene	Lp-PLA2 enzyme	CHD	Small
Drug:	Lp-PLA2 inhibitor	Lp-PLA2 enzyme	CHD	Large

5. **Dissection of biological pathways**

Although the genetic evidence is encouraging regarding a possible causal effect of Hcy in stroke, particularly in populations with low levels of folic acid consumption, genetic studies do not provide evidence on the reversibility of the risk that is associated with the exposure. Such evidence when possible, should come from randomised interventions. Evidence published to date on trials using Hyc-lowering interventions is almost entirely based on Europe, America and Australia, regions for which (if genetic studies are unbiased) the expected benefit for Hcy-lowering interventions is small or negligible. In contrast, no large trials using Hcy-lowering therapies is being conducted in areas with folic acid deficiency, despite genetic evidence (if free from bias) suggesting that the benefit is likely to be greater.

Discussions of the differences between the randomised trials and genetic studies with particular emphasis to the Hcy-stroke example were presented in Chapter 3. A more general discussion of the differences between genetic studies and trials is provided in the section 8.5 of this Chapter.

8.2 CRP and cardiovascular disease

Despite more than a decade of epidemiological studies and perhaps several more from incorporating basic scientific research focussed on CRP, its functions and the nature of its associations with cardiovascular disorders still remains to be established.

The findings from Chapters 4 and 5 confirmed the multiplicity of associations of plasma CRP, not only with established, but also with emerging risk factors in CVD that restrain the utility of classical observational epidemiology to evaluate the nature of its associations.[167] Other strategies to complement the observational evidence would therefore be ideal, see Figure 8.2.

Figure 8.2. Strategies often used to investigate the causal role of a biomarker/exposure in disease development and the applicability of such strategies for the case of CRP. APR: Acute phase reactant. RCT: randomised clinical trial.

Strategies to evaluate the role of CRP in CHD development

Approach	Limitation	CRP
Manipulation in vitro	Isolated system	No known receptor
Genetically modified strain	Extrapolation to human disease	CRP not an APR in mice
Human mendelian disorder		None described
RCT	Requires specific drug	None until recently

As described in Figure 8.2 however, such additional strategies, in the case of CRP, are absent or of poor utility.[115] A genetic approach therefore offers a unique opportunity to evaluate the role of CRP while overcoming the limitations from observational epidemiology. Findings from a Mendelian randomisation applied to plasma CRP, whether using a single-variant or a tagging-haplotype from the *CRP* gene were generally concordant. Apart from the effect on its cognate protein (plasma CRP), no

single association was observed either for the *CRP*-variant +1444/C>T or for the *CRP* tagging-haplotypes. Since CRP is a protein with a gene that encodes for it, the utilisation of this gene, *CRP* (described in Chapter 7 as a pQTG) is therefore *the* most specific instrument that could be used to distinguish causal from non-causal associations of plasma CRP. This CRP example lies in the category of gene-direct intermediate phenotype (Figure 8.1) from the uses of Mendelian randomisation, and contrasts with the example of *MTHFR*-Hcy-stroke which lies in the category of gene-indirect intermediate phenotype.

There are several possibilities that may explain the null effect of the genetic approach used for CRP. Firstly, the selected genetic instrument may not confer a sufficient difference in the concentrations of CRP. However, the effects on CRP concentrations of the *CRP* variant +1444/C>T (0.68 mg/L for TT vs. CC-genotype) and those from a *CRP* tagging-haplotype (42% relative difference for CGG vs. CAT-haplotype) confirmed the adequate selection of tools. These findings have received support from published studies evaluating individuals of European ancestry.[68] Secondly, the existence of a gene-environment interaction, that if not averted, may reduce or abolish the presence of a genetic signal. The observational evidence on CRP does not however support the existence of a gene-environment interaction. This scenario contrasts with that observed for the *MTHFR*-Hcy-Stroke example. Thirdly, a false negative result due to limited power to detect a small but causally important effect of CRP in cardiovascular disease. Results from this document (Chapters 4 and 5), plus the available evidence from several CRP-genetic studies, now including several thousands of CHD events indicates that this is perhaps unlikely.[111,112,114,122,123,168,169] However, the undergoing effort of the CCGC collaboration, which will include over 30,000 CHD events and 110,000 controls using individual patient data will provide a unique opportunity to confirm or refute CRP as valid drug target for CHD.[110]

8.3 Lp-PLA2 and CHD

Findings from the Chapter 6 showed a graded and continuous association of Lp-PLA2 activity with CHD. The existence of multiple associations of Lp-PLA2 activity with cardiovascular risk factors however limits the utility of observational studies to infer a causal role on CHD. Randomised trials using a specific Lp-PLA2 inhibitor (darapladib) are underway and will be the final arbiter to evaluate the reversibility of the risk associated with Lp-PLA2 from observational studies.

However, as way to increase the confidence in the Lp-PLA2 inhibitor, and reduce risk of a possible failure in phase-III trials, I presented how to use a genetic approach to evaluate the validity of a drug target. For this, I utilised the Lp-PLA2 gene (known as *PLA2G7*) to evaluate the utility of the Lp-PLA2 enzyme as a valid therapeutic target. This example lies in the category of proxies of drug intervention from the uses of Mendelian randomisation (Figure 8.1).

This new extension of Mendelian randomisation will only be possible when the intervention and the gene, target the same molecule (numeral 4, Figure 8.1). Since these two instruments, also, share the characteristic of being randomly allocated (drug by trial randomisation process and gene due the Mendel's 2nd law) a "like with like" comparison is generated. However, a main feature that distinguishes these two approaches is the difference in the potency of their instruments (drug vs. gene) to modify the target molecule, Lp-PLA2 enzyme. While a dose of 160 mg of the Lp-PLA2 inhibitor darapladib reduced the Lp-PLA2 activity by 66%, the best of seven common tagging-SNPs of the *PLA2G7* gene (rs1051931) only reduced Lp-PLA2 activity by 7%. Yet, despite the differences in potency, neither the Lp-PLA2 inhibitor nor the genetic variant modified the concentrations of lipid particles. These findings contrast with the associations of Lp-PLA2 activity with lipid particles from observational studies.

This gene-drug approach for Lp-PLA2 nonetheless serves to highlight two practical issues to be considered when conducting this analysis. Firstly, to consider the utility of gene-resequencing for the detection of rare variants with a greater genetic effect that could be used to mimic the effect of the intervention of interest. Secondly, the use of other ethnic groups with genetic variants (not present in European descent individuals) that possess an adequate genetic signal, such as the V279F variant in the *PLA2G7* gene detected in subjects from Chinese and Japanese ancestry.

8.4 Mendelian randomisation and biological pathways

As described throughout this document for the Hcy, CRP, and Lp-PLA2 examples, the judgment of the nature of the multiple associations of these three biomarkers with other established and emerging risk factors is a challenging issue. The current strategy (non-randomised observations mainly from laboratory research in animals, cell and tissues) used to establish the nature of these associations is also prone to important biases (Figure 7.1). The magnitude of this problem, distinguishing causal from non-causal associations, is expected to grow exponentially with the development and refinement of OMICs technology.

Capitalising on the unique properties of allelic variants (randomly allocated, unmodifiable and being the initial point of a unidirectional flow of information; see Figure 7.2), they are ideal instruments to distinguish causal from non-causal associations and therefore dissecting biological pathways that will finally lead to a more effective process of drug-target discovery. To describe a basic proof of principle, I focused on six blood proteins and utilised *the* most specific genetic instruments, *cis*-acting variants in gene encoding the blood protein of interest (pQTSNPs).

Findings from Chapter 7, also provide an immediate translation, which is to inform observational epidemiology if it is or is not appropriate to adjust for blood proteins when evaluating their role in disease development. The topical example of CRP, fibrinogen and CHD was utilised to exemplify this benefit.

The theoretical limitations of this approach are not that different to those already described for Mendelian randomisation experiments (e.g. canalization, interactions, large-sample sizes).[22,23] Whilst, I restricted this approach to blood proteins, it could be applied to non-single protein, such as lipid fractions and metabolites, although the specificity of the gene-variant as a dissecting instrument would not be as high as it was in the pQTSNPs-cognate protein examples presented in Chapter 7.

The breadth and depth of this genetic strategy to map biological pathways in humans will growth exponentially when current observational studies start to obtain information at large-scale on other biological levels, such as mRNA from micro-arrays, protein levels from proteomics or protein activity from metabolomics. It is envisaged that methodological developments will be necessary in order to exploit the maximum potential of this approach, especially when the different levels of biological information are considered to be integrated with one another.

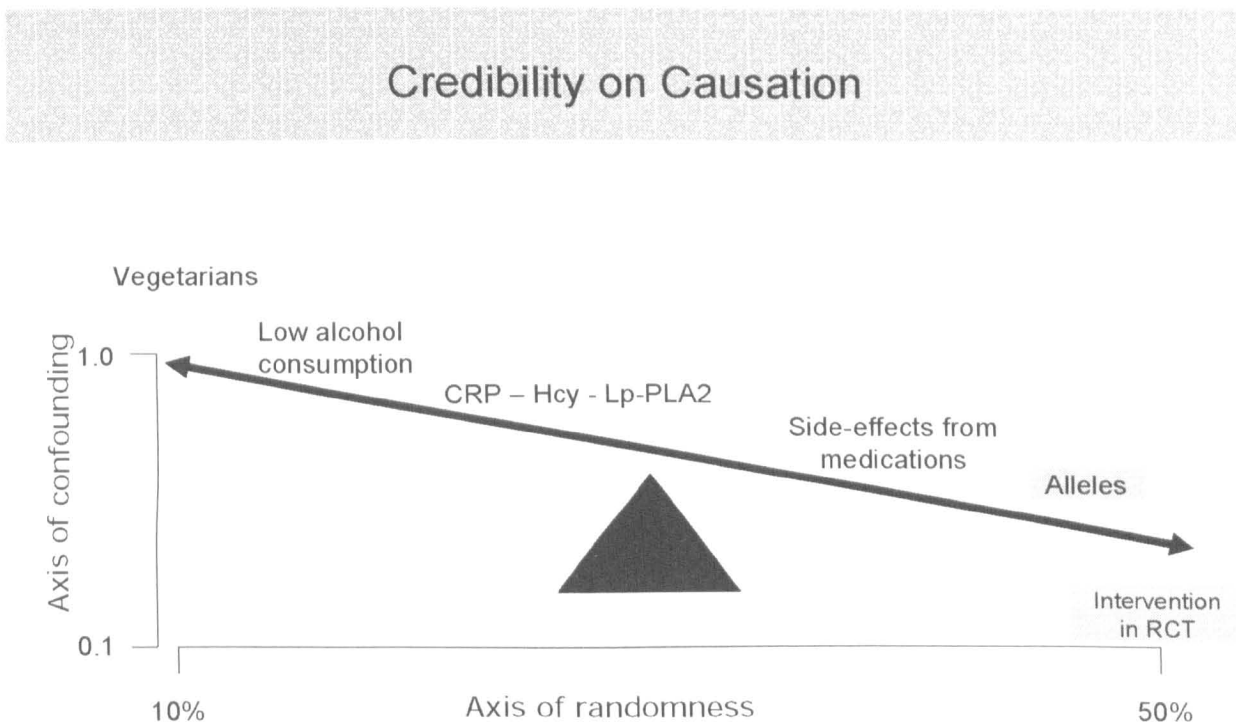
8.5 Mendelian randomisation as an inferential instrument

Finally, I would like to conclude this document with a more general discussion of the Mendelian randomisation approach as an additional inferential instrument to those already in place for obtaining evidence on disease aetiology.

Firstly, although the biomarker-outcome associations utilised as examples in this document may suggest a futility of observational epidemiology in providing evidence on causality, due to reverse causation and residual confounding, I would like to clarify that this will not necessarily be the case for all exposure and diseases as illustrated in Figure 8.3. The location of the exposure under investigation on the axis of randomness will subsequently affect its location on the axis of confounding. An exposure located in the extreme left of the axis of randomness (e.g. being vegetarian) indicates that to be allocated to that exposure (or the probability of acquiring this exposure) is far removed from being a random process, therefore the probability of confounding is extremely high. In consequence, results from an observational study investigating this exposure will have very low levels of credibility. On the contrary, in the extreme right appears an intervention used in randomised trial, a totally random process with very low (or almost absent) probability of confounding to occur. Thus, the credibility of these experiments is of the highest. This simplistic

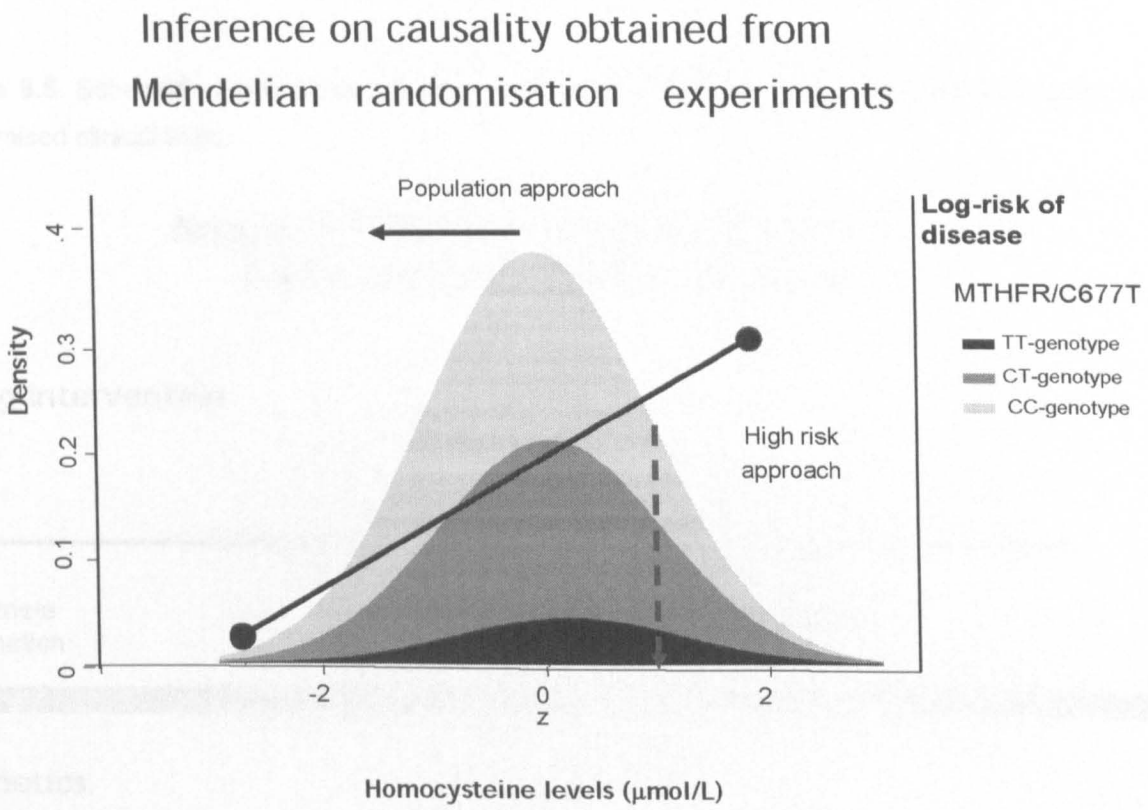
example illustrates the fact that not all exposure-outcome associations using an observational design are precluded as an instrument to infer causation. Thus, Mendelian randomisation should be conceived as an "additional" instrument to infer causation, with certain scenarios (exposure-outcome association) in which its application will have greater benefits than others.

Figure 8.3. Schematic representation of the utility of a genetic approach as an inferential instrument to distinguish causal from non-causal associations.



Secondly, although Mendelian randomisation experiments use the most personalised characteristic (genetic background), the inference on causation achieved by this approach are for all the individuals in a population regardless of their genetic background (Figure 8.4). Therefore, if (hypothetically) Hcy, is proved to be causal using a genetic approach, the benefit will be obtained through the well known strategies for prevention of disease: a population approach, high-risk approach (often with medications) or both in all individuals regardless of their genetic background (Figure 8.4).

Figure 8.4. Schematic representation of the inference obtained from Mendelian randomisation experiments in terms of the strategies for disease prevention. The *x* axis shows the Hcy concentrations in a hypothetical population. The right *y* axis shows the risk of disease (log-scale). The left *y* axis is the probability density. Light grey is the Hcy distribution in a population for subjects with CC-genotype of the *MTHFR*/C677T variant. Dark grey is the distribution of Hcy for subjects heterozygous and in black the Hcy distribution for subjects with TT genotype. This figure is a modification from that originally described by Prof. Geoffrey Rose in "Strategies for Disease Prevention".



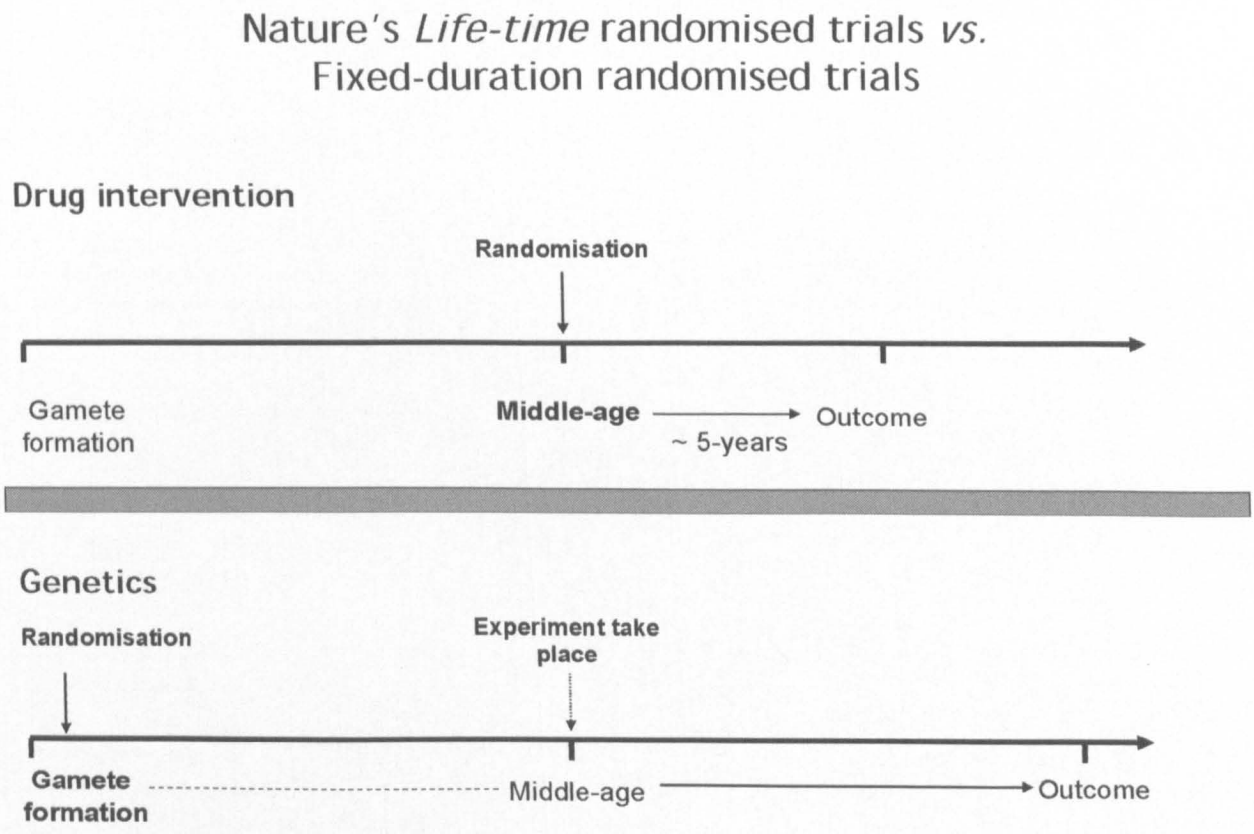
Thirdly, it has been mentioned repeatedly throughout this document the similarities of genetic studies with randomised clinical trials, but there are some important issues that merits further discussion (see Figure 8.5). In particular I will mention several potential scenarios in which these two strategies may arrive at different results. For this I will assume that the sample size in both trials and genetic studies, as well as the potency of the genetic variant, are not an issue for consideration.

Scenario-1 (trial-null and genetic-positive): randomised clinical trials are designed to test the efficacy (and safety) of an intervention (e.g. statins) on an outcome (e.g. CHD risk) and by inference provide

evidence on the reversibility of the risk associated to the exposure they intend to modify (e.g. LDL-cholesterol). In contrast, gene studies are by nature studies that evaluate the potential causality of a risk factor, but not the possibility of reversing the risk associated with the exposure. Thus, if a given risk factor is causal but not reversible, the trial and the genetic studies will obtain different findings.

Scenario-2 (trial-null and genetic-positive): genetic studies can be conceived as lifetime randomised interventions contrasting with randomised clinical trials which often last for 5 years. Therefore, the benefit of certain interventions, in theory, may become more evident after longer periods of use.

Figure 8.5. Schematic representation of the parallels and differences between genetic studies and randomised clinical trials.



Scenario-3 (trial-null and genetic-positive): randomised clinical trials are designed to test the efficacy and safety of an intervention. A wrong selection of the intervention may render the trial null and in the meantime the genetic study may give a positive result. Such could have been the scenario observed in the 1980's when available interventions to modify cholesterol were ineffective and their associated (using these interventions) were null.[170] In contrast, the genetic evidence from individuals suffering from familial hypercholesterolemia indicated the opposite.[22]

Scenario-4 (trial-positive and genetic-null): even if a risk factor is causal, a genetic study may be null due to the process of canalization or compensation that may occur between the actual moment of randomisation (gamete formation) and the experiment commencing (often several decades later), see Figure 8.5. In contrast, the randomised trial will give a positive result.

References

1. World Health Organization. The World Health Report 2002: reducing risks, promoting healthy life. Geneva: WHO. 2002.

2. European Cardiovascular Disease Statistics, 2000 Edition, British Heart Foundation

3. http://www.heartstats.org/2004CHDstatisticspublication (Accessed May, 2006).

4. De Backer G, Ambrosioni E, Borch-Johnsen K, et al. European guidelines on cardiovascular disease prevention in clinical practice: third joint task force of European and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of eight societies and by invited experts). Eur J Cardiovasc Prev Rehabil. 2003;10:S1-S10.

5. Pearson TA, Blair SN, Daniels SR, et al. AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update: Consensus Panel Guide to Comprehensive Risk Reduction for Adult Patients Without Coronary or Other Atherosclerotic Vascular Diseases. American Heart Association Science Advisory and Coordinating Committee. Circulation. 2002;106:388-91.

6. Goldstein LB, Adams R, Becker K, et al. Primary prevention of ischemic stroke: A statement for healthcare professionals from the Stroke Council of the American Heart Association. Circulation. 2001;103:163-82.

7. Stamler J, Stamler R, Neaton JD, Wentworth D, Daviglus ML, Garside D, Dyer AR, Liu K, Greenland P. Low risk-factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy: findings for 5 large cohorts of young adult and middle-aged men and women. JAMA. 1999;282:2012-8.

8. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, Wilson PW. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. JAMA. 2003;290:891-7.

9. Nieto FJ. Infections and atherosclerosis: new clues from an old hypothesis? Am J Epidemiol 1998;148:937-48.

10. Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. JAMA. 2002;288:2015-22.

11. Pearson TA, Mensah GA, Alexander RW, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation. 2003;107:499-511.

12. Danesh J, Lewington S, Thompson SG, et al. Fibrinogen Studies Collaboration. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. JAMA. 2005;294:1799-809.

13. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? Lancet. 2004;363:1724-7.

14. Petitti DB, Freedman DA. Invited commentary: how far can epidemiologists get with statistical adjustment? Am J Epidemiol. 2005;162:415-8.

15. Smith GD, Phillips AN. Confounding in epidemiological studies: why "independent" effects may not be all they seem. BMJ. 1992;305:757-9.

16. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet. 2001;357:373-80.

17. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO, Hennekens CH. 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;335:827-38.

18. Rothman K, Greenland S. Modern Epidemiology. Philadelphia: Lippincott Williams & Wilkins, 1998.

19. Hackam DG, Anand SS. Emerging risk factors for atherosclerotic vascular disease: a critical review of the evidence. JAMA. 2003;290:932-40.

20. Szklo M, Nieto FJ. Epidemiology: Beyond the basics. Maryland: Aspen publishers. 2000.

21. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, Peto R. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am J Epidemiol. 1999;150:341-53.

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

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

24. Keavney B. Genetic epidemiological studies of coronary heart disease. Int J Epidemiol. 2002;31:730-6.

25. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J. Natl. Cancer Inst. 2000;92, 1151-1158.

26. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet. 2003;361:865-72.

27. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet. 2004;36:512-7.

28. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM. Control of confounding of genetic associations in stratified populations. Am J Hum Genet. 2003;72:1492-1504.

29. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986;1:507-8.

30. Kivimäki M, Smith GD, Timpson NJ, Lawlor DA, Batty GD, Kähönen M, Juonala M, Rönnemaa T, Viikari JS, Lehtimäki T, Raitakari OT. Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study. Eur Heart J. 2008;20:2552-60.

31. Chen L, Davey Smith G, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. PLoS Med. 2008;5:e52.

32. Sacerdote C, Guarrera S, Smith GD, Grioni S, Krogh V, Masala G, Mattiello A, Palli D, Panico S, Tumino R, Veglia F, Matullo G, Vineis P. Lactase persistence and bitter taste response: instrumental variables and mendelian randomization in epidemiologic studies of dietary factors and cancer risk. Am J Epidemiol. 2007;166:576-81.

33. Palmer TM, Thompson JR, Tobin MD. Meta-analysis of Mendelian randomization studies incorporating all three genotypes. Stat Med. 2008;27:6570-82.

34. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. Int J Epidemiol. 2008;37:1161-8.

35. Warlow CP. Epidemiology of stroke. Lancet. 1998;352(Suppl.3):SIII1-4.

36. Goldstein LB, Adams R, Becker K, et al. Primary Prevention of Ischemic Stroke. A Statement for Healthcare Professionals From the Stroke Council of the American Heart Association. Circulation. 2001;103:163-82.

37. Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. BMJ. 2002;325:1202-6.

38. Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. JAMA. 2002;288:2015-22.

39. Bautista LE, Arenas IA, Penuela A, Martinez LX. Total plasma homocysteine level and risk of cardiovascular disease: a meta-analysis of prospective cohort studies. J Clin Epidemiol. 2002; 55:882-7.

40. Brattstrom L, Wilcken DE. Homocysteine and cardiovascular disease: cause or effect? Am J Clin Nutr. 2000;72:315-23

41. Ueland PM, Refsum H, Beresford SAA, Vollset SE. The controversy over homocysteine and cardiovascular risk. Am J Clin Nutr. 2000;72:324-32

42. Refsum H, Smith AD, Ueland PM, Nexo E, Clarke R, McPartlin J, Johnston C, Engbaek F, Schneede J, McPartlin C, Scott JM. Facts and recommendations about total homocysteine determinations: an expert opinion. Clin Chem. 2004;50:3-32.

43. Gibson JB, Carson NA, Neill DW. Pathological findings in homocystinuria. J Clin Pathol. 1964;17:427-37.

44. Mudd SH, Skovby F, Levy HL, Pettigrew KD, Wilcken B, Pyeritz RE, Andria G, Boers GH, Bromberg IL, Cerone R, et al. The natural history of homocystinuria due to cystathionine beta-synthase deficiency. Am J Hum Genet. 1985;37:1-31.

45. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995;10:111-3.

46. Klerk M, Verhoef P, Clarke R, et al; MTHFR Studies Collaboration Group. MTHFR 677C-->T polymorphism and risk of coronary heart disease: a meta-analysis. JAMA. 2002;288:2023-31.

47. Brattstrom L, Wilcken DE, Ohrvik J, Brudin L. Common methylenetetrahydrofolate reductase gene mutation leads to hyperhomocysteinemia but not to vascular disease: the result of a meta-analysis. Circulation. 1998;98:2520-6.

48. Morgan TH. Heredity and sex. Columbia University Press, New York, , USA. 1913.

49. Clarke R, Armitage J. Vitamin supplements and cardiovascular risk: review of the randomized trials of homocysteine-lowering vitamin supplements. Semin Thromb Hemost .2000;26:341-8.

50. Yap S, Boers GH, Wilcken B, Wilcken DE, Brenton DP, Lee PJ, Walter JH, Howard PM, Naughten ER. Vascular outcome in patients with homocystinuria due to cystathionine beta-synthase deficiency treated chronically: a multicenter observational study. Arterioscler Thromb Vasc Biol. 2001;21:2080-5.

51. Kluijtmans LA, Boers GH, Kraus JP, van den Heuvel LP, Cruysberg JR, Trijbels FJ, Blom HJ. The molecular basis of cystathionine beta-synthase deficiency in Dutch patients with homocystinuria: effect of CBS genotype on biochemical and clinical phenotype and on response to treatment. Am J Hum Genet. 1999; 65:59-67.

52. Yang Q, Botto LD, Erickson JD, Berry RJ, Sambell C, Johansen H, Friedman JM. Improvement in stroke mortality in Canada and the United States, 1990 to 2002. Circulation. 2006; 113:1335-43.

53. DerSimonian R, Laird NM. Meta-analysis in clinical trials. Controlled Clinical Trials. 1986; 7:177-88.

54. Robin JS, Greenland S, Breslow NE. A general estimator of the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol. 1986;124:719-23.

55. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in a meta-analysis. In: Egger M, Davey Smith G, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. BMJ Publications; 2001:285-312.

56. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34.

57. Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. Lancet. 1998;351:123–7.

58. Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in nongenetic observational studies Using "Mendelian triangulation" Ann Epidemiol. 2006;16:675-80.

59. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003; 326:219.

60. Keavney BD, Youngman LD, Palmer A, et al. Large-scale test of hypothesized associations between polymorphisms of lipid-related genes and myocardial infarction in about 5000 cases and 6000 controls [Abstract]. Circulation. 2000;102(Suppl.II):852.

61. Youngman LD, Keavney BD, Palmer A, et al. Plasma fibrinogen and fibrinogen genotypes in 4865 cases of myocardial infarction and 6002 controls: Test of causality by "Mendelian randomization"[Abstract]. Circulation. 2000;102(Suppl.II):32.

62. Russo GT, Friso S, Jacques PF, et al. Age and gender affect the relation between methylenetetrahydrofolate reductase C677T genotype and fasting plasma homocysteine concentrations in the Framingham Offspring Study Cohort. J Nutr 2003;133:3416-21.

63. Nallamothu BK, Fendrick AM, Rubenfire M, Saint S, Bandekar RR, Omenn GS. Potential clinical and economic effects of homocyst(e)ine lowering. Arch Intern Med. 2000;160:3406-12.

64. Lewis SJ, Ebrahim S, Davey Smith G. Meta-analysis of MTHFR 677C->T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? BMJ. 2005;331:1053.

65. Hustad S, Midttun Ø, Schneede J, Vollset SE, Grotmol T, Ueland PM. The methylenetetrahydrofolate reductase 677C-->T polymorphism as a modulator of a B vitamin network with major effects on homocysteine metabolism. Am J Hum Genet. 2007;80:846-55.

66. Eichholzer M, Tönz O, Zimmermann R. Folic acid: a public-health challenge. Lancet. 2006; 367:1352-61.

67. Wang X, Qin X, Demirtas H, Li J, Mao G, Huo Y, Sun N, Liu L, Xu X. Efficacy of folic acid supplementation in stroke prevention: a meta-analysis. Lancet. 2007;369:1876-82.

68. Verzilli C, Shah T, Casas JP, Chapman J, Sandhu M, Debenham SL, Boekholdt MS, Khaw KT, Wareham NJ, Judson R, Benjamin EJ, Kathiresan S, Larson MG, Rong J, Sofat R, Humphries SE, Smeeth L, Cavalleri G, Whittaker JC, Hingorani AD. Bayesian meta-analysis of genetic association studies with different sets of markers. Am J Hum Genet. 2008;82:859-72.

69. Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, Hingorani AD. Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. Am J Epidemiol. 2006;164:921-35.

70. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol. 2004;61:1652-61.

71. Bazzano LA, Reynolds K, Holder KN, He J. Effect of folic acid supplementation on risk of cardiovascular diseases: a meta-analysis of randomized controlled trials. JAMA. 2006;296:2720-6.

72. Little J, Higgins JPT (editors). The HuGENet HuGE Review Handbook, version 1.0. http://www.hugenet.ca (accessed 28 February 2006).

73. Zhu L, Ling H. National Neural Tube Defects Prevention Program in China. Food Nutr Bull. 2008;29(2 Suppl):S196-204.

74. Liem A, Reynierse-Buitenwerf GH, Zwinderman AH, Jukema JW, van Veldhuisen DJ. Secondary prevention with folic acid: results of the Goes extension study. Heart. 2005;91:1213-4.

75. Righetti M, Serbelloni P, Milani S, Ferrario G. Homocysteine-lowering vitamin B treatment decreases cardiovascular events in hemodialysis patients. Blood Purif. 2006;24:379-86.

76. Toole JF, Malinow MR, Chambless LE, Spence JD, Pettigrew LC, Howard VJ, Sides EG, Wang CH, Stampfer M. Lowering homocysteine in patients with ischemic stroke to prevent recurrent stroke, myocardial infarction, and death: the Vitamin Intervention for Stroke Prevention (VISP) randomized controlled trial. JAMA. 2004;291:565-75.

77. Lonn E, Yusuf S, Arnold MJ, Sheridan P, Pogue J, Micks M, McQueen MJ, Probstfield J, Fodor G, Held C, Genest J Jr; Heart Outcomes Prevention Evaluation (HOPE) 2 Investigators. Homocysteine lowering with folic acid and B vitamins in vascular disease. N Engl J Med. 2006;354:1567-77.

78. Bønaa KH, Njølstad I, Ueland PM, Schirmer H, Tverdal A, Steigen T, Wang H, Nordrehaug JE, Arnesen E, Rasmussen K; NORVIT Trial Investigators. Homocysteine lowering and cardiovascular events after acute myocardial infarction. N Engl J Med. 2006;354:1578-88.

79. Zoungas S, McGrath BP, Branley P, Kerr PG, Muske C, Wolfe R, Atkins RC, Nicholls K, Fraenkel M, Hutchison BG, Walker R, McNeil JJ. Cardiovascular morbidity and mortality in the Atherosclerosis and Folic Acid Supplementation Trial (ASFAST) in chronic renal failure: a multicenter, randomized, controlled trial. J Am Coll Cardiol. 2006;47:1108-16.

80. Wrone EM, Hornberger JM, Zehnder JL, McCann LM, Coplon NS, Fortmann SP. Randomized trial of folic acid for prevention of cardiovascular events in end-stage renal disease. J Am Soc Nephrol. 2004;15:420-6.

81. Mark SD, Wang W, Fraumeni JF Jr, Li JY, Taylor PR, Wang GQ, Guo W, Dawsey SM, Li B, Blot WJ. Lowered risks of hypertension and cerebrovascular disease after vitamin/mineral supplementation: the Linxian Nutrition Intervention Trial. Am J Epidemiol. 1996;143:658-64.

82. Albert CM, Cook NR, Gaziano JM, Zaharris E, MacFadyen J, Danielson E, Buring JE, Manson JE. Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial. JAMA. 2008;299:2027-36.

83. Ebbing M, Bleie Ø, Ueland PM, Nordrehaug JE, Nilsen DW, Vollset SE, Refsum H, Pedersen EK, Nygård O. Mortality and cardiovascular events in patients treated with homocysteine-lowering B vitamins after coronary angiography: a randomized controlled trial. JAMA. 2008;300:795-804.

84. Clarke R, Armitage J, Lewington S, Collins R; B-Vitamin Treatment Trialists' Collaboration. Homocysteine-lowering trials for prevention of vascular disease: protocol for a collaborative meta-analysis. Clin Chem Lab Med. 2007;45:1575-81.

85. Strong K, Mathers C, Bonita R. Preventing stroke: saving lives around the world. Lancet Neurol. 2007;6:182-7.

86. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. N Engl J Med. 1997;336:973-79.

87. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. N Engl J Med. 2000;342:836-43.

88. Kuller LH, Tracy RP, Shaten J, Meilahn EN. Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study. Multiple Risk Factor Intervention Trial. Am J Epidemiol. 1996;144:537-47.

89. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest. 2003;111:1805-12.

90. Danesh J, Whincup P, Walker M, et al. Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. BMJ. 2000;321:199-204.

91. Freeman DJ, Norrie J, Caslake MJ, et al; West of Scotland Coronary Prevention Study. C-reactive protein is an independent predictor of risk for the development of diabetes mellitus in the West of Scotland Coronary Prevention Study. Diabetes. 2002;51:1596-600.

92. Sattar N, McConnachie A, O'Reilly D, et al. Inverse association between birthweight and C-reactive protein concentrations in the MIDSPAN family study. Arterioscler Thromb Vasc Biol. 2004;24:583-87.

93. Balk EM, Lau J, Goudas LC, et al. Effects of statins on nonlipid serum markers associated with cardiovascular disease: a systematic review. Ann Intern Med 2003;139:670-82.

94. Sjoholm A, Nystrom T. Endothelial inflammation in insulin resistance. Lancet. 2005;365:610-12.

95. MacGregor AJ, Gallimore JR, Spector TD, Pepys MB. Genetic effects on baseline values of C-reactive protein and serum amyloid a protein: a comparison of monozygotic and dizygotic twins. Clin Chem 2004;50:130-34.

96. Pankow JS, Folsom AR, Cushman M, et al. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis. 2001;154:681-89.

97. D'Aiuto F, Casas JP, Shah T, Humphries SE, Hingorani AD, Tonetti M. C-reactive Protein (+1444C>T) Polymorphism Is Associated with Enhanced Acute Release of Inflammatory Markers. Atherosclerosis. 2005;179:413-17.

98. Brull DJ, Serrano N, Zito F, et al. A polymorphism in the human CRP gene influences basal and stimulated CRP levels: implications for the prediction and pathogenesis of coronary heart disease. Arterioscler Thromb Vasc Biol. 2003;23:2063-69.

99. Miller GJ, Bauer KA, Barzegar S, et al. The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. Thromb Haemost. 1995;73:82-6.

100. The WOSCOPS group. Baseline risk factors and their association with outcome in the West of Scotland Coronary Prevention Study. The West of Scotland Coronary Prevention Study Group. Am J Cardiol. 1997;79:756-62.

101. Meade T, Zuhrie R, Cook C, Cooper J, on behalf of MRC General Practice Research Framework. Bezafibrate in men with lower extremity arterial disease: randomised controlled trial. BMJ. 2002;325:1139.

102. Juhan-Vague I, Morange PE, Aubert H, et al; HIFMECH Study Group. Plasma thrombin-activatable fibrinolysis inhibitor antigen concentration and genotype in relation to myocardial infarction in the north and south of Europe. Arterioscler Thromb Vasc Biol. 2002;22:867-73.

103. Dhamrait SS, Stephens JW, Cooper JA, et al. Cardiovascular risk in healthy men and markers of oxidative stress in Diabetics men are associated with common variation in the gene for uncoupling protein 2. Eur Heart J. 2004;25:468-75.

104. Colhoun HM, Rubens MB, Underwood SR, Fuller JH. The effect of type 1 diabetes mellitus on the gender difference in coronary artery calcification. J Am Coll Cardiol. 2000;36:2160-67.

105. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. Circulation. 1994;90:583-12.

106. Shadish WR, Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges LV, eds. The handbook of Research Synthesis. New York: Rusell Sage Foundation; 1994:261-84.

107. Danesh J, Wheeler JG, Hirschfield GM, et al. C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. N Engl J Med. 2004;350:1387-97.

108. Davey Smith G, Lawlor DA, Harbord R, et al. Association of C-Reactive Protein With Blood Pressure and Hypertension. Life Course Confounding and Mendelian Randomization Tests of Causality. Arterioscler Thromb Vasc Biol. 2005;25:1051-56.

109. Russell AI, Cunninghame Graham DS, Shepherd C, et al. Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. Hum Mol Genet. 2004;13:137-47.

110. CRP CHD Genetics Collaboration. Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomisation. Eur J Epidemiol. 2008;23:531-40.

111. Kovacs A, Green F, Hansson LO, et al. A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. Atherosclerosis. 2005;178:193-98.

112. Zee RY, Ridker PM. Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. Atherosclerosis. 2002;162:217-19.

113. KWJ Lee, J Hill, J Frohlich. C-reactive protein as a predictor of clinical outcomes in patients with coronary atherosclerosis. (Oral session young investigator award). Canadian Cardiovascular Congress. 2002.

114. Miller DT, Zee RY, Danik JS, et al. Association of common CRP gene variants with CRP levels and cardiovascular events. Ann Hum Genet. 2005;69:623-38.

115. Casas JP, Shah T, Hingorani AD, Danesh J, Pepys MB. C-reactive protein and coronary heart disease: a critical review. J Intern Med. 2008;264:295-314.

116. Pepys MB, Hirschfield GM, Tennent GA, Gallimore JR, Kahan MC, Bellotti V, Hawkins PN, Myers RM, Smith MD, Polara A, Cobb AJ, Ley SV, Aquilina JA, Robinson CV, Sharif I, Gray GA, Sabin CA, Jenvey MC, Kolstoe SE, Thompson D, Wood SP. Targeting C-reactive protein for the treatment of cardiovascular disease. Nature. 2006;440:1217-21.

117. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. JAMA. 2001;286:327-34

118. Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM.C-reactive protein and the risk of developing hypertension. JAMA. 2003;290:2945-51.

119. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A, Davey Smith G. C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. Lancet. 2005;366:1954-9.

120. Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005;366:1906-8.

121. Kivimäki M, Lawlor DA, Smith GD, Kumari M, Donald A, Britton A, Casas JP, Shah T, Brunner E, Timpson NJ, Halcox JP, Miller MA, Humphries SE, Deanfield J, Marmot MG, Hingorani AD. Does high C-reactive protein concentration increase atherosclerosis? The Whitehall II Study. PLoS ONE. 2008;3:e3013.

122. Lawlor DA, Harbord RM, Timpson NJ, Lowe GD, Rumley A, Gaunt TR, Baker I, Yarnell JW, Kivimäki M, Kumari M, Norman PE, Jamrozik K, Hankey GJ, Almeida OP, Flicker L, Warrington N, Marmot MG, Ben-Shlomo Y, Palmer LJ, Day IN, Ebrahim S, Smith GD. The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants. PLoS ONE. 2008;3:e3011

123. 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:e1395.

124. Brunner EJ, Kivimäki M, Witte DR, Lawlor DA, Davey Smith G, Cooper JA, Miller M, Lowe GD, Rumley A, Casas JP, Shah T, Humphries SE, Hingorani AD, Marmot MG, Timpson NJ, Kumari M. Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream. PLoS Med. 2008;5:e155.

125. Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, Cushman M, Bis JC, Zeng D, Lin D, Kuller LH, Nickerson DA, Psaty BM, Tracy RP, Reiner AP. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. JAMA. 2006;296:2703-11.

126. Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GD, Fowkes FG. C-reactive protein, interleukin-6, and soluble adhesion molecules as predictors of progressive peripheral atherosclerosis in the general population: Edinburgh Artery Study. Circulation. 2005;112:976-83.

127. Ridker PM, Buring JE, Cook NR, Rifai N. C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: an 8-year follow-up of 14 719 initially healthy American women. Circulation. 2003;107:391-7.

128. Karabina SA, Ninio E.. Plasma PAF-acetylhydrolase: an unfulfilled promise? Biochim Biophys Acta. 2006;1761:1351-8.

129. Garza CA, Montori VM, McConnell JP, Somers VK, Kullo IJ, Lopez-Jimenez F. Association between lipoprotein-associated phospholipase A2 and cardiovascular disease: a systematic review. Mayo Clin Proc. 2007;82:159-65.

130. Packard CJ, O'Reilly DS, Caslake MJ, McMahon AD, Ford I, Cooney J, Macphee CH, Suckling KE, Krishna M, Wilkinson FE, Rumley A, Lowe GD. N Engl J Med. 2000;343:1148-55.

131. Serruys PW, García-García HM, Buszman P, Erne P, Verheye S, Aschermann M, Duckers H, Bleie O, Dudek D, Bøtker HE, von Birgelen C, D'Amico D, Hutchinson T, Zambanini A, Mastik F, van Es GA, van der Steen AF, Vince DG, Ganz P, Hamm CW, Wijns W, Zalewski A; Integrated Biomarker and Imaging Study-2 Investigators. Effects of the direct lipoprotein-associated phospholipase A(2) inhibitor darapladib on human coronary atherosclerotic plaque. Circulation. 2008;118:1172-82.

132. Mohler ER 3rd, Ballantyne CM, Davidson MH, Hanefeld M, Ruilope LM, Johnson JL, Zalewski A; Darapladib Investigators. The effect of darapladib on plasma lipoprotein-associated phospholipase A2 activity and cardiovascular biomarkers in patients with stable coronary heart disease or coronary heart disease risk equivalent: the results of a multicenter, randomized, double-blind, placebo-controlled study. J Am Coll Cardiol. 2008;51:1632-41.

133. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, Lopez-Sendon J, Mosca L, Tardif JC, Waters DD, Shear CL, Revkin JH, Buhr KA, Fisher MR, Tall AR, Brewer B; ILLUMINATE Investigators. N Engl J Med. 2007;357:2109-22.

134. Nissen SE, Wolski K, Topol EJ. Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. JAMA. 2005;294:2581-6.

135. Guerra R, Zhao B, Mooser V, Stafforini D, Johnston JM, Cohen JC. Determinants of plasma platelet-activating factor acetylhydrolase: heritability and relationship to plasma lipoproteins. J Lipid Res. 1997;38:2281-8.

136. Stafforini DM, Satoh K, Atkinson DL, Tjoelker LW, Eberhardt C, Yoshida H, Imaizumi T, Takamatsu S, Zimmerman GA, McIntyre TM, Gray PW, Prescott SM. Platelet-activating factor acetylhydrolase deficiency. A missense mutation near the active site of an anti-inflammatory phospholipase. J Clin Invest. 1996; 97:2784-2791.

137. Ninio E, Tregouet D, Carrier JL, Stengel D, Bickel C, Perret C, Rupprecht HJ, Cambien F, Blankenberg S, Tiret L. Platelet-activating factor-acetylhydrolase and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. Hum Mol Genet. 2004;13:1341-1351.

138. Abuzeid AM, Hawe E, Humphries SE, Talmud PJ; HIFMECH Study Group. Association between the Ala379Val variant of the lipoprotein associated phospholipase A2 and risk of myocardial infarction in the north and south of Europe. Atherosclerosis. 2003;168:283-8.

139. Sutton BS, Crosslin DR, Shah SH, Nelson SC, Bassil A, Hale AB, Haynes C, Goldschmidt-Clermont PJ, Vance JM, Seo D, Kraus WE, Gregory SG, Hauser ER. Comprehensive genetic analysis of the platelet activating factor acetylhydrolase (PLA2G7) gene and cardiovascular disease in case-control and family datasets. Hum Mol Genet. 2008;17:1318-28.

140. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, Wareham N. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer. 1999;80 (Suppl 1):95-103.

141. Marmot MG, Smith GD, Stansfeld S, Patel C, North F, Head J, White I, Brunner E, Feeney A. Health inequalities among British civil servants: the Whitehall II study. Lancet. 1991;337:1387-1393

142. Yang IA, Holloway JW, Ye S; Southampton Atherosclerosis Study (SAS) Group. TLR4 Asp299Gly polymorphism is not associated with coronary artery stenosis. Atherosclerosis. 2003;170:187-190.

143. Hoffmann MM, Winkler K, Renner W, Winkelmann BR, Seelhorst U, Wellnitz B, Boehm BO, Marz W. Genetic variants and haplotypes of lipoprotein associated phospholipase A2 and their influence on cardiovascular disease (The Ludwigshafen Risk and Cardiovascular Health Study). J Thromb Haemost. 2009;7:41-48.

144. Panayiotou A, Griffin M, Georgiou N, Bond D, Tyllis T, Tziakouri-Shiakalli C, Fessas C, Nicolaides A. ApoB/ApoA1 ratio and subclinical atherosclerosis. Int Angiol. 2008;27:74-80.

145. Corsetti JP, Rainwater DL, Moss AJ, Zareba W, Sparks CE. High lipoprotein-associated phospholipase A2 is a risk factor for recurrent coronary events in postinfarction patients. Clin Chem. 2006;52:1331-1338.

146. Tselepis AD, Dentan C, Karabina SA, Chapman MJ, Ninio E. PAF-degrading acetylhydrolase is preferentially associated with dense LDL and VHDL-1 in human plasma. Catalytic characteristics and relation to the monocyte-derived enzyme. Arterioscler Thromb Vasc Biol. 1995;15:1764-1773.

147. Blankenberg S, Stengel D, Rupprecht HJ, Bickel C, Meyer J, Cambien F, Tiret L, Ninio E. Plasma PAF-acetylhydrolase in patients with coronary artery disease: results of a cross-sectional analysis. J Lipid Res. 2003;44:1381-1386.

148. Kosaka T, Yamaguchi M, Soda Y, Kishimoto T, Tago A, Toyosato M, Mizuno K. Spectrophotometric assay for serum platelet-activating factor acetylhydrolase activity. Clin Chim Acta. 2000;296:151-161.

149. Wilensky RL, Shi Y, Mohler ER, III, Hamamdzic D, Burgert ME, Li J, Postle A, Fenning RS, Bollinger JG, Hoffman BE, Pelchovitz DJ, Yang J, Mirabile RC, Webb CL, Zhang L, Zhang P, Gelb MH, Walker MC, Zalewski A, Macphee CH.Inhibition of lipoprotein-associated phospholipase A2 reduces complex coronary atherosclerotic plaque development. Nat Med. 2008;14:1059-1066.

150. Miwa M, Miyake T, Yamanaka T, Sugatani J, Suzuki Y, Sakata S, Araki Y, Matsumoto M. Characterization of serum platelet-activating factor (PAF) acetylhydrolase. Correlation between deficiency of serum PAF acetylhydrolase and respiratory symptoms in asthmatic children. J Clin Invest. 1998;82:1983-1991.

151. Yamada Y, Ichihara S, Izawa H, Tanaka M, Yokota M. Association of a G994 --> T (Val279 --> Phe) polymorphism of the plasma platelet-activating factor acetylhydrolase gene with myocardial damage in Japanese patients with nonfamilial hypertrophic cardiomyopathy.J Hum Genet. 2001;46:436-441.

152. Yamada Y, Yoshida H, Ichihara S, Imaizumi T, Satoh K, Yokota M. Correlations between plasma platelet-activating factor acetylhydrolase (PAF-AH) activity and PAF-AH genotype, age, and atherosclerosis in a Japanese population. Atherosclerosis. 2000;150:209-216.

153. Ballantyne C, Cushman M, Psaty B, Furberg C, Khaw KT, Sandhu M, Oldgren J, Rossi GP, Maiolino G, Cesari M, Lenzini L, James SK, Rimm E, Collins R, Anderson J, Koenig W, Brenner H, Rothenbacher D, Berglund G, Persson M, Berger P, Brilakis E, McConnell JP, Koenig W, Sacco R, Elkind M, Talmud P, Rimm E, Cannon CP, Packard C, Barrett-Connor E, Hofman A, Kardys I, Witteman JC, Criqui M, Corsetti JP, Rainwater DL, Moss AJ, Robins S, Bloomfield H, Collins D, Packard C, Wassertheil-Smoller S, Ridker P, Ballantyne C, Cannon CP, Cushman M, Danesh J, Gu D, Hofman A, Nelson JJ, Thompson S, Zalewski A, Zariffa N, Di Angelantonio E, Kaptoge S, Thompson A, Thompson S, Walker M, Watson S, Wood A. Collaborative meta-analysis of individual participant data from observational studies of Lp-PLA2 and cardiovascular diseases. Eur J Cardiovasc Prev Rehabil. 2007;14:3-11.

154. Rader DJ, Daugherty A. Translating molecular discoveries into new therapies for atherosclerosis. Nature. 2008;451:904-913.

155. Hopkins PN, Williams RR. A survey of 246 suggested coronary risk factors. Atherosclerosis. 1981;40:1-52.

156. Brotman DJ, Walker E, Lauer MS, O'Brien RG. In search of fewer independent risk factors. Arch Intern Med. 2005;165:138-145.

157. Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol. 1981;114:593-603.

158. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, Simes R. 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: 1267-1278.

159. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, Lauretani F, Murray A, Gibbs JR, Paolisso G, Rafiq S, Simon-Sanchez J, Lango H, Scholz S, Weedon MN, Arepalli S, Rice N, Washecka N, Hurst A, Britton A, Henley W, van de Leemput J, Li R, Newman AB, Tranah G, Harris T, Panicker V, Dayan C, Bennett A, McCarthy MI, Ruokonen A, Jarvelin MR, Guralnik J, Bandinelli S, Frayling TM, Singleton A, Ferrucci L. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet. 2008;4:e1000072.

160. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004;430:743-747.

161. McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, Steyn K, Sanderson JE, Hasani M, Volkova E, Kazmi K, Yusuf S; INTERHEART study investigators. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study. Lancet. 2008;372:224-33.

162. Miller GJ, Ireland HA, Cooper JA, Bauer KA, Morrissey JH, Humphries SE, Esnouf MP. Relationship between markers of activated coagulation, their correlation with inflammation, and association with coronary heart disease (NPHSII). J Thromb Haemost. 2008;6:259-67.

163. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. Nature. 2006;444:444-454.

164. Rockman MV. Reverse engineering the genotype-phenotype map with natural genetic variation. Nature. 2008;456:738-744.

165. Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. PLoS Med. 2005;2:e334.

166. Li Z, Sun L, Zhang H, Liao Y, Wang D, Zhao B, Zhu Z, Zhao J, Ma A, Han Y, Wang Y, Shi Y, Ye J, Hui R; Multicenter Case-Control Study in China. Elevated plasma homocysteine was associated with hemorrhagic and ischemic stroke, but methylenetetrahydrofolate reductase gene C677T polymorphism was a risk factor for thrombotic stroke: a Multicenter Case-Control Study in China. Stroke. 2003;34:2085-90.

167. Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study.Am J Epidemiol. 2007;166:646-55.

168. Kardys I, de Maat MP, Uitterlinden AG, Hofman A, Witteman JC.C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. Eur Heart J. 2006;27:1331-7.

169. 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;359:1897-908.

170. Steinberg D, Gotto AM Jr.Preventing coronary artery disease by lowering cholesterol levels: fifty years from bench to bedside. JAMA. 1999 1;282:2043-50.

Appendix A: References of studies included in the meta-analyses from Chapter 2

w1. Caudill MA, Le T, Moonie SA, Esfahani ST, Cogger EA. Folate status in women of childbearing age residing in Southern California after folic acid fortification. J Am Coll Nutr. 2001;20:129-34

w2. Thogersen AM, Nilsson TK, Dahlen G, Jansson JH, Boman K, Huhtasaari F, et al. Homozygosity for the C677–>T mutation of 5,10-methylenetetrahydrofolate reductase and total plasma homocysteine are not associated with greater than normal risk of a first myocardial infarction in northern Sweden. Coron Artery Dis. 2001;12:85-90.

w3. Schmitz C, Lindpaintner K, Verhoef P, Gaziano JM, Buring J. Genetic polymorphism of methylenetetrahydrofolate reductase and myocardial infarction. A case-control study. Circulation. 1996;94:1812-4.

w4. Chambers JC, Ireland H, Thompson E, Reilly P, Obeid OA, Refsum H, et al. Methylenetetrahydrofolate reductase 677 C-->T mutation and coronary heart disease risk in UK Indian Asians. Arterioscler Thromb Vasc Biol. 2000;20:2448-52.

w5. Fohr IP, Prinz-Langenohl R, Bronstrup A, Bohlmann AM, Nau H, Berthold HK, et al. 5,10-Methylenetetrahydrofolate reductase genotype determines the plasma homocysteine-lowering effect of supplementation with 5-methyltetrahydrofolate or folic acid in healthy young women. Am J Clin Nutr. 2002;75:275-82.

w6. Guinotte CL, Burns MG, Axume JA, Hata H, Urrutia TF, Alamilla A, et al. Methylenetetrahydrofolate reductase 677C-->T variant modulates folate status response to controlled folate intakes in young women. J Nutr. 2003;133:1272-80.

w7. Meisel C, Cascorbi I, Gerloff T, Stangl V, Laule M, Muller JM, et al. Identification of six methylenetetrahydrofolate reductase (MTHFR) genotypes resulting from common polymorphisms: impact on plasma homocysteine levels and development of coronary artery disease. Atherosclerosis. 2001;154:651-8.

w8. Ho CH. The influence of age, sex, vitamin B(12), folate levels and methylenetetrahydrofolate reductase C677T genetic mutations on plasma homocysteine in the Chinese population. Haematologica. 2000;85:1051-4.

w9. Dekou V, Whincup P, Papacosta O, Ebrahim S, Lennon L, Ueland PM, et al. The effect of the C677T and A1298C polymorphisms in the methylenetetrahydrofolate reductase gene on homocysteine levels in elderly men and women from the British regional heart study. Atherosclerosis. 2001;154:659-66.

w10. Voutilainen S, Lakka TA, Hamelahti P, Lehtimaki T, Poulsen HE, Salonen JT. Plasma total homocysteine concentration and the risk of acute coronary events: the Kuopio Ischaemic Heart Disease Risk Factor Study. J Intern Med. 2000;248:217-22.

w11. Thuillier L, Chadefaux-Vekemans B, Bonnefont JP, Kara A, Aupetit J, Rochette C, et al. Does the polymorphism 677C-T of the 5,10-methylenetetrahydrofolate reductase gene contribute to homocysteine-related vascular disease? J Inherit Metab Dis. 1998;21:812-22.

w12. Somekawa Y, Kobayashi K, Tomura S, Aso T, Hamaguchi H. Effects of hormone replacement therapy and methylenetetrahydrofolate reductase polymorphism on plasma folate and homocysteine levels in postmenopausal Japanese women. Fertil Steril. 2002;77:481-6.

w13. Jacques PF, Kalmbach R, Bagley PJ, Russo GT, Rogers G, Wilson PW, et al. The relationship between riboflavin and plasma total homocysteine in the Framingham Offspring cohort is influenced by folate status and the C677T transition in the methylenetetrahydrofolate reductase gene. J Nutr. 2002;132:283-8.

w14. Silaste ML, Rantala M, Sampi M, Alfthan G, Aro A, Kesaniemi YA. Polymorphisms of key enzymes in homocysteine metabolism affect diet responsiveness of plasma homocysteine in healthy women. J Nutr. 2001;131:2643-7.

w15. Bailey LB, Duhaney RL, Maneval DR, Kauwell GP, Quinlivan EP, Davis SR, et al. Vitamin B-12 status is inversely associated with plasma homocysteine in young women with C677T and/or A1298C methylenetetrahydrofolate reductase polymorphisms. J Nutr. 2002;132:1872-8.

w16. Domagala TB, Adamek L, Nizankowska E, Sanak M, Szczeklik A. Mutations C677T and A1298C of the 5,10-methylenetetrahydrofolate reductase gene and fasting plasma homocysteine levels are not associated with the increased risk of venous thromboembolic disease. Blood Coagul Fibrinolysis 2002;13:423-31.

w17. Chango A, Potier De Courcy G, Boisson F, Guilland JC, Barbe F, Perrin MO, et al. 5,10-methylenetetrahydrofolate reductase common mutations, folate status and plasma homocysteine in healthy French adults of the Supplementation en Vitamines et Mineraux Antioxydants (SU.VI.MAX) cohort. Br J Nutr. 2000;84:891-6.

w18. Cappuccio FP, Bell R, Perry IJ, Gilg J, Ueland PM, Refsum H, et al. Homocysteine levels in men and women of different ethnic and cultural background living in England. Atherosclerosis. 2002;164:95-102.

w19. Castro R, Rivera I, Ravasco P, Jakobs C, Blom HJ, Camilo ME, et al. 5,10-Methylenetetrahydrofolate reductase 677C-->T and 1298A-->C mutations are genetic determinants of elevated homocysteine. QJM. 2003;96:297-303.

w20. Chango A, Boisson F, Barbe F, Quilliot D, Droesch S, Pfister M, et al. The effect of 677C-->T and 1298A-->C mutations on plasma homocysteine and 5,10-methylenetetrahydrofolate reductase activity in healthy subjects. Br J Nutr. 2000;83:593-6.

w21. Ma J, Stampfer MJ, Hennekens CH, Frosst P, Selhub J, Horsford J, et al. Methylenetetrahydrofolate reductase polymorphism, plasma folate, homocysteine, and risk of myocardial infarction in US physicians. Circulation. 1996;94:2410-6.

w22. Friedman G, Goldschmidt N, Friedlander Y, Ben-Yehuda A, Selhub J, Babaey S, et al. A common mutation A1298C in human methylenetetrahydrofolate reductase gene: association with plasma total homocysteine and folate concentrations. J Nutr. 1999;129:1656-61.

w23. Husemoen LL, Thomsen TF, Fenger M, Jorgensen HL, Jorgensen T. Contribution of thermolabile methylenetetrahydrofolate reductase variant to total plasma homocysteine levels in healthy men and women. Inter99 (2). Genet Epidemiol. 2003;24:322-30.

w24. Reyes-Engel A, Munoz E, Gaitan MJ, Fabre E, Gallo M, Dieguez JL, et al. Implications on human fertility of the 677C-->T and 1298A-->C polymorphisms of the MTHFR gene: consequences of a possible genetic selection. Mol Hum Reprod. 2002;8:952-7.

w25. Schwartz SM, Siscovick DS, Malinow MR, Rosendaal FR, Beverly RK, Hess DL, et al. Myocardial infarction in young women in relation to plasma total homocysteine, folate, and a common variant in the methylenetetrahydrofolate reductase gene. Circulation. 1997;96:412-7.

w26. Christensen B, Frosst P, Lussier-Cacan S, Selhub J, Goyette P, Rosenblatt DS, et al. Correlation of a common mutation in the methylenetetrahydrofolate reductase gene with plasma homocysteine in patients with premature coronary artery disease. Arterioscler Thromb Vasc Biol. 1997;17:569-73.

w27. Meleady R, Ueland PM, Blom H, Whitehead AS, Refsum H, Daly LE, et al; EC Concerted Action Project: Homocysteine and Vascular Disease. Thermolabile methylenetetrahydrofolate reductase, homocysteine, and cardiovascular disease risk: the European Concerted Action Project. Am J Clin Nutr. 2003;77:63-70.

w28. Hustad S, Ueland PM, Vollset SE, Zhang Y, Bjorke-Monsen AL, Schneede J. Riboflavin as a determinant of plasma total homocysteine: effect modification by the methylenetetrahydrofolate reductase C677T polymorphism. Clin Chem. 2000;46:1065-71.

w29. Yoo JH, Park SC. Low plasma folate in combination with the 677 C-->T methylenetetrahydrofolate reductase polymorphism is associated with increased risk of coronary artery disease in Koreans. Thromb Res. 2000;97:77-84.

w30. Passaro A, Vanini A, Calzoni F, Alberti L, Zamboni PF, Fellin R, et al. Plasma homocysteine, methylenetetrahydrofolate reductase mutation and carotid damage in elderly healthy women. Atherosclerosis. 2001;157:175-80.

w31. Pullin CH, Ashfield-Watt PA, Burr ML, Clark ZE, Lewis MJ, Moat SJ, et al. Optimization of dietary folate or low-dose folic acid supplements lower homocysteine but do not enhance endothelial function in healthy adults, irrespective of the methylenetetrahydrofolate reductase (C677T) genotype. J Am Coll Cardiol. 2001;38:1799-805.

w32. Jee SH, Song KS, Shim WH, Kim HK, Suh I, Park JY, et al. Major gene evidence after MTHFR-segregation analysis of serum homocysteine in families of patients undergoing coronary arteriography. Hum Genet. 2002;111:128-35.

w33. Litynski P, Loehrer F, Linder L, Todesco L, Fowler B. Effect of low doses of 5-methyltetrahydrofolate and folic acid on plasma homocysteine in healthy subjects with or without the 677C-->T polymorphism of methylenetetrahydrofolate reductase. Eur J Clin Invest. 2002;32:662-8.

w34. de Bree A, Verschuren WM, Bjorke-Monsen AL, van der Put NM, Heil SG, Trijbels FJ, et al. Effect of the methylenetetrahydrofolate reductase 677C-->T mutation on the relations among folate intake and plasma folate and homocysteine concentrations in a general population sample. Am J Clin Nutr. 2003;77:687-93.

w35. Ou T, Yamakawa-Kobayashi K, Arinami T, Amemiya H, Fujiwara H, Kawata K, et al. Methylenetetrahydrofolate reductase and apolipoprotein E polymorphisms are independent risk factors for coronary heart disease in Japanese: a case-control study. Atherosclerosis. 1998;137:23-8.

w36. Vizcaino G, Diez-Ewald M, Herrmann FH, Schuster G, Perez-Requejo JL. Relationships between homocysteine, folate and vitamin B12 levels with the methylenetetrahydrofolate reductase polymorphism, in Indians from Western Venezuela. Thromb Haemost. 2001;85:186-7.

w37. D'Angelo A, Coppola A, Madonna P, Fermo I, Pagano A, Mazzola G, et al. The role of vitamin B12 in fasting hyperhomocysteinemia and its interaction with the homozygous C677T mutation of the methylenetetrahydrofolate reductase (MTHFR) gene. A case-control study of patients with early-onset thrombotic events. Thromb Haemost. 2000;83:563-70.

w38. Zittoun J, Tonetti C, Bories D, Pignon JM, Tulliez M. Plasma homocysteine levels related to interactions between folate status and methylenetetrahydrofolate reductase: a study in 52 healthy subjects. Metabolism. 1998;47:1413-8.

w39. Gemmati D, Previati M, Serino ML, Moratelli S, Guerra S, Capitani S, et al. Low folate levels and thermolabile methylenetetrahydrofolate reductase as primary determinant of mild hyperhomocystinemia in normal and thromboembolic subjects. Arterioscler Thromb Vasc Biol. 1999;19:1761-7.

w40. Yingdong Z, Zhigang Z; Yang L. Association of plasma homocysteine level and N^5, N^{10}-methylenetetrahydrofolate reductase gene polymorphism with cerebral infarction. Chinese Medical Sciences Journal. 2002;17:231-5.

w41. Varela ML, Adamczuk YP, Forastiero RR, Martinuzzo ME, Cerrato GS, Pombo G, et al. Major and potential prothrombotic genotypes in a cohort of patients with venous thromboembolism. Thromb Res. 2001;104:317-24.

w42. Lalouschek W, Aull S, Korninger L, Mannhalter C, Pabinger-Fasching I, Schmid RW, et al. 677C to T mutation in the 5,10-methylenetetrahydrofolate reductase (MTHFR) gene and plasma homocysteine levels in patients with TIA or minor stroke. J Neurol Sci. 1998;155:156-62.

w43. Kim CH, Hwang KY, Choi TM, Shin WY, Hong SY. The methylenetetrahydrofolate reductase gene polymorphism in Koreans with coronary artery disease. Int J Cardiol. 2001;78:13-7.

w44. Kosokabe T, Okumura K, Sone T, Kondo J, Tsuboi H, Mukawa H, et al. Relation of a common methylenetetrahydrofolate reductase mutation and plasma homocysteine with intimal hyperplasia after coronary stenting. Circulation. 2001;103:2048-54.

w45. Dunn J, Title LM, Bata I, Johnstone DE, Kirkland SA, O'Neill BJ, et al. Relation of a common mutation in methylenetetrahydrofolate reductase to plasma homocysteine and early onset coronary artery disease. Clin Biochem. 1998;31:95-100.

w46. Kluijtmans LA, Kastelein JJ, Lindemans J, Boers GH, Heil SG, Bruschke AV, et al. Thermolabile methylenetetrahydrofolate reductase in coronary artery disease. Circulation. 1997;96:2573-7.

w47. Morita H, Kurihara H, Tsubaki S, Sugiyama T, Hamada C, Kurihara Y, et al. Methylenetetrahydrofolate reductase gene polymorphism and ischemic stroke in Japanese. Arterioscler Thromb Vasc Biol. 1998;18:1465-9.

w48. Anderson JL, Muhlestein JB, Horne BD, Carlquist JF, Bair TL, Madsen TE, et al. Plasma homocysteine predicts mortality independently of traditional risk factors and C-reactive protein in patients with angiographically defined coronary artery disease. Circulation. 2000;102:1227-32.

w49. Deloughery TG, Evans A, Sadeghi A, McWilliams J, Henner WD, Taylor LM Jr, et al. Common mutation in methylenetetrahydrofolate reductase. Correlation with homocysteine metabolism and late-onset vascular disease. Circulation. 1996;94:3074-8.

w50. Gaustadnes M, Rudiger N, Moller J, Rasmussen K, Bjerregaard Larsen T, Ingerslev J. Thrombophilic predisposition in stroke and venous thromboembolism in Danish patients. Blood Coagul Fibrinolysis. 1999;10:251-9.

w51. Morelli VM, Lourenco DM, D'Almeida V, Franco RF, Miranda F, Zago MA, et al. Hyperhomocysteinemia increases the risk of venous thrombosis independent of the C677T mutation of the methylenetetrahydrofolate reductase gene in selected Brazilian patients. Blood Coagul Fibrinolysis. 2002;13:271-5.

w52. Tokgozoglu SL, Alikasifoglu M, Unsal, Atalar E, Aytemir K, Ozer N, Ovunc K, Usal O, Kes S, Tuncbilek E. Methylene tetrahydrofolate reductase genotype and the risk and extent of coronary artery disease in a population with low plasma folate. Heart. 1999;81:518-22.

w53. Markus HS, Ali N, Swaminathan R, Sankaralingam A, Molloy J, Powell J. A common polymorphism in the methylenetetrahydrofolate reductase gene, homocysteine, and ischemic cerebrovascular disease. Stroke. 1997;28:1739-43.

w54. Candito M, Bedoucha P, Gibelin P, Jambou D, de Franchis R, Sadoul JL, et al. Fasting, postprandial, and post-methionine-load homocysteinaemia and methylenetetrahydrofolate reductase polymorphism in vascular disease. J Inherit Metab Dis. 1999;22:588-92.

w55. Andreassi MG, Botto N, Cocci F, Battaglia D, Antonioli E, Masetti S, et al. Methylenetetrahydrofolate reductase gene C677T polymorphism, homocysteine, vitamin B12, and DNA damage in coronary artery disease. Hum Genet. 2003;112:171-7.

w56. Kaye JM, Stanton KG, McCann VJ, Vasikaran SD, Burke V, Taylor RR, et al. Homocysteine, folate, methylene tetrahydrofolate reductase genotype and vascular morbidity in diabetic subjects. Clin Sci (Lond). 2002;102:631-7.

w57. Hanson NQ, Aras O, Yang F, Tsai MY. C677T and A1298C polymorphisms of the methylenetetrahydrofolate reductase gene: incidence and effect of combined genotypes on

plasma fasting and post-methionine load homocysteine in vascular disease. Clin Chem. 2001;47:661-6.

w58. Malinow MR, Nieto FJ, Kruger WD, Duell PB, Hess DL, Gluckman RA, et al. The effects of folic acid supplementation on plasma total homocysteine are modulated by multivitamin use and methylenetetrahydrofolate reductase genotypes. Arterioscler Thromb Vasc Biol. 1997;17:1157-62.

w59. Eikelboom JW, Hankey GJ, Anand SS, Lofthouse E, Staples N, Baker RI. Association between high homocyst(e)ine and ischemic stroke due to large- and small-artery disease but not other etiologic subtypes of ischemic stroke. Stroke. 2000;31:1069-75.

w60. Tsai MY, Welge BG, Hanson NQ, Bignell MK, Vessey J, Schwichtenberg K, et al. Genetic causes of mild hyperhomocysteinemia in patients with premature occlusive coronary artery diseases. Atherosclerosis. 1999;143:163-70.

w61. Mazza A, Motti C, Nulli A, Marra G, Gnasso A, Pastore A, et al. Lack of association between carotid intima-media thickness and methylenetetrahydrofolate reductase gene polymorphism or serum homocysteine in non-insulin-dependent diabetes mellitus. Metabolism. 2000;49:718-23.

w62. Lalouschek W, Aull S, Serles W, Wolfsberger M, Deecke L, Pabinger-Fasching I, et al. The relation between erythrocyte volume and folate levels is influenced by a common mutation in the methylenetetrahydrofolate reductase (MTHFR) gene (C677T). J Investig Med. 2000;48:14-20.

w63. Gonzalez-Ordonez AJ, Fernandez-Alvarez CR, Rodriguez JM, Garcia EC, Alvarez MV. Genetic polymorphism of methylenetetrahydrofolate reductase and venous thromboembolism: a case-control study. Haematologica. 1999;84:190-1.

w64. Tsai AW, Cushman M, Tsai MY, Heckbert SR, Rosamond WD, Aleksic N, et al. Serum homocysteine, thermolabile variant of methylene tetrahydrofolate reductase (MTHFR), and venous thromboembolism: Longitudinal Investigation of Thromboembolism Etiology (LITE). Am J Hematol. 2003;72:192-200.

w65. Chao CL, Tsai HH, Lee CM, Hsu SM, Kao JT, Chien KL, et al. The graded effect of hyperhomocysteinemia on the severity and extent of coronary atherosclerosis. Atherosclerosis. 1999;147:379-86.

w66. Lievers KJ, Boers GH, Verhoef P, den Heijer M, Kluijtmans LA, van der Put NM, et al. A second common variant in the methylenetetrahydrofolate reductase (MTHFR) gene and its relationship to MTHFR enzyme activity, homocysteine, and cardiovascular disease risk. J Mol Med. 2001;79:522-8.

w67. Nakai K, Fusazaki T, Suzuki T, Ohsawa M, Ogiu N, Kamata J, et al. Genetic polymorphism of 5,10-methylenetetrahydrofolate increases risk of myocardial infarction and is correlated to elevated levels of homocysteine in the Japanese general population. Coron Artery Dis. 2000;11:47-51.

w68. Kluijtmans LA, van den Heuvel LP, Boers GH, Frosst P, Stevens EM, van Oost BA, et al. Molecular genetic analysis in mild hyperhomocysteinemia: a common mutation in the methylenetetrahydrofolate reductase gene is a genetic risk factor for cardiovascular disease. Am J Hum Genet. 1996;58:35-41.

w69. Spence MS, McGlinchey PG, Patterson CC, Belton C, Murphy G, McMaster D, et al. Family-based investigation of the C677T polymorphism of the methylenetetrahydrofolate reductase gene in ischaemic heart disease. Atherosclerosis. 2002;165:293-9.

w70. Scaglione L, Gambino R, Rolfo E, Lillaz E, Gai M, Cassader M, et al. Plasma homocysteine, methylenetetrahydrofolate reductase gene polymorphism and carotid intima-media thickness in Italian type 2 diabetic patients. Eur J Clin Invest. 2002;32:24-8.

w71. McIlroy SP, Dynan KB, Lawson JT, Patterson CC, Passmore AP. Moderately elevated plasma homocysteine, methylenetetrahydrofolate reductase genotype, and risk for stroke, vascular dementia, and Alzheimer disease in Northern Ireland. Stroke. 2002;33:2351-6.

w72. Fujimura H, Kawasaki T, Sakata T, Ariyoshi H, Kato H, Monden M, et al. Common C677T polymorphism in the methylenetetrahydrofolate reductase gene increases the risk for deep vein thrombosis in patients with predisposition of thrombophilia. Thromb Res. 2000;98:1-8.

w73. Arai K, Yamasaki Y, Kajimoto Y, Watada H, Umayahara Y, Kodama M, et al. Association of methylenetetrahydrofolate reductase gene polymorphism with carotid arterial wall thickening and myocardial infarction risk in NIDDM. Diabetes. 1997;46:2102-4.

w74. Raslova K, Smolkova B, Vohnout B, Gasparovic J, Frohlich JJ. Risk factors for atherosclerosis in survivors of myocardial infarction and their spouses: comparison to controls without personal and family history of atherosclerosis. Metabolism. 2001;50:24-9.

w75. Legnani C, Palareti G, Grauso F, Sassi S, Grossi G, Piazzi S, et al. Hyperhomocyst(e)inemia and a common methylenetetrahydrofolate reductase mutation (Ala223Val MTHFR) in patients with inherited thrombophilic coagulation defects. Arterioscler Thromb Vasc Biol. 1997;17:2924-9.

w76. Girelli D, Martinelli N, Pizzolo F, Friso S, Olivieri O, Stranieri C, et al. The interaction between MTHFR 677 C-->T genotype and folate status is a determinant of coronary atherosclerosis risk. J Nutr. 2003;133:1281-5.

w77. Rassoul F, Richter V, Janke C, Purschwitz K, Klotzer B, Geisel J, et al. Plasma homocysteine and lipoprotein profile in patients with peripheral arterial occlusive disease. Angiology. 2000;51:189-96.

w78. Yoo JH. A thermolabile variant of methylenetetrahydrofolate reductase is a determinant of hyperhomocyst(e)inemia in the elderly. Ann N Y Acad Sci 2001;928:344.

w79. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995;10:111-3.

w80. Kawashiri M, Kajinami K, Nohara A, Yagi K, Inazu A, Koizumi J, et al. Effect of common methylenetetrahydrofolate reductase gene mutation on coronary artery disease in familial hypercholesterolemia. Am J Cardiol. 2000;86:840-5.

w81. Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, et al. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proc Natl Acad Sci U S A. 2002;99:5606-11.

w82. Giles WH, Kittner SJ, Ou CY, Croft JB, Brown V, Buchholz DW, et al. Thermolabile methylenetetrahydrofolate reductase polymorphism (C677T) and total homocysteine concentration among African-American and white women. Ethn Dis. 1998;8:149-57.

w83. Girelli D, Friso S, Trabetti E, Olivieri O, Russo C, Pessotto R, et al. Methylenetetrahydrofolate reductase C677T mutation, plasma homocysteine, and folate in subjects from northern Italy with or without angiographically documented severe coronary atherosclerotic disease: evidence for an important genetic-environmental interaction. Blood. 1998;91:4158-63.

w84. Harmon DL, Woodside JV, Yarnell JW, McMaster D, Young IS, McCrum EE, et al. The common 'thermolabile' variant of methylene tetrahydrofolate reductase is a major determinant of mild hyperhomocysteinaemia. QJM. 1996;89:571-7.

w85. Harrington DJ, Malefora A, Schmeleva V, Kapustin S, Papayan L, Blinov M, et al. Genetic variations observed in arterial and venous thromboembolism: relevance for therapy, risk prevention and prognosis. Clin Chem Lab Med. 2003;41:496-500.

w86. Herrmann W, Obeid R, Schorr H, Zarzour W, Geisel J. Homocysteine, methylenetetrahydrofolate reductase C677T polymorphism and the B-vitamins: a facet of nature-nurture interplay. Clin Chem Lab Med. 2003;41:547-53.

w87. Jacques PF, Bostom AG, Williams RR, Ellison RC, Eckfeldt JH, Rosenberg IH, et al. Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. Circulation. 1996;93:7-9.

w88. Kluijtmans LA, Young IS, Boreham CA, Murray L, McMaster D, McNulty H, et al. Genetic and nutritional factors contributing to hyperhomocysteinemia in young adults. Blood. 2003;101:2483-8.

w89. Kruger WD, Evans AA, Wang L, Malinow MR, Duell PB, Anderson PH, et al. Polymorphisms in the CBS gene associated with decreased risk of coronary artery disease and increased responsiveness to total homocysteine lowering by folic acid. Mol Genet Metab. 2000;70:53-60.

w90. Madsen JS, Kristensen SR, Klitgaard NA, Bladbjerg EM, Abrahamsen B, Stilgren L, et al. Effect of long-term hormone replacement therapy on plasma homocysteine in postmenopausal women: a randomized controlled study. Am J Obstet Gynecol. 2002;187:33-9.

w91. McQuillan BM, Beilby JP, Nidorf M, Thompson PL, Hung J. Hyperhomocysteinemia but not the C677T mutation of methylenetetrahydrofolate reductase is an independent risk determinant of carotid wall thickening. The Perth Carotid Ultrasound Disease Assessment Study (CUDAS). Circulation. 1999;99:2383-8.

w92. Moriyama Y, Okamura T, Kajinami K, Iso H, Inazu A, Kawashiri M, et al. Effects of serum B vitamins on elevated plasma homocysteine levels associated with the mutation of methylenetetrahydrofolate reductase gene in Japanese. Atherosclerosis. 2002;164:321-8.

w93. Motti C, Gnasso A, Bernardini S, Massoud R, Pastore A, Rampa P, et al. Common mutation in methylenetetrahydrofolate reductase. Correlation with homocysteine and other risk factors for vascular disease. Atherosclerosis. 1998;139:377-83.

w94. Rothenbacher D, Fischer HG, Hoffmeister A, Hoffmann MM, Marz W, Bode G, et al. Homocysteine and methylenetetrahydrofolate reductase genotype: association with risk of coronary heart disease and relation to inflammatory, hemostatic, and lipid parameters. Atherosclerosis. 2002;162:193-200.

w95. Todesco L, Angst C, Litynski P, Loehrer F, Fowler B, Haefeli WE. Methylenetetrahydrofolate reductase polymorphism, plasma homocysteine and age. Eur J Clin Invest. 1999;29:1003-9.

w96. Verhoef P, Kok FJ, Kluijtmans LA, Blom HJ, Refsum H, Ueland PM, et al. The 677C-->T mutation in the methylenetetrahydrofolate reductase gene: associations with plasma total homocysteine levels and risk of coronary atherosclerotic disease. Atherosclerosis. 1997;132:105-13.

w97. Verhoeff BJ, Trip MD, Prins MH, Kastelein JJ, Reitsma PH. The effect of a common methylenetetrahydrofolate reductase mutation on levels of homocysteine, folate, vitamin B12 and on the risk of premature atherosclerosis. Atherosclerosis. 1998;141:161-6.

w98. Ashfield-Watt PA, Pullin CH, Whiting JM, Clark ZE, Moat SJ, Newcombe RG, et al. Methylenetetrahydrofolate reductase 677C-->T genotype modulates homocysteine responses to a folate-rich diet or a low-dose folic acid supplement: a randomized controlled trial. Am J Clin Nutr. 2002;76:180-6.

w99. Pullin CH, Wilson JF, Ashfield-Watt PA, Clark ZE, Whiting JM, Lewis MJ, et al. Influence of methylenetetrahydrofolate reductase genotype, exercise and other risk factors on endothelial function in healthy individuals. Clin Sci (Lond). 2002;102:45-50.

w100. Trinh BN, Ong CN, Coetzee GA, Yu MC, Laird PW. Thymidylate synthase: a novel genetic determinant of plasma homocysteine and folate levels. Hum Genet. 2002;111:299-302.

w101. Saw SM, Yuan JM, Ong CN, Arakawa K, Lee HP, Coetzee GA, et al. Genetic, dietary, and other lifestyle determinants of plasma homocysteine concentrations in middle-aged and older Chinese men and women in Singapore. Am J Clin Nutr. 2001;73:232-9.

w102. Mager A, Battler A, Birnbaum Y, Magal N, Shohat M. Plasma homocysteine, methylenetetrahydrofolate reductase genotypes, and age at onset of symptoms of myocardial ischemia. Am J Cardiol. 2002;89:919-23.

w103. Soriente L, Coppola A, Madonna P, Cerbone AM, Di Minno G, Orefice G, et al. Homozygous C677T mutation of the 5,10 methylenetetrahydrofolate reductase gene and hyperhomocysteinemia in Italian patients with a history of early-onset ischemic stroke. Stroke. 1998;29:869-71.

w104. Morita H, Taguchi J, Kurihara H, Kitaoka M, Kaneda H, Kurihara Y, et al. Genetic polymorphism of 5,10-methylenetetrahydrofolate reductase (MTHFR) as a risk factor for coronary artery disease. Circulation. 1997;95:2032-6.

w105. Grossmann R, Geisen U, Merati G, Mullges W, Schambeck CM, Walter U, et al. Genetic risk factors in young adults with 'cryptogenic' ischemic cerebrovascular disease. Blood Coagul Fibrinolysis. 2002;13:583-90.

w106. Nakata Y, Katsuya T, Takami S, Sato N, Fu Y, Ishikawa K, et al. Methylenetetrahydrofolate reductase gene polymorphism: relation to blood pressure and cerebrovascular disease. Am J Hypertens. 1998;11:1019-23.

w107. Reuner KH, Ruf A, Kaps M, Druschky KF, Patscheke H. The mutation C677-->T in the methylene tetrahydrofolate reductase gene and stroke. Thromb Haemost. 1998;79:450-1.

w108. Zhang G, Dai C. Gene polymorphisms of homocysteine metabolism-related enzymes in Chinese patients with occlusive coronary artery or cerebral vascular diseases. Thromb Res. 2001;104:187-95.

w109. Yoo JH, Choi GD, Kang SS. Pathogenicity of thermolabile methylenetetrahydrofolate reductase for vascular dementia. Arterioscler Thromb Vasc Biol. 2000;20:1921-5.

w110. Zheng YZ, Tong J, Do XP, Pu XQ, Zhou BT. Prevalence of methylenetetrahydrofolate reductase C677T and its association with arterial and venous thrombosis in the Chinese population. Br J Haematol. 2000;109:870-4.

w111. Salooja N, Catto A, Carter A, Tudenham EG, Grant PJ. Methylene tetrahydrofolate reductase C677T genotype and stroke. Clin Lab Haematol. 1998;20:357-361.

w112. Kostulas K, Crisby M, Huang WX, Lannfelt L, Hagenfeldt L, Eggertsen G, et al. A methylenetetrahydrofolate reductase gene polymorphism in ischaemic stroke and in carotid artery stenosis. Eur J Clin Invest. 1998;28:285-9.

w113. Lopaciuk S, Bykowska K, Kwiecinski H, Mickielewicz A, Czlonkowska A, Mendel T, et al. Factor V Leiden, prothrombin gene G20210A variant, and methylenetetrahydrofolate reductase C677T genotype in young adults with ischemic stroke. Clin Appl Thromb Hemost. 2001;7:346-50.

w114. Lalouschek W, Aull S, Serles W, Schnider P, Mannhalter C, Lang T, et al. Genetic and nongenetic factors influencing plasma homocysteine levels in patients with ischemic cerebrovascular disease and in healthy control subjects. J Lab Clin Med. 1999;133:575-82.

w115. De Stefano V, Chiusolo P, Paciaroni K, Casorelli I, Rossi E, Molinari M, et al. Prothrombin G20210A mutant genotype is a risk factor for cerebrovascular ischemic disease in young patients. Blood. 1998;91:3562-5.

w116. Szolnoki Z, Somogyvari F, Kondacs A, Szabo M, Fodor L, Bene J, et al. Evaluation of the modifying effects of unfavourable genotypes on classical clinical risk factors for ischaemic stroke. J Neurol Neurosurg Psychiatry. 2003;74:1615-20.

w117. Notsu Y, Nabika T, Park HY, Masuda J, Kobayashi S. Evaluation of genetic risk factors for silent brain infarction. Stroke. 1999;30:1881-6.

w118. Li Z, Sun L, Zhang H, Liao Y, Wang D, Zhao B, et al; Multicenter Case-Control Study in China. Elevated plasma homocysteine was associated with hemorrhagic and ischemic stroke, but methylenetetrahydrofolate reductase gene C677T polymorphism was a risk factor for thrombotic stroke: a Multicenter Case-Control Study in China. Stroke. 2003;34:2085-90.

w119. Voetsch B, Damasceno BP, Camargo EC, Massaro A, Bacheschi LA, Scaff M, et al. Inherited thrombophilia as a risk factor for the development of ischemic stroke in young adults. Thromb Haemost. 2000;83:229-33.

w120. Margaglione M, D'Andrea G, Giuliani N, Brancaccio V, De Lucia D, Grandone E, et al. Inherited prothrombotic conditions and premature ischemic stroke: sex difference in the association with factor V Leiden. Arterioscler Thromb Vasc Biol. 1999;19:1751-6.

w121. Pezzini A, Del Zotto E, Magoni M, Costa A, Archetti S, Grassi M, et al. Inherited thrombophilic disorders in young adults with ischemic stroke and patent foramen ovale. Stroke. 2003;34:28-33.

w122. Harmon DL, Doyle RM, Meleady R, Doyle M, Shields DC, Barry R, et al. Genetic analysis of the thermolabile variant of 5, 10-methylenetetrahydrofolate reductase as a risk factor for ischemic stroke. Arterioscler Thromb Vasc Biol. 1999;19:208-11.

w123. Choi BO, Kim NK, Kim SH, Kang MS, Lee S, Ahn JY, et al. Homozygous C677T mutation in the MTHFR gene as an independent risk factor for multiple small-artery occlusions. Thromb Res. 2003;111:39-44.

w124. Kristensen B, Malm J, Nilsson TK, Hultdin J, Carlberg B, Dahlen G, et al. Hyperhomocysteinemia and hypofibrinolysis in young adults with ischemic stroke. Stroke. 1999;30:974-80.

w125. Pezzini A, Del Zotto E, Archetti S, Negrini R, Bani P, Albertini A, et al. Plasma homocysteine concentration, C677T MTHFR genotype, and 844ins68bp CBS genotype in young adults with spontaneous cervical artery dissection and atherothrombotic stroke. Stroke. 2002;33:664-9.

w126. Wu Y, Tomon M, Sumino K. Methylenetetrahydrofolate reductase gene polymorphism and ischemic stroke: sex difference in Japanese. Kobe J Med Sci. 2001;47:255-62.

w127. Press RD, Beamer N, Evans A, DeLoughery TG, Coull BM. Role of a common mutation in the homocysteine regulatory enzyme methylenetetrahydrofolate reductase in ischemic stroke. Diagn Mol Pathol. 1999;8:54-8.

w128. Topic E, Timundic AM, Ttefanovic M, Demarin V, Vukovic V, Lovrencic-Huzjan A, et al. Polymorphism of apoprotein E (APOE), methylenetetrahydrofolate reductase (MTHFR) and paraoxonase (PON1) genes in patients with cerebrovascular disease. Clin Chem Lab Med. 2001;39:346-50.

w129. Pongracz E, Tordai A, Csornai M, Nagy Z. [Genetics of blood coagulation in young stroke patients]. Ideggyogy Sz. 2002;55:111-7.

w130. Li C, Zhang C, Qiu S, Lu X, Zeng Y, Wu H, et al. Polymorphisms of ACE-1 and MTHFR genes and genetic susceptibility of ischemic strok. Zhonghua Yi Xue Za Zhi. 2002;82:1046-9.

w131. Huang Y, Zhao YI Y, Li S. [Hyperhomocysteine, methylenetetrahydrofolate reductase gene, and other risk factors in ischemic stroke]. Zhonghua Yi Xue Za Zhi. 2002;82:119-22.

w132. Kim NK, Choi BO, Jung WS, Choi YJ, Choi KG. Hyperhomocysteinemia as an independent risk factor for silent brain infarction. Neurology. 2003 Dec 9;61(11):1595-9.

w133. Kohara K, Fujisawa M, Ando F, Tabara Y, Niino N, Miki T, et al; NILS-LSA Study. MTHFR gene polymorphism as a risk factor for silent brain infarcts and white matter lesions in the Japanese general population: The NILS-LSA Study. Stroke. 2003 May;34(5):1130-5.

w134. Duca F, Sacchi E, Tagliabue L, Tajoli E. C677T methylenetetrahydrofolate reductase (MTHFR) mutation in stroke [Abstract]. Thromb Haemost. 1997;78(suppl):102.

w135. Gross B. Antebi A, Cassel A, Honigman S. Is a mutation in the enzyme MTHFR a risk factor for stroke in young adults? [Abstract]. Neurology. 2000;54:(suppl 3):142.

w136. De Stefano V, Chiusolo P, Casorelli I, et al. Inborn putative causes of mild hyperhomocysteinemia and risk for ischemic stroke in the young. [Abstract]. Thromb Haemost 2001;85(suppl):P2266.

w137. Pratnicka M, Lewandowski L, Lukasik M, et al. High frequency of combined genetic prothrombotic defects in young patients with ischemic stroke. [Abstract]. Journal of Thrombosis and Haemostasis 2003;1(suppl 1):P0995

w138. Kalashnikova L, Patrusheva L, Kovalenko T, Dobrynina L, Alexandrova E, Nassonov E. Factor V Leiden, prothrombin G20210A, and MTHFR/C677T gene mutations in patients with primary anti-phospolipid syndrome and cerebrovsacular disorder.[Abstract]. Journal of Thrombosis and Haemostasis 2003;1(suppl 1):P0994

w139. Szolnoki Z, Somogyvari F, Kondacs A, Szabo M, Fodor L. Evaluation of the interactions of common genetic mutations in stroke subtypes. J Neurol. 2002;249:1391-7.

w140. Lalouschek W, Aull S, Serles W, Schnider P, Mannhalter C, Pabinger-Fasching I, et al. C677T MTHFR mutation and factor V Leiden mutation in patients with TIA/minor stroke: a case-control study. Thromb Res. 1999;93:61-9.

w141. Morita H, Kurihara H, Sugiyama T, Kitamura K, Suzuki S, Sumiyoshi T, et al. [Genetic polymorphisms of methylenetetrahydrofolate reductase and methionine synthase: association with homocysteine metabolism and late-onset vascular diseases in the Japanese population]. J Cardiol. 1999;33:106-7.

w142. Voetsch B, Benke KS, Damasceno BP, Siqueira LH, Loscalzo J. Paraoxonase 192 Gln-->Arg polymorphism: an independent risk factor for nonfatal arterial ischemic stroke among young adults. Stroke. 2002;33:1459-64.

w143. Kelly PJ, Shih VE, Kistler JP, Barron M, Lee H, Mandell R, et al. Low vitamin B6 but not homocysteine is associated with increased risk of stroke and transient ischemic attack in the era of folic acid grain fortification. Stroke. 2003;34:e51-4.

w144. Madonna P, de Stefano V, Coppola A, Cirillo F, Cerbone AM, Orefice G, et al. Hyperhomocysteinemia and other inherited prothrombotic conditions in young adults with a history of ischemic stroke. Stroke. 2002;33:51-6.

w145. Austin H, Chimowitz MI, Hill HA, Chaturvedi S, Wechsler LR, Wityk RJ, et al; Genetics and Stroke in the Young Study Group. Cryptogenic stroke in relation to genetic variation in clotting factors and other genetic polymorphisms among young men and women. Stroke. 2002 Dec;33(12):2762-8.

Appendix B: References of studies included in the meta-analyses from Chapter 3

s1. Caudill MA, Le T, Moonie SA, Esfahani ST, Cogger EA. Folate status in women of childbearing age residing in Southern California after folic acid fortification. J Am Coll Nutr. 2001;20:129-34

s2. Thogersen AM, Nilsson TK, Dahlen G, Jansson JH, Boman K, Huhtasaari F, et al. Homozygosity for the C677-->T mutation of 5,10-methylenetetrahydrofolate reductase and total plasma homocysteine are not associated with greater than normal risk of a first myocardial infarction in northern Sweden. Coron Artery Dis. 2001;12:85-90.

s3. Schmitz C, Lindpaintner K, Verhoef P, Gaziano JM, Buring J. Genetic polymorphism of methylenetetrahydrofolate reductase and myocardial infarction. A case-control study. Circulation. 1996;94:1812-4.

s4. Chambers JC, Ireland H, Thompson E, Reilly P, Obeid OA, Refsum H, et al. Methylenetetrahydrofolate reductase 677 C-->T mutation and coronary heart disease risk in UK Indian Asians. Arterioscler Thromb Vasc Biol. 2000;20:2448-52.

s5. Fohr IP, Prinz-Langenohl R, Bronstrup A, Bohlmann AM, Nau H, Berthold HK, et al. 5,10-Methylenetetrahydrofolate reductase genotype determines the plasma homocysteine-lowering effect of supplementation with 5-methyltetrahydrofolate or folic acid in healthy young women. Am J Clin Nutr. 2002;75:275-82.

s6. Guinotte CL, Burns MG, Axume JA, Hata H, Urrutia TF, Alamilla A, et al. Methylenetetrahydrofolate reductase 677C-->T variant modulates folate status response to controlled folate intakes in young women. J Nutr. 2003;133:1272-80.

s7. Meisel C, Cascorbi I, Gerloff T, Stangl V, Laule M, Muller JM, et al. Identification of six methylenetetrahydrofolate reductase (MTHFR) genotypes resulting from common polymorphisms: impact on plasma homocysteine levels and development of coronary artery disease. Atherosclerosis. 2001;154:651-8.

s8. Ho CH. The influence of age, sex, vitamin B(12), folate levels and methylenetetrahydrofolate reductase C677T genetic mutations on plasma homocysteine in the Chinese population. Haematologica. 2000;85:1051-4.

s9. Dekou V, Whincup P, Papacosta O, Ebrahim S, Lennon L, Ueland PM, et al. The effect of the C677T and A1298C polymorphisms in the methylenetetrahydrofolate reductase gene on homocysteine levels in elderly men and women from the British regional heart study. Atherosclerosis. 2001;154:659-66.

s10. Voutilainen S, Lakka TA, Hamelahti P, Lehtimaki T, Poulsen HE, Salonen JT. Plasma total homocysteine concentration and the risk of acute coronary events: the Kuopio Ischaemic Heart Disease Risk Factor Study. J Intern Med. 2000;248:217-22.

s11. Thuillier L, Chadefaux-Vekemans B, Bonnefont JP, Kara A, Aupetit J, Rochette C, et al. Does the polymorphism 677C-T of the 5,10-methylenetetrahydrofolate reductase gene contribute to homocysteine-related vascular disease? J Inherit Metab Dis. 1998;21:812-22.

s12. Somekawa Y, Kobayashi K, Tomura S, Aso T, Hamaguchi H. Effects of hormone replacement therapy and methylenetetrahydrofolate reductase polymorphism on plasma folate and homocysteine levels in postmenopausal Japanese women. Fertil Steril. 2002;77:481-6.

s13. Jacques PF, Kalmbach R, Bagley PJ, Russo GT, Rogers G, Wilson PW, et al. The relationship between riboflavin and plasma total homocysteine in the Framingham Offspring cohort is influenced by folate status and the C677T transition in the methylenetetrahydrofolate reductase gene. J Nutr. 2002;132:283-8.

s14. Silaste ML, Rantala M, Sampi M, Alfthan G, Aro A, Kesaniemi YA. Polymorphisms of key enzymes in homocysteine metabolism affect diet responsiveness of plasma homocysteine in healthy women. J Nutr. 2001;131:2643-7.

s15. Bailey LB, Duhaney RL, Maneval DR, Kauwell GP, Quinlivan EP, Davis SR, et al. Vitamin B-12 status is inversely associated with plasma homocysteine in young women with C677T and/or A1298C methylenetetrahydrofolate reductase polymorphisms. J Nutr. 2002;132:1872-8.

s16. Domagala TB, Adamek L, Nizankowska E, Sanak M, Szczeklik A. Mutations C677T and A1298C of the 5,10-methylenetetrahydrofolate reductase gene and fasting plasma homocysteine levels are not associated with the increased risk of venous thromboembolic disease. Blood Coagul Fibrinolysis 2002;13:423-31.

s17. Chango A, Potier De Courcy G, Boisson F, Guilland JC, Barbe F, Perrin MO, et al. 5,10-methylenetetrahydrofolate reductase common mutations, folate status and plasma homocysteine in healthy French adults of the Supplementation en Vitamines et Mineraux Antioxydants (SU.VI.MAX) cohort. Br J Nutr. 2000;84:891-6.

s18. Cappuccio FP, Bell R, Perry IJ, Gilg J, Ueland PM, Refsum H, et al. Homocysteine levels in men and women of different ethnic and cultural background living in England. Atherosclerosis. 2002;164:95-102.

s19. Castro R, Rivera I, Ravasco P, Jakobs C, Blom HJ, Camilo ME, et al. 5,10-Methylenetetrahydrofolate reductase 677C-->T and 1298A-->C mutations are genetic determinants of elevated homocysteine. QJM. 2003;96:297-303.

s20. Chango A, Boisson F, Barbe F, Quilliot D, Droesch S, Pfister M, et al. The effect of 677C-->T and 1298A-->C mutations on plasma homocysteine and 5,10-methylenetetrahydrofolate reductase activity in healthy subjects. Br J Nutr. 2000;83:593-6.

s21. Ma J, Stampfer MJ, Hennekens CH, Frosst P, Selhub J, Horsford J, et al. Methylenetetrahydrofolate reductase polymorphism, plasma folate, homocysteine, and risk of myocardial infarction in US physicians. Circulation. 1996;94:2410-6.

s22. Friedman G, Goldschmidt N, Friedlander Y, Ben-Yehuda A, Selhub J, Babaey S, et al. A common mutation A1298C in human methylenetetrahydrofolate reductase gene: association with plasma total homocysteine and folate concentrations. J Nutr. 1999;129:1656-61.

s23. Husemoen LL, Thomsen TF, Fenger M, Jorgensen HL, Jorgensen T. Contribution of thermolabile methylenetetrahydrofolate reductase variant to total plasma homocysteine levels in healthy men and women. Inter99 (2). Genet Epidemiol. 2003;24:322-30.

s24. Reyes-Engel A, Munoz E, Gaitan MJ, Fabre E, Gallo M, Dieguez JL, et al. Implications on human fertility of the 677C-->T and 1298A-->C polymorphisms of the MTHFR gene: consequences of a possible genetic selection. Mol Hum Reprod. 2002;8:952-7.

s25. Schwartz SM, Siscovick DS, Malinow MR, Rosendaal FR, Beverly RK, Hess DL, et al. Myocardial infarction in young women in relation to plasma total homocysteine, folate, and a common variant in the methylenetetrahydrofolate reductase gene. Circulation. 1997;96:412-7.

s26. Christensen B, Frosst P, Lussier-Cacan S, Selhub J, Goyette P, Rosenblatt DS, et al. Correlation of a common mutation in the methylenetetrahydrofolate reductase gene with plasma homocysteine in patients with premature coronary artery disease. Arterioscler Thromb Vasc Biol. 1997;17:569-73.

s27. Meleady R, Ueland PM, Blom H, Whitehead AS, Refsum H, Daly LE, et al; EC Concerted Action Project: Homocysteine and Vascular Disease. Thermolabile methylenetetrahydrofolate reductase, homocysteine, and cardiovascular disease risk: the European Concerted Action Project. Am J Clin Nutr. 2003;77:63-70.

s28. Hustad S, Ueland PM, Vollset SE, Zhang Y, Bjorke-Monsen AL, Schneede J. Riboflavin as a determinant of plasma total homocysteine: effect modification by the methylenetetrahydrofolate reductase C677T polymorphism. Clin Chem. 2000;46:1065-71.

s29. Yoo JH, Park SC. Low plasma folate in combination with the 677 C-->T methylenetetrahydrofolate reductase polymorphism is associated with increased risk of coronary artery disease in Koreans. Thromb Res. 2000;97:77-84.

s30. Passaro A, Vanini A, Calzoni F, Alberti L, Zamboni PF, Fellin R, et al. Plasma homocysteine, methylenetetrahydrofolate reductase mutation and carotid damage in elderly healthy women. Atherosclerosis. 2001;157:175-80.

s31. Pullin CH, Ashfield-Watt PA, Burr ML, Clark ZE, Lewis MJ, Moat SJ, et al. Optimization of dietary folate or low-dose folic acid supplements lower homocysteine but do not enhance endothelial function in healthy adults, irrespective of the methylenetetrahydrofolate reductase (C677T) genotype. J Am Coll Cardiol. 2001;38:1799-805.

s32. Jee SH, Song KS, Shim WH, Kim HK, Suh I, Park JY, et al. Major gene evidence after MTHFR-segregation analysis of serum homocysteine in families of patients undergoing coronary arteriography. Hum Genet. 2002;111:128-35.

s33. Litynski P, Loehrer F, Linder L, Todesco L, Fowler B. Effect of low doses of 5-methyltetrahydrofolate and folic acid on plasma homocysteine in healthy subjects with or without the 677C-->T polymorphism of methylenetetrahydrofolate reductase. Eur J Clin Invest. 2002;32:662-8.

s34. de Bree A, Verschuren WM, Bjorke-Monsen AL, van der Put NM, Heil SG, Trijbels FJ, et al. Effect of the methylenetetrahydrofolate reductase 677C-->T mutation on the relations among folate intake and plasma folate and homocysteine concentrations in a general population sample. Am J Clin Nutr. 2003;77:687-93.

s35. Ou T, Yamakawa-Kobayashi K, Arinami T, Amemiya H, Fujiwara H, Kawata K, et al. Methylenetetrahydrofolate reductase and apolipoprotein E polymorphisms are independent risk factors for coronary heart disease in Japanese: a case-control study. Atherosclerosis. 1998;137:23-8.

s36. Vizcaino G, Diez-Ewald M, Herrmann FH, Schuster G, Perez-Requejo JL. Relationships between homocysteine, folate and vitamin B12 levels with the methylenetetrahydrofolate reductase polymorphism, in Indians from Western Venezuela. Thromb Haemost. 2001;85:186-7.

s37. D'Angelo A, Coppola A, Madonna P, Fermo I, Pagano A, Mazzola G, et al. The role of vitamin B12 in fasting hyperhomocysteinemia and its interaction with the homozygous C677T mutation of the methylenetetrahydrofolate reductase (MTHFR) gene. A case-control study of patients with early-onset thrombotic events. Thromb Haemost. 2000;83:563-70.

s38. Zittoun J, Tonetti C, Bories D, Pignon JM, Tulliez M. Plasma homocysteine levels related to interactions between folate status and methylenetetrahydrofolate reductase: a study in 52 healthy subjects. Metabolism. 1998;47:1413-8.

s39. Gemmati D, Previati M, Serino ML, Moratelli S, Guerra S, Capitani S, et al. Low folate levels and thermolabile methylenetetrahydrofolate reductase as primary determinant of mild hyperhomocystinemia in normal and thromboembolic subjects. Arterioscler Thromb Vasc Biol. 1999;19:1761-7.

s40. Yingdong Z, Zhigang Z; Yang L. Association of plasma homocysteine level and N^5, N^{10}-methylenetetrahydrofolate reductase gene polymorphism with cerebral infarction. Chinese Medical Sciences Journal. 2002;17:231-5.

s41. Varela ML, Adamczuk YP, Forastiero RR, Martinuzzo ME, Cerrato GS, Pombo G, et al. Major and potential prothrombotic genotypes in a cohort of patients with venous thromboembolism. Thromb Res. 2001;104:317-24.

s42. Lalouschek W, Aull S, Korninger L, Mannhalter C, Pabinger-Fasching I, Schmid RW, et al. 677C to T mutation in the 5,10-methylenetetrahydrofolate reductase (MTHFR) gene and plasma homocysteine levels in patients with TIA or minor stroke. J Neurol Sci. 1998;155:156-62.

s43. Kim CH, Hwang KY, Choi TM, Shin WY, Hong SY. The methylenetetrahydrofolate reductase gene polymorphism in Koreans with coronary artery disease. Int J Cardiol. 2001;78:13-7.

s44. Kosokabe T, Okumura K, Sone T, Kondo J, Tsuboi H, Mukawa H, et al. Relation of a common methylenetetrahydrofolate reductase mutation and plasma homocysteine with intimal hyperplasia after coronary stenting. Circulation. 2001;103:2048-54.

s45. Dunn J, Title LM, Bata I, Johnstone DE, Kirkland SA, O'Neill BJ, et al. Relation of a common mutation in methylenetetrahydrofolate reductase to plasma homocysteine and early onset coronary artery disease. Clin Biochem. 1998;31:95-100.

s46. Kluijtmans LA, Kastelein JJ, Lindemans J, Boers GH, Heil SG, Bruschke AV, et al. Thermolabile methylenetetrahydrofolate reductase in coronary artery disease. Circulation. 1997;96:2573-7.

s47. Morita H, Kurihara H, Tsubaki S, Sugiyama T, Hamada C, Kurihara Y, et al. Methylenetetrahydrofolate reductase gene polymorphism and ischemic stroke in Japanese. Arterioscler Thromb Vasc Biol. 1998;18:1465-9.

s48. Anderson JL, Muhlestein JB, Horne BD, Carlquist JF, Bair TL, Madsen TE, et al. Plasma homocysteine predicts mortality independently of traditional risk factors and C-reactive protein in patients with angiographically defined coronary artery disease. Circulation. 2000;102:1227-32.

s49. Deloughery TG, Evans A, Sadeghi A, McWilliams J, Henner WD, Taylor LM Jr, et al. Common mutation in methylenetetrahydrofolate reductase. Correlation with homocysteine metabolism and late-onset vascular disease. Circulation. 1996;94:3074-8.

s50. Gaustadnes M, Rudiger N, Moller J, Rasmussen K, Bjerregaard Larsen T, Ingerslev J. Thrombophilic predisposition in stroke and venous thromboembolism in Danish patients. Blood Coagul Fibrinolysis. 1999;10:251-9.

s51. Morelli VM, Lourenco DM, D'Almeida V, Franco RF, Miranda F, Zago MA, et al. Hyperhomocysteinemia increases the risk of venous thrombosis independent of the C677T mutation of the methylenetetrahydrofolate reductase gene in selected Brazilian patients. Blood Coagul Fibrinolysis. 2002;13:271-5.

s52. Tokgozoglu SL, Alikasifoglu M, Unsal, Atalar E, Aytemir K, Ozer N, Ovunc K, Usal O, Kes S, Tuncbilek E. Methylene tetrahydrofolate reductase genotype and the risk and extent of coronary artery disease in a population with low plasma folate. Heart. 1999;81:518-22.

s53. Markus HS, Ali N, Swaminathan R, Sankaralingam A, Molloy J, Powell J. A common polymorphism in the methylenetetrahydrofolate reductase gene, homocysteine, and ischemic cerebrovascular disease. Stroke. 1997;28:1739-43.

s54. Candito M, Bedoucha P, Gibelin P, Jambou D, de Franchis R, Sadoul JL, et al. Fasting, postprandial, and post-methionine-load homocysteinaemia and methylenetetrahydrofolate reductase polymorphism in vascular disease. J Inherit Metab Dis. 1999;22:588-92.

s55. Andreassi MG, Botto N, Cocci F, Battaglia D, Antonioli E, Masetti S, et al. Methylenetetrahydrofolate reductase gene C677T polymorphism, homocysteine, vitamin B12, and DNA damage in coronary artery disease. Hum Genet. 2003;112:171-7.

s56. Eikelboom JW, Hankey GJ, Anand SS, Lofthouse E, Staples N, Baker RI. Association between high homocyst(e)ine and ischemic stroke due to large- and small-artery disease but not other etiologic subtypes of ischemic stroke. Stroke. 2000;31:1069-75.

s57. Nakata Y, Katsuya T, Takami S, Sato N, Fu Y, Ishikawa K, et al. Methylenetetrahydrofolate reductase gene polymorphism: relation to blood pressure and cerebrovascular disease. Am J Hypertens. 1998;11:1019-23.

s58. Li Z, Sun L, Zhang H, Liao Y, Wang D, Zhao B, et al; Multicenter Case-Control Study in China. Elevated plasma homocysteine was associated with hemorrhagic and ischemic stroke, but methylenetetrahydrofolate reductase gene C677T polymorphism was a risk factor for thrombotic stroke: a Multicenter Case-Control Study in China. Stroke. 2003;34:2085-90.

s59. Choi BO, Kim NK, Kim SH, Kang MS, Lee S, Ahn JY, et al. Homozygous C677T mutation in the MTHFR gene as an independent risk factor for multiple small-artery occlusions. Thromb Res. 2003;111:39-44.

s60. Huang Y, Zhao YI Y, Li S. [Hyperhomocysteine, methylenetetrahydrofolate reductase gene, and other risk factors in ischemic stroke]. Zhonghua Yi Xue Za Zhi. 2002;82:119-22.

s61. Kim NK, Choi BO, Jung WS, Choi YJ, Choi KG. Hyperhomocysteinemia as an independent risk factor for silent brain infarction. Neurology. 2003 Dec 9;61(11):1595-9.

s62. Kohara K, Fujisawa M, Ando F, Tabara Y, Niino N, Miki T, et al; NILS-LSA Study. MTHFR gene polymorphism as a risk factor for silent brain infarcts and white matter lesions in the Japanese general population: The NILS-LSA Study. Stroke. 2003 May;34(5):1130-5.

s63. Dikmen M, Ozbabalik D, Gunes HV, Degirmenci I, Bal C, Ozdemir G, Basaran A. Acute stroke in relation to homocysteine and methylenetetrahydrofolate reductase gene polymorphisms. Acta Neurol Scand. 2006 May;113(5):307-14.

s64. Fang X, Namba H, Akamine S, Sugiyama K. Methylenetetrahydrofolate reductase gene polymorphisms in patients with cerebral hemorrhage. Neurol Res. 2005 Jan;27(1):73-6.

s65. Fowkes FG, Lee AJ, Hau CM, Cooke A, Connor JM, Lowe GD. Methylene tetrahydrofolate reductase (MTHFR) and nitric oxide synthase (ecNOS) genes and risks of peripheral arterial disease and coronary heart disease: Edinburgh Artery Study. Atherosclerosis. 2000 May;150(1):179-85.

s66. Frederiksen J, Juul K, Grande P, Jensen GB, Schroeder TV, Tybjaerg-Hansen A, Nordestgaard BG. Methylenetetrahydrofolate reductase polymorphism (C677T), hyperhomocysteinemia, and risk of ischemic cardiovascular disease and venous thromboembolism: prospective and case-control studies from the Copenhagen City Heart Study. Blood. 2004 Nov 15;104(10):3046-51.

s67. Hassan A, Hunt BJ, O'Sullivan M, Bell R, D'Souza R, Jeffery S, Bamford JM, Markus HS. Homocysteine is a risk factor for cerebral small vessel disease, acting via endothelial dysfunction. Brain. 2004 Jan;127(Pt 1):212-9.

s68. Kim OJ, Hong SP, Ahn JY, Hong SH, Hwang TS, Kim SO, Yoo W, Oh D, Kim NK. Influence of combined methionine synthase (MTR 2756A > G) and methylenetetrahydrofolate reductase (MTHFR 677C > T) polymorphisms to plasma homocysteine levels in Korean patients with ischemic stroke. Yonsei Med J. 2007 Apr 30;48(2):201-9.

s69. Miyaki K, Omae K, Murata M, Tanahashi N, Saito I, Watanabe K. High throughput multiple combination extraction from large scale polymorphism data by exact tree method. J Hum Genet. 2004;49(9):455-62.

s70. Pezzini A, Grassi M, Del Zotto E, Assanelli D, Archetti S, Negrini R, Caimi L, Padovani A.Interaction of homocysteine and conventional predisposing factors on risk of ischaemic stroke in young people: consistency in phenotype-disease analysis and genotype-disease analysis. J Neurol Neurosurg Psychiatry. 2006 Oct;77(10):1150-6.

s71. Slooter AJ, Rosendaal FR, Tanis BC, Kemmeren JM, van der Graaf Y, Algra A. Prothrombotic conditions, oral contraceptives, and the risk of ischemic stroke. J Thromb Haemost. 2005 Jun;3(6):1213-7.

s72. Miller GJ, Bauer KA, Barzegar S, et al. The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. Thromb Haemost 1995;73:82-6.

s73. Shmeleva, Kasputin S. Genetic polymorphisms and environmental influences in arterial occlusive diseases. Journal of Thrombosis and Haemostasis. 2005;3(Suppl 1): Abstract number: P2273.

s74. Kim IG, Kim JT, Kim MK. Methylenetetrahydrofolate reductase gene single nucleotide polymorphism in korean population with ischemic stroke. Chonnam Medical Journal 2007;43:25-32

s75. Panigrahi I, Chatterjee T, Biswas A, Behari M, Choudhry PV, Saxena R. Role of MTHFR C677T polymorphism in ischemic stroke. Neurol India. 2006 Mar;54(1):48-50.

s76. Karakus Z , Gurkan E. Annals of Medical Sciences. 2005; 14(3):31-36

s77. Zee RY, Mora S, Cheng S, Erlich HA, Lindpaintner K, Rifai N, Buring JE, Ridker PM. Homocysteine, 5,10-methylenetetrahydrofolate reductase 677C>T polymorphism, nutrient intake, and incident cardiovascular disease in 24,968 initially healthy women. Clin Chem. 2007 May;53(5):845-51.

s78. Linnebank M, Moskau S, Farmand S, Fliessbach K, Kölsch H, Bös M, Grothe C, Becker D, Harbrecht U, Pohl C, Wüllner U, Klockgether T. Homocysteine and carotid intima-media thickness in a german population: lack of clinical relevance. Stroke. 2006 Nov;37(11):2840-2.

s79. Marcucci R, Sofi F, Fedi S, Lari B, Sestini I, Cellai AP, Pulli R, Pratesi G, Pratesi C, Gensini GF, Abbate R. Thrombophilic risk factors in patients with severe carotid atherosclerosis. J Thromb Haemost. 2005 Mar;3(3):502-7.

s80. Motti C, Gnasso A, Bernardini S, Massoud R, Pastore A, Rampa P, et al. Common mutation in methylenetetrahydrofolate reductase. Correlation with homocysteine and other risk factors for vascular disease. Atherosclerosis. 1998;139:377-83.

s81. Rothenbacher D, Fischer HG, Hoffmeister A, Hoffmann MM, Marz W, Bode G, et al. Homocysteine and methylenetetrahydrofolate reductase genotype: association with risk of coronary heart disease and relation to inflammatory, hemostatic, and lipid parameters. Atherosclerosis. 2002;162:193-200.

s82. Todesco L, Angst C, Litynski P, Loehrer F, Fowler B, Haefeli WE. Methylenetetrahydrofolate reductase polymorphism, plasma homocysteine and age. Eur J Clin Invest. 1999;29:1003-9.

s83. Verhoef P, Kok FJ, Kluijtmans LA, Blom HJ, Refsum H, Ueland PM, et al. The 677C-->T mutation in the methylenetetrahydrofolate reductase gene: associations with plasma total homocysteine levels and risk of coronary atherosclerotic disease. Atherosclerosis. 1997;132:105-13.

s84. Verhoeff BJ, Trip MD, Prins MH, Kastelein JJ, Reitsma PH. The effect of a common methylenetetrahydrofolate reductase mutation on levels of homocysteine, folate, vitamin B12 and on the risk of premature atherosclerosis. Atherosclerosis. 1998;141:161-6.

s85. Ashfield-Watt PA, Pullin CH, Whiting JM, Clark ZE, Moat SJ, Newcombe RG, et al. Methylenetetrahydrofolate reductase 677C-->T genotype modulates homocysteine responses to a folate-rich diet or a low-dose folic acid supplement: a randomized controlled trial. Am J Clin Nutr. 2002;76:180-6.

s86. Pullin CH, Wilson JF, Ashfield-Watt PA, Clark ZE, Whiting JM, Lewis MJ, et al. Influence of methylenetetrahydrofolate reductase genotype, exercise and other risk factors on endothelial function in healthy individuals. Clin Sci (Lond). 2002;102:45-50.

s87. Trinh BN, Ong CN, Coetzee GA, Yu MC, Laird PW. Thymidylate synthase: a novel genetic determinant of plasma homocysteine and folate levels. Hum Genet. 2002;111:299-302.

s88. Saw SM, Yuan JM, Ong CN, Arakawa K, Lee HP, Coetzee GA, et al. Genetic, dietary, and other lifestyle determinants of plasma homocysteine concentrations in middle-aged and older Chinese men and women in Singapore. Am J Clin Nutr. 2001;73:232-9.

s89. Mager A, Battler A, Birnbaum Y, Magal N, Shohat M. Plasma homocysteine, methylenetetrahydrofolate reductase genotypes, and age at onset of symptoms of myocardial ischemia. Am J Cardiol. 2002;89:919-23.

s90. Soriente L, Coppola A, Madonna P, Cerbone AM, Di Minno G, Orefice G, et al. Homozygous C677T mutation of the 5,10 methylenetetrahydrofolate reductase gene and hyperhomocysteinemia in Italian patients with a history of early-onset ischemic stroke. Stroke. 1998;29:869-71.

s91. Morita H, Taguchi J, Kurihara H, Kitaoka M, Kaneda H, Kurihara Y, et al. Genetic polymorphism of 5,10-methylenetetrahydrofolate reductase (MTHFR) as a risk factor for coronary artery disease. Circulation. 1997;95:2032-6.

s92. Grossmann R, Geisen U, Merati G, Mullges W, Schambeck CM, Walter U, et al. Genetic risk factors in young adults with 'cryptogenic' ischemic cerebrovascular disease. Blood Coagul Fibrinolysis. 2002;13:583-90.

s93. Reuner KH, Ruf A, Kaps M, Druschky KF, Patscheke H. The mutation C677-->T in the methylene tetrahydrofolate reductase gene and stroke. Thromb Haemost. 1998;79:450-1.

s94. Zhang G, Dai C. Gene polymorphisms of homocysteine metabolism-related enzymes in Chinese patients with occlusive coronary artery or cerebral vascular diseases. Thromb Res. 2001;104:187-95.

s95. Yoo JH, Choi GD, Kang SS. Pathogenicity of thermolabile methylenetetrahydrofolate reductase for vascular dementia. Arterioscler Thromb Vasc Biol. 2000;20:1921-5.

s96. Zheng YZ, Tong J, Do XP, Pu XQ, Zhou BT. Prevalence of methylenetetrahydrofolate reductase C677T and its association with arterial and venous thrombosis in the Chinese population. Br J Haematol. 2000;109:870-4.

s97. Salooja N, Catto A, Carter A, Tudenham EG, Grant PJ. Methylene tetrahydrofolate reductase C677T genotype and stroke. Clin Lab Haematol. 1998;20:357-361.

s98. Kostulas K, Crisby M, Huang WX, Lannfelt L, Hagenfeldt L, Eggertsen G, et al. A methylenetetrahydrofolate reductase gene polymorphism in ischaemic stroke and in carotid artery stenosis. Eur J Clin Invest. 1998;28:285-9.

s99. Lopaciuk S, Bykowska K, Kwiecinski H, Mickielewicz A, Czlonkowska A, Mendel T, et al. Factor V Leiden, prothrombin gene G20210A variant, and methylenetetrahydrofolate reductase C677T genotype in young adults with ischemic stroke. Clin Appl Thromb Hemost. 2001;7:346-50.

s100. Lalouschek W, Aull S, Serles W, Schnider P, Mannhalter C, Lang T, et al. Genetic and nongenetic factors influencing plasma homocysteine levels in patients with ischemic cerebrovascular disease and in healthy control subjects. J Lab Clin Med. 1999;133:575-82.

s101. De Stefano V, Chiusolo P, Paciaroni K, Casorelli I, Rossi E, Molinari M, et al. Prothrombin G20210A mutant genotype is a risk factor for cerebrovascular ischemic disease in young patients. Blood. 1998;91:3562-5.

s102. Szolnoki Z, Somogyvari F, Kondacs A, Szabo M, Fodor L, Bene J, et al. Evaluation of the modifying effects of unfavourable genotypes on classical clinical risk factors for ischaemic stroke. J Neurol Neurosurg Psychiatry. 2003;74:1615-20.

s103. Notsu Y, Nabika T, Park HY, Masuda J, Kobayashi S. Evaluation of genetic risk factors for silent brain infarction. Stroke. 1999;30:1881-6.

s104. Voetsch B, Damasceno BP, Camargo EC, Massaro A, Bacheschi LA, Scaff M, et al. Inherited thrombophilia as a risk factor for the development of ischemic stroke in young adults. Thromb Haemost. 2000;83:229-33.

s105. Margaglione M, D'Andrea G, Giuliani N, Brancaccio V, De Lucia D, Grandone E, et al. Inherited prothrombotic conditions and premature ischemic stroke: sex difference in the association with factor V Leiden. Arterioscler Thromb Vasc Biol. 1999;19:1751-6.

s106. Pezzini A, Del Zotto E, Magoni M, Costa A, Archetti S, Grassi M, et al. Inherited thrombophilic disorders in young adults with ischemic stroke and patent foramen ovale. Stroke. 2003;34:28-33.

s107. Harmon DL, Doyle RM, Meleady R, Doyle M, Shields DC, Barry R, et al. Genetic analysis of the thermolabile variant of 5, 10-methylenetetrahydrofolate reductase as a risk factor for ischemic stroke. Arterioscler Thromb Vasc Biol. 1999;19:208-11.

s108. Kristensen B, Malm J, Nilsson TK, Hultdin J, Carlberg B, Dahlen G, et al. Hyperhomocysteinemia and hypofibrinolysis in young adults with ischemic stroke. Stroke. 1999;30:974-80.

s109. Pezzini A, Del Zotto E, Archetti S, Negrini R, Bani P, Albertini A, et al. Plasma homocysteine concentration, C677T MTHFR genotype, and 844ins68bp CBS genotype in young adults with spontaneous cervical artery dissection and atherothrombotic stroke. Stroke. 2002;33:664-9.

s110. Wu Y, Tomon M, Sumino K. Methylenetetrahydrofolate reductase gene polymorphism and ischemic stroke: sex difference in Japanese. Kobe J Med Sci. 2001;47:255-62.

s111. Press RD, Beamer N, Evans A, DeLoughery TG, Coull BM. Role of a common mutation in the homocysteine regulatory enzyme methylenetetrahydrofolate reductase in ischemic stroke. Diagn Mol Pathol. 1999;8:54-8.

s112. Topic E, Timundic AM, Ttefanovic M, Demarin V, Vukovic V, Lovrencic-Huzjan A, et al. Polymorphism of apoprotein E (APOE), methylenetetrahydrofolate reductase (MTHFR) and paraoxonase (PON1) genes in patients with cerebrovascular disease. Clin Chem Lab Med. 2001;39:346-50.

s113. Pongracz E, Tordai A, Csornai M, Nagy Z. [Genetics of blood coagulation in young stroke patients]. Ideggyogy Sz. 2002;55:111-7.

s114. Li C, Zhang C, Qiu S, Lu X, Zeng Y, Wu H, et al. Polymorphisms of ACE-1 and MTHFR genes and genetic susceptibility of ischemic strok. Zhonghua Yi Xue Za Zhi. 2002;82:1046-9.

s115. Duca F, Sacchi E, Tagliabue L, Tajoli E. C677T methylenetetrahydrofolate reductase (MTHFR) mutation in stroke [Abstract]. Thromb Haemost. 1997;78(suppl):102.

s116. Gross B. Antebi A, Cassel A, Honigman S. Is a mutation in the enzyme MTHFR a risk factor for stroke in young adults? [Abstract]. Neurology. 2000;54:(suppl 3):142.

s117. De Stefano V, Chiusolo P, Casorelli I, et al. Inborn putative causes of mild hyperhomocysteinemia and risk for ischemic stroke in the young. [Abstract]. Thromb Haemost 2001;85(suppl):P2266.

s118. Pratnicka M, Lewandowski L, Lukasik M, et al. High frequency of combined genetic prothrombotic defects in young patients with ischemic stroke. [Abstract]. Journal of Thrombosis and Haemostasis 2003;1(suppl 1):P0995

s119. Kalashnikova L, Patrusheva L, Kovalenko T, Dobrynina L, Alexandrova E, Nassonov E. Factor V Leiden, prothrombin G20210A, and MTHFR/C677T gene mutations in patients with primary anti-phospolipid syndrome and cerebrovsacular disorder.[Abstract]. Journal of Thrombosis and Haemostasis 2003;1(suppl 1):P0994

s120. Szolnoki Z, Somogyvari F, Kondacs A, Szabo M, Fodor L. Evaluation of the interactions of common genetic mutations in stroke subtypes. J Neurol. 2002;249:1391-7.

s121. Lalouschek W, Aull S, Serles W, Schnider P, Mannhalter C, Pabinger-Fasching I, et al. C677T MTHFR mutation and factor V Leiden mutation in patients with TIA/minor stroke: a case-control study. Thromb Res. 1999;93:61-9.

s122. Morita H, Kurihara H, Sugiyama T, Kitamura K, Suzuki S, Sumiyoshi T, et al. [Genetic polymorphisms of methylenetetrahydrofolate reductase and methionine synthase: association with homocysteine metabolism and late-onset vascular diseases in the Japanese population]. J Cardiol. 1999;33:106-7.

s123. Voetsch B, Benke KS, Damasceno BP, Siqueira LH, Loscalzo J. Paraoxonase 192 Gln-->Arg polymorphism: an independent risk factor for nonfatal arterial ischemic stroke among young adults. Stroke. 2002;33:1459-64.

s124. Kelly PJ, Shih VE, Kistler JP, Barron M, Lee H, Mandell R, et al. Low vitamin B6 but not homocysteine is associated with increased risk of stroke and transient ischemic attack in the era of folic acid grain fortification. Stroke. 2003;34:e51-4.

s125. Madonna P, de Stefano V, Coppola A, Cirillo F, Cerbone AM, Orefice G, et al. Hyperhomocysteinemia and other inherited prothrombotic conditions in young adults with a history of ischemic stroke. Stroke. 2002;33:51-6.

s126. Austin H, Chimowitz MI, Hill HA, Chaturvedi S, Wechsler LR, Wityk RJ, et al; Genetics and Stroke in the Young Study Group. Cryptogenic stroke in relation to genetic variation in clotting factors and other genetic polymorphisms among young men and women. Stroke. 2002 Dec;33(12):2762-8.

s127. Alluri RV, Mohan V, Komandur S, Chawda K, Chaudhuri JR, Hasan Q. MTHFR C677T gene mutation as a risk factor for arterial stroke: a hospital based study. Eur J Neurol. 2005;12(1):40-4.

s128. Bosco P, Guéant-Rodriguez RM, Anello G, Spada R, Romano A, Fajardo A, Caraci F, Ferri R, Guéant JL. Association of homocysteine (but not of MTHFR 677 C>T, MTR 2756 A>G, MTRR 66 A>G and TCN2 776 C>G) with ischaemic cerebrovascular disease in Sicily. Thromb Haemost. 2006 Aug;96(2):154-9.

s129. Hermans MP, Gala JL, Buysschaert M. The MTHFR CT polymorphism confers a high risk for stroke in both homozygous and heterozygous T allele carriers with Type 2 diabetes. Diabet Med. 2006 May;23(5):529-36.

s130. Lewis SJ, Lawlor DA, Davey Smith G, Araya R, Timpson N, Day IN, Ebrahim S.The thermolabile variant of MTHFR is associated with depression in the British Women's Heart and Health Study and a meta-analysis. Mol Psychiatry. 2006 Apr;11(4):352-60.

s131. Ye H, Yan JT, Shao JM, Zhang F, Hong ML, Wang DW. [A case-control study on the relationship between stroke and plasma homocysteine level and the mutation of MTHFR gene.] Zhonghua Liu Xing Bing Xue Za Zhi. 2004 Nov;25(11):958-61.

s132. Baum L, Wong KS, Ng HK, Tomlinson B, Rainer TH, Chan DK, Thomas GN, Chen X, Poon P, Cheung WS, Woo KS. Methylenetetrahydrofolate reductase gene A222V polymorphism and risk of ischemic stroke. Clin Chem Lab Med. 2004;42(12):1370-6.

s133. Boncoraglio G, Carriero MR, Chiapparini L, Ciceri E, Ciusani E, Erbetta A, Parati EA. Hyperhomocysteinemia and other thrombophilic risk factors in 26 patients with cerebral venous thrombosis. Eur J Neurol. 2004 Jun;11(6):405-9.

s134. Gutiérrez JI, Pérez F, Tamparillas M, Grasa JM, Calvo MT, Osuna C, Sánchez B, Torres M. [Polymorphisms in cystathionine beta-synthase and methylenetetrahydrofolate reductase genes as risk factors for cerebral vascular disease]. Med Clin (Barc). 2005 Mar 5;124(8):281-4.

s135. Kawamoto R, Kohara K, Tabara Y, Miki T, Doi T, Tokunaga H, Konishi I. An association of 5,10-methylenetetrahydrofolate reductase (MTHFR) gene polymorphism and common carotid atherosclerosis. J Hum Genet. 2001;46(9):506-10.

s136. Lalouschek W, Endler G, Schillinger M, Hsieh K, Lang W, Cheng S, Bauer P, Wagner O, Mannhalter C. Candidate genetic risk factors of stroke: results of a multilocus genotyping assay. Clin Chem. 2007 Apr;53(4):600-5.

s137. Linnebank M, Montenarh M, Kölsch H, Linnebank A, Schnez K, Schweichel D, Pohl C, Urbach H, Heun R, Harbrecht U, Klockgether T, Wüllner U. Common genetic variants of homocysteine metabolism in ischemic stroke: a case-control study. Eur J Neurol. 2005 Aug;12(8):614-8.

s138. Pezzini A, Grassi M, Del Zotto E, Archetti S, Spezi R, Vergani V, Assanelli D, Caimi L, Padovani A. Cumulative effect of predisposing genotypes and their interaction with modifiable factors on the risk of ischemic stroke in young adults. Stroke. 2005 Mar;36(3):533-9.

s139. Pezzini A, Grassi M, Iacoviello L, Del Zotto E, Archetti S, Giossi A, Padovani A. Inherited thrombophilia and stratification of ischaemic stroke risk among users of oral contraceptives. J Neurol Neurosurg Psychiatry. 2007 Mar;78(3):271-6.

s140. Roest M, van der Schouw YT, Grobbee DE, Tempelman MJ, de Groot PG, Sixma JJ, Banga JD. Methylenetetrahydrofolate reductase 677 C/T genotype and cardiovascular disease mortality in postmenopausal women. Am J Epidemiol. 2001 Apr 1;153(7):673-9.

s141. Sánchez-Marín B, Grasa JM, Torres M, Calvo MT, Martínez-Jarreta B, García-Erce JA, Giralt M. [Prevalence of methylenetetrahydrofolate reductase C677T mutation among patients with acute ischemic cerebrovascular disease in Aragon] An Med Interna. 2006 Apr;23(4):153-5.

s142. Sazci A, Ergul E, Tuncer N, Akpinar G, Kara I. Methylenetetrahydrofolate reductase gene polymorphisms are associated with ischemic and hemorrhagic stroke: Dual effect of MTHFR polymorphisms C677T and A1298C. Brain Res Bull. 2006 Dec 11;71(1-3):45-50.

s143. Szolnoki Z, Havasi V, Bene J, Komlósi K, Szöke D, Somogyvári F, Kondacs A, Szabó M, Fodor L, Bodor A, Gáti I, Wittman I, Melegh B. Endothelial nitric oxide synthase gene interactions and the risk of ischaemic stroke. Acta Neurol Scand. 2005 Jan;111(1):29-33.

s144. Szolnoki Z, Somogyvári F, Szabó M, Kondacs A, Fodor L, Melegh B. [Interactions between the MTHFR C677T and MTHFR A1298C mutations in ischaemic stroke] Ideggyogy Sz. 2006 Mar 20;59(3-4):107-12.

s145. Uçar F, Sönmez M, Ovali E, Ozmenoglu M, Karti SS, Yilmaz M, Pakdemir A. MTHFR C677T polymorphism and its relation to ischemic stroke in the Black Sea Turkish population. Am J Hematol. 2004 May;76(1):40-3.

s146. Voetsch B, Benke KS, Panhuysen CI, Damasceno BP, Loscalzo J. The combined effect of paraoxonase promoter and coding region polymorphisms on the risk of arterial ischemic stroke among young adults. Arch Neurol. 2004 Mar;61(3):351-6.

s147. Gao X, Yang H, ZhiPing T. Association studies of genetic polymorphism, environmental factors and their interaction in ischemic stroke. Neurosci Lett. 2006 May 8;398(3):172-7.

s148. Yamada Y, Metoki N, Yoshida H, Satoh K, Ichihara S, Kato K, Kameyama T, Yokoi K, Matsuo H, Segawa T, Watanabe S, Nozawa Y. Genetic risk for ischemic and hemorrhagic stroke. Arterioscler Thromb Vasc Biol. 2006 Aug;26(8):1920-5.

s149. Kostulas K, Brophy VH, Moraitis K, Manolescu A, Kostulas V, Gretarsdottir S, Cheng S, Hillert J. Genetic profile of ischemic cerebrovascular disease and carotid stenosis. Acta Neurol Scand. 2008 Sep;118(3):146-52.

s150. Semmler A, Linnebank M, Krex D, Götz A, Moskau S, Ziegler A, Simon M. Polymorphisms of homocysteine metabolism are associated with intracranial aneurysms. Cerebrovasc Dis. 2008;26(4):425-9.

s151. Fang L, Wu YQ, Wang TW. Correlations of polymorphism of methylenetetrahydrofolate reductase and cystathionine beta-synthase genes with heredity of cerebral infarction and cerebral hemorrhage in northern Chinese Han people. Zhongguo Linchuang Kangfu (2004), 8(22), 4654-4656.

s152. Gao HF, Liu JG, Zhang ZC, et al. The relationship between hyperhomocysteinemia and MTHFR gene in young adult cerebral infarction patients. Tian Jin Yi Yao. 2003;10:627-629

s153. Li YJ, Huo Y, Gao XG, et al. A statistical study on the relationship between gene polymorphism of b-fibrinogen -148C/T, N5,N10-methylene tetrahydrofolic acid reductase 677C/T and cerebral infarction. Zhong Guo Shen Jing Mian Yi Xue He Shen Jing Bing Xue Za Zhi. 2005;12:363-367.

s154. Qin MZ, Qin YZ. Correlation between Hcy and polymorphism of gene methylenetetrahydrofolate reductase and ischemic cerebral stroke. Zhing Guo Mei Tan Gong Ye Yi Xue Za Zhi. 2005;8:474-475.

s155. Sun WP, Wan Qi, Su MQ. Genetic polymorphism of 5,10-methylenetetrahydrofolate reductase (MTHFR) in patients with cerebral infarction. Zhong Hua Lao Nian Xin Nao Xue Guan Bing Za Zhi. 2003;5:36-38.

s156. Teng L, Zhao YB, Wang XC, et al. Relation between MTHFR gene polymorphism and arteriosclerotic cerebral infarction. Yi Chuan. 1999; 21:13-16.

s157. Tu CQ, Huang MQ, Wu JZ, et al. The relation of methylenetetrahydrofolate reductase gene mutation and cystathionine B-synthase gene mutation exist in cerebral infarction. Xue Shuan Yu Zhi Xue Xue. 2002; 8:149-151.

s158. Xiao YQ, Jiang LL, Lu Q, et al. Relationship between gene polymorphism of methylenetetrahydrofolate reductase and plasma homocysteine level with cerebrovascular disease. Jian Yan Yi Xue. 2006;21:201-204.

s159. Zhang YD, Zhu ZG, Liu Y. The association of plasma Hcy level and MTHFR gene polymorphism with cerebral infarction. Zhong Feng Yu Shen Jing Ji Bing Za Zhi. 2001; 18:198-200

s160. Zhang Y, Xie RP, Wang YH, et al. Study about association between MTHFR gene polymorphism and cerebral infarction. Zhong Hua Shen Jing Ke Za Zhi. 2003; 36:73-74.

s161. Fang L. et al. Journal of Jilin University (Medicine Edition). 2007; 33(2):310-313

s162. McIlroy SP, Dynan KB, Lawson JT, Patterson CC, Passmore AP. Moderately elevated plasma homocysteine, methylenetetrahydrofolate reductase genotype, and risk for stroke, vascular dementia, and Alzheimer disease in Northern Ireland. Stroke. 2002;33:2351-6.

s163. Li CM, et al. Chin Crit Care Med. 2006; 18(5):264-267

s164. Nan GX, et al. Chinese Journal of Clinical Rehabilitation. 2006; 10(4):184-186

s165. NG KM, et al. Hereditas (Beijing). 2004; 26(3):298-302

s166. Gadelha TB, Nogueira MA, Nucci M. Thrombophilic predisposition to ischemic stroke in Brazilian young adults. Thrombosis and Haemostasis. 2001 (suppl) P2202 (abstr.).

s167. Toulon P, Zuber M, Mas JL. Genetic polymorphisms and the risk of development of ischemic stroke. Results of a case control study. Journal of Thrombosis and Haemostasis. 2003;1(Suppl 1): Abstract P0998.

Appendix C: Published papers

1. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. Lancet. 2005;365:224-32.

2. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, Packard CJ, O'Reilly DS, Juhan-Vague I, Yudkin JS, Tremoli E, Margaglione M, Di Minno G, Hamsten A, Kooistra T, Stephens JW, Hurel SJ, Livingstone S, Colhoun HM, Miller GJ, Bautista LE, Meade T, Sattar N, Humphries SE, Hingorani AD. Insight into the nature of the CRP-coronary event association using Mendelian randomization. Int J Epidemiol. 2006;35:922-31.

3. Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in nongenetic observational studies using "mendelian triangulation". Ann Epidemiol. 2006;16:675-80.

4. Hingorani AD, Shah T, Casas JP. Linking observational and genetic approaches to determine the role of C-reactive protein in heart disease risk. European Heart Journal 2006; 27:1261-1263.

5. Casas JP, Shah T, Hingorani AD, Danesh J, Pepys MB. C-reactive protein and coronary heart disease: a critical review. J Intern Med. 2008;264:295-314.

Homocysteine and stroke: evidence on a causal link from mendelian randomisation

Juan P Casas, Leonelo E Bautista, Liam Smeeth, Pankaj Sharma*, Aroon D Hingorani*

Lancet 2005; 365: 224–32

See Comment page 194

Centre for Clinical Pharmacology, Department of Medicine, British Heart Foundation Laboratories, Royal Free and University College Medical School, University College London, Rayne Building, 5 University Street, London WC1E 6JJ, UK (J P Casas MD, A D Hingorani FRCP); Population Health Sciences, University of Wisconsin Medical School, Madison, WI, USA (L E Bautista DrPH); Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK (L Smeeth MRCGP); and Hammersmith Hospitals Acute Stroke Unit and Department of Cellular and Molecular Neuroscience, Imperial College London, London, UK (P Sharma PhD)

Correspondence to: Dr Aroon D Hingorani a.hingorani@ucl.ac.uk or Dr Pankaj Sharma psharma@imperial.ac.uk

*These authors contributed equally to the study

Summary

Background Individuals homozygous for the T allele of the *MTHFR* C677T polymorphism have higher plasma homocysteine concentrations (the phenotype) than those with the CC genotype, which, if pathogenetic, should put them at increased risk of stroke. Since this polymorphism is distributed randomly during gamete formation, its association with stroke should not be biased or confounded. We investigated consistency between the expected odds ratio for stroke among TT homozygotes, extrapolated from genotype–phenotype and phenotype–disease studies, and the observed odds ratio from a meta-analysis of genotype–disease association studies.

Methods We searched MEDLINE and EMBASE up to June, 2003, for all relevant studies on the association between homocysteine concentration and the *MTHFR* polymorphism, and until December, 2003, for those on the association between the polymorphism and the risk of stroke. Pooled odds ratios and 95% CI were calculated by random-effects and fixed-effects models. Consistency between expected and observed odds ratios was assessed by interaction test.

Findings 111 studies met the selection criteria. Among 15 635 people without cardiovascular disease, the weighted mean difference in homocysteine concentration between TT and CC homozygotes was 1·93 μmol/L (95% CI 1·38 to 2·47). The expected odds ratio for stroke corresponding to this difference based on previous observational studies was 1·20 (1·10 to 1·31). In our genetic meta-analysis (n=13 928) the odds ratio for stroke was 1·26 (1·14 to 1·40) for TT versus CC homozygotes, similar to the expected odds ratio (p=0·29). Consistency between the odds ratios was preserved in analyses by age-group, ethnic background, and geographical location.

Interpretation The observed increase in risk of stroke among individuals homozygous for the *MTHFR* T allele is close to that predicted from the differences in homocysteine concentration conferred by this variant. This concordance is consistent with a causal relation between homocysteine concentration and stroke.

Introduction

Stroke is the third most common cause of death in more developed countries.[1] About 80% of strokes are thromboembolic (ischaemic) in origin and the remainder are haemorrhagic.[1–3] In the UK, stroke is the largest single cause of severe disability with more than 125 000 incident strokes and about 60 000 deaths due to stroke each year.[2] Because treatments for stroke are limited, the best approach to reduce mortality and morbidity is primary prevention through modification of acquired risk factors (eg, high blood pressure, smoking, diabetes, and atrial fibrillation).[3]

Data from cohort and case-control studies suggest that a raised circulating concentration of homocysteine is associated with a higher risk of stroke.[4–6] However, homocysteine concentration is also related to smoking status, blood pressure, and social class and is higher in people with existing atherosclerosis than in those without.[7,8] Therefore, this relation could be subject to residual confounding, reverse-causality bias, or both.[7,8]

A common functional polymorphism, C677T, in the gene encoding methylenetetrahydrofolate reductase (MTHFR), an enzyme involved in homocysteine metabolism, has been associated with differences in homocysteine concentration.[4,9,10] Since carriage of this variant is subject to the random assortment of maternal and paternal alleles at the time of gamete formation, according to Mendel's second law[11] associations between *MTHFR* genotype and homocysteine concentration or stroke should not be subject to reverse-causality bias and should also be largely free from confounding by other determinants of homocysteine concentration or risk factors for stroke.[9] Moreover, genotype is a fixed characteristic, so there is unlikely to be regression-dilution bias, which results from measurement error and biological variability of the exposure under assessment and leads to underestimation of the association between a risk factor and disease.[12,13] Therefore, if homocysteine increases the risk of stroke, carriage of the *MTHFR* polymorphism that exposes individuals to an increased homocysteine concentration should confer an increased risk of stroke proportional to the difference in homocysteine concentration attributable to variant and to the relative risk observed in non-genetic observational studies.

The investigation of consistency between risk estimates obtained from genotype–disease studies and those from phenotype–disease studies, to provide insight into the nature of the observed associations, has been referred to as "mendelian randomisation".[12,13] This approach has been used to test the nature of the association between homocysteine concentration and coronary heart disease, venous thromboembolism, and stroke.[4,9] However, in the previous genetic analyses of

the *MTHFR* polymorphism and stroke, the number of available studies was small. Also, information on the relative effect of the *MTHFR* polymorphism on homocysteine concentration in different geographical locations, ethnic groups, and age-groups was limited.[4] Demonstration that homocysteine concentration is causally related to the development of stroke would have important implications in primary prevention, because administration of folic acid is known to lower homocysteine concentrations,[14] and a policy of fortification of cereal with folic acid to lower the incidence of neural-tube defects has already been undertaken in North America.

We have carried out two updated and comprehensive meta-analyses. In the first, we investigated the extent to which homocysteine concentrations are determined by the *MTHFR* C677T polymorphism. We also assessed the effect of pre-existing cardiovascular disease, geographical location, ethnic background, and age on this association. In the second meta-analysis, we estimated the odds ratio of stroke conferred by the TT genotype. Our meta-analysis of *MTHFR* C677T polymorphism and homocysteine concentration includes data from 48 more studies than a previous meta-analysis[4] and extends the observations to people without cardiovascular disease. Our meta-analysis of the *MTHFR* C677T polymorphism and stroke includes data from 23 more studies than the previous meta-analysis[4] and assessed the robustness of the results by sensitivity analysis.

Methods

Two electronic databases (MEDLINE and EMBASE) were searched up to June, 2003, for all studies on the association between the *MTHFR* C677T polymorphism and homocysteine concentrations and up to December, 2003, for studies on the association between the *MTHFR* C677T polymorphism and stroke. For the first search, we used the text words, which were also MeSH terms, "polymorphism", "mutation", "genes", and "cardiovascular disease" in combination with "homocysteine". For the second search, the terms used were "stroke", "brain infarction", "cerebral isch(a)emia", "h(a)emorrhagic stroke" and "silent brain infarction" in combination with "genetic", "polymorphism", "mutation", or "genes". Both literature searches were limited to "human" and "English language". We searched for any additional studies in the references of all identified publications, including previous relevant meta-analyses.[4,15,16]

Selection criteria

For inclusion in the meta-analysis on homocysteine concentrations and the *MTHFR* polymorphism, studies had to have an analytical design (case-control, cohort, or cross-sectional) and had to examine the association between homocysteine concentrations and the polymorphism. Studies were included only if they were published as full-length articles or letters in peer-reviewed journals. For duplicate publications the smaller dataset was excluded.

For inclusion in the meta-analysis on *MTHFR* and stroke, studies had to involve unrelated individuals and to examine the associations between ischaemic or haemorrhagic stroke and the presence of the polymorphism. For the main comparison, only studies published as full-length articles or letters in peer-reviewed journals in English were included. However, to assess the robustness of the association, we did a sensitivity analysis that included silent brain infarction as an outcome, non-full-text papers, and papers published in languages other than English. In all searches, when relevant information was not reported or there was doubt about duplicate publications, we contacted the authors to obtain the required information.

Data extraction

Data for analysis (country of origin, study design, mean age of participants, frequency of genotypes and alleles, homocysteine and folate concentrations, ethnic background, and frequency of cardiovascular risk factors) were extracted and entered into databases by two of us (JPC and PS). The results were compared and disagreements resolved by consensus.

Statistical analysis

We obtained a summary estimate of the effect of raised plasma homocysteine concentrations on risk of stroke from a recently published meta-analysis by Wald and others of eight prospective studies, which included 676 stroke cases, mainly ischaemic in aetiology, in white people, carried out in European and North American countries.[4] In that analysis, a difference of 5 μmol/L in plasma homocysteine was associated with an odds ratio for stroke of 1·59 (95% CI 1·29 to 1·96) adjusted for confounding variables and for regression-dilution bias.[4] We did a meta-analysis to obtain the weighted mean difference in plasma homocysteine concentrations between individuals homozygous for the T allele and those homozygous for the C allele. The weighted mean difference was obtained separately for people with and without known cardiovascular disease (ischaemic heart disease, stroke, or venous thrombosis) and for both groups combined. For these analyses, a random-effects model[17] was used to allow for any heterogeneity across studies. Estimates of the weighted mean difference were obtained for different age-groups, by geographical location, and by ethnic background. We then did a second meta-analysis of all published studies to obtain a summary odds ratio for all strokes for individuals homozygous for the T allele compared with those homozygous for the C allele. Fixed-effects summary odds ratios and 95% CI were calculated by the Mantel-Haenszel method,[18,19] and DerSimonian and Laird's method[17] was used to calculate random-effects summary odds ratios and

their 95% CI. Also, to test the robustness of our findings, we calculated different odds ratios according to outcome (ischaemic stroke confirmed by MRI or CT, haemorrhagic stroke, or silent brain infarction), ethnic background (white and non-white), publication language (English and other), and type of publication (full text or abstract).

We used the DerSimonian and Laird Q test[20] to assess the degree of heterogeneity between studies, and funnel plots and Egger's regression asymmetry test to assess small-study bias, of which publication bias is one potential cause.[21] In addition, the influence of individual studies on the summary odds ratio was investigated by re-estimation and plotting of the summary odds ratio in the absence of each study. Meta-regression was used to assess the extent to which different variables explained heterogeneity in the weighted mean difference and in the summary odds ratios.[22] Finally, we used the weighted mean difference in homocysteine concentration by *MTHFR* C677T polymorphism to estimate an expected increase in the risk of stroke assuming that an increase of 5 μmol/L in plasma homocysteine would be associated with an odds ratio for stroke of 1·59 (1·26–1·96) and that this association follows a log-linear relation.[4] The uncertainty surrounding the expected odds ratio is a function of the variability of the weighted mean difference in homocysteine by genotype and the variability of the summary odds ratio from Wald's meta-analysis[4] and cannot be directly calculated. Therefore, to obtain a 95% CI for this odds ratio, we generated a million values from a normal distribution with mean and SD equal to the weighted mean difference and its SE, and a million values from a normal distribution with mean and SD equal to the natural logarithm of the summary odds ratio from Wald's meta-analysis[4] and its SE (calculated from its 95% CI). We used the simulated values to calculate a million estimates of this odds ratio and took the 2·5% and 97·5% centile values of the created empirical distribution as 95% confidence limits. Then the expected odds ratio was compared with the summary odds ratio obtained from the meta-analysis of genetic studies by means of an interaction test.[23] Consistency between the two odds ratios would indicate that the association between plasma homocysteine concentration and stroke seen in non-genetic epidemiological studies is unlikely to be severely affected by residual or reverse-causality bias. Data were analysed by use of the Review Manager software (version 4.2) from the Cochrane Collaboration 2003 and Stata (version 8.0).

Role of the funding source

No funding source had any role in study design; collection, analysis, or interpretation of data; or the writing of the report. All authors had full access to all the data in the study, and all took full responsibility for the decision to submit the paper for publication.

Results

The primary search for studies on homocysteine and the *MTHFR* C677T polymorphism generated 104 potentially relevant studies (see webreferences, numbered w1–w145, at http://image.thelancet.com/extras/03art11437webreferences.pdf) of which 81 met the selection criteria (w1–w81). Of the 23 articles excluded, 16 (w82–w97) did not provide sufficient data for us to calculate the weighted mean difference between the genotypes, and the relevant information could not be obtained from the authors. Two (w98, w99) were discarded as probable duplication, and five (w100–w104) reported only the homocysteine concentrations for TT homozygotes and C-allele carriers but not for the CC genotype alone.

Of the 81 studies included (31 355 individuals), information for the main comparison (TT *vs* CC genotype) was available for 15 635 people without known cardiovascular disease, 6312 with cardiovascular disease, and for 9408 reported only as the combination of individuals with and without cardiovascular disease.

Among individuals without cardiovascular disease (webreferences w1–w41), the weighted mean difference in plasma homocysteine between those homozygous for the T allele and those homozygous for the C allele was 1·93 μmol/L (95% CI 1·38 to 2·47; p<0·0001, figure 1). There was significant inter-study heterogeneity (p for heterogeneity <0·0001). From the variables examined in a meta-regression analysis, the only major source of heterogeneity detected was the mean concentration of serum folate. The weighted mean difference in plasma homocysteine concentrations comparing the TT and CC genotypes was 0·048 μmol/L less for each 1 nmol/L increase in mean serum folate (p=0·035). The weighted mean difference in homocysteine concentration was lower in studies done in North America than in those done in Europe or other continents (table 1). Conversely, the weighted mean serum folate concentration was higher in North American (25·3 nmol/L [7·3 to 43·3]) than in European studies (13·6 nmol/L [11·4 to 15·6]) and those conducted in other continents (15·3 nmol/L [9·0 to 21·6]). No other sources of heterogeneity by age (p=0·94), ethnic background (white *vs* other; p=0·62), smoking status (p=0·41), or sex (p=0·57) were observed in a meta-regression analysis of studies in which these variables were reported.

The distribution of the weighted mean difference in relation to its SD in the funnel plot was symmetrical, and the result of Egger's test was not significant (p=0·72) providing no positive evidence for small-study bias. Visual assessment of a graph of the individual weighted mean difference for each study showed that none of the studies had an undue influence on the overall weighted mean difference.

Among people with cardiovascular disease (w2, w4, w7, w11, w16, w21, w26, w27, w29, w32, w37, w39– w55), the

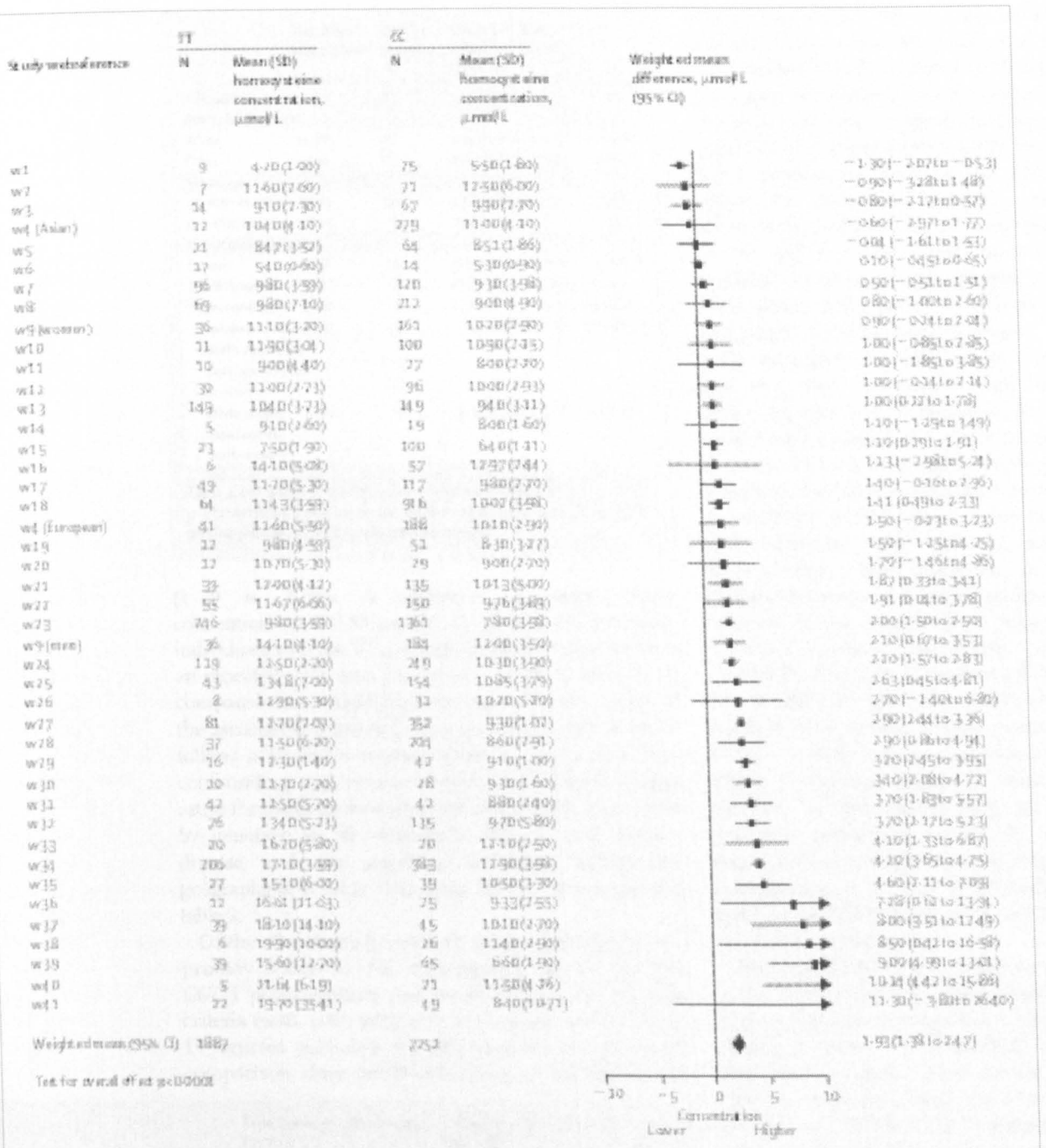


Figure 1: **Weighted mean differences in plasma homocysteine concentration according to the *MTHFR* C677T genotype (TT *vs* CC) among people without known cardiovascular disease**

For webreferences see http://image.thelancet.com/extras/03art11437webreferences.pdf. The total number of people without known cardiovascular disease included in the meta-analysis was 15 635 (TT 1887, CT 6496, CC 7252).

mean homocysteine concentration was 4·35 μmol/L (3·22 to 5·49; p<0·0001) higher for those homozygous for the T allele than for those homozygous for the C allele. A meta-regression analysis indicated that the greater difference in homocysteine by genotype among people with cardiovascular disease was explained partly by lower serum folate concentrations (the crude β coefficient for disease status was 3·04; after adjustment for serum folate it was 2·73).

When the data from all studies, including those that investigated individuals with and without cardiovascular disease without distinction, were combined (w1–w81), the weighted mean difference for homozygotes for the T allele against homozygotes for the C allele was 3·10 μmol/L (2·54 to 3·65; p<0·0001).

According to a previous meta-analysis of prospective studies,[4] a plasma homocysteine concentration higher by 5 μmol/L corresponds to an odds ratio for stroke of 1·59

	Number of participants	Number of studies	Weighted mean difference, μmol/L (95% CI)	p
All studies	9139	43	1·93 (1·38 to 2·47)	<0·0001
Ethnic background				
White	6948	32	1·96 (1·41 to 2·51)	<0·0001
Other	2191	11	1·83 (0·58 to 3·09)	0·004
Mean age				
40 years or older	8575	36	1·96 (1·46 to 2·47)	<0·0001
Less than 40 years	564	7	1·66 (0 to 3·31)	0·05
Geographical location				
Europe	7031	26	2·04 (1·45 to 2·64)	<0·0001
North America	1027	8	0·57 (−0·28 to 1·42)	0·19
Other continents	1081	9	2·97 (1·72 to 4·22)	<0·0001
European and North-American, mean age 40 years or older	7742	29	1·84 (1·29 to 2·39)	<0·0001
White people mean age 40 years or older	6747	29	1·95 (1·39 to 2·51)	<0·0001

Table 1: **Weighted mean difference in plasma homocysteine concentrations according to the *MTHFR* C677T genotype (TT vs CC) among people without cardiovascular disease**

(1·29 to 1·96). A difference in homocysteine concentration of 1·93 μmol/L (1·38 to 2·47) in healthy individuals with the TT genotype would therefore result in an expected odds ratio for stroke of 1·20 (1·10 to 1·31) compared with individuals homozygous for the C allele, if the association between homocysteine and risk of stroke follows a log-linear relation (table 2), and is free from confounding and reverse-causality bias. Expected odds ratios for stroke derived from differences in homocysteine by genotype in all individuals without cardiovascular disease, and also separately by ethnic background, geographical location, and mean age are summarised in table 2.

Of the 49 potentially relevant studies identified in the primary search for the meta-analysis on the *MTHFR* C677T polymorphism and stroke, 30 met the selection criteria (w40, w47, w53, w59, w71, w105–w128). Of the 19 articles excluded for the purpose of the main comparison, three (w129–w131) were published in non-English journals, two (w132, w133) reported the outcome silent brain infarction, and five (w134–w138) were published only as abstracts. Though excluded from the main comparison, these studies were used in a sensitivity analysis. Of the remaining nine publications excluded, five studies (w62, w139–w142) encompassed some duplication. Four (w103, w143–w145) did not report the genotype frequency, and the relevant information could not be obtained from the authors. Of the 30 studies (a total of 6324 cases and 7604 controls) included, 19 were in white people, ten involved Asian participants, and one included both white individuals and people of Afro-Caribbean origin.

The summary odds ratio, under a fixed-effects model, indicated that individuals with the TT genotype compared with those homozygous for the C allele had an odds ratio for stroke of 1·26 (1·14 to 1·40; p<0·0001; figure 2). There was significant heterogeneity among the results of individual studies (p for heterogeneity=0·034). A sensitivity analysis showed that the study by Morita and colleagues (w47) was the main cause of the heterogeneity. After exclusion of this study, the heterogeneity was no longer significant (p=0·32) but the estimate of the overall effect changed very little and remained significant (odds ratio 1·20 [1·08 to 1·34]; p=0·0006). Similarly, a random-effects model that took into account the variability within and between studies resulted in a similar overall estimate (1·26 [1·07 to 1·47]; p=0·004). A meta-regression analysis showed that ethnic background (white *vs* other, p=0·23) and the presence of risk factors such as age (p=0·38), sex (p=0·46), hypertension (p=0·15), smoking (p=0·43), and diabetes (p=0·10) were not significant sources of heterogeneity in a group of 20 studies (w47, w53, w59, w105, w108, w109, w111, w116–w127) with information on all these variables.

The distribution of the odds ratio in relation to its SD in the funnel plot was symmetrical, and the result of Egger's test was not significant (p=0·32) providing no positive evidence of small-study bias. No individual study had an undue influence on the summary odds ratio. A sensitivity analysis showed a robust association between the *MTHFR* C677T polymorphism and stroke (figure 3). No significant changes in the summary odds ratio were detected after addition of studies published in languages other than English (w129–w131) or in abstract form (w134–w138), or those that used the outcome silent brain infarction (w132, w133). Similarly, no differences in the summary odds ratio were observed when the analysis was restricted according to ethnic background or to studies in which the outcome was solely ischaemic stroke (figure 3). Data from five studies (w106, w110, w111, w118, w129; 611 cases and 2405 controls) for which information was available on the association of the *MTHFR* C677T polymorphism and haemorrhagic stroke (TT *vs* CC) gave a summary odds ratio of 1·16 (0·90 to 1·50; p=0·25), under a fixed-effects model.

	Expected odds ratio for stroke (95% CI)*	Observed genetic odds ratio (95% CI)†	p‡
Main comparison			
All studies	1·20 (1·10 to 1·31)	1·26 (1·14 to 1·40)	0·29
Secondary comparisons			
Studies with mean age 40 years or older	1·20 (1·10 to 1·30)	1·26 (1·14 to 1·40)§	0·30
Europe and North America only	1·17 (1·09 to 1·27)	1·21 (1·02 to 1·43)	0·37
White people only	1·20 (1·10 to 1·31)	1·19 (1·02 to 1·39)	0·39

All odds ratios are based on comparisons of TT *vs* CC genotypes. *Odds ratios were calculated only for individuals without cardiovascular disease. Mean expected odds ratios were calculated with the formula: expected odds ratio=1·59 raised to the power of d/5, where d=weighted mean difference in homocysteine by genotype, on the assumption that a 5 μmol/L increase in homocysteine is associated with an odds ratio for stroke of 1·59.[4] †Odds ratios obtained from the meta-analysis of genotype–disease association studies of *MTHFR* C677T and stroke. ‡For the comparison by use of an interaction test. §The weighted mean age was 58 years for cases and 53 years for controls.

Table 2: **Assessment of consistency between odds ratios derived by extrapolation from phenotype–disease studies and those derived from meta-analysis of genetic studies**

Estimates of odds ratios for TT homozygous individuals in comparison with those who had the CC genotype, extrapolated from the genotype effect on homocysteine and the homocysteine–stroke association, were mathematically very similar to those from our genotype–disease meta-analysis (table 2). The p value for interaction for the main comparison of the study, the non-genetic and genetic odds ratios for stroke in TT versus CC individuals, derived from all the data, was 0·29. Therefore, there was no evidence of a significant difference between the expected odds ratio estimated from the increment in the homocysteine concentrations by genotype (*MTHFR* C677T) and that observed from the genotype–disease meta-analysis. Moreover, no significant differences between estimated odds ratios were observed when similar comparisons were done among more homogeneous groups defined by geographical location, ethnic background, or mean age (table 2). In all the comparisons, the expected odds ratios were within the 95% CI of the corresponding observed odds ratio obtained from the genetic association studies.

Discussion

The main finding of these meta-analyses was that the odds ratio for stroke conferred by the *MTHFR* TT genotype was similar to that estimated by use of the homocysteine difference by genotype and homocysteine–stroke odds ratio from phenotype–disease studies. Indeed, after exclusion of the study of Morita and colleagues (w47) which was the cause of much of the heterogeneity in the genotype–stroke analysis, the summary odds ratio for the *MTHFR* TT genotype was 1·20 (95% CI 1·08 to 1·34), identical to the predicted effect of the polymorphism estimated from the studies of genotype–homocysteine differences and homocysteine–stroke risk. Because of the random allocation of genotype in advance of disease development, these results imply that the relation between homocysteine concentration and stroke seen in phenoype–disease studies is not subject to substantial residual confounding or reverse-causality bias. Thus our study provides evidence for a role for homocysteine in stroke pathogenesis.

Our first meta-analysis of studies examining the association between homocysteine concentration and *MTHFR* genotype, involving more than 31 000 people, allowed us to refine the size estimate of the effect of the *MTHFR* C677T polymorphism on plasma homocysteine concentration and to explore potential sources of heterogeneity. The absolute difference in homocysteine concentration conferred by the genotype, though consistent in direction, was greater in people with established atherosclerosis than in individuals who were healthy at the time of measurement. A meta-analysis of individual patients' data had similar findings.[9] In that study, among individuals with coronary heart disease, there was a difference in homocysteine concentration between TT and CC individuals of 2·2 μmol/L; the difference among healthy individuals was smaller, but still significant (1·5 μmol/L).[9] Therefore, our data suggest the presence of an additional effect of disease status on the association between homocysteine

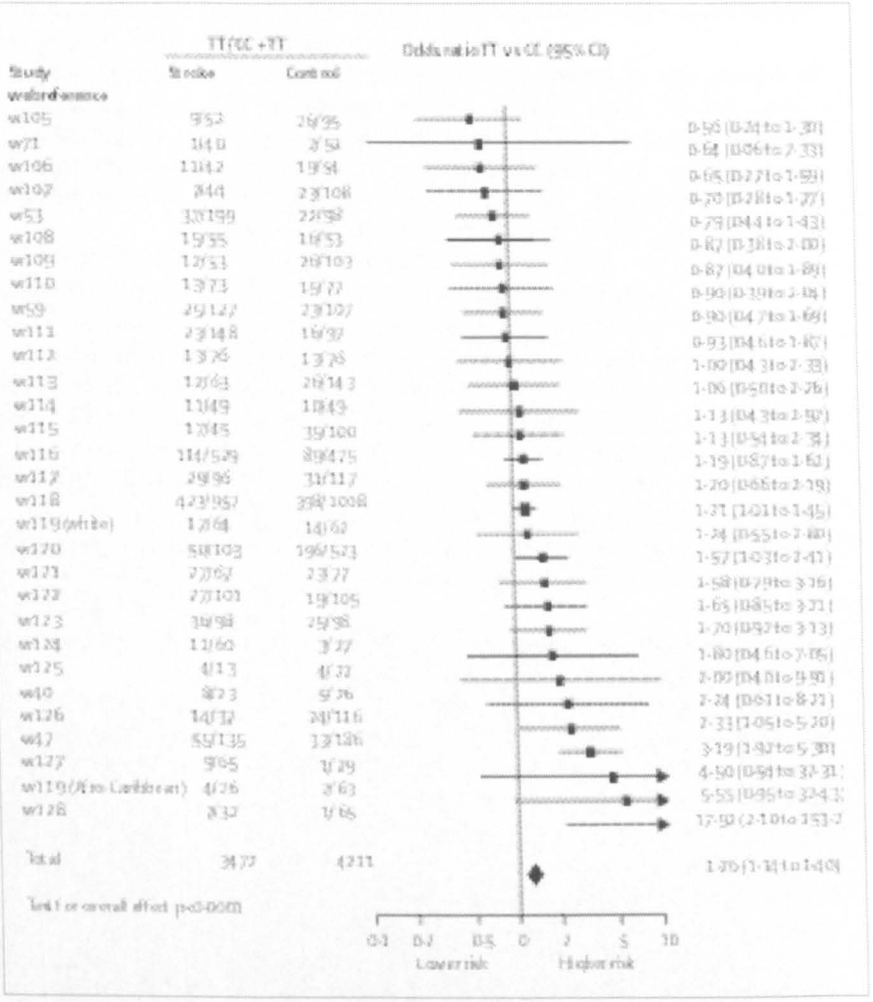


Figure 2: **Odds ratio for stroke in individuals with the TT genotype compared with those homozygous for the C allele of the *MTHFR* C677T polymorphism**
For webreferences see http://image.thelancet.com/extras/03art11437webreferences.pdf. The total number of stroke cases was 6324 (TT 1041, CT 2852, CC 2431) and the total number of controls was 7604 (TT 1140, CT 3393, CC 3071).

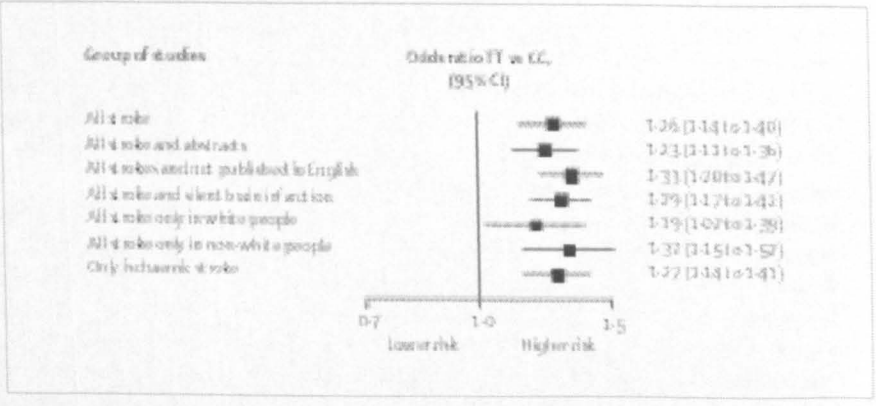


Figure 3: **Sensitivity analysis of the *MTHFR* C677T polymorphism and stroke association**
Summary odds ratio of stroke for TT INDIVIDUALS compared WITH CC INDIVIDUALS by publication type, stroke classification or type, and ethnic group.

concentration and *MTHFR* genotype, partly explained by lower serum folate concentrations in people with atherosclerotic disease. We therefore used estimates from healthy individuals only when calculating the predicted odds ratio for stroke among TT homozygotes. In estimation of an association between genotype and intermediate phenotype, the potential modifier effect of disease status should be examined.[24,25]

The second genetic meta-analysis (*MTHFR* C677T and stroke), which included about 14 000 individuals, allowed us to obtain a precise estimate of the effect of *MTHFR* genotype on stroke risk. Individuals homozygous for the T allele had an odds ratio for stroke of 1·26 (1·14 to 1·40) compared with those homozygous for the C allele. This difference in risk is similar to that expected from the difference in homocysteine concentrations by genotype in healthy individuals (1·20 [1·10 to 1·31]). Moreover, when the comparison was restricted to white individuals or to studies done in Europe or North America, within a similar age range to the population included in the meta-analysis of prospective studies of homocysteine and stroke risk,[4] similar results were obtained.

Clinical studies have shown that supplementation with folic acid and vitamin B12 lowers homocysteine concentrations by about 3 μmol/L.[14,26] If homocysteine is causally associated with an increased risk of stroke, nutritional interventions to lower concentrations might be expected to produce a relative-risk reduction in the incidence of stroke of about 23%.[3] Several randomised clinical trials are currently investigating the effects on cardiovascular outcomes of lowering homocysteine concentration by administration of B vitamins.[27,28] One clinical trial in North America[29] did not detect a beneficial effect of high versus low folate doses in secondary prevention of stroke. However, as suggested by the trial investigators and the results of our study, a larger sample size, longer periods of intervention, and targeting of the intervention to populations with low folate concentrations might all be required to show any potential benefit of this intervention. Such randomised intervention trials of the effect of lowering homocysteine concentrations are important for several reasons: the type of study we have done has several potential limitations; intervention trials are necessary to assess reversibility and could establish precisely the magnitude of any treatment effect; and they could ascertain whether there are any unexpected adverse effects of such therapies.

Despite the association found between the *MTHFR* C677T polymorphism, homocysteine concentration, and stroke risk, the size of the effect is modest compared with those of classic cardiovascular risk factors and does not necessarily provide a rational basis for screening for the polymorphism or for the measurement of homocysteine concentration in isolation, in the prediction of stroke. Whether either of these measurements would add useful predictive information to more established risk prediction tools (eg, Framingham risk equation) will require further investigation.

Our analyses must be interpreted in the context of the limitations of the available data. In the meta-analysis of the *MTHFR* variant and homocysteine concentrations, we found significant heterogeneity. This finding is perhaps not surprising, because expression of the mutation is likely to depend on environmental factors such as folate concentrations, as detected in our meta-regression analysis. This idea is supported by the smaller effect of the *MTHFR* C677T polymorphism on the homocysteine concentration in studies in North America, in which the mean concentrations of serum folate were higher. However, an overview of data from individual patients will be required for more precise quantification of this potential gene–environment interaction.[9,10] For this reason, only the estimates of weighted mean difference derived from a random-effects model were used in this report.

Publication bias or small-study bias was considered as an explanation for the observed associations between the *MTHFR* polymorphism, homocysteine concentration, and stroke risk, but the results obtained from the funnel-plot analysis and Egger's tests did not provide positive evidence for such bias. Although confounding is less likely in analyses of an association of a genotype with disease, some imbalance in the distribution of cardiovascular risk factors by *MTHFR* genotype cannot be totally excluded. However, previous studies that have investigated the effect of the *MTHFR* variant on coronary heart disease have suggested that there was no major systematic confounding from other cardiovascular risk factors (eg, age, sex, hypertension, diabetes, obesity, or alcohol intake).[7,30]

The mendelian randomisation approach used here is a potentially useful tool to assess the nature of the observed associations between putative risk factors and disease. This approach overcomes some potential limitations of observational studies, such as confounding, reverse-causality bias, and regression-dilution bias.[12,13,31,32] However, it also has theoretical limitations, such as the potential for confounding of the association between genotype and intermediate phenotype by linkage disequilibrium with other genes. Similarly, associations of the genotype with environmental exposures that regulate the concentrations of the intermediate phenotype could lead to spurious genotype–disease association. Finally, population stratification (confounding of gene–disease association by ethnicity), multiple disparate (pleiotropic) effects of gene polymorphisms on more than one biological system, or compensatory biological adaptation to its effects on disease risk (canalisation) could distort the observed genotype–disease associations in one or other direction.[12,13,33] There are other practical limitations. Since the estimated relative risk arising from genotype

(eg, *MTHFR* variant) on the intermediate phenotype (eg, homocysteine concentration) will in many cases be small, very large genetic association studies or meta-analysis of smaller studies will be needed to assess the nature of such associations. Unless careful consideration is given to the inclusion or exclusion of studies used in any meta-analyses, and to the potential for publication bias, the comparison of genetic and non-genetic odds ratio estimates could become distorted.

Adequately powered randomised controlled trials of supplementation with folic acid (with or without B vitamins) will be necessary to validate the therapeutic approach of lowering homocysteine concentrations to prevent stroke and other cardiovascular events.[34] The findings of this study and previous analyses of the same type emphasise the potential importance of such trials. The mendelian randomisation approach we and others have used for homocysteine might also be useful in assessment of whether other "novel" risk factors for cardiovascular disease could have aetiological roles.

Contributors

Juan P Casas contributed to protocol design, data extraction, and statistical analysis. Leonelo E Bautista helped with design, statistical analysis, and interpretation of data. Liam Smeeth helped with statistical analysis. Pankaj Sharma contributed to protocol design and data extraction, and Aroon D Hingorani contributed to protocol design and interpretation of data; both these authors contributed equally. All the authors contributed to the writing and revision of the report.

Conflict of interest statement

ADH has received fees for lecturing on features of the assessment and management of cardiovascular-disease risk at meetings organised by a medical conference company with a pharmaceutical sponsor. PS has received honoraria for lecturing in industry-sponsored meetings, industry funding for attending national and international meetings, and research grants from pharmaceutical companies, and has been a paid consultant to the biotech industry. The other authors declare no conflicts of interest.

Acknowledgments

We thank W Lalouschek, T Katsuya, T Ogihara, E Pongracz, A Pezzini, R Grossmann, B Voetsch, J Toscalzo, L L Huseman, G S Zhang, C Meisel, S Hustad, J Hung, B McQuillan, H Iso, and Y Moriyama for providing relevant information about the genotype frequencies, homocysteine concentrations, and data on cardiovascular risk factors from their studies. LS holds a Medical Research Council Clinician Scientist Fellowship. ADH holds a British Heart Foundation Senior Fellowship.

References

1 WHO. The World Health Report 2002: reducing risks, promoting healthy life. Geneva: WHO, 2002. Also available at http://www.who.int/whr/2002/en/ (accessed Oct 4, 2003).

2 Warlow CP. Epidemiology of stroke. *Lancet* 1998; **352** (suppl 3): SIII1–4.

3 Goldstein LB, Adams R, Becker K, et al. Primary prevention of ischemic stroke: a statement for healthcare professionals from the Stroke Council of the American Heart Association. *Circulation* 2001; **103**: 163–82.

4 Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* 2002; **325**: 1202–06.

5 Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA* 2002; **288**: 2015–22.

6 Bautista LE, Arenas IA, Penuela A, Martinez LX. Total plasma homocysteine level and risk of cardiovascular disease: a meta-analysis of prospective cohort studies. *J Clin Epidemiol* 2002; **55**: 882–87.

7 Brattstrom L, Wilcken DE. Homocysteine and cardiovascular disease: cause or effect? *Am J Clin Nutr* 2000; **72**: 315–23.

8 Ueland PM, Refsum H, Beresford SAA, Vollset SE. The controversy over homocysteine and cardiovascular risk. *Am J Clin Nutr* 2000; **72**: 324–32.

9 Klerk M, Verhoef P, Clarke R, et al. MTHFR 677C→T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA* 2002; **288**: 2023–31.

10 Brattstrom L, Wilcken DE, Ohrvik J, Brudin L. Common methylenetetrahydrofolate reductase gene mutation leads to hyperhomocysteinemia but not to vascular disease: the result of a meta-analysis. *Circulation* 1998; **98**: 2520–26.

11 Morgan TH. Heredity and sex. New York: Columbia University Press, 1913.

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

13 Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33**: 30–42.

14 Clarke R, Armitage J. Vitamin supplements and cardiovascular risk: review of the randomized trials of homocysteine-lowering vitamin supplements. *Semin Thromb Hemost* 2000; **26**: 341–48.

15 Kelly PJ, Rosand J, Kistler JP, et al. Homocysteine, MTHFR 677C→T polymorphism, and risk of ischemic stroke: results of a meta-analysis. *Neurology* 2002; **59**: 529–36.

16 Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol* 2004; **61**: 1652–61.

17 DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.

18 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–48.

19 Robin JS, Greenland S, Breslow NE. A general estimator of the variance of the Mantel-Haenszel odds ratio. *Am J Epidemiol* 1986; **124**: 719–23.

20 Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in a meta-analysis. In: Egger M, Davey Smith G, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publications, 2001: 285–312.

21 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.

22 Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351**: 123–27.

23 Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003; **326**: 219.

24 Keavney BD, Youngman LD, Palmer A, et al. Large-scale test of hypothesized associations between polymorphisms of lipid-related genes and myocardial infarction in about 5000 cases and 6000 controls. *Circulation* 2000; **102** (suppl II): 852 (abstr).

25 Youngman LD, Keavney BD, Palmer A, et al. Plasma fibrinogen and fibrinogen genotypes in 4865 cases of myocardial infarction and 6002 controls: test of causality by "Mendelian randomization". *Circulation* 2000; **102** (suppl II): 32 (abstr).

26 Wald DS, Bishop L, Wald NJ, et al. Randomized trial of folic acid supplementation and serum homocysteine levels. *Arch Intern Med* 2001; **161**: 695–700.

27 Schnyder G, Roffi M, Flammer Y, Pin R, Hess OM. Effect of homocysteine-lowering therapy with folic acid, vitamin B12, and vitamin B6 on clinical outcome after percutaneous coronary intervention: the Swiss Heart study—a randomized controlled trial. *JAMA* 2002; **288**: 973–79.

28 VITATOPS Trial Study Group. The VITATOPS (Vitamins to Prevent Stroke) Trial: rationale and design of an international, large, simple, randomised trial of homocysteine-lowering multivitamin therapy in patients with recent transient ischaemic attack or stroke. *Cerebrovasc Dis* 2002; **13**: 120–26.

29 Toole JF, Malinow MR, Chambless LE, et al. Lowering homocysteine in patients with ischemic stroke to prevent recurrent stroke, myocardial infarction, and death: the Vitamin Intervention for Stroke Prevention (VISP) randomized controlled trial. *JAMA* 2004; **291:** 565–75.

30 Russo GT, Friso S, Jacques PF, et al. Age and gender affect the relation between methylenetetrahydrofolate reductase C677T genotype and fasting plasma homocysteine concentrations in the Framingham Offspring Study Cohort. *J Nutr* 2003; **133:** 3436–41.

31 Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002; **31:** 730–36.

32 Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358:** 1356–60.

33 Little J, Khoury MJ. Mendelian randomisation: a new spin or real progress? *Lancet* 2003; **362:** 930–31.

34 Nallamothu BK, Fendrick AM, Rubenfire M, Saint S, Bandekar RR, Omenn GS. Potential clinical and economic effects of homocysteine lowering. *Arch Intern Med* 2000; **160:** 3406–12.

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2006;**35**:922–931
doi:10.1093/ije/dyl041

MENDELIAN RANDOMIZATION

Insight into the nature of the CRP–coronary event association using Mendelian randomization

Juan P Casas,[1,2†] Tina Shah,[1,3†] Jackie Cooper,[3] Emma Hawe,[3] Alex D McMahon,[4] Dairena Gaffney,[5] Christopher J Packard,[5] Denis S O'Reilly,[5] Irene Juhan-Vague,[6] John S Yudkin,[7] Elena Tremoli,[8,9] Maurizio Margaglione,[10] Giovanni Di Minno,[10] Anders Hamsten,[11] Teake Kooistra,[12] Jeffrey W Stephens,[3] Steven J Hurel,[13] Shona Livingstone,[14] Helen M Colhoun,[14] George J Miller,[15] Leonelo E Bautista,[16] Tom Meade,[2] Naveed Sattar,[5] Steve E Humphries[3] and Aroon D Hingorani[1]*

Accepted 15 February 2006

Background It is unclear wheather the association between C-reactive protein (CRP) and incident coronary events is free from bias and confounding. Individuals homozygous for a +1444C>T polymorphism in the CRP gene have higher circulating concentrations of CRP. Since the distribution of this polymorphism occurs at random during gamete formation, its association with coronary events should not be biased or confounded.

Methods We calculated the weighted mean difference in CRP between individuals with variants of the +1444C>T polymorphism in the CRP gene among 4659 European men from six studies (genotype-intermediate phenotype studies). We used this difference together with data from previous observational studies to compute an expected odds ratio (OR) for non-fatal myocardial infarction (MI) among individuals homozygous for the T allele. We then performed four new genetic association studies (6201 European men) to obtain a summary OR for the association between the −1444C>T polymorphism and non-fatal MI (genotype-disease studies).

Results CRP was 0.68 mg/l [95% confidence interval (95% CI) 0.31–1.10; $P = 0.0001$] higher among subjects homozygous for the +1444-T allele, with no confounding by a range of covariates. The expected ORs among TT subjects for non-fatal MI corresponding to this difference in CRP was 1.20 (95% CI 1.07–1.38) using the Reykjavik Heart study data and 1.25 (1.09–1.43)

[1] Centre for Clinical Pharmacology, Department of Medicine, BHF Laboratories at University College London (UCL), London, UK.
[2] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK
[3] Centre for Cardiovascular Genetics, Department of Medicine, BHF Laboratories at UCL, London, UK
[4] Robertson Centre for Biostatistics, University of Glasgow, Glasgow, Scotland, UK.
[5] Department of Pathological Biochemistry, Glasgow Royal Infirmary, Glasgow, Scotland, UK
[6] Laboratoire d'Hématologie, CHU Timone, Inserm Unit 626, Marseille, France
[7] Diabetes and Cardiovascular Disease Academic Unit, Archway Campus, Royal Free and UCL Medical School, London, UK.
[8] Department of Pharmacological Sciences, University of Milan, Milan, Italy.
[9] Monzino Cardiologu Center, Milan, Italy
[10] Instituto di Ricovero e Cura a Carattere Scientifico, Ospedale Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy.
[11] King Gustaf V Research Institute Department of Medicine, Karolinska Hospital, Karolinska Institute, Stockholm, Sweden.
[12] The Gaubius Laboratory, The Netherlands Organization for Applied Scientific Research-Prevention and Health, Leiden, The Netherlands.
[13] Department of Diabetes, Endocrinology and Metabolism, Middlesex Hospital, London, UK.
[14] Department of Epidemiology at UCL, London, UK.
[15] Wolfson Institute of Preventive Medicine at the Medical College of St Bartholomew's Hospital, London, UK.
[16] Department of Population Health Sciences, University of Wisconsin Medical School, Madison, WI, USA.
* Corresponding author. Centre for Clinical Pharmacology, BHF Laboratories at UCL, Rayne Building, 5 University Street, London WC1E 6JJ, UK. E-mail: a.hingorani@ucl.ac.uk
† Both authors contributed equally to this article.

for all observational studies to 2004. The estimate for the observed adjusted-OR for non-fatal MI among TT subjects was 1.01 (95% CI 0.74–1.38), lower than both expected ORs.

Conclusions A common CRP gene polymorphism is associated with important differences in CRP concentrations, free from confounding. The null association of this variant with coronary events suggests possible residual confounding (or reverse causation) in the CRP–coronary event association in observational studies, though the confidence limits are still compatible with a modest causal effect. Additional studies of genotype (or haplotype) and coronary events would help clarify whether or not the link between CRP and coronary events in observational studies is causal.

Keywords C-reactive protein, cardiovascular disease, genetics, polymorphism, meta-analysis

Prospective studies indicate a robust association, in healthy individuals, between levels of the acute phase reactant C-reactive protein (CRP) and later coronary events,[1–3] and it has been proposed that measurement of CRP may be a useful adjunct to coronary risk assessment and that CRP could be causally involved in atherosclerosis.[4] However, CRP concentrations are correlated with smoking status, blood pressure, obesity, diabetes, physical activity, social class, low birth weight, and other products of the inflammatory response and are also higher in individuals with clinical cardiovascular disease.[5–8] Despite statistical adjustment, this association may, therefore, be subject to residual confounding or bias due to reverse causation, where the disease even in its sub-clinical state leads to elevation in the level of CRP.[5–8]

A randomized controlled clinical trial of a selective intervention to reduce CRP would provide an unbiased insight into the nature of the association. Unfortunately, no such selective intervention currently exists. Although inhibitors of HMG-CoA reductase (statins) and glitazones reduce CRP,[9,10] they also have major effects on lipid profile and glucose metabolism. An alternative approach to control for confounding and reverse causality is to identify a common polymorphism in the CRP gene reliably associated with differences in circulating CRP concentration. The inheritance of such a variant should be subject to the random assortment of maternal and paternal alleles at the time of gamete formation, according to Mendel's second law.[11] If CRP actually increases the risk of coronary events then carriage of an allele that exposes individuals to a long-term elevation in CRP should confer an increased risk of coronary events proportional to the difference in CRP attributable to the allele. This relationship should be largely unconfounded and free of reverse causality bias.[12,13] Therefore, if non-genetic observational studies were unbiased, the increase in risk estimated from these studies should be consistent with the increase in risk conferred by carriage of the allele.[12,13] This approach, known as 'Mendelian randomization', has been used recently to investigate the link between homocysteine, fibrinogen, and cardiovascular disease.[14–17]

CRP concentration is a heritable trait.[18,19] In three small studies we found that homozygosity for the rarer T allele of a +1444C>T polymorphism in the 3′-untranslated region of the CRP gene was associated with higher basal and/or stimulated CRP concentrations.[20,21] We have now genotyped a large number of Caucasian men to test whether the increase in risk of coronary events in individuals homozygous for this polymorphism is close to that predicted under the assumption that the CRP–coronary event association from previous non-genetic observational studies is free from residual confounding and bias due to reverse causation.

Methods

Study populations

After obtaining ethical approvals, genotyping for the CRP polymorphism and measurement of plasma CRP concentration were conducted in samples obtained from a number of cross-sectional or prospective studies, or randomized controlled trials summarized in Table 1 and in supplementary materials available online.[21–27] To test the association of +1444C>T genotype with CRP concentration, male subjects without clinically evident cardiovascular disease from six studies with available DNA and plasma samples were studied (Table 1).[21–23,25–27] A second analysis was conducted to evaluate the association between genotype and risk of myocardial infarction (MI) in male subjects from four studies (Table 1).[22–25]

Data collection

Data on sex, mean age, systolic and diastolic blood pressure, body mass index (BMI), smoking status, glucose, lipid profile, alcohol consumption, fibrinogen, and plasma CRP values were obtained from the original studies. Subjects were classified using unified definitions of hypertension, hypercholesterolaemia, type-2 diabetes mellitus, and obesity from the guidelines on primary prevention of the American Heart Association.[28] For the studies relating genotype and coronary events,[22–25] non-fatal MI according to the WHO criteria[29] was considered the primary outcome, as this end-point had been uniformly used across all studies. Analyses of genotype and CRP were limited to male subjects to preserve consistency with available studies of genotype and coronary events, which were all conducted in men.

Table 1 Design characteristics of the original studies included in the present analysis

Name of the study	Study design and median follow-up (years)	Sample size Original/Genotyped	Country(ies)	Study population [a]	Main exclusion criteria	Primary analysis	
						Genotype and CRP level	Genotype and non-fatal MI
NPHS-II[22]	Prospective cohort (10.6)	3012/2676	UK[b]	Healthy middle-aged men	Pre-existing cardiovascular disease Coronary surgery Aspirin or anticoagulant therapy Malignant disease	Yes (2221)	Yes (2676)
LEADER[24]	Nested case–control study from a clinical trial of bezafibrate treatment (4.6)	1568/1066	UK	Men with lower extremity arterial disease	Unstable angina Total cholesterol <3.5 or >8.0 mmol/l Significant renal or hepatic disease or malignant disease	No	Yes (1066)
WOSCOPS[23]	Nested case–control in a clinical trial of pravastatin treatment (4.9)	6595/1451	UK	Moderately hypercholesterolaemic men	MI. or angina pectoris requiring hospitalization Life-threatening non-cardiac illness	Yes (1334)	Yes (1451)
HIFMECH[25]	Case–control (N/A)	Cases: 533/491 Controls: 575/517	UK Sweden France Italy	Cases: male MI survivors Controls: men matched by age and regional areas	Familial hypercholesterolaemia Insulin-dependent diabetes mellitus	Yes (465 controls)	Yes (1008)
Army[21]	Cross-sectional (N/A)	250/219	UK	Healthy UK-Army recruits	N/A	Yes (219)	No
UDACS[27]	Case–control (N/A)	Controls: 449/348	UK	Cases: men and women with DM[c] and cardiovascular disease Controls: men and women with DM but without cardiovascular disease	N/A	Yes (348 Male controls)	No
EBCT[26]	Case–control (N/A)	Controls: 94/72	UK	Cases: men and women with type-1 DM Controls: healthy men and women matched by age and gender	N/A	Yes (72 Male controls)	No

[a] All subjects evaluated were Caucasians.
[b] United Kingdom.
[c] DM: diabetes mellitus.

Laboratory analyses

All studies included in the present report used high sensitivity assays to measure plasma CRP concentrations. For specific details of the assays used please see supplementary materials provided online.

CRP genotyping and examination of linkage disequilibrium

The CRP/+1444C>T single nucleotide polymorphism (SNP; rs1130864) was genotyped by PCR and RFLP analysis using primer pairs described previously[21] and the restriction enzymes *Sdu*I or *Bsp*1286I that cleave the 181 bp PCR product into 23bp and 158bp fragments only in the presence of the common C allele. All DNA analysis was performed by staff unaware of the clinical status of the subjects. A public domain resequencing resource (Seattle SNPs; http://pga.mbt.washington.edu/) was accessed to investigate associations between the rs1130864 and other SNPs in the CRP gene, and to ascertain haplotype structure. MEDLINE was also searched to identify published studies reporting linkage disequilibrium (LD) between CRP SNPs and haplotype structure.

Statistical analysis

Genotype and CRP concentration

To quantify the effect of the CRP gene polymorphism on CRP concentration, we genotyped 4659 men from six studies (Table 1).[21–23,25–27] We calculated the within-study mean difference in CRP concentration between individuals homozygous for the T allele and carriers of the C allele and then weighted each mean by the inverse of its variance to obtain an overall weighted mean difference (WMD). In the calculation of the WMD, we limited the analysis to subjects without known coronary or peripheral artery disease at the time of blood sampling in order to avoid the potential for established disease to modify the size of the genotype–CRP association.[16] However, an analysis including all subjects was also conducted. Therefore, for prospective studies, baseline CRP data were used from all available subjects without clinically evident atherosclerosis. For case–control studies, genotype–CRP associations were analysed solely using control subjects. Because of its skewed distribution, CRP values were log-transformed before the analysis. For methods used to calculate the absolute geometric-WMD in CRP concentrations by genotype please see supplementary materials available at *IJE* online. Random and fixed effect models were used for these analyses.[30,31]

Genotype and non-fatal myocardial infarction

To examine the effect of CRP genotype on risk of non-fatal MI, we genotyped 6201 men from four studies (Table 1).[22–25] We obtained the adjusted odds ratios (ORs) and 95% confidence intervals (95% CIs) for subjects homozygous for the T allele compared with carriers of the C allele from each study. In addition, for the two intervention trials[23,24] the potential interaction of the genotype–MI association with the active therapy was also evaluated. We then pooled the within-study ORs to obtain a summary adjusted-OR and 95% CI for non-fatal MI, under both fixed and random effect models. Fixed effect summary-ORs were calculated using the inverse variance-weighted method,[32] and the DerSimonian and Laird Q test[33] was used to evaluate the degree of heterogeneity between studies.

Consistency between the ORs from genetic and from non-genetic observational studies

We estimated the expected OR for non-fatal MI, corresponding to the WMD in CRP between TT subjects and C allele carriers, based on data from prior non-genetic observational studies that examined the CRP–coronary event association. To do this we used information from the most recent meta-analysis of observational studies of CRP and coronary events as well as data from the Reykjavik Heart Study, a very large study in which more extensive control for potential confounders was made. The first estimate, based on an OR of 1.58 (95% CI 1.48–1.68), for the top vs bottom tertiles of the CRP distribution, came from the recent meta-analysis[34] of 22 prospective studies with 7068 cases with different degrees of adjustment for traditional cardiovascular risk factors across the studies included and no adjustment for regression dilution bias. The second estimate was based on an OR of 1.45 (95% CI 1.26–1.68) between top and bottom tertiles of the CRP distribution reported in the Reykjavik Heart Study,[34] the largest single prospective observational study of CRP, which contributed 2459 cases to the meta-analysis and also undertook a comprehensive adjustment for potential confounders (age, sex, period of recruitment, smoking, systolic blood pressure, total cholesterol level, triglyceride levels, BMI, forced expiratory volume in one second, diabetes, and socioeconomic status). The third estimate from the Reykjavik Heart Study[34] was based on an OR of 1.92 (95% CI 1.68–2.18) for top vs bottom tertile of CRP, based on a more limited degree of adjustment (age, sex and period of recruitment). The final estimate also from the Reykjavik study[34] was based on the adjusted standard 10 year follow-up risk of 1.84 (95% CI 1.49–2.28) for top vs bottom tertile of CRP.

For these estimates we assumed equivalence between relative risk and OR, that the usual mean difference in CRP between the individuals in top and bottom tertiles was 1.4 mg/l[6], and that the CRP–coronary event relationship was log-linear. The expected OR for TT homozygous subjects with reference to C allele carriers was calculated using the formula: expected OR $= \mathrm{OR}_{\text{non genetic studies}}^{(\mathrm{WMD}/1.4)}$. The 95% CI for this expected OR was obtained by simulation,[16] and it was compared with that observed by an interaction test.[35] For details, please see supplementary materials available at *IJE* online. Consistency between the observed genetic and expected (non-genetic) ORs would suggest that the association between CRP and coronary events identified in prior observational epidemiological studies would be unlikely to be the result of residual confounding or reverse causality bias. Data were analysed using the Review Manager software (version 4.2) from the Cochrane Collaboration 2003 and Stata 8.2 (Stata Corporation, College Station, TX, 2003).

Results

Allele and genotype frequencies, and LD with other CRP SNPs

The allele and genotype frequencies of the +1444C>T polymorphism were in Hardy–Weinberg equilibrium for all

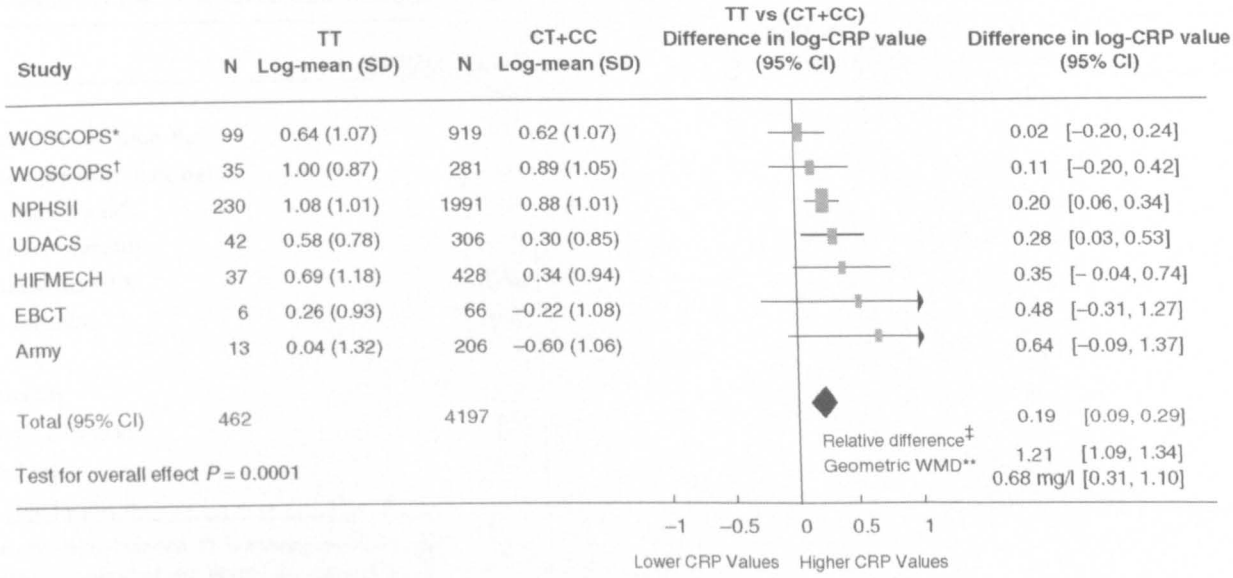


Figure 1 Mean log-CRP and mean difference in log-CRP (mg/l) by study among subjects without known cardiovascular disease. *, represents data for control subjects; †, represents data for subjects with later non-fatal MI; ‡, represents relative difference was obtained by antilog of the difference in log-CRP values; and **, represents Geometric WMD = [(relative difference in CRP concentration between TT homozygotes and C-allele carriers × mean-CRP concentration in C-allele carriers) − mean-CRP concentration in C-allele carriers]

studies included in the present report, apart from a marginal distortion in the NPHS-II study (please see Table 1 in supplementary materials available at *IJE* online). The frequencies of the rare allele in disease-free subjects from all studies were very similar (range: 26–33%). Data from the Seattle SNPs resequencing resource (http://pga.gs.washington.edu/data/crp) indicates that four common CRP haplotypes (approximate frequencies 0.3, 0.3, 0.3, and 0.05) account for >90% of chromosomes among subjects of European descent. The +1444T allele lies on one of the three common haplotypes (approximate frequency 0.3), which also contains the T allele of a common tri-allelic polymorphism (−286C>T/A; rs3091244). Prior studies also report a strong LD between minor alleles at +1444 and −286 polymorphisms; reported D' values in Caucasians being 0.97,[36] 0.91,[37] 0.91 (unpublished data) and with an r^2-value of 0.97.[38] In functional studies *in vitro*, the −286C>T/A SNP appears to account for increased transcriptional activity in studies using promoter–reporter assays. The unique haplotype containing minor alleles at −286 and +1444 has been associated with higher CRP values in two published studies,[36,39] and we have identified concordant population haplotype frequencies and associations with CRP in our own recent studies in the NPHS dataset of ~2700 men (unpublished data).

CRP/+1444C>T polymorphism and CRP concentrations

Using data from 4659 men from six studies, the weighted mean CRP concentration in C-allele carriers without known cardiovascular disease was 2.01 mg/l (95% CI 1.94–2.07). Under a fixed effect model, the geometric-WMD in CRP concentration between individuals homozygous for the T allele compared with carriers of the C allele was 0.68 mg/l (95% CI 0.31–1.10; $P = 0.0001$). There was no significant between-study heterogeneity [P-value for heterogeneity (P_{Het}) = 0.47; Figure 1]. When alternative models of the effect of genotype on CRP were evaluated, only the TT vs CC comparison was significant [WMD = 0.78 mg/l (0.41–1.20); $P < 0.0001$], while heterozygosity (CT vs CC) was not [WMD = 0.06 (−0.24 − 0.44); $P = 0.66$]. When subjects with coronary or peripheral atherosclerosis at the time of blood sampling were included in this analysis ($n = 5658$), the WMD between individuals homozygous for the T allele compared with carriers of C was very similar [0.74 mg/l (95% CI 0.37–1.10); $P < 0.0001$]. The values we have obtained are concordant with those obtained from other studies examining associations of CRP SNPs with CRP concentration, taking into account known LD.[36–40]

To evaluate potential confounding of the genotype–CRP association, we quantified the association between TT genotype and age, systolic and diastolic blood pressure, BMI, current smoking, glucose, alcohol intake, fibrinogen, triglyceride, total and HDL cholesterol. No significant difference was observed for any of these risk factors, with the exception of a slightly higher BMI among TT subjects (mean difference = 0.34 kg/m^2; $P = 0.02$; Table 2). In view of the multiple comparisons made, however, this significance level is not particularly extreme. Moreover, this apparent association was not significant when subjects with and without coronary heart disease were considered together, and it is likely, but not yet certain, to reflect the play of chance.

OR estimated from observational studies and the expected mean difference in plasma CRP

Expected ORs for non-fatal MI among TT homozygotes compared with C allele carriers based on a between-genotype WMD difference in CRP of 0.68 mg/l (95% CI 0.31–1.10) were 1.37 (95% CI 1.14–1.68), 1.20 (95% CI 1.07–1.38), 1.34 (95% CI 1.12–1.67), and 1.25 (95% CI 1.09–1.43) based, respectively, on the Reykjavik Heart Study minimally and maximally

Table 2 Cardiovascular risk factor distribution according to the CRP/+1444C>T polymorphism

Variable	Subjects (studies)	Weighted mean difference (TT minus C-carriers) (95% CI)	*P*-value
Age (years)	6358 (7)	0.17 (−0.19 to 0.52)	0.35
Systolic blood pressure (mm Hg)	6356 (7)	−0.31 (−1.78 to 1.16)	0.68
Diastolic blood pressure (mm Hg)	6356 (7)	−0.70 (−1.56 to 0.16)	0.11
Body mass index (kg/m^2)	6359 (7)	0.34 (0.06–0.62)	0.02
Total cholesterol (mmol/l)	6115 (6)	−0.01 (−0.08 to 0.06)	0.87
HDL-cholesterol (mmol/l)	4714 (5)	−0.01 (−0.04 to 0.01)	0.32
Triglyceride (mmol/l)	6041 (5)	−0.03 (−0.11 to 0.06)	0.51
Fibrinogen (g/l)	5656 (4)	0.03 (−0.02 to 0.08)	0.20
Glucose (mmol/l)	1913 (3)	−0.01 (−0.13 to 0.12)	0.90
Alcohol intake (U/week)	4549 (4)	0.01 (−1.38 to 1.39)	0.99
Current smoking[a]	6132 (6)	1.00 (0.84–1.20)	0.98
C-reactive protein (mg/l)	4659 (6)	0.68 (0.31–1.10)	0.0001

Comparisons are made between TT homozygotes and C-allele carriers.
[a] For this variable instead of the WMD, the value reported is the weighted odds ratio.

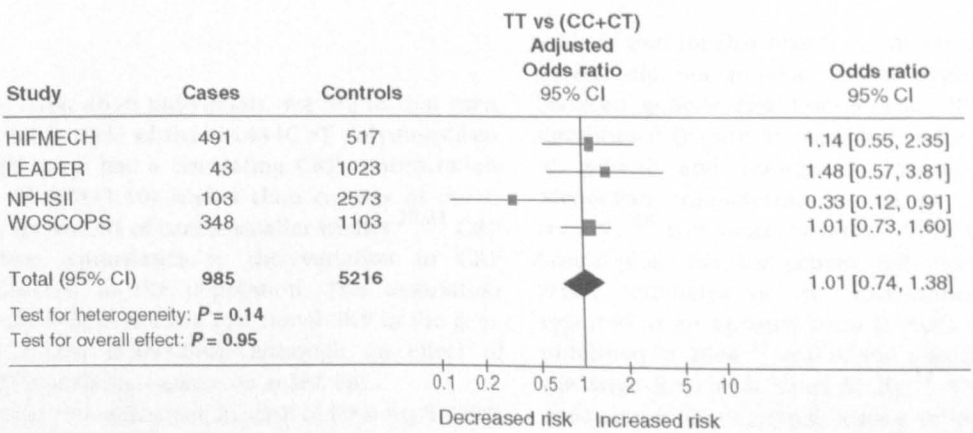


Figure 2 Adjusted odds ratio (a multivariate logistic regression model was used to adjust the OR from each study by age, total cholesterol, alcohol intake, body mass index, hypertension, diabetes, and smoking) for non-fatal MI among subjects with the +1444TT-genotype in comparison with carriers of the +1444C allele

adjusted models, the Reykjavik study 10 year estimate, and on the meta-analysis of studies up to 2004.[34]

OR of non-fatal MI for the CRP/+1444C>T polymorphism

Data from four new studies involving 985 subjects with non-fatal MI and 5216 control subjects were pooled to obtain a summary adjusted OR. After combining the studies under a fixed effect model, subjects homozygous for the T allele compared with C allele carriers had no significant increase in the risk of non-fatal MI [summary adjusted-OR= 1.01 (95% CI 0.74–1.38); P = 0.95] (Figure 2). No significant inter-study heterogeneity was observed (P_{Het} = 0.14). When alalysis of the risk of non-fatal MI conferred by homozygosity for the T allele was restricted to prospective genetic studies, there was still no significant association with non-fatal MI [OR = 0.98 (95% CI 0.70–1.38); P = 0.93]. No significant increase in risk was observed in the individual studies, and there was no significant interaction between treatment and genotype–outcome association in the two clinical trials involved in this analysis: for LEADER, P = 0.53 and WOSCOPS, P = 0.44.

Consistency between the ORs estimated from the mean difference in CRP level and the OR from the genetic study

The estimate of the observed OR from the genetic study [1.01 (95% CI 0.74–1.38)] was mathematically smaller than the expected OR [1.37 (95% CI 1.14–1.68) or 1.34 (95% CI 1.12–1.67)], calculated using data from the minimally adjusted estimate or the 10 year estimate from the Reykjavik Heart Study.[34] It was, however, closer to estimates based on data from the maximally adjusted model of the large Reykjavik Heart Study [1.20 (95% CI 1.07–1.38)] and to an estimate based on an updated meta-analysis of studies to 2004 [1.25 (95% CI 1.09–1.43)]. Indeed, most of the 95% confidence limits of these latter two estimates lay within the upper bound of the 95% confidence limit of the genetic estimate (Figure 3). The P-value for interaction for the comparisons of the minimally and maximally adjusted estimates from the Reykjavik study, and from an updated meta-analysis of studies to 2004 with the observed genetic estimate is outlined in Figure 3.

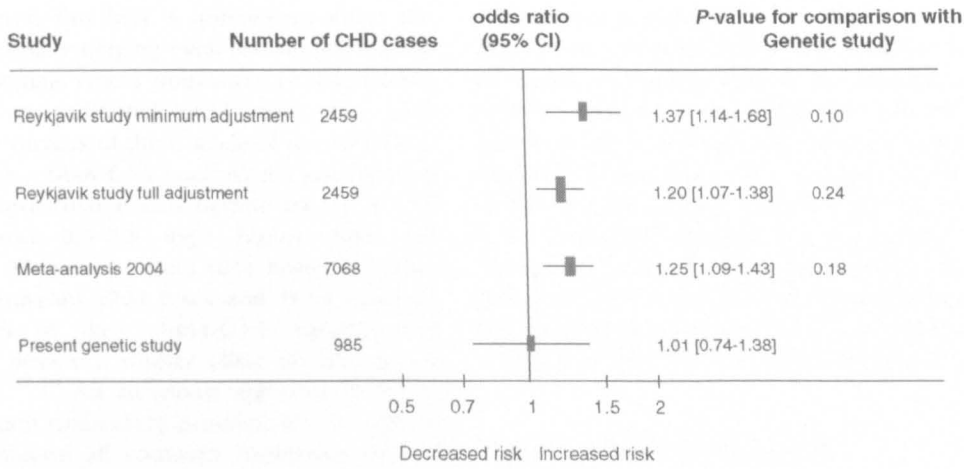


Figure 3 Comparison of expected and observed odds ratios for non-fatal MI for individuals homozygous for the CRP +1444C>T polymorphism. Expected ORs were calculated assuming a log-linear association between CRP and non-fatal MI, and that a 1.4 mg/l difference in CRP corresponds to an OR for coronary events of 1.92 (95% CI 1.68–2.18) and 1.45 (95% CI 1.25–1.68) based on the minimum and fully adjusted models from the Reykjavik Heart Study published in 2004 and 1.58 (95% CI 1.48–1.68) based on an updated meta-analysis published in 2004. [Data derived from references (6) and (34), respectively]

Discussion

Using pooled data from 4659 individuals, we found that men homozygous for the T allele of the +1444C>T polymorphism of the human CRP gene had a circulating CRP concentration 0.68 mg/l (95% CI 0.31–1.10) higher than carriers of the C allele, confirming the results of earlier smaller studies.[20,21] CRP genotype, therefore, contributes to the variation in CRP concentration observed in the population. This association probably reflects LD with a putative functional SNP in the gene promoter (−286C>T/A; rs3091244), although an effect of +1444C>T on RNA stability cannot be ruled out.

We estimated that this difference in CRP of 0.68 mg/l (95% CI 0.31–1.10) would confer an OR for non-fatal MI of 1.25 (95% CI 1.09–1.43) or 1.20 (95% CI 1.07–1.38) using data from a meta-analysis of prospective cohort studies published in 2004,[34] and from the maximally adjusted model of the Reykjavik Heart Study,[34] respectively, assuming the observational associations have been unbiased and free from residual confounding. However, despite possessing a genotype that would have exposed them to a long-term elevation in CRP concentration, for men with the TT genotype, the point estimate of the OR for non-fatal MI was 1.01 (95% CI 0.74–1.38). This finding is important since, in contrast to the reported association between CRP and coronary events in previous studies, the association between CRP *genotype* and events should not be subject to reverse causality bias, regression dilution bias, or to confounding by other cardiovascular risk factors related to CRP, though some doubt about a marginal effect of the genotype on BMI remains to be evaluated.[7,8,34,41–47] This is supported by the data in Table 2, which indicate that certain established cardiovascular risk factors and some inflammatory markers, that are known to be associated with CRP, were not associated with +1444C>T genotype.

The estimate of risk derived from the genetic association studies of CRP reported here should not be biased or confounded, but it is currently much less precise than the estimates obtained from classical observational epidemiology. It is likely that for this reason an interaction test, which has low power, did not provide formal evidence for heterogeneity between genetic and non-genetic ORs at the 5% level of significance (Figure 3). Approaches for evaluating consistency of genetic and non-genetic disease risk estimates from Mendelian randomization studies have been reviewed recently.[48] It is clear, however, from visual inspection of the Forest plot, that the genetic risk estimate is closer to more recent estimates of the CRP–coronary event association reported in an updated meta-analysis of observational studies published in 2004,[34] and in the maximally adjusted model of the large Reykjavik Heart Study.[34] Thus, recent results from prospective observational studies in which the random error has been reduced substantially by considering more studies in a meta-analysis, and in which more extensive statistical adjustment has been conducted, as exemplified in the Reykjavik Heart Study, indicate that the association of CRP with coronary heart disease risk has decreased substantially in comparison to earlier estimates.[34] The strong potential for confounding in observational studies of the measured phenotype (CRP) and the utility of genotype as a proxy for CRP was illustrated by a recent study[40] in which a range of covariates that exhibited strong correlations with CRP were distributed evenly among >3500 British women separated according to genotypes for a 1059G>C polymorphism that is in LD with the variant reported here.[36]

Although a null association was identified between CRP genotype and coronary events in this study the width of the confidence limits is such that our observation could be compatible with a modest but potentially important causal link. It is estimated that a dataset of ~10 000 cases and similar number of controls would be required in a genetic association study to exclude a small effect of CRP on coronary heart disease aetiology.[49] However, evidence from different sources indicates that, if anything, the discrepancy observed between the genetic and non-genetic ORs is likely to be greater, rather than smaller. First, the observational estimate of the CRP–coronary event association[34] on which we based the non-genetic OR, did not adjust for regression dilution bias and so may have been

underestimated. However, this bias is unlikely to affect the genetic estimate.[12] Second, emerging evidence supporting the current null genetic estimate comes from three genetic studies in Caucasians, with an additional 1858 cases and 1347 controls,[37,50,51] where carriers of the C allele of the 1059G>C variant (in LD with the +1444/C>T variant) did not have an increased risk of cardiovascular events despite having a CRP concentration that was 0.3–0.6 mg/l higher than GG homozygous subjects. Moreover, recent data from two other genetic studies in Caucasians (703 cases and 1053 controls) found no association of the +1444/C>T variant with cardiovascular events despite a similar effect of this variant on CRP concentration.[37,38] An additional approach in future Mendelian randomization studies is to genotype several SNPs at the CRP locus to generate all common haplotypes and to ensure inclusion of SNPs with demonstrated functionality from *in vitro* studies in such analyses. Since there is extensive LD across the CRP gene, only a small number of 'tagging' SNPs (3–5) require typing to capture most of the genetic variation at this locus.[52,53] In addition to −286C>T/A, published and public domain data indicate that four common haplotypes identified by these tagging-SNPs account for >90% of chromosomes in subjects of European descent.[53]

It is important to note that there is evidence for an additional functional CRP promoter SNP found at high frequency in subjects of African origin.[36,54]

It is important to make clear that our study was limited to male Caucasian subjects and that additional studies will be required to ascertain whether similar findings pertain to women and also to other ethnic groups. However, recent evidence suggests that a similar effect of the +1444C>T polymorphism on CRP concentrations would be expected to occur in women as recently reported in British and American women.[38,55] This is also supported by the effect of the 1059/G>C variant on CRP levels, and the 0.42 mg/l difference in CRP concentrations among C carriers compared with women homozygous to the G allele is very similar to the effect observed here in men.[38,40]

Our finding using a genetic approach also needs to be set in the context of recent mechanistic studies that have investigated a potential pro-atherogenic effect of CRP on vascular cells and tissues.[56–61] It is becoming clear now that many of the pro-atherogenic effects demonstrated *in vitro* in initial studies could have been due to the presence of pro-inflammatory substances contaminating commercially sourced recombinant CRP from bacterial sources, or to the sodium azide preservative.[62–70] More recent experimental studies, using pure CRP preparations, have failed to confirm the proposed pro-atherogenic actions of CRP on vascular tissues and would be concordant with our own data from genetic epidemiology,[62–70] as would the recent experiments in atherosclerosis-prone mice, which showed that in contrast to earlier studies transgenic overexpression of human CRP did not enhance development of atherosclerotic lesions.[71–73]

In conclusion, Caucasian men with a genotype that would have exposed them to a long-term elevation in CRP concentration [0.68 mg/l (95% CI 0.31–1.10)] were not at increased risk of non-fatal MI (OR 1.01; 95% CI 0.74–1.38), though the confidence limits currently encompass substantial uncertainty. Taking this into account, it is at the very least likely that unbiased and non-confounded estimates of the effect of CRP on coronary events are smaller than earlier studies estimated, a finding that is in agreement with the latest evidence derived from prospective observational studies and also with more refined and carefully conducted mechanistic studies.[74] If our null point estimate of the genetic effect is stable with the addition of further studies, it might lead to the re-evaluation of CRP both as a risk marker[4] and as a potential therapeutic target.[5] Additional studies with data on CRP genotype, CRP levels, and coronary events will be essential to obtain more precise genetic risk estimates and to help clarify whether or not CRP plays a causal role in atherothrombosis.

Declaration of interest

The authors have declared no conflict of interest.

Acknowledgements

This work was partially supported by a BHF-PhD studentship [FS/02/086/14760] awarded to T.S. J.W.S. was supported by Diabetes UK. A.D.H. holds a British Heart Foundation Senior Fellowship. S.H. holds the British Heart Foundation Chair of Cardiovascular Genetics (RG 2000/015). The HIFMECH study was supported by the Commission of the European Community: Hifmech Study contract BMH4-CT96-0272. The LEADER trial was supported by grants from the British Heart Foundation and the Medical Research Council. The NPHS-II study was supported by the British Medical Research Council, the US National Institute of Health (NHLBI 33014), and DuPont Pharma, Wilmington, USA. The WOSCOPS genotype analysis was partly funded by a research grant from Chest, Heart and Stroke Scotland. The different funding organizations named above had no role in the design and conduct, data management and analysis, manuscript preparation, or in the review or authorization for submission. The authors acknowledge the critical reading and suggestions of Prof George Davey-Smith.

References

1 Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med* 1997;**336**:973–79.

2 Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med* 2000;**342**:836–43.

3 Kuller LH, Tracy RP, Shaten J, Meilahn EN. Relation of C-reactive protein and coronary heart disease in the MRFIT nested case–control study. Multiple Risk Factor Intervention Trial. *Am J Epidemiol* 1996;**144**:537–47.

4 Pearson TA, Mensah GA, Alexander RW, *et al.* Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 2003;**107**:499–511.

5 Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003;**111**:1805–12.

6 Danesh J, Whincup P, Walker M, *et al.* Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ* 2000;**321**:199–204.

[7] Freeman DJ, Norrie J, Caslake MJ, Gaw A, Ford I, Lowe GD *et al.* West of Scotland Coronary Prevention Study. C-reactive protein is an independent predictor of risk for the development of diabetes mellitus in the West of Scotland Coronary Prevention Study. *Diabetes* 2002;**51**:1596–600.

[8] Sattar N, McConnachie A, O'Reilly D, Upton MN, Greer IA, Davey Smith G *et al.* Inverse association between birthweight and C-reactive protein concentrations in the MIDSPAN family study. *Arterioscler Thromb Vasc Biol* 2004;**24**:583–87.

[9] Balk EM, Lau J, Goudas LC, Jordan HS, Kupelnick B, Kim LU *et al.* Effects of statins on nonlipid serum markers associated with cardiovascular disease: a systematic review. *Ann Intern Med* 2003;**139**:670–82.

[10] Sjoholm A, Nystrom T. Endothelial inflammation in insulin resistance. *Lancet* 2005;**365**:610–12.

[11] Morgan TH. *Heredity and Sex*. New York, USA: Columbia University Press, 1913.

[12] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.

[13] Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002;**31**:730–36.

[14] Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, Schouten EG *et al.* MTHFR 677C →T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA* 2002;**288**:2023–31.

[15] Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* 2002;**325**:1202.

[16] Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from Mendelian randomisation. *Lancet* 2005;**365**:224–32.

[17] Davey Smith G, Harbord R, Milton J, Ebrahim S, Sterne JA. Does elevated plasma fibrinogen increase the risk of coronary heart disease? Evidence from a meta-analysis of genetic association studies. *Arterioscler Thromb Vasc Biol* 2005;**25**:2228–33.

[18] MacGregor AJ, Gallimore JR, Spector TD, Pepys MB. Genetic effects on baseline values of C-reactive protein and serum amyloid a protein: a comparison of monozygotic and dizygotic twins. *Clin Chem* 2004;**50**:130–34.

[19] Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH *et al.* Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* 2001;**154**:681–89.

[20] D'Aiuto F, Casas JP, Shah T, Humphries SE, Hingorani AD, Tonetti M. C-reactive protein (+1444C >T) polymorphism is associated with enhanced acute release of inflammatory markers. *Atherosclerosis* 2005;**179**:413–17.

[21] Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A *et al.* A polymorphism in the human CRP gene influences basal and stimulated CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arterioscler Thromb Vasc Biol* 2003;**23**:2063–69.

[22] Miller GJ, Bauer KA, Barzegar S, Foley AJ, Mitchell JP, Cooper JA *et al.* The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. *Thromb Haemost* 1995;**73**:82–86.

[23] The WOSCOPS Group. Baseline risk factors and their association with outcome in the West of Scotland Coronary Prevention Study. The West of Scotland Coronary Prevention Study Group. *Am J Cardiol* 1997;**79**:756–62.

[24] Meade T, Zuhric R, Cook C, Cooper J on behalf of MRC General Practice Research Framework. Bezafibrate in men with lower extremity arterial disease: randomised controlled trial. *BMJ* 2002;**325**:1139.

[25] Juhan-Vague I, Morange PE, Aubert H, Henry M, Aillaud MF, Alessi MC *et al.* Plasma thrombin-activatable fibrinolysis inhibitor antigen concentration and genotype in relation to myocardial infarction in the north and south of Europe. *Arterioscler Thromb Vasc Biol* 2002;**22**:867–73.

[26] Colhoun HM, Rubens MB, Underwood SR, Fuller JH. The effect of type I diabetes mellitus on the gender difference in coronary artery calcification. *J Am Coll Cardiol* 2000;**36**:2160–67.

[27] Dhamrait SS, Stephens JW, Cooper JA, Acharya J, Mani AR, Moore K *et al.* Cardiovascular risk in healthy men and markers of oxidative stress in Diabetics men are associated with common variation in the gene for uncoupling protein 2. *Eur Heart J* 2004;**25**:468–75.

[28] Pearson TA, Blair SN, Daniels SR, Eckel RH, Fair JM, Fortmann SP *et al.* AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update: Consensus Panel Guide to Comprehensive Risk Reduction for Adult Patients Without Coronary or Other Atherosclerotic Vascular Diseases. American Heart Association Science Advisory and Coordinating Committee. *Circulation* 2002;**106**:388–91.

[29] Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 1994;**90**:583–12.

[30] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**: 719–48.

[31] DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.

[32] Shadish WR, Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges LV (eds). *The Handbook of Research Synthesis*. New York: Rusell Sage Foundation, 1994, pp. 261–84.

[33] Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in a meta-analysis. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Publications, 2001, pp. 285–312.

[34] Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A *et al.* C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med* 2004;**350**:1387–97.

[35] Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;**326**:219.

[36] Russell AI, Cunninghame Graham DS, Shepherd C, Roberton CA, Whittaker J, Meeks J *et al.* Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Hum Mol Genet* 2004;**13**:137–47.

[37] Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S *et al.* A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 2005;**178**:193–98.

[38] Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R *et al.* Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann Hum Genet* 2005;**69**:623–38.

[39] Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M *et al.* Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am J Hum Genet* 2005;**77**:64–77.

[40] Davey Smith G, Lawlor DA, Harbord R, et al. Association of C-reactive protein with blood pressure and hypertension. Life course confounding and Mendelian randomization tests of causality. *Arterioscler Thromb Vasc Biol* 2005;**25**:1051–56.

[41] Woodward M, Rumley A, Lowe GD, Tunstall-Pedoe H. C-reactive protein: associations with haematological variables, cardiovascular risk factors and prevalent cardiovascular disease. *Br J Haematol* 2003;**122:**135–41.

[42] Hackam DG, Anand SS. Emerging risk factors for atherosclerotic vascular disease: a critical review of the evidence. *JAMA* 2003;**290:**932–40.

[43] Primatesta P, Falaschetti E, Poulter NR. Influence of hormone replacement therapy on C-reactive protein: population-based data. *J Cardiovasc Risk* 2003;**10:**57–60.

[44] Bautista LE. Inflammation, endothelial dysfunction, and the risk of high blood pressure: epidemiologic and biological evidence. *J Hum Hypertens* 2003;**17:**223–30.

[45] Saito M, Ishimitsu T, Minami J, Ono H, Ohrui M, Matsuoka H. Relations of plasma high-sensitivity C-reactive protein to traditional cardiovascular risk factors. *Atherosclerosis* 2003;**167:**73–79.

[46] Visser M, Bouter LM, McQuillan GM, Wener MH, Harris TB. Elevated C-reactive protein levels in overweight and obese adults. *JAMA* 1999;**282:**2131–35.

[47] Stewart SH, Mainous AG III, Gilbert G. Relation between alcohol consumption and C-reactive protein levels in the adult US population. *J Am Board Fam Pract* 2002;**15:**437–42.

[48] Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in non-genetic observational studies using 'Mendelian triangulation'. *Ann Epidemiol* 2006. In Press.

[49] Davey Smith G, Harbord R, Ebrahim S. Fibrinogen, C-reactive protein and coronary heart disease: does Mendelian randomization suggest the associations are non-causal? *QJM* 2004;**97:**163–66.

[50] Zee RY, Ridker PM. Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. *Atherosclerosis* 2002;**162:**217–19.

[51] KWJ Lee, J Hill, J Frohlich. C-reactive protein as a predictor of clinical outcomes in patients with coronary atherosclerosis. (Oral session young investigator award). Canadian Cardiovascular Congress, 2002.

[52] Goldstein DB, Ahmadi KR, Weale ME, Wood NW. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet* 2003;**19:**615–22.

[53] SeattleSNPs. NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA. (http://pga.gs.washington.edu) (Accessed May 1, 2005).

[54] Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A *et al.* Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J Mol Med* 2005;**83:**440–47.

[55] Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ *et al.* C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet* 2005;**366:** 1954–59.

[56] Verma S, Kuliszewski MA, Li SH, Szmitko PE, Zucco L, Wang CH *et al.* C-reactive protein attenuates endothelial progenitor cell survival, differentiation, and function: further evidence of a mechanistic link between C-reactive protein and cardiovascular disease. *Circulation* 2004;**109:**2058–67.

[57] Li SH, Szmitko PE, Weisel RD, Wang CH, Fedak PW, Li RK *et al.* C-reactive protein upregulates complement-inhibitory factors in endothelial cells. *Circulation* 2004;**109:**833–36.

[58] Wang CH, Li SH, Weisel RD, Fedak PW, Dumont AS, Szmitko P *et al.* C-reactive protein upregulates angiotensin type 1 receptors in vascular smooth muscle. *Circulation* 2003;**107:**1783–90.

[59] Verma S, Wang CH, Li SH, Dumont AS, Fedak PW, Badiwala MV *et al.* A self-fulfilling prophecy: C-reactive protein attenuates nitric oxide production and inhibits angiogenesis. *Circulation* 2002;**106:**913–19.

[60] Pasceri V, Willerson JT, Yeh ET. Direct proinflammatory effect of C-reactive protein on human endothelial cells. *Circulation* 2000;**102:**2165–68.

[61] Pasceri V, Cheng JS, Willerson JT, Yeh ET. Modulation of C-reactive protein-mediated monocyte chemoattractant protein-1 induction in human endothelial cells by anti-atherosclerosis drugs. *Circulation* 2001;**103:**2531–34.

[62] Taylor KE, Giddings JC, van den Berg CW. C-Reactive protein-induced in vitro endothelial cell activation is an artefact caused by azide and lipopolysaccharide. *Arterioscler Thromb Vasc Biol* 2005;**25:**1225–30.

[63] Lafuente N, Azcutia V, Matesanz N, Cercas E, Rodriguez-Manas L, Sanchez-Ferrer CF *et al.* Evidence for sodium azide as an artifact mediating the modulation of inducible nitric oxide synthase by C-reactive protein. *J Cardiovasc Pharmacol* 2005;**45:**193–96.

[64] van den Berg CW, Taylor KE, Lang D. C-reactive protein-induced *in vitro* vasorelaxation is an artefact caused by the presence of sodium azide in commercial preparations. *Arterioscler Thromb Vasc Biol* 2004;**24:**e168–71.

[65] Swafford AN Jr, Bratz IN, Knudson JD, Rogers PA, Timmerman JM, Tune JD *et al.* C-reactive protein does not relax vascular smooth muscle: effects mediated by sodium azide in commercially available preparations. *Am J Physiol Heart Circ Physiol* 2005;**288:**H1786–95.

[66] Clapp BR, Hirschfield GM, Storry C, Gallimore JR, Stidwill RP, Singer M *et al.* Inflammation and endothelial function: direct vascular effects of human C-reactive protein on nitric oxide bioavailability. *Circulation* 2005;**111:**1530–36.

[67] van den Berg CW, Taylor KE. Letter Regarding Article by Li et al, "C-Reactive Protein Upregulates Complement-Inhibitory Factors in Endothelial Cells". *Circulation* 2004;**110:**e542.

[68] Liu C, Wang S, Deb A, Nath KA, Katusic ZS, McConnell JP *et al.* Proapoptotic, antimigratory, antiproliferative, and antiangiogenic effects of commercial C-reactive protein on various human endothelial cell types in vitro: implications of contaminating presence of sodium azide in commercial preparation. *Circ Res* 2005;**97:**135–43.

[69] Nerurkar SS, McDevitt PJ, Scott GF, Johanson KO, Willette RN, Yue TL. Lipopolysaccharide (LPS) contamination plays the real role in C-reactive protein-induced IL-6 secretion from human endothelial cells in vitro. *Arterioscler Thromb Vasc Biol* 2005;**25:**e136.

[70] Pepys MB, Hawkins PN, Kahan MC, Tennent GA, Gallimore JR, Graham D *et al.* Proinflammatory effects of bacterial recombinant human C-reactive protein are caused by contamination with bacterial products, not by C-reactive protein itself. *Circ Res* 2005;**97:**e97–103.

[71] Hirschfield GM, Gallimore JR, Kahan MC, Hutchinson WL, Sabin CA, Benson GM *et al.* Transgenic human C-reactive protein is not proatherogenic in apolipoprotein E-deficient mice. *Proc Natl Acad Sci USA* 2005;**102:**8309–14.

[72] Reifenberg K, Lehr H-A, Baskal D, Wiese E, Schaefer SC, Black S *et al.* Role of C-reactive protein in atherogenesis. Can the apolipoprotein E knockout mouse provide the answer? *Arterioscler Thromb Vasc Biol* 2005;**25:**1641–46.

[73] Trion A, de Maat MP, Jukema JW, van der Laarse A, Maas MC, Offerman EH *et al.* No effect of C-reactive protein on early atherosclerosis development in apolipoprotein E*3-Leiden/human C-reactive protein transgenic mice. *Arterioscler Thromb Vasc Biol* 2005;**25:**1635–40.

[74] Pepys MB. CRP or not CRP? That is the question. *Arterioscler Thromb Vasc Biol* 2005;**25:**1091–94.

ELSEVIER

Estimation of Bias in Nongenetic Observational Studies Using "Mendelian Triangulation"

LEONELO E. BAUTISTA, MD, MPH, DRPH, LIAM SMEETH, MRCGP, MSC, PHD, AROON D. HINGORANI, FRCP, PHD, AND JUAN P. CASAS, MD

PURPOSE: Phenotype–disease odds ratios calculated from the effect of a genotype on its phenotype and on disease risk ("Mendelian triangulation") can be used as a standard to assess bias on the corresponding odds ratio from nongenetic studies. Statistical tests are commonly used to compare these odds ratios. We propose a method to estimate the magnitude of the bias and judge the validity of the phenotype–disease association.

METHODS: For four published examples, we obtained 10,000 random values from distributions of the odds ratios from both genetic and nongenetic studies. A range of values compatible with an unbiased odds ratio was then calculated from the empirical distribution of the differences between both odds ratios.

RESULTS: We show that estimating a range of likely values for an unbiased odds ratio is useful to judge the effect of the phenotype and identify cases for which information from genetic studies adds little to the evaluation of the phenotype–disease association. Conversely, statistical tests could be misleading.

CONCLUSIONS: Estimating a range of values for an unbiased odds ratio is more informative and appropriate than statistical tests when using the Mendelian triangulation approach for assessment of bias in phenotype–disease association studies.

Ann Epidemiol 2006;16:675–680.

KEY WORDS: Mendelian Randomization, Genetics, Observational Epidemiology.

INTRODUCTION

Observational epidemiologic studies are a useful tool to establish causal associations and have had a considerable impact on public health and clinical practice. However, recent discrepancies with results from randomized clinical trials have brought to focus the need for careful design to control bias in observational studies (1). Spurious associations in these studies are frequently the result of uncontrolled confounders and reverse-causality bias. Assessing the likelihood and magnitude of such biases is complicated because some confounding factors may be unknown or difficult to measure, and ascertaining the starting time for exposure and disease may not be possible.

Assuming that the effect of a gene variant on disease is mediated exclusively through its influence on the intermediate phenotype (IP) of interest, the odds ratio from studies of the effect of the genotype on disease risk (ORgd) and the between-genotype difference in IP (ΔIP) can be used to calculate an unconfounded odds ratio ($ORgpd = ORgd^{(k/\Delta IP)}$) for the effect of k units of the IP on the risk of disease (Fig. 1) (2). The fundamental concept behind this approach is that the inheritance of genetic variants is subject to the random assortment of maternal and paternal alleles at the time of gamete formation, according to Mendel's second law (3). Because the presence of the genotype is unlikely to be related to other risk factors for the disease, genotype–disease and genotype–IP studies should be largely free of confounding. They also should be free of reverse-causality bias because genotype is a hereditary fixed characteristic (2, 4). Therefore, the ORgpd also should be unbiased and could be used as a standard to assess whether odds ratios from nongenetic observational studies of the association between an IP and a disease (ORpd) are biased. If ORgpd and ORpd are of a similar magnitude, this offers support for the validity of results from nongenetic observational studies. However, if ORgpd and ORpd differ to a substantial degree, this suggests that results from nongenetic observational studies may be biased.

Although the ORgpd is supposedly unbiased, in some cases, it would not provide a conclusive picture of the phenotype–disease association. This would happen mostly when genetic studies are not large enough to produce a precise estimate of ORgpd. In those cases, the ORgpd still could be used to assess the validity of ORpd and improve our estimates of the strength of the phenotype–disease association.

From the Department of Population Health Sciences, Medical School, University of Wisconsin, Madison, WI (L.E.B.); Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine (L.S., J.P.C.); and Centre for Clinical Pharmacology, Department of Medicine, BHF Laboratories at University College London, UK (A.D.H.).

Address correspondence to: Leonelo E. Bautista, MD, MPH, DrPH, University of Wisconsin, Medical School, Department of Population Health Sciences, 610 Walnut Street, 703 WARF, Madison, WI 53726-2397. Tel.: (608) 265-6176; fax: (608) 263-2820. E-mail: lebautista@wisc.edu.

Received September 8, 2005; accepted February 13, 2006.

1047-2797/06/$ see front matter
doi:10.1016/j.annepidem.2006.02.001

AEP Vol. 16, No. 9
September 2006: 675–680

Selected Abbreviations and Acronyms

IP = intermediate phenotype
ORgd = odds ratio from studies of the effect of the genotype on disease risk
ΔIP = between-genotype difference in intermediate phenotype
ORgpd = unconfounded odds ratio of the intermediate phenotype on risk of disease
ORpd = odds ratios from nongenetic studies of the association between an intermediate phenotype and a disease
MTHFR = methylenetetrahydrofolate reductase
CI = confidence interval
CRP = C-reactive protein
ApoB = apolipoprotein B

A proper assessment of the phenotype–disease association in these cases can be conducted only if both the random variability of the ORgpd and of the ORpd is taken into account when comparing both estimates.

Calculation of the ORgpd has been called "Mendelian randomization" (2), an acknowledgment of the underlying biologic process of random allocation of genetic traits during gamete formation. Because the approach is basically an indirect calculation of the association between an IP and risk of disease, we prefer to call it "Mendelian triangulation" (2, 4). This term has been used previously (2, 4) and conveys the fact that a "triangulation" is applied to data from studies of the effect of the genotype on risk of disease and on IP levels to indirectly obtain ORgpd, an unbiased estimate of the effect of phenotype levels on disease incidence (ORpd). For example, a common polymorphism, C677T, in the gene encoding methylenetetrahydrofolate reductase (MTHFR) has been associated with significant differences in homocysteine levels. We studied the effect of this polymorphism and found that subjects homozygous for the T allele had greater homocysteine levels (ΔIP = 1.93 μmol/L) and

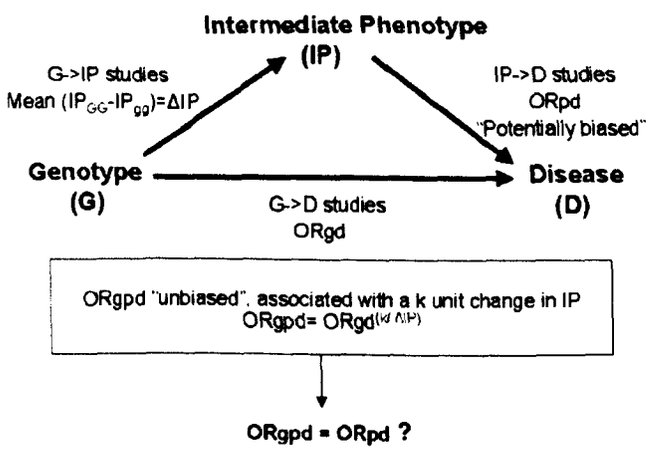


FIGURE 1. Components of the Mendelian triangulation strategy.

greater risk of stroke (ORgd = 1.26) compared with subjects homozygous for the C allele (5). A meta-analysis of prospective studies showed that an increase of 5.0 μmol/L in plasma homocysteine level increases the risk of stroke by 59% (ORpd = 1.59) (6). Based on results from our genetic study, the increase in risk for stroke corresponding to a difference of 5.0 μmol/L in homocysteine levels would be ORgpd = $1.26^{(5/1.93)}$ = 1.81. By comparing ORpd = 1.59 with ORgpd = 1.81, one could evaluate whether results from nongenetic observational studies are distorted by confounding or reverse-causality bias.

In this report, we consider two aspects of Mendelian triangulation that have received relatively little attention to date. First, we consider the random variation in the various estimates used in the Mendelian triangulation strategy, in particular, variability in the estimate of the effect of the genotype on its IP. Second, we ponder the nature of the comparison between the estimate of effect of the IP on disease risk obtained from nongenetic observational studies (ORpd) and the estimate derived from genetic studies (ORgpd). Novel approaches are presented that use reliable estimates of uncertainty derived from multiple simulations and integrate the information from both genetic and nongenetic studies in the evaluation of the IP–disease association.

Uncertainty in Estimates of the Effect of Genotype on IP

To date, most studies using Mendelian triangulation have paid little attention to the sources of random error in the estimation of the ORgpd. In particular, uncertainty in the measurement of the between-genotype difference in phenotype level (ΔIP) has been ignored (7). A failure to incorporate the variability in ΔIP results in overestimation of the precision of the ORgpd and may lead to misleading inferences. Because discordance between ORgpd and ORpd could result from random and/or systematic error, a sound evaluation of bias can be conducted only when random error has been properly accounted for. Assuming a log-linear relationship between the IP and risk for disease, ORgpd can be calculated as $ORgd^{(k/\Delta IP)}$ (7). If ORgd and ΔIP come from large studies, it would be reasonable to assume that log(ORgd) and ΔIP itself are both normally distributed. If ORgd and ΔIP come from different studies, a practical approach to estimate the standard error of ORgpd would be to generate a large number of random values from the distribution of log(ORgd) and ΔIP and use those values to create an empirical distribution of ORgpd values. Although the formula of Fieller (8) for the variance in the ratio of two normally distributed variables could be used to obtain valid confidence intervals (CIs) and statistical tests for the ORgpd, a simulation approach has the additional advantage of providing a simple way to estimate the magnitude of bias

in ORpd and combine data from genetic and nongenetic studies.

Comparing the Estimate of the Effect of IP on Disease Risk Obtained From Nongenetic Studies With That From Genetic Studies: Judging Agreement versus Hypothesis Testing

Judgments on the magnitude of residual confounding and reverse-causality bias in nongenetic IP–disease association studies have relied largely on hypothesis tests of the difference between ORpd and ORgpd (2, 5, 9, 10). Considering solely the size of the ORpd – ORgpd difference can be misleading because large or small differences could result from random variability in either of the two quantities. As suggested previously (2, 9), a *t*-test of the hypothesis ORpd = ORgpd can be constructed as (11):

$$(\text{ORpd} \quad \text{ORgpd})/[(\text{standard error of ORpd})^2 + (\text{standard error of ORgpd})^2]^{1/2}$$

It is possible to incorporate the variability of ΔIP into such a test and thus avoid a spuriously precise result. However, even if correctly calculated, use of a statistical test to judge the magnitude of bias in observational studies can be questioned on theoretical and practical grounds (12). When using the Mendelian triangulation strategy, we take ORgpd as a standard and compare it with ORpd. The underlying principle is that departures of ORpd from ORgpd are caused by *systematic errors*, largely confounding and reverse causality. However, statistical tests aim to evaluate whether the difference between ORpd and ORgpd is caused by *random error*. Because *p* values depend on the magnitude of the difference (bias), as well as on the precision of ORpd and ORgpd, small differences may be statistically significant ($p < 0.05$) if estimates of ORpd and ORgpd come from studies with large sample size. This may lead to the erroneous conclusion that ORpd is significantly biased, even if its value is close to that of the ORgpd. Conversely, large differences between ORpd and ORgpd could be statistically nonsignificant ($p \geq 0.05$) if estimates of ORpd and/or ORgpd come from small studies. In this case, one would erroneously conclude that ORpd is unbiased, even if it is very different from ORgpd. Thus, the significance of a statistical test, regardless of the chosen significance level, is not the most appropriate measure of degree of bias in ORpd.

The likely validity of IP–disease association studies can be assessed by how close the potentially "biased" estimate (ORpd) is to the theoretically "unbiased" estimate (ORgpd). We propose that the *degree* of consistency between ORpd and ORgpd is a matter of estimation and judgment, rather than hypothesis testing. This is analogous to deciding whether material confounding is likely to be present when comparing crude and adjusted odds ratios in an epidemiologic study of a putative risk factor and a disease. Significance testing in such a setting is inappropriate and unhelpful. However, when comparing different estimates of effect, it remains important to consider their precision. Thus, methods to assess the degree of uncertainty caused by random error in these estimates are needed, although hypothesis testing is inappropriate.

An average estimate of the bias can be obtained by generating a large number of random values from the distribution of both log(ORgpd) and log(ORpd). Also, a range of values compatible with an unbiased ORpd can be calculated from the empirical distribution of calculated ORpd – ORgpd differences. This approach is similar to the method of Bland and Altman (13) of limits of agreement used to compare clinical tests. An assessment based on the size of the bias (or agreement for the Bland-Altman method) and the range of reasonably unbiased ORpd would depend on the particular phenotype–disease association being studied.

METHODS

We used data from a study of the association between homocysteine level and stroke that used Mendelian triangulation for the C677T/MTHFR genotype (5) to illustrate the consequence of ignoring variability in ΔIP in the estimation of the precision of the ORgpd. We drew a random sample of 10,000 values from the distribution of log(ORgd) and ΔIP and used them to calculate an equal number of values for ORgpd. A 95% CI for ORgpd was constructed by excluding values greater than the upper and less than the lower 2.5% percentiles of its distribution. The standard error of ORgpd is given approximately by: (upper 95% CI limit – lower 95% CI limit) ÷ (2 * 1.96). We compared results obtained depending on whether variability in ΔIP is taken into account. To show the effect of variability in ΔIP, results were obtained for cumulative results of studies published between 1998 and 2003.

Using four other published examples of studies that used the Mendelian triangulation strategy (2, 9, 10), we randomly drew 10,000 values from the distribution of log(ORpd) and log(ORgpd) (sample means and standard errors). We used these values to estimate the likely degree of bias in the ORpd compared with the ORgpd by calculating:

1. Mean relative bias = mean of $(\text{ORpd}_i - \text{ORgpd}_i) \div \text{ORgpd}_i$
2. Smallest and largest relative bias = 2.5% lowest and upper percentile values for the relative bias $(\text{ORpd}_i - \text{ORgpd}_i) \div \text{ORgpd}_i$
3. Range of reasonable unbiased ORpd = mean $\text{ORpd}_i \div$ lowest and largest relative bias

AEP Vol. 16, No. 9
September 2006: 675–680

These provide a range of reasonable values for ORpd if it were unbiased.

For comparison purposes, we also present for each example results of standard statistical hypothesis tests of a difference between ORpd and ORgpd that do not take into account variability in ΔIP. All analyses were performed using Stata 8.2 (Stata Corp., College Station, TX).

RESULTS

In our previous study of the association between homocysteine level and stroke incidence using Mendelian triangulation for the C677T/MTHFR genotype (5), ΔIP and its standard error varied with the accumulation of data from new studies (Table 1). The phenotype difference corresponding to ORpd was 5 μmol/L. With data available up to 1998, a CI for ORgpd that correctly accounts for random error in ΔIP was almost three times wider than an interval calculated ignoring the variability of ΔIP. By 2003, the sample size was more than 12 times that in 1998, and the standard error of ΔIP was considerably reduced. However, a correct CI was still 9% wider than the interval obtained ignoring the variability of ΔIP.

Results of using repeated simulations to estimate the magnitude of bias in ORpd are listed in Table 2. In example 1, coronary heart disease is the outcome of interest, homocysteine is the IP, and the MTHFR/C677T variant is the polymorphism associated with both phenotype levels and disease risk (2). Mean relative bias was 0.975, i.e., ORpd was underestimated on average by 2.5%. However, according to the data at hand, ORpd could be underestimated by 13% or overestimated by 9%, and the range of reasonable values for an unbiased ORpd was 1.04 to 1.29. Because this range of values for ORpd does not include 1, we are inclined to conclude that greater plasma homocysteine level increases the risk for coronary heart disease. In other words, even if ORpd is biased, the bias is small and it is unlikely that the observed phenotype–disease association is caused mostly by confounding or reverse-causality bias. Therefore, results of genotype–disease and genotype–phenotype association studies provide additional support for the association between phenotype and disease quantified in nongenetic observational studies. In this example, p value from a t-test of a difference between ORpd and ORgpd that does not take into account variability in ΔIP was 0.69. This nonsignificant p suggests that the difference between the two estimates could be attributed to random error. However, it provides no insight into evidence underlying the judgment of lack of bias because it could be the result of a small ORpd – ORgpd difference and/or large variability in either of the two estimates (i.e., lack of power). The test is conducted under the assumption that the difference between ORpd and ORgpd is not the consequence of bias.

TABLE 1. Mendelian triangulation estimates of the odds ratio for stroke for a 5-μmol/L difference in plasma homocysteine level with and without correcting the variance for uncertainty in the association between MTHFR/C677T genotype and homocysteine level

Studies up to	Sample size	ΔIP[a]	SE (ΔIP)[b]	Odds ratio	95% CI ignoring SE(ΔIP)[b]	95% CI accounting for SE(ΔIP)[b]	Ratio of width[c]
1998	730	2.10	0.5996	1.73	1.37–2.23	1.31–3.82	2.92
1999	935	2.72	0.3725	1.53	1.27–1.86	1.25–2.00	1.27
2000	2515	2.18	0.2282	1.70	1.35–2.16	1.33–2.28	1.17
2001	3619	1.94	0.2237	1.81	1.40–2.38	1.38–2.55	1.19
2002	5847	1.86	0.1620	1.86	1.42–2.47	1.40–2.57	1.11
2003	9139	1.93	0.1477	1.82	1.40–2.39	1.38–2.46	1.09

MTHFR = methylenetetrahydrofolate reductase; CI = confidence interval.
[a]Difference in plasma homocysteine levels between individuals with genotype TT and CC for the MTHFR/C677T polymorphism.
[b]Standard error of the between-genotype phenotype difference in plasma homocysteine levels.
[c]Ratio of the width of the 95% CIs that takes into account and the one that ignores the standard error of the between-genotype phenotype difference in plasma homocysteine levels.

Example 2 is of coronary heart disease, fibrinogen level, and the –148 *Hind*III variant (2). There was an average positive bias of 17%. Taking random error into account, ORpd was overestimated by 7% (smallest relative bias) to 27% (largest relative bias). These figures correspond to values of 0.94 to 1.12 for a range of unbiased ORpd. Therefore, available data from both nongenetic observational studies and genotype–IP and genotype–disease studies point to a very small or a null effect of fibrinogen level on risk for coronary heart disease. The effect reflected by the ORpd seems to be mostly a result of bias. The p from a t-test for a difference between ORpd and ORgpd that does not take into account the variability in ΔIP was highly significant ($p < 0.001$). This small p value results from a large difference between ORpd and ORgpd and very precise measurements of both estimates. However, contrary to the estimation approach, the p value does not allow us to explicitly combine all data from genetic and nongenetic studies to make a judgment on the phenotype–disease association.

Example 3 is of coronary heart disease, fibrinogen level, and the –148 *Hind*III variant (9). The relative mean bias was very small (−0.6%), suggesting the ORpd may be unbiased. However, the range of plausible unbiased ORpd values was very wide (0.70 to 1.64), reflecting the large uncertainty in estimation of ORgpd. In this case, a conclusion regarding the effect of plasma fibrinogen level on coronary heart disease is not warranted because the range of values for unbiased ORpd is too wide. Briefly, if ORgpd from the Mendelian triangulation cannot be calculated with sufficient precision, it will be of limited use to assess whether ORpd is biased. The p value from a t-test for a difference between ORpd and ORgpd was nonsignificant ($p = 0.80$). This could have led to the inappropriate conclusion that ORpd was

TABLE 2. Examples of comparison of odds ratios from nongenetic observational studies (ORpd) and odds ratios from genotype–phenotype disease studies (ORgpd)

Exposure (phenotype/ disease/gene variant)	ORpd	ORgpd	Relative mean bias	Smallest bias (%)	Largest bias (%)	Smallest ORpd[a]	Largest ORpd[a]	p^b
1. Homocysteine/CHD/MTHFR/C677T[2]	1.13 (1.08–1.19)	1.16 (1.05–1.28)	0.975	0.874	1.09	1.04	1.29	0.69
2. Fibrinogen/CHD/-148 Hind III[2]	1.20 (1.13–1.26)	1.03 (0.96–1.10)	1.17	1.07	1.27	0.94	1.12	<0.001
3. Fibrinogen/CHD/-148 Hind III[2,9]	1.07 (1.01–1.11)	1.08 (0.71–1.65)	0.994	0.653	1.52	0.70	1.64	0.80
4. Apo B/Apo A1/MI Apo-E ε2/ε3/ε4[10]	1.63 (1.58–1.68)	1.11 (1.07–1.36)	1.36	1.20	1.54	1.06	1.36	<0.001

CHD = coronary heart disease; Apo B = apolipoprotein B; MTHFR = methylenetetrahydrofolate reductase.
[a]Calculated using the 95% confidence limits of the relative bias.
[b]p from a t-test for a difference between ORpd and ORgpd that does not take into account the variability of ΔIP.

unbiased. However, the estimation approach clearly illustrated that there was absence of evidence of bias, rather than clear evidence of absence of bias.

Example 4 is of myocardial infarction, apolipoprotein B (ApoB)/ApoA1, and Apo-E ε2/ε3/ε4 variants (10). The ORpd was biased substantially, with a relative mean bias of 36%. Despite the large bias, the range of unbiased values for the ORpd (1.06 to 1.35) suggests that the ratio of plasma Apo B/Apo A1 is associated with increased risk for myocardial infarction. However, the increase in risk is likely smaller than that reported in nongenetic observational studies. The p value from a t-test for a difference between ORpd and ORgpd was highly significant ($p < 0.001$). This would lead to a conclusion that ORpd was biased. However, the more subtle interpretation that whereas there was likely to be substantial bias in the estimate of ORpd, the association was still likely to be present could not be drawn from the p value.

DISCUSSION

We have shown that simulations can help correctly interpret combined results from nongenetic studies and Mendelian triangulation studies. Random variation in ΔIP and thus ORgpd can be estimated reliably, avoiding the problem of spurious precision that arises when such variability is ignored. More important, simulations also are useful to obtain an estimate of the likely degree of bias in ORpd, which cannot be accomplished with statistical tests. We propose that the comparison between ORpd and ORgpd be based on their degree of consistency and is a matter of estimation and judgment about the level of systematic error in the ORpd, rather than of hypothesis testing. The most reasonable way to evaluate bias in this context is to estimate its magnitude and judge whether it is sufficiently large to make us believe that confounding or reverse-causality bias were so substantial as to shed doubts on the validity of ORpd. Even if random error is negligible, proper quantification of the magnitude of the bias is essential to assess whether the nature of the phenotype–disease association is likely valid and possibly causal. By calculating a range of plausible unbiased values for the phenotype–disease association, taking into account results from genotype–disease and genotype–IP studies, one can make full use of the information available to learn what should be reasonably expected in the case of extreme underestimation or overestimation bias. This may help identify cases in which the IP is still likely associated with the disease, even if considerable bias is present. This cannot be accomplished by using the "yes" or "no" decision rules attached to statistical tests and the use of single cutoff values. In addition, the range of plausible values for an unbiased ORpd allows us to identify studies in which the ORgpd obtained by Mendelian triangulation is so uncertain that it is practically of little use to assess the validity of ORpd. In these cases, the range of values for the ORpd could be too wide to reach a conclusion, and inspection of the CI for ORgd, ΔIP, and ORpd would help us define what type of additional information should be gathered in future studies. This cannot be accomplished with statistical tests.

Alternative approaches to the problems we highlight include the use of bioequivalence tests and instrumental variables. Bioequivalence tests have been proposed as a method to measure the degree of residual confounding in ORpd (7). Because these tests evaluate the probability of getting a difference equal to or larger than a preset value, they pertain to the evaluation of random error, and their use in evaluating systematic error (residual confounding and reverse-causality bias) is questionable. Large differences between ORpd and ORgpd could be statistically nonsignificant simply because of lack of study power. This could happen if the estimates of ORpd and/or ORgd come from small studies. Conversely, a small ORpd – ORgpd difference, which will indicate little, if any, systematic error, could be statistically significant if both ORpd and ORgpd are measured with high precision, such as would be expected in large studies. Briefly, p value from a statistical test by itself will tell very little about the magnitude of the bias in ORpd regardless of the level of significance used for interpretation.

Another approach to control for unmeasured confounders in the context of genetic studies is the use of instrumental-variable methods (14,15). For example, the instrumental variable approach was used to estimate an unbiased effect of plasma C-reactive protein (CRP) level on

AEP Vol. 16, No. 9
September 2006: 675–680

hypertension, taking advantage of the association between the 1059G/C polymorphism in the human CRP gene with CRP levels (16). Genetic variants may be treated as instrumental variables if they are unrelated to other risk factors for the disease under study, correlate with the IP, and are related to disease risk only through their association with the IP. An estimate of the effect of IP on disease risk can be obtained as the ratio of the genotype–disease effect/genotype–IP effect. Hausman's (17) specification error test frequently is used to evaluate whether the regression coefficient from a model including the instrumental variable differs from a model not including the instrumental variable. Because the goal of the instrumental variable approach is largely to evaluate and control for systematic error, not for random error, use of statistical tests in this context is as questionable as its use in Mendelian triangulation. In addition, the instrumental variable adjustment for confounding can be particularly unreliable if the genotype–IP association is weak and when the genotype–disease and the genotype–IP associations have been measured with considerable uncertainty (14, 15). When results from an instrumental variable analysis differ from those of a standard analysis, the judgment of bias becomes subjective (18). An approach based on estimation, rather than statistical testing, such as the one we propose, may be useful in these cases.

We recognize that our approach should be used with caution and tested with additional data. If data from overlapping studies are used to estimate ORgd, ΔIP, and ORpd, the nonindependence of the observations must be taken into account in the calculation of the variability of the ORgpd. Also, when any of these estimates come from small studies, the assumption of normal distribution may not apply. In such situations, the values of ORgpd and ORpd should be generated by sampling random values from the proper distributions of ORgd, ΔIP, and ORpd.

Despite its potential value, the Mendelian triangulation strategy should be considered as an emerging additional source of evidence about the effects of environmental exposures on disease, not as proof of causality. A number of caveats may apply. First, if the studied genotype has alternative splicing leading to the production of two different IPs, the Mendelian triangulation approach may not be appropriate. Second, it is assumed that the lifetime genetically determined difference in an IP does not lead to compensatory changes in other systems that may confound the genetic estimate. Third, the genotype–IP and genotype–disease associations may be confounded by linkage disequilibrium with other causal genes or by ethnicity, population stratification. Fourth, cases and controls in genetic studies are rarely random samples from their population, and relatively small samples frequently are used. In consequence, associations between a genotype and risk factors for a disease can occur in genetic studies and can result in selection bias and confounding, even if they are not expected to occur in the entire population (19).

In conclusion, Mendelian triangulation is a promising strategy for the study of disease cause, but adequate methods to compare ORgpd and ORpd are needed to avoid misleading assessment of the causal role of an IP on disease risk. The simulation methods presented here provide a useful approach to attain this goal.

REFERENCES

1. Lawlor DA, Davey SG, Ebrahim S. Commentary: The hormone replacement-coronary heart disease conundrum: Is this the death of observational epidemiology? Int J Epidemiol. 2004;33:464–467.
2. Davey SG, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1–22.
3. Morgan TH. Heredity and Sex. New York: Columbia University; 1913.
4. Davey SG, Ebrahim S. Mendelian randomization: Prospects, potentials, and limitations. Int J Epidemiol. 2004;33:30–42.
5. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: Evidence on a causal link from Mendelian randomisation. Lancet. 2005;365:224–232.
6. Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: Evidence on causality from a meta-analysis. BMJ. 2002;325:1202–1206.
7. Thompson JR, Tobin MD, Minelli C. On the Accuracy of Estimates of the Effect of Phenotype on Disease Derived From Mendelian Randomisation Studies. Technical report 2003-GEI. 8-11-2003. Leicester: University of Leicester. Available at: http://www.prw.le.ac.uk/research/HCG/research.html. Accessed
8. Fieller E. Some problems in interval estimation. J R Stat Soc. 1954;16:175–178.
9. Davey SG, Harbord R, Ebrahim S. Fibrinogen, C-reactive protein and coronary heart disease: Does Mendelian randomization suggest the associations are non-causal? QJM. 2004;97:163–166.
10. Keavney B, Palmer A, Parish S, Clark S, Youngman L, Danesh J, et al. Lipid-related genes and myocardial infarction in 4685 cases and 3460 controls: Discrepancies between genotype, blood lipid concentrations, and coronary disease risk. Int J Epidemiol. 2004;33:1002–1013.
11. Altman DG, Bland JM. Interaction revisited: The difference between two estimates. BMJ. 2003;326:219.
12. Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman N, Greenland S, eds. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven; 1998.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–310.
14. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29:722–729.
15. Thomas DC, Conti DV. Commentary: The concept of 'Mendelian randomization'. Int J Epidemiol. 2004;33:21–25.
16. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GD, et al. Association of C-reactive protein with blood pressure and hypertension. Life course confounding and Mendelian randomization tests of causality. Arterioscler Thromb Vasc Biol. 2005;25:1051–1056.
17. Hausman JA. Specification tests in econometrics. Econometrics. 1978;46:1251–1271.
18. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: Relating endogeneity and unobserved heterogeneity to confounding. Ann Epidemiol. 1997;7:251–257.
19. Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005;366:1906–1908.

European Heart Journal (2006) 27, 1261–1263
doi:10.1093/eurheartj/ehi852

Editorial

Linking observational and genetic approaches to determine the role of C-reactive protein in heart disease risk

Aroon D. Hingorani[1]*, Tina Shah[1], and Juan P. Casas[2]

[1]Department of Medicine, Centre for Clinical Pharmacology, BHF Laboratories, University College London, Rayne Building, 5 University Street, London WC1E 6JJ, UK; and [2]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Online publish-ahead-of-print 9 May 2006

This editorial refers to 'C-reactive protein gene haplotypes and risk of coronary heart disease. The Rotterdam Study'† by I. Kardys *et al.*, on page 1331

In the last decade, there has been a burgeoning interest in associations between concentrations of various circulating biomarkers and later coronary disease.[1] Of these, the acute phase reactant, C-reactive protein, has garnered the greatest attention. This spotlight on C-reactive protein reflects the strength and consistency of the observed associations, the fact that C-reactive protein is a stable and readily measured analyte, and the focus on the link between inflammation and atherosclerosis. The findings led initially to the proposal that its measurement might provide useful predictive information on coronary disease risk.[2] Later, the proposal was made that C-reactive protein might even play a contributory role in atherogenesis.[2] Associations identified between C-reactive protein and incident blood pressure, diabetes, and metabolic syndrome were used in support of this, and experimental data from studies of vascular cells and tissues *in vitro* and from infusions of C-reactive protein provided preliminary evidence for pro-inflammatory, pro-adhesive, and pro-thrombotic actions.[3] These data are exciting because it was through a similar mix of observational and experimental work that cholesterol was first identified as a potential causal factor in atherosclerosis, a link which randomized clinical trials of cholesterol-lowering drugs subsequently confirmed to be causal. If the same is true for C-reactive protein, it would represent a new therapeutic target for coronary disease prevention.

More recently, however, evidence has emerged, which questions the role of C-reactive protein in both the prediction of clinical events and in the pathogenesis of atherosclerosis. A meta-analysis of the prospective cohort studies, including a very large new study, indicated that the strength of the C-reactive protein-coronary event association may be smaller than initially supposed and that C-reactive protein may offer limited predictive information, particularly after levels of traditional risk factors are taken into account.[4] This is because C-reactive protein is itself associated with a range of established risk factors for cardiovascular disease including smoking, blood pressure, diabetes, as well as HDL-cholesterol. Once these are measured, it has been argued, there is little additional predictive information in a C-reactive protein measurement.

Associations among established risk factors and biomarkers also limit inference on causation in observational studies because associations of any one factor such as C-reactive protein with incident disease could be affected by confounding (where the link arises because of the common association of the biomarker and the disease outcome with other causal risk factors like smoking, or with an enlarging list of other biomarkers of less certain causal relevance). Statistical adjustment can limit confounding, but rarely overcomes it completely. Adequate adjustment can only be made for those risk factors or biomarkers that have been measured. In a recent review, perhaps 100 biomarkers or exposures (including C-reactive protein) were listed as having shown association with coronary risk in observational studies.[1] Importantly, most of these correlate to a greater or lesser degree with one another. However, no study has adjusted for them all. Even if all were assayed, the inherent biological and analytical variability in any measure generally limits the ability to make complete adjustment. Despite this, the persisting association of C-reactive protein with disease risk, after adjustment for a generally limited range of risk factors or other biomarkers, referred to as 'independent association' in statistical parlance, has sometimes been over-interpreted as equating with causation. The possibility of residual confounding by unmeasured exposures remains however, and the link with the disease may not be causal. There is an additional problem. Observational studies may not well define the direction of causation, should it exist. Reverse causation (where atheroma leads to an elevation of C-reactive protein rather than vice versa) could provide an alternative explanation for the link between C-reactive protein concentration and disease risk. Although this is more of a problem in case-control studies, it may still

The opinions expressed in this article are not necessarily those of the Editors of the *European Heart Journal* or of the European Society of Cardiology.

* Corresponding author.
E-mail address: a.hingorani@ucl.ac.uk
† doi:10.1093/eurheartj/ehl018

affect prospective studies of initially healthy subjects, because atherosclerosis has such a long subclinical phase.

Against this setting, another prospective observational study demonstrating the association of C-reactive protein with coronary events is reported by Witteman *et al.*[5] This association, which diminishes but persists after adjustment for a range of potential confounders, is insistently concordant with prior work. So what is different about this study? The beauty of the current work is that, in line with a growing number of similar studies, it takes the observational analysis, with its inherent limitations, a stage further by incorporating a genetic approach. Taking advantage of its recent resequencing (http://pga.gs.washington.edu/data/crp/ and Carlson *et al.*,[6]), Kardys *et al.* type three common genetic variants (single nucleotide polymorphisms; SNPs) in the gene that encodes C-reactive protein. Through their association with the 12 or so additional untyped variants in the gene, these three 'tagging' SNPs adequately capture almost all the genetic diversity at this locus in European subjects. All the SNPs, as well as the haplotypes inferred from them, are associated with differences in C-reactive protein concentration of about 0.6 mg/L between the most extreme categories, representing nearly 25% of the interquartile range of the C-reactive protein distribution in this population. Seasoned readers of the journal may have become wary of positive findings from genetic association studies with disease end-points, because of their inconsistency. Nevertheless, they can be reassured that, in this case, the association between C-reactive protein genotype and C-reactive protein concentration (an intermediate phenotype) is almost certainly correct. The genotype-C-reactive protein associations identified by the Rotterdam study are closely concordant with other studies of C-reactive protein SNPs.[6–8] In general, genotype biomarker associations appear to be more readily established than genotype disease associations perhaps because: (a) a change in level or property of the biomarker is the immediate consequence of the genetic variation, so the 'signal-to-noise ratio' is greater in such studies; (b) a given biomarker is likely influenced by a smaller range of genes and exposures than the disease; (c) biomarkers are continuous traits, so power is enhanced greatly over studies where the outcome is categorical (e.g. disease event); and (d) biomarker levels are less prone to misclassifcation error than commonly used disease end-points such as 'myocardial infarction' or 'angiographic coronary artery disease (CAD)'.

So why bother to study gene variants when one can measure C-reactive protein itself? Even if genotype contributes to differences in C-reactive protein concentration, it is only one of the many factors to do so, and the genetic contribution to the variance in C-reactive protein is likely to be much smaller than that due to the combined effect of acquired exposures such as abdominal obesity, smoking, blood pressure, and so on. The answer is that, there is one particular property of genotype that distinguishes it from all the acquired exposures that influence C-reactive protein. Genotype is determined at random at conception according to Mendel's Laws, because of the independent assortment of alleles during meiotic cell division and the reconstitution of the diploid genome during fusion of gametes. In effect, allocation to a low-C-reactive protein genotype (or haplotype) at conception is analogous to being randomized to a specific C-reactive protein-lowering drug in a clinical trial. This natural Mendelian randomization should ensure that factors that could confound associations of C-reactive protein itself with coronary disease should be distributed evenly among individuals with differing C-reactive protein genotype (or haplotype).[9,10] It is clear from Kardys's study that this is indeed the case. Although a whole range of measured (and by extrapolation unmeasured) exposures relevant to coronary risk varied by quartile of measured C-reactive protein, none differed among the four common haplotypes or genotypes of the C-reactive protein gene, despite a substantial genetic influence on C-reactive protein concentration. Moreover, as genotype/haplotype is determined prior to the disease, reverse causation should also be overcome. Thus, studying the effect of C-reactive protein indirectly through its gene, counter-intuitively, provides a less biased and confounded assessment of its influence on disease risk than studying C-reactive protein itself. If a high C-reactive protein actually contributed causally to coronary disease risk, individuals with genotypes that raise C-reactive protein concentration should have a higher risk of coronary disease, commensurate with the effect of genotype on C-reactive protein level. In the current study, however, no such increase in coronary risk was observed.

So is the case now closed? Has a causal role for C-reactive protein in atherosclerosis been unequivocally refuted? The answer is no, but evidence such as this is forcing a more cautious and critical approach. The null point estimate from the current genetic study is concordant with genetic risk estimates from other work[7,8] and consistent with recent experimental studies *in vitro*, which have indicated that the apparently pro-atherogenic actions of C-reactive protein observed in earlier studies may have been mediated by bacterial peptides and sodium azide present in commercial recombinant C-reactive protein preparations rather than by C-reactive protein itself.[11] Because the genetic effect on C-reactive protein is small, a very large genetic association study with about 10 000 cases would be required to exclude a small but, in population terms, extremely important causal effect of C-reactive protein on coronary outcomes. Accumulating this number of cases may be beyond the capacity of individual research groups, but could be achieved through a collaborative effort to pool studies of the type reported here. The development of networks of genetic investigators may help facilitate this and, in the absence of a selective C-reactive protein-lowering drug, may provide the only currently practical way of providing randomized evidence on the effect of C-reactive protein-lowering on disease risk.[12] Although statins lower C-reactive protein, this is neither their primary nor main action and it is unclear if this effect is achieved independently of or through cholesterol reduction.

What of C-reactive protein and disease prediction? C-reactive protein might still provide useful information on coronary risk even if it does not play a causal role in coronary disease, because its concentration might capture information on the extent of exposure to orthodox risk factors, inflammatory signals from adipose, or the burden or inflammatory state of atheromatous plaques. However, the finding that common variation in the C-reactive protein gene influences C-reactive protein concentration both in health and during acute inflammation has potentially

important implications for risk prediction. Two individuals with contrasting C-reactive protein genotype but the same measured C-reactive protein concentration might have different underlying levels of risk factor exposure or inflammatory activity. If it is these that influence disease outcome, the prognostic value of a given C-reactive protein measure may need to be considered in conjunction with genotype or haplotype. Additional work will be required to assess if the predictive utility of a C-reactive protein measurement is modified by genotype.

Combining genetic and observational approaches as Kardys *et al.* have done, offers the possibility of better distinguishing causal from confounded associations of biomarkers and coronary disease and may, in future, help prioritize new therapeutic targets. Rich and expanding databases of human genome variation are providing the genetic tools for this type of work, but of no lesser importance are the existence of well-phenotyped cohorts, of which the Rotterdam study is an excellent example.

Acknowledgement

A.D.H. acknowledges the support of the British Heart Foundation Senior Research Fellowship.

Conflict of interest: No financial conflicts of interest.

References

1. Brotman DJ, Walker E, Lauer MS, O'Brien RG. In search of fewer independent risk factors. *Arch Intern Med* 2005;**165**:138–145.
2. Verma S, Szmitko PE, Ridker PM. C-reactive protein comes of age. *Nat Clin Pract Cardiovasc Med* 2005;**2**:29–36.
3. Bisoendial RJ, Kastelein JJ, Levels JH, Zwaginga JJ, van den Bogaard B, Reitsma PH, Meijers JC, Hartman D, Levi M, Stroes ES. Activation of inflammation and coagulation after infusion of C-reactive protein in humans. *Circ Res* 2005;**96**:714–716.
4. Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med* 2004;**350**:1387–1397.
5. Kardys I, de Maat MPM, André G, Uitterlinden AG, Albert Hofman A, Witteman JCM. C-reactive protein gene haplotypes and risk of coronary heart disease. The Rotterdam Study. *Eur Heart J* 2006;**27**:1331–1337. First published on May 3, 2006, doi:10.1093/eurheartj/ehl018.
6. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, Fornage M, Boerwinkle E, Gross M, Jaquish C, Nickerson DA, Myers RM, Siscovick DS, Reiner AP. Polymorphisms within the C-reactive protein promoter region are associated with plasma CRP levels. *Am J Hum Genet* 2005;**77**:64–77.
7. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffrey D, Pockard CJ, O'Reilly DS, Juhea-Vague I, Yudkin JS, Tremoli E, Margaglione M, Di Pinno G, Hamster A, Kooistra T, Stephens JW, Hurel SJ, Livingstone S, Colhan HM, Miller GJ, Bautista LE, Mecde T, Sattar N, Humphries SE, Hingorani AD. Insight into the CRP-coronary event association using Mendelian randomization. *Int J Epidemiol* 2006, in press.
8. Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM, Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann Hum Genet* 2005;**69**:623–638.
9. Davey SG, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
10. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005; **366**:1906–1908.
11. Pepys MB, Hawkins PN, Kahan MC, Tennent GA, Gallimore JR, Graham D, Sabin CA, Zychlinsky A, de Diego J. Proinflammatory effects of bacterial recombinant human C-reactive protein are caused by contamination with bacterial products, not by C-reactive protein itself. *Circ Res* 2005; **97**:e97–e103. Published online ahead of print, 2005.
12. Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ; Human Genome Epidemiology Network and the Network of Investigator Networks. A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006;**38**:3–5.

Review

Journal of INTERNAL MEDICINE

doi: 10.1111/j.1365-2796.2008.02015.x

C-reactive protein and coronary heart disease: a critical review

J. P. Casas[1,2], T. Shah[2], A. D. Hingorani[2], J. Danesh[3] & M. B. Pepys[4]

From the [1]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine; [2]Division of Medicine, Centre for Clinical Pharmacology, British Heart Foundation Laboratories, University College London, London; [3]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge; and [4]Division of Medicine, Centre for Amyloidosis and Acute Phase Proteins, University College London, London, UK

Abstract. Casas JP, Shah T, Hingorani AD, Danesh J, Pepys MB (London School of Hygiene and Tropical Medicine, University College London, London; Institute of Public Health, University of Cambridge, Cambridge and University College London, London; UK). C-reactive protein and coronary heart disease: a critical review (Review). *J Intern Med 2008*; **264**: 295–314.

Modestly elevated baseline concentrations of C-reactive protein (CRP), the classical acute phase protein, are associated with the long-term risk of coronary heart disease in general populations, whilst the major acute phase response of CRP following myocardial infarction is associated with death and cardiac complications. The pathogenic and clinical significance of these associations is controversial. Here we critically review the evidence and describe large-scale epidemiological studies, novel experiments and possible specific therapies which will rigorously inform the debate. We distinguish between the potential pathogenicity of high acute phase circulating CRP concentrations in individuals with substantial tissue damage and modest but persistent increases in baseline values in generally healthy subjects.

Keywords: coronary heart disease, C-reactive protein.

Introduction

Inflammation has been proposed to contribute to different stages in the pathogenesis of coronary heart disease (CHD), including the lifelong process of atherogenesis, the acute atherothrombotic event, which causes ischaemic necrosis in acute myocardial infarction (MI) and the myocardial damage following ischaemia [1–5]. C-reactive protein (CRP), the classical acute phase protein, is the most extensively studied systemic marker of inflammation [6]. Since the original demonstration of an association between even modest increases in baseline circulating CRP and subsequent cardiovascular outcomes in patients with unstable angina at the initial examination [7, 8], CRP has been the focus of intense investigation [5].

Several population-based prospective studies of CHD (defined in this paper as non fatal MI or coronary death) have reported on associations of subtle, prolonged increases in baseline CRP levels with CHD risk [9]. Also, CRP has been known for more than 25 years to bind to LDL [10, 11] and it has been detected in atherosclerotic plaques [12]. These observations have collectively raised the possibility that CRP may play a direct causal role in CHD (and, by implication, could be an important therapeutic target), although there is debate about the interpretation of both epidemiological and experimental observations [13–15]. There is also uncertainty about the separate issue of the value of measurement of CRP in the long-term prediction of cardiovascular outcomes [16, 17]. A completely distinct possibility is that, in contrast to modest long-term differences in baseline CRP values, the major acute phase response of CRP triggered by MI may contribute acutely to severity and outcome of the ischaemic lesion [18]. The objective of this article is to provide a critical review of the

available evidence on associations of CRP with CHD and to highlight emerging approaches that may help clarify the uncertainties described above.

Structure of CRP

C-reactive protein is a nonglycosylated circulating plasma protein which together with the distinct but closely related protein, serum amyloid P component (SAP), comprises the pentraxin family of proteins [19, 20], which belong to the lectin fold superfamily [21]. Each pentraxin is composed of five identical noncovalently associated subunits arranged with cyclic pentameric symmetry in a disc-like configuration and has characteristic calcium-dependent binding to specific ligands (Fig. 1). Each CRP protomer has the flattened β-jellyroll lectin fold and bears on one face, the B or binding face, a pocket which contains two calcium ions bound just 4 Å apart by coordination with protein carboxylate and amide side chains derived from loops that congregate on one face of the protomer core. These calcium atoms are essential for all physiological ligand binding by CRP and also markedly stabilize both the structure of the protomer [22] and the integrity of the native pentamer. Calcium is, of course, always present in the extracellular environment *in vivo* in which CRP exists, and inappropriate omission of calcium in handling CRP for *in vitro* experiments produces misleading observations.

In the absence of calcium, even if pH and ionic strength are physiological, CRP readily aggregates, loses its physiological ligand binding and the normal secondary effects thereof, dissociates into protomers which then readily misfold, and becomes susceptible to proteolytic cleavage. In marked contrast, native CRP in the presence of physiological conditions of calcium, pH and ionic strength is remarkably stable. Although isolated pure CRP can aggregate if subjected to violent turbulence, it is otherwise very stable, unaffected by repeated freeze-thawing, remains in its normal pentameric assembly and is highly resistant to proteolysis [23]. It does not dissociate into protomers even in the presence of the very strong anionic detergent, sodium dodecyl sulphate, unless boiled. CRP is also stable in serum which enhances greatly its utility as a clinical marker. In view of these robust observations, speculations about physiological or pathophysiological effects of denatured CRP subunits, so-called modified CRP, or of proteolytic cleavage fragments of CRP are unlikely to be correct and the evidence for existence of such moieties *in vivo* is unconvincing.

The critical importance of appropriate handling of CRP is strikingly exemplified by recent observations

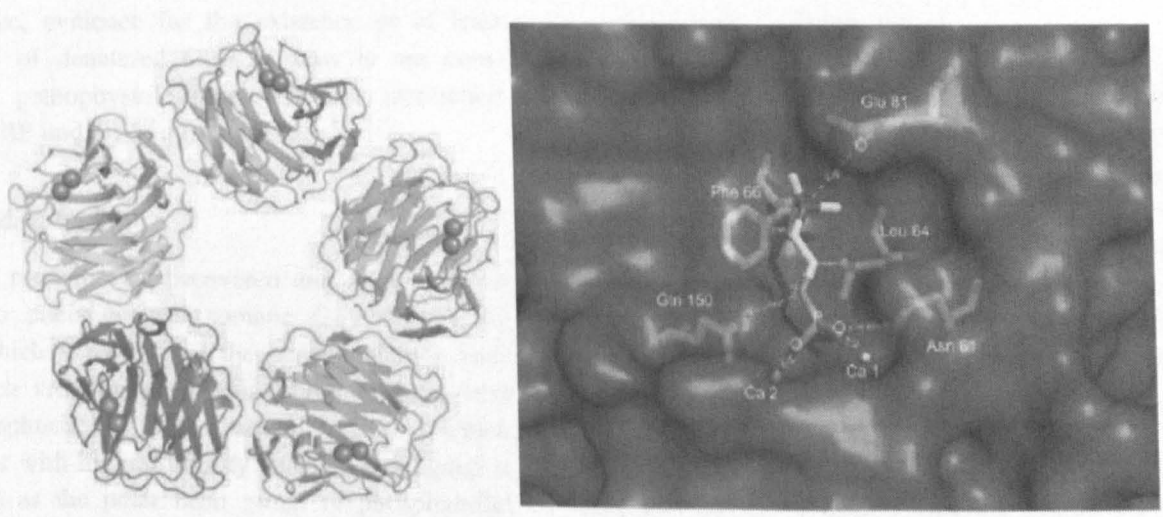


Fig. 1 3D X-ray crystal structure of the human C-reactive protein (CRP) molecule with bound calcium (left) and of a phosphocholine molecule bound in the ligand binding pocket of one CRP protomer (right), showing the intermolecular interactions responsible for binding [187]. X-ray crystallographic structures provided by Dr. Simon Kolstoe.

on the putative binding interaction between CRP and complement regulatory protein, factor H (fH). One of the two major polymorphic variants of fH is a highly significant risk factor for age-related macular degeneration [24–26], and the amino acid substitution (Y402H) is in the seventh short consensus repeat, the region of fH reported to bind to CRP [27]. Recent reports of binding between CRP and fH were all based on solid phase studies in which CRP was immobilized, usually from calcium-free solutions and thereby destabilized, by nonspecific adherence to plastic surfaces, conditions in which such immobilized proteins are inevitably at least partly denatured [28]. When these studies were repeated using CRP in the presence of calcium, no interaction with fH was detected (S. Hakobyan, C.L. Harris, C.W. van den Berg, M.C. Fernandez Alonso, E. Goicoechea de Jorge, S. Rodriguez de Cordoba, G. Rivas, P. Mangione, M.B. Pepys and B.P. Morgan, unpublished data). Further extensive controlled studies with fluid phase proteins and different configurations and methods of immobilization of CRP or fH confirmed that there was no detectable interaction between fH and native CRP either in solution or on a solid phase (S. Hakobyan, C.L. Harris, C.W. van den Berg, M.C. Fernandez Alonso, E. Goicoechea de Jorge, S. Rodriguez de Cordoba, G. Rivas, P. Mangione, M.B. Pepys and B.P. Morgan, unpublished data). Binding was only seen when the CRP was denatured and since, as noted above, evidence for the existence or at least persistence of denatured CRP *in vivo* is not compelling, a pathophysiologically relevant interaction between CRP and fH is questionable.

Ligand binding by CRP

C-reactive protein was discovered and named for its binding to pneumococcal somatic C-polysaccharide [29] in which it recognizes the phosphocholine residues which are present in this ribitol teichoic acid [30]. Phosphocholine is the natural ligand to which CRP binds with highest affinity and this key ligand is ubiquitous as the polar head group of phosphatidlyl choline in cell membranes and plasma lipoproteins. Phosphocholine is also present in constituents of many bacteria, fungi and parasites and plants and the importance for mammalian biology of its recognition is exemplified by the fact that a significant proportion of the germline antibody specificities are directed at it. However, CRP does not bind to all materials containing phosphocholine as the residues must be 'available' or in an appropriate sterochemical configuration. Thus CRP binds to dead or damaged cells in which significant amounts of lysophosphatidyl choline are present, but not the surface of living healthy cells [31]. Binding of CRP to apoptotic cells is controversial and the most rigorous evidence suggests that CRP only binds to so-called late apoptotic cells which are effectively necrotic [32–34]. CRP also binds to oxidized phospholipids [33], platelet activating factor [35], modified LDL [36], β-VLDL, concentrated normal VLDL [37] and to small nuclear ribonucleoprotein particles (which do not contain phosphocholine) when these are exposed in dead or damaged cells [38, 39].

Binding of CRP to its macromolecular ligands has many of the same effects as binding of antibodies to antigens, thus CRP precipitates soluble ligands, aggregates particulate ligands and activates the classical complement pathway [40–42]. By analogy with antibodies, it is therefore possible that CRP might contribute both to host defence against infection and enhancement of inflammatory tissue damage.

Functions of human CRP

Despite many claims and assertions in the literature, neither the normal functions of human CRP nor its possible role in disease is known. This is because no deficiency or even structural polymorphism of human CRP has yet been reported, nor is any drug or other therapeutic manoeuvre yet available which specifically inhibits or depletes human CRP *in vivo*. The effects of absence, lack of function or inhibition of human CRP have thus so far not been tested.

The notable evolutionary and phylogenetic conservation of both the existence and the structure of the CRP molecule [20, 43, 44] suggest that it is likely to have a beneficial role and in view of its ligand binding specificity and secondary effects, this role could

be related to both innate immunity against infection and the appropriate safe handling and disposal of damaged autologous cells and lipids. However, testing these possibilities experimentally is extremely challenging because despite considerable structural similarity, the CRP molecules in different species show major differences in normal concentration, in behaviour as acute phase proteins, fine ligand binding specificity and secondary effects after ligand binding. Extrapolation from animals to man is therefore difficult and requires that putatively conserved functions must be compatible with these major differences. Furthermore, effects of heterologous CRP *in vivo*, following administration or transgenic expression of human CRP in animals, are not necessarily genuine effects of human CRP in humans. Finally, any function proposed for human CRP must be consistent with the remarkable speed and dynamic range of its plasma concentration, which can rise by over 1000-fold in 24–48 h after a strong acute stimulus such as sepsis or acute MI, and can fall with a half time of about 24 h when the stimulus is removed. These dramatic changes are not associated with any local or systemic vascular or inflammatory effects in patients other than those related to the pathology or treatment, which respectively triggered or alleviated the acute phase response. The fact that injection of even huge doses of isolated pure authentic human CRP into healthy animals has no adverse, inflammatory or tissue damaging effects is consistent with these long well-established clinical observations.

Synthesis of CRP

The circulating CRP is produced by hepatocytes. Reports of extra-hepatic CRP synthesis by some other cell types have not always been reproducible [45, 46], and although possibly relevant to potential local effects of CRP, the contribution to plasma concentrations can only be minute with respect to that from the liver. Hepatic CRP synthesis is under exquisitely sensitive transcriptional regulation via pro-inflammatory cytokines including interleukin (IL)-6, IL-1 and tumour necrosis factor-α so that almost any form of tissue injury, infection or inflammation, and indeed also most forms of adverse nonphysiological 'stress', are associated with increased circulating CRP values [47]. CRP is apparently cleared from the plasma and catabolized exclusively by hepatocytes [48] and the plasma half-life in humans of about 19 h is the same in all individuals regardless of the presence of disease or the circulating concentration of CRP. The sole determinant of the plasma concentration is therefore the synthesis rate [49].

It is critically important for any interpretation of the significance and utility of CRP values to distinguish between modest increments in the very low normal baseline values of CRP and the massive rises of several orders of magnitude which occur very rapidly following major stimuli, such as sepsis, acute trauma or tissue necrosis, and which may persist for months or years in individuals with chronic active inflammatory and tissue damaging diseases, such as chronic infections, rheumatoid arthritis, Crohn's disease, lymphoma and many others. Furthermore the almost completely nonspecific nature of the CRP response means that it is neither possible nor medically appropriate to interpret CRP values in an individual in the absence of full clinical information on that person, including history, physical examination and full results of all available investigations [6].

Concentration of circulating CRP in healthy normal individuals and in general populations

In healthy normal subjects CRP is a trace plasma protein. In the first substantial study of such individuals, the median value in 468 British Caucasian volunteer blood donors aged 18–63 years was 0.8 mg L^{-1}, the ninetieth percentile of the distribution was 3 mg L^{-1} and the ninety-ninth percentile was 10 mg L^{-1} [50]. Similar distributions have been reported from the general adult population of USA [51] and Northern Europe [52] although values are slightly higher overall in these unselected populations, more extreme high values are present and concentrations tend to increase slightly with age. The markedly skewed distribution of CRP values is typically normalized by log transformation. In USA 50% of individuals have baseline CRP greater than 2 mg L^{-1} and 33% are between 3 and 10 mg L^{-1} [51]. Compared with people of

Northern European ancestry, higher baseline CRP values of CRP have been observed in people of Afro-Caribbean origin residing in North America and lower values in people of Japanese or Chinese ancestry [53, 54]. In the indigenous Japanese, baseline CRP values are about one-tenth of those seen in individuals of European ancestry [55] but the reasons for these differences are uncertain. In healthy individuals followed serially with monthly samples, most CRP values cluster at a level typical for that individual and generally within the range of 0.1–3 mg L^{-1}, but with occasional spikes unrelated to any obvious clinical pathology. About 50% of the individual variance in baseline CRP concentration is genetic and largely attributable to noncoding polymorphisms in the CRP gene (see below). The other major determinant, independent of genetic factors, is the level of adiposity, especially central abdominal obesity [56], reflecting production of pro-inflammatory cytokines by macrophages associated with such adipose tissue and by the adipocytes themselves [57] and possibly even some CRP production by adipocytes [58]. The year-to-year within-person variability of CRP in populations of initially apparent healthy adults is comparable with those seen for levels of blood pressure or serum cholesterol (Fig. 2a) and some other circulating markers of inflammation (e.g. fibrinogen, leucocyte count) [59–61]. For example, the intraclass correlation coefficient for CRP levels (calculated from two different measurements obtained from the same individuals a few years apart) is about 0.6 [62].

The term 'high-sensitivity' or 'highly sensitive' CRP, abbreviated as hs-CRP, has been widely adopted in recent literature. It refers to measurement of CRP in serum or plasma samples using immunoassay methods with sufficient sensitivity to quantify CRP throughout its normal range in contrast to older less sensitive commercial assays which had detection limits in the range 2–10 mg L^{-1} and were suitable for measurement of acute phase responses of CRP rather than baseline values. It is very important to recognize that the analyte designated as hs-CRP is just CRP itself, not anything new or different and in particular is not a novel analyte with any special relationship to cardiovascular disease. hs-CRP is the same exquisitely

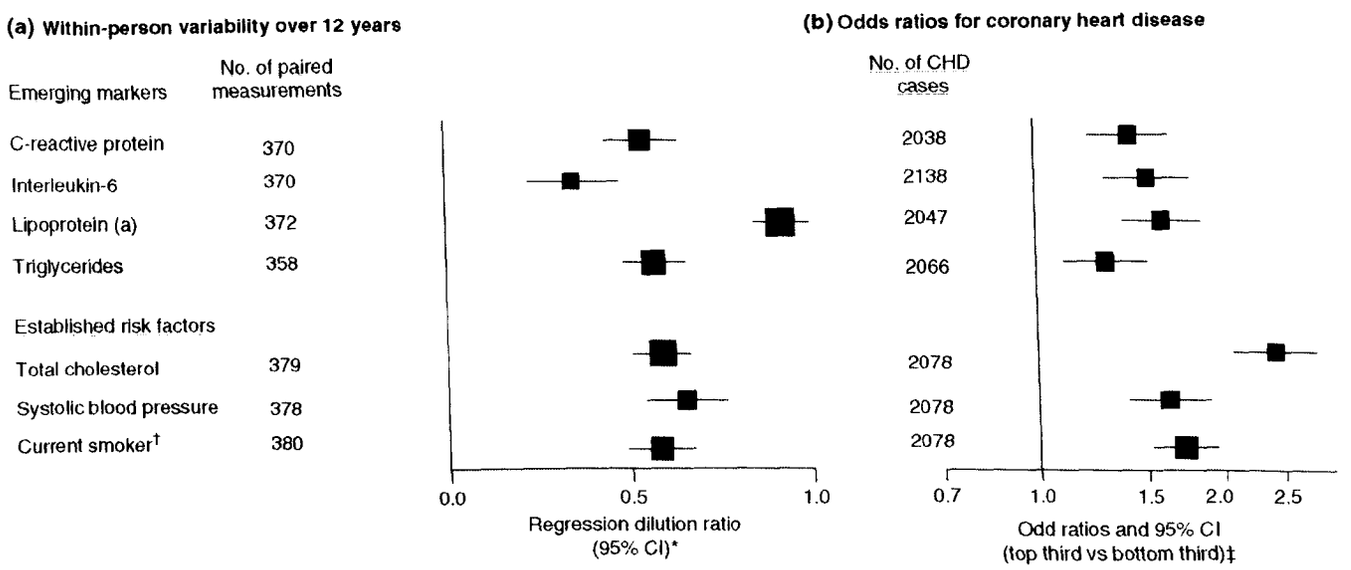


Fig. 2 Direct comparisons of C-reactive protein, several established cardiovascular risk factors and emerging markers in relation to: (a) within-person variability over 12 years (expressed as the regression dilution ratio*) and (b) odds ratios for coronary heart disease (adjusted for established risk factors‡). *Regression dilution ratio calculated using Rosner's multivariate regression method, adjusted for baseline age, sex, smoking history, diabetes history, total cholesterol, log triglycerides, systolic blood pressure, and body mass index. †Compared with not current smokers. ‡Odds ratios (top third vs. bottom third) are adjusted for cohort, age, sex, period of recruitment, smoking status, history of diabetes, total cholesterol, log triglycerides, systolic blood pressure and body mass index. Data source: data have been derived from the 19 000-participants Reykjavik cohort study and collated from the following publications: [62, 102, 104, 105].

sensitive and entirely nonspecific systemic marker of infection, inflammation, tissue damage and/or almost any form of adverse nonphysiological stress as the CRP, which has been extensively studied and used clinically for over 75 years. Precisely the same constraints apply to use of its values, notably that these cannot be interpreted in the absence of full clinical information on the individual patient at the time the sample was taken [6]. An individual with an apparent baseline CRP value of, say 7 mg L^{-1}, should not just be told that they have an increased risk of cardiovascular disease and be advised to alter their lifestyle accordingly, when the increased CRP concentration might reflect any of a very wide range of other serious diseases which have not yet declared themselves, for example, Hodgkin's disease or renal carcinoma or any other condition which is often occult for some time before clinical presentation [6].

Long-term prospective epidemiological studies of CRP and CHD

In 1996, the first population-based prospective study reported on associations of 'baseline' CRP levels with CHD outcomes [63]. By 2008, over 40 further such studies had reported [62–101], typically involving middle-aged populations of European ancestry. A literature-based meta-analysis was published in 2004 involving 22 population-based prospective studies with a total of 7068 incident CHD cases with a mean follow-up of 12 years. The odds ratio for CHD adjusted for several established risk factors was 1.6 (95% CI: 1.5–1.7) in a comparison of people with baseline CRP levels in the top third compared with those in the bottom third of the population distribution (corresponding to values of about 2.4 mg L^{-1} vs. 1 mg L^{-1}) [62]. This odds ratio is similar in magnitude to those reported for some other nonspecific circulating markers of inflammation (e.g. fibrinogen [60], IL-6 [102], leucocyte count [103]) and some other emerging biochemical risk markers (e.g. triglycerides [104], lipoprotein (a) [105]) (Fig. 2b). By contrast, an odds ratio of 2.0 (1.6–2.5) for CHD was reported in a previous review of the 11 initial prospective studies of CRP [71] and even more extreme odds ratios have been reported in some earlier individual studies. This attenuation in CHD risk associated with CRP levels during the course of its study is likely, at least in part, to reflect the decrease in scope for random error and for selective publication in larger and later hypothesis-testing studies. It is also possible, however, that there may be important differences across studies in features related to population characteristics, assay procedures and statistical methods used. To help address such uncertainties more reliably, the Emerging Risk Factors Collaboration (ERFC) has collated and harmonized individual data from participants in 48 population-based prospective studies of CRP involving a total of about 120 000 participants and about 9500 CHD outcomes [9]. In contrast with literature-based reviews (which have access to only published or limited aggregated data), detailed pooled analyses involving primary data should provide: (i) precise analyses of CRP-CHD associations under a range of various circumstances (such as at different ages, in women and men and at different levels of risk factors), (ii) reliable characterization of the shape of any dose–response relationship curves, (iii) consistent approaches to adjust for possible confounding factors, (iv) correction for within-person variability in levels of CRP and of possible confounding factors and (v) detailed investigation of potential sources of heterogeneity.

Even analyses in the ERFC may not, however, distinguish reliably whether CRP is mainly a causal risk factor in CHD or mainly a marker of established cardiovascular risk factors to which it is correlated, or mainly a marker of subclinical disease or some combination of these possibilities. The potential scope for confounding in studies of CRP is suggested by observations that individuals with higher than average levels of CRP tend to have higher values of blood pressure, pulse-wave velocity, central arterial blood pressure, proatherogenic lipid fractions (LDL, apolipoprotein-B), higher body mass index and levels of abdominal obesity and a greater prevalence of diabetes and metabolic syndrome and smoking and lower socio-economic status and birth weight [106–109]. CRP is inversely associated with a range of potentially protective risk factors in CHD, such as physical activity, HDL cholesterol, apolipoprotein-AI and

consumption of fruits and vegetables [100, 110]. CRP levels are also associated with levels of several other emerging biochemical risk markers of uncertain relevance to CHD, such as fibrinogen, adiponectin and triglycerides [88, 100]. The third National Health and Nutrition Examination Survey reported that in up to about two-thirds of Americans, a CRP level greater than 3 mg L^{-1} could be attributable to elevations in at least one established cardiovascular risk factor such as those named above [111]. The many potentially relevant correlates of CRP make it difficult, therefore, to determine to what extent the observed associations of CRP with CHD risk are independent from established and emerging coronary risk markers. Statistical adjustment for confounding factors is potentially limited because not all relevant confounders have been (or can be) measured in a study. Moreover, even measured confounders may be incompletely adjusted for because allowances are typically not made for within-person variability in levels of confounders (e.g. blood pressure, serum lipid concentrations). Alternatively, statistical overadjustment – the correction for markers in any causal pathway between CRP levels and CHD risk – could in principle obscure a potentially important actiological relationship. In practice, however, it is difficult to judge the likelihood of over adjustment given current uncertainties about whether CRP exerts any vascular effects, as described in the next section.

Experimental studies and animal models of CRP and atherogenesis

The possibility that CRP might have proatherogenic actions was first suggested in 1982 by the discovery of its specific binding to LDL and VLDL [10, 11] and was supported by its detection in atherosclerotic plaque [12]. Even before this it was well established that CRP can activate the classical complement pathway [40, 41] and is therefore potentially pro-inflammatory. However, compelling evidence for a role of CRP in atherosclerosis has not emerged despite many reports describing a very large range of pro-inflammatory, pro-thrombotic, vasoactive and thus potentially pro-atherogenic and pro-atherothrombotic effects of CRP preparations on various cell types *in vitro*. Almost none of these reports, most of which use commercial preparations of CRP, have described any characterization of the integrity or purity of the protein, and very few included any controls. None of the early studies removed either the toxic sodium azide, present in all commercial preparations as a bacteriostatic, or considered the inevitable presence of bacterial endotoxin (lipopolysaccharide) in CRP produced by recombinant *Escherchia coli*. Careful studies with authentic pure CRP isolated from human material and with recombinant CRP produced by mammalian cells, and thus free of such biologically active contamination, do not confirm claims for wide-ranging direct cellular actions of CRP [13, 112–119].

There is now an extensive and controversial literature which extends to *in vivo* studies involving either injection of CRP in different species or transgenic expression of CRP in mice. Injection of even enormous doses (40 mg kg^{-1}) of purified authentic human CRP into mice and rats neither elicited inflammation nor produced any clinical ill effects (including changes in blood pressure) [120]. This important observation is consistent with the fact that human CRP concentrations can cover a 10 000-fold range from less than 50 $\mu g\ L^{-1}$ to more than 500 mg L^{-1} in the acute phase response and this is very unlikely to be compatible with significant effects of CRP on vascular tone, activation of inflammatory cells, triggering of coagulation or any of the other purported signalling functions lately ascribed to CRP on the basis of *in vitro* studies with commercial CRP preparations. Although there is one report of enhanced atherosclerosis in apolipoprotein E knockout mice expressing transgenic human CRP [121] other and larger studies show no such effect [122–124] nor any pro-inflammatory or pro-thrombotic action even in aged atherosclerotic animals [125]. In the more humanized model of atherosclerosis in LDL receptor knockout mice expressing apoB100, transgenic human CRP was atheroprotective [126]. Indeed, the presence of CRP in human atheroma is no less likely *a priori* to be atheroprotective than it is to be atherogenic [36]. Extrapolation to humans from *in vivo* experimental animal studies of CRP function is not possible because, despite the considerable phylogenetic conservation of CRP structure, there are very substantial differences

between CRPs of different species with respect to fine details of ligand recognition, secondary effects of ligand binding including complement activation, normal concentrations and behaviour as acute phase reactants [44, 127, 128]. Introduction of human CRP into animals, wherein the protein is interacting with xenogeneic molecules, cells, physiological and pathophysiological processes, cannot be assumed to be a robust test for functions of human CRP in humans. Despite the claims to the contrary, the only rigorous conclusion at present is that neither the physiological nor the pathophysiological functions of human CRP are yet known.

To address these questions, work is in progress in the Pepys laboratory in three areas:

1 C-reactive protein knockout mice have been created using C57BL/6 embryonic stem cells in order to have genetically homogenous pure line C57BL/6 mice deficient in CRP. These mice are fertile and so far have an apparently normal life span. Characterization of their phenotype in relation to atherogenesis and other challenges will elucidate the functions of mouse CRP and may be informative about the human protein, even though the two have radically different concentrations and behaviour as acute phase proteins. Others have also lately reported production of a mouse CRP knockout but not its genetic background [129].

2 Clinical 'good manufacturing practice' grade human CRP has been isolated and purified from human plasma pooled from over 6000 comprehensively validated normal healthy donors. This is a considerable undertaking in view of the normal plasma concentration of CRP of less than 1 mg L^{-1} and will be only the second reported human CRP preparation made from healthy individuals [49] rather than from diseased individuals mounting an acute phase response or produced in recombinant bacteria or cells. It is thus entirely physiological in addition to being highly purified, completely free of exposure to bacterial endotoxin or other contamination, sterile, nonpyrogenic, fully structurally and functionally intact and approved by the UK regulatory authorities for administration to human subjects. This preparation will be used to test for effects of human CRP itself, on cells *in vitro*, on animals *in vivo* and finally in human volunteers, unrelated to any possible contamination or alteration in the native protein. This unique reagent will be made available for replication studies by others and should make a major contribution to resolution of the present controversies.

3 The first specific CRP inhibitor drug, 1,6 (bis)-phosphocholine hexane [120] and related compounds are being developed for clinical testing. Although they are not orally bioavailable, have a very short half-life, and will thus be suitable only for short-term acute use, they should be informative about the role of CRP following acute MI or in other acute clinical conditions characterized by tissue damage and a major acute phase CRP response. Development of new drugs able to specifically block CRP function for sustained periods will be required to establish robustly the role of human CRP in health and disease, as described in the final section of this review.

Randomized trials of statins in relation to CRP and CHD

The Justification for the Use of Statins in Primary Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) randomized about 18 000 participants without evidence of cardiovascular disease, who had relatively high baseline levels of CRP (2 mg L^{-1} or greater) and relatively low levels of LDL cholesterol (i.e. <3.36 mmol L^{-1}), to receive either rosuvastatin 20 mg once daily or placebo [130, 131]. Based on a recommendation from an independent data and safety monitoring board, the trial was stopped early in February 2008 because there was unequivocal evidence of a reduction in cardiovascular morbidity and mortality in patients treated with rosuvastatin compared with placebo (http://www.cardiosource.com/rapidnews summaries/summary.asp?SumID=318). The JUPITER findings extend the known benefits of statins beyond those previously demonstrated in a meta-analysis of individual results on about 90 000 patients that showed cardioprotective effects across a wide range of predominantly intermediate-to-high risk individuals [132].

The findings of the JUPITER trial, which have not yet been published in detail, have raised several questions. First, is the measurement of CRP and LDL cholesterol an optimum approach to target statin medication in individuals at low-to-intermediate risk of cardiovascular disease? This topic is considered in a subsequent section of this review in relation to measurement of CRP for the purposes of risk stratification. Second, might statins confer at least part of their benefits through lowering CRP (or related anti-inflammatory activities)? This question arises because statins are known to lower baseline CRP concentrations by about 20% [133]. Unfortunately, this hypothesis is difficult to confirm or refute reliably in randomized trials of statins because these agents so potently lower the plasma concentration of LDL cholesterol, a major known causative factor in CHD. A third question suggested by the JUPITER findings is: might the cardioprotective benefits of statins be proportionally greater in people with higher baseline CRP levels than those with lower CRP levels (i.e. effect-modification of statin effectiveness), given the possibility that statins may have relevant anti-inflammatory activities? [133–136]. The JUPITER trial itself may not be the optimum study to test this hypothesis because, as noted above, it was restricted to participants with CRP levels 2 mg L^{-1} or greater. This question can, however, be readily addressed by measurement of CRP levels in the stored serum samples of several existing large statin trials that recruited patients without reference to baseline CRP levels. For example, such analyses of CRP are expected to emerge from the Heart Protection Study, which randomized about 20 000 high-risk patients to 40 mg simvastatin or placebo and has recorded several thousand cardiovascular outcomes during follow-up [137, 138].

Genetic studies

An emerging approach to evaluate any causal relevance of subtle, prolonged increases in CRP levels to CHD involves genetic epidemiology [139–141]. 'Mendelian randomization' experiments attempt to minimize confounding and avoid reverse association bias by measurement of common polymorphisms or haplotypes in regulatory regions of the CRP gene that have been reliably associated with differences in circulating CRP concentration (but not with any known change in CRP function). According to Mendel's second law [142], the inheritance of genetic variants should be subject to the random assortment of maternal and paternal alleles at the time of gamete formation. So, if CRP levels actually increase the risk of CHD, then carriage of alleles (or haplotypes) that expose individuals to a long-term elevation of CRP should confer an increased risk of CHD outcomes in proportion to the difference in CRP levels attributable to the allele. Because of the randomized allocation of alleles from parents to offspring, potential confounders should be distributed amongst the genotypic classes, and any bias because of reverse causation should be avoided because genotypes are fixed at conception and not prone to modification by the onset of disease [143, 144] (Fig. 3). This approach has been applied to the study of other emerging risk markers, including plasma levels of fibrinogen and homocysteine [139, 141]. The potential limitations of Mendelian randomization analyses include the need for very large sample sizes because most genotypes have only

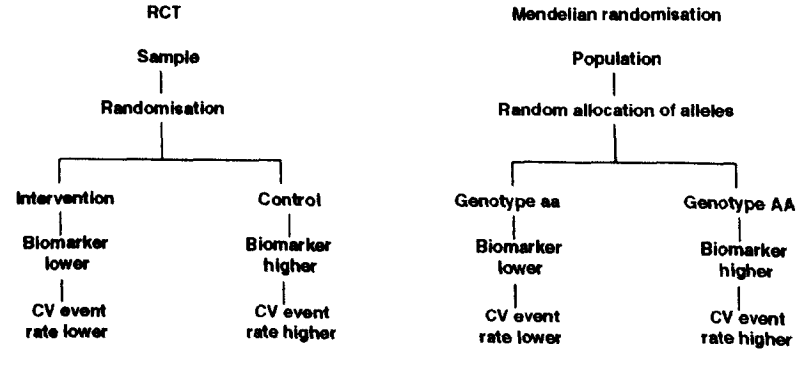


Fig. 3 Conceptual parallels between a randomized controlled trial and a hypothetical Mendelian randomization experiment to judge the causal relevance of a biomarker associated with coronary heart disease (CHD) risk.

modest effects on concentrations of biochemical markers and, at least in principle, the scope for residual confounding by unrecognized pleiotropic effects of genotypes and by developmental adaptation ('canalization') [140].

C-reactive protein concentration is a heritable trait [56, 145–148] and several single nucleotide polymorphisms (SNPs) within the CRP gene have now been identified (Fig. 4) [149]. Twelve common SNPs with minor allele frequency of >5% are present in a 6 kb region encompassing the gene in people of European ancestry and there are around 22 SNPs in subjects of African descent. In Europeans, the 12 SNPs are distributed amongst just four common haplotypes that account for approximately 95% of all chromosomes. A number of studies have reported consistent associations with CRP concentration of individual SNPs or subsets of SNPs, some of which have been typed for their ability to capture haplotype diversity (tagging SNPs) [108, 150–152]. For example, a combined analysis of 25 reports involving a total of around 31 000 participants of European descent has yielded a difference in CRP levels of 0.12–0.96 mg L^{-1} per allele, depending on the SNP studied, a difference that equates to between 0.3 and 0.8 standard deviations of the CRP distribution. A meta-analysis of genetic studies involving a new Bayesian method (enabling integration of information across studies that had typed a partially overlapping set of CRP SNPs) has provided evidence of four functional SNPs at the CRP locus that influence its circulating concentration [149]. Several initial studies have reported that although these CRP genotypes (and haplotypes) are clearly associated with CRP levels, they are not materially associated with any of a large panel of established or emerging cardiovascular risk markers, suggesting that these variants may well be unbiased proxies for CRP and suitable for use in Mendelian randomization analyses (Table 1) [108, 151, 153–155]. Although such studies (which have collectively involved a few thousand CHD cases) have generally reported lack of strong association of these genetic variants with CHD risk, they have not been sufficiently powerful to realistically confirm or refute any moderate effect of CRP levels on CHD such as a 10–20% increase in risk per standard deviation increase in levels. As sample size calculations suggest that at least 15 000 CHD cases and a similar number of controls may well be needed to provide

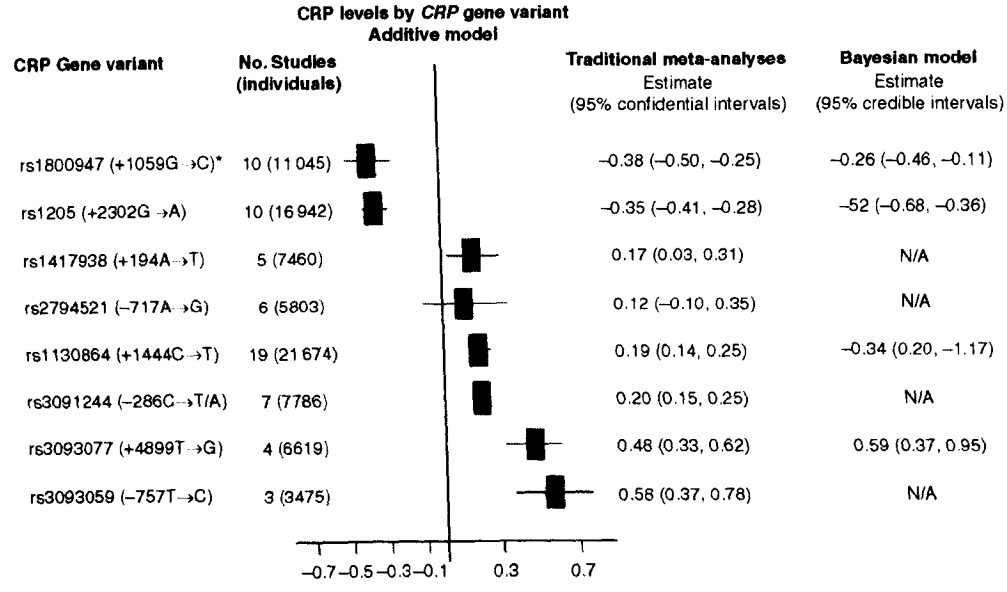


Fig. 4 The effect of eight common polymorphisms in the human *CRP* gene on CRP concentration (mg L^{-1}). Information was obtained from published and unpublished studies, and analyses were based on traditional meta-analysis using an additive model as well as Bayesian meta-analysis. The figure is reproduced from Fig. 4 in Verzilli *et al.* Am J Hum Genet, 2008; 82: 859–872, p 865. N/A refers to SNPs excluded from the Bayesian model.

Table 1 Cardiovascular risk factor distribution according to the CRP/+1444C>T polymorphism

Variable	Subjects (studies)	Weighted mean difference [TT – C-carriers] (95%CI)	*P*-value
Age (years)	6358 (7)	0.17 (−0.19, 0.52)	0.35
Systolic blood pressure (mmHg)	6356 (7)	−0.31 (−1.78, 1.16)	0.68
Diastolic blood pressure (mmHg)	6356 (7)	−0.70 (−1.56, 0.16)	0.11
Body mass index (kg m^{-2})	6359 (7)	0.34 (0.06, 0.62)	0.02
Total cholesterol (mmol L^{-1})	6115 (6)	−0.01 (−0.08, 0.06)	0.87
HDL cholesterol (mmol L^{-1})	4714 (5)	−0.01 (−0.04, 0.01)	0.32
Triglyceride (mmol L^{-1})	6041 (5)	−0.03 (−0.11, 0.06)	0.51
Fibrinogen (g L^{-1})	5656 (4)	0.03 (−0.02, 0.08)	0.20
Glucose (mmol L^{-1})	1913 (3)	−0.01 (−0.13, 0.12)	0.90
Alcohol intake (U per week)	4549 (4)	0.01 (−1.38, 1.39)	0.99
Current smoking[a]	6132 (6)	1.00 (0.84, 1.20)	0.98
C-reactive protein (mg L^{-1})	4659 (6)	0.68 (0.31, 1.10)	0.0001

CRP, C-reactive protein.
[a]For this variable instead of the WMD, the value reported is the weighted odds ratio.
Data source: Table taken from Casas *et al.*, 2006 [154].
Comparisons are made between TT homozygotes and C-allele carriers.

such informative analyses, the CRP Coronary Disease Genetics Collaboration (CCGC) has been established [156]. This collaboration currently involves over 30 epidemiological studies with a total of about 37 000 cases of MI or coronary death and about 120 000 controls. The objective is to conduct pooled analyses of individual participant data from these studies to examine the association between relevant CRP genetic variants and CHD risk under various circumstances, thereby helping to assess causality.

CRP and risk stratification in acute coronary syndromes

Several scores have been proposed to assess the short-term risk for recurrent events and adverse prognosis in patients presenting with acute coronary syndromes, such as the thrombolysis in myocardial infarction (TIMI) risk score, which is based on clinical and electrocardiogram variables [157]. As noted above, several studies have reported on associations between CRP levels and recurrent cardiovascular events in patients presenting with acute coronary syndrome [7, 8, 158, 159]. In 2003, the American Heart Association/Centres for Disease Control (AHA/CDC) recommended CRP measurement as an additional marker of prognosis in patients with acute coronary syndromes [16] on top of characteristics in the TIMI risk score [160, 161]. Further studies have proposed the addition of troponin T and N-terminal pro-brain natriuretic peptide to the TIMI risk score plus CRP [162]. Interpretation of this evidence would be enhanced by pooled analyses of individual data from each of the available studies of acute coronary syndromes (which is currently outside the scope of the ERFC) and by consideration of how addition of CRP (and other markers) to risk alogrithms might lead to any change in the management of such patients.

CRP and the evaluation of long-term risk of coronary disease

Several risk algorithms have been proposed to help stratify risk of cardiovascular disease in general Western populations, such as Framingham, PROCAM, SCORE, Reynolds and QRISK [163–167]. Each of these algorithms involves a core set of the same established risk factors (e.g. smoking, blood pressure, total cholesterol) but differ in their inclusion of various other characteristics, such as HDL-C (in Framingham), triglycerides (in PROCAM only), CRP (in Reynolds only) and body mass index or

markers of socio-economic status (in QRISK only). As CRP levels display relatively limited biological variability (outside acute phase fluctuations) and can be accurately assessed using assays with internationally agreed standards, their measurement for risk prediction has been regarded as particularly practicable by some workers (although such advantages also apply to the measurement of some other nonspecific inflammatory markers that may be measured more cost effectively than CRP such as the leucocyte count and serum albumin). In 2003, a report of the AHA/CDC recommended that 'CRP may be used at the discretion of the physician as part of a global coronary risk assessment in adults without known cardiovascular disease', concluding that a CRP value above a cut point of 3 mg L^{-1} was indicative of high risk [16]. The report acknowledged, however, the need for further research to address several uncertainties in the evidence. By contrast, European authorities have not recommended measurement of CRP to enhance cardiovascular risk assessment [17].

Part of the uncertainty stems from the fact that although many published prospective studies have commented on the potential value of CRP in risk prediction, they have often reported only measures of association (e.g. odds ratios, hazard ratios), which do not directly address the issue of the utility of a marker in prediction or stratification. Furthermore, even studies that have involved statistics relevant to the assessment of risk prediction have involved different metrics, including measures of discrimination (e.g. the C index [168, 169] and the measure D [170], with the former related to the area under the receiver operating characteristic curve) and reclassification methods that aim to summarize the potential of a marker to re-assign individuals into more appropriate risk groups [171]. Each of these approaches may impart somewhat different information. As recommended by the US National Heart Lung and Blood Institute 2006 workshop report on CRP, further work is needed to compare and contrast the strengths and limitations of each of these approaches [172] to enable their use in the multiple study situation (as in the ERFC database) and to incorporate health economic analyses to help judge the value of any such measurements (http://www.nhlbi.nih.gov/meetings/workshops/crp/report.htm).

CRP in acute ischaemic tissue damage

There is always a substantial CRP acute phase response following acute MI [173, 174]. An association between sustained high values of CRP following acute MI and adverse outcomes was first reported in 1982 [175] and subsequent large studies have shown that increased peak and postinfarct CRP concentrations are significantly associated with increased incidence of cardiac complications including heart failure and cardiac death, apparently independently of other predictors [14, 159, 176–181]. Acute MI is invariably associated with acute inflammation around the lesion [1] and complement makes a substantial contribution to the size of experimental acute myocardial infarcts produced by coronary artery ligation in mammals [182], including old world primates [183]. All human acute MI lesions contain CRP and activated complement co-deposited in and around the infarct [2, 184]. The rat provides an excellent model for testing the possible pathogenicity of human CRP in this situation because although rats have abundant CRP, their protein does not activate rat complement whereas human CRP potently activates both rat and human complement [185]. Effects of human CRP injected into rats can therefore reflect what human CRP may do in humans. Indeed, administration of authentic pure human CRP to rats after they have undergone coronary artery ligation or ischaemia-reperfusion injury causes significant worsening of cardiac function and increased mortality, which reflect increased infarct size [18]. The effect is specific for human CRP and absolutely complement-dependent, and human CRP is co-deposited with rat complement in the infarct [18]. The same pathogenic effect of human CRP is seen in the rat model of cerebral infarction produced by middle cerebral artery occlusion [186]. The adverse effect of human CRP in the acute MI situation is completely abrogated by administration of the novel small molecule specific CRP inhibitor drug, 1, 6 (bis)-phosphocholine hexane [120] (Fig. 5). Importantly, this drug has

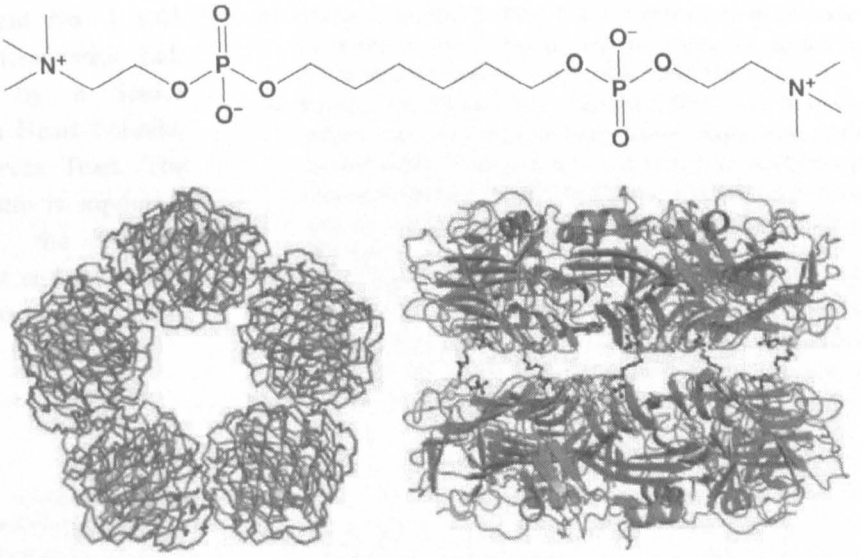


Fig. 5 Chemical formula of bis(phosphocholine)-hexane (above) and 3D crystal structure of the drug-C-reactive protein (CRP) complex (below); face on view (left) and side view (right) of two CRP molecules cross-linked by five drug molecules [120]. X-ray crystallographic structures provided by Dr. Simon Kolstoe.

no effect at all in rats subjected to coronary artery ligation which do not receive human CRP and is thus not cardioprotective *per se* but only by virtue of its inhibition of CRP [120].

The likely mechanism of CRP pathogenicity is therefore binding of abundant CRP to the ligands exposed in dead and damaged cells, triggering substantial complement activation with release of chemotactic factors and opsonization of cells in and around the lesion, leading to enhanced infiltration by inflammatory cells and consequent bystander damage. The terminal complement sequence may also directly kill cells which would otherwise survive and the end result is death of more myocardial tissue than would be killed by ischaemia alone. Development is now in progress towards clinical testing of CRP inhibition in patients with acute MI.

Conclusion

Until specific CRP inhibitor drugs suitable for long-term administration become available, the question of whether low level chronically increased CRP concentration is a risk factor for CHD, may be addressed by pooled analyses of data from prospective studies of circulating CRP levels and from studies of CRP genetic variants. Uncertainties related to the separate issue of risk prediction in the general population can also be addressed in the near future by collaborative approaches involving large combined datasets. Meanwhile, it is very important to distinguish clearly between (i) the potential pathogenicity of high circulating CRP concentrations in individuals with a substantial volume of damaged tissue in whom CRP binding to exposed ligands and consequent major complement activation can exacerbate injury and (ii) the possible actions of baseline or rather modestly increased CRP concentrations which together with complement may assist in clearance and resolution of minor tissue injury and thus be beneficial. The experimental evidence from animals indicates that human CRP can definitely be pathogenic after ischaemic infarction, but any relevance to the slow process of atherogenesis remains unproved.

Conflict of interest statement

Pentraxin Therapeutics Ltd is a UCL spin out company of which Professor Pepys is the founding director and which owns his proprietary knowledge and patents on CRP as a therapeutic target and bis-phosphocholine compounds as CRP inhibitor drugs.

Acknowledgements

The work of Professor Pepys reported here is supported by the Medical Research Council (Programme Grant PG 7900510) and the National Heart Lung and

Blood Institute of the US NIH (Grant No. 1 R01 HL078578-01) and by Pentraxin Therapeutics Ltd. Professor Hingorani is supported by a Senior Research Fellowship from the British Heart Foundation (FS/05/125) and by the Rosetrees Trust. The work of Professor Danesh reported here is supported by the British Heart Foundation, the Medical Research Council, the Wellcome Trust and the BUPA Foundation and by unrestricted educational grants from GlaxoSmithKline.

References

1 Entman ML, Michael L, Rossen RD *et al.* Inflammation in the course of early myocardial ischemia. *FASEB J* 1991; 5: 2529–37.
2 Lagrand WK, Niessen HWM, Wolbink G-J *et al.* C-reactive protein colocalizes with complement in human hearts during acute myocardial infarction. *Circulation* 1997; 95: 97–103.
3 Libby P, Ridker PM, Maseri A. Inflammation and atherosclerosis. *Circulation* 2002; 105: 1135–43.
4 Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004; 351: 2611–8.
5 Lowe GDO, Pepys MB. C-reactive protein and cardiovascular disease: weighing the evidence. *Curr Atheroscler Rep* 2006; 8: 421–8.
6 Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003; 111: 1805–12.
7 Berk BC, Weintraub WS, Alexander RW. Elevation of C-reactive protein in "active" coronary artery disease. *Am J Cardiol* 1990; 65: 168–72.
8 Liuzzo G, Biasucci LM, Gallimore JR *et al.* The prognostic value of C-reactive protein and serum amyloid A protein in severe unstable angina. *N Engl J Med* 1994; 331: 417–24.
9 Emerging Risk Factors Collaboration, Danesh J, Erqou S *et al.* The Emerging Risk Factors Collaboration: analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *Eur J Epidemiol* 2007; 22: 839–69.
10 de Beer FC, Soutar AK, Baltz ML, Trayner I, Feinstein A, Pepys MB. Low density and very low density lipoproteins are selectively bound by aggregated C-reactive protein. *J Exp Med* 1982; 156: 230–42.
11 Pepys MB, Rowe IF, Baltz ML. C-reactive protein: binding to lipids and lipoproteins. *Int Rev Exp Pathol* 1985; 27: 83–111.
12 Zhang YX, Cliff WJ, Schoefl GI, Higgins G. Coronary C-reactive protein distribution: its relation to development of atherosclerosis. *Atherosclerosis* 1999; 145: 375–9.
13 Pepys MB. CRP or not CRP? That is the question. *Arterioscler Thromb Vasc Biol* 2005; 25: 1091–4.
14 Scirica BM, Morrow DA. Is C-reactive protein an innocent bystander or proatherogenic culprit? The verdict is still out. *Circulation* 2006; 113: 2128–34.
15 Verma S, Devaraj S, Jialal I. Is C-reactive protein an innocent bystander or proatherogenic culprit? C-reactive protein promotes atherothrombosis. *Circulation* 2006; 113: 2135–50.
16 Pearson TA, Mensah GA, Alexander RW *et al.* Markers of inflammation and cardiovascular disease. Application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 2003; 107: 499–511.
17 De Backer G, Ambrosioni E, Borch-Johnsen K *et al.* European guidelines on cardiovascular disease prevention in clinical practice: third joint task force of European and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of eight societies and by invited experts). *Eur J Cardiovasc Prev Rehabil* 2003; 10: S1–10.
18 Griselli M, Herbert J, Hutchinson WL *et al.* C-reactive protein and complement are important mediators of tissue damage in acute myocardial infarction. *J Exp Med* 1999; 190: 1733–9.
19 Osmand AP, Friedenson B, Gewurz H, Painter RH, Hofmann T, Shelton E. Characterization of C-reactive protein and the complement subcomponent C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). *Proc Natl Acad Sci USA* 1977; 74: 739–43.
20 Pepys MB, Baltz ML. Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv Immunol* 1983; 34: 141–212.
21 Srinivasan N, Rufino SD, Pepys MB, Wood SP, Blundell TL. A superfamily of proteins with the lectin fold. *Chemtracts-Biochem Mol Biol* 1996; 6: 149–64.
22 Kinoshita CM, Ying S-C, Hugli TE *et al.* Elucidation of a protease-sensitive site involved in the binding of calcium to C-reactive protein. *Biochemistry* 1989; 28: 9840–8.
23 Tennent GA, Lovat LB, Pepys MB. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer's disease and systemic amyloidosis. *Proc Natl Acad Sci USA* 1995; 92: 4299–303.
24 Edwards AO, Ritter R III, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. *Science* 2005; 308: 421–4.
25 Hageman GS, Anderson DH, Johnson LV *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci USA* 2005; 102: 7227–32.
26 Haines JL, Hauser MA, Schmidt S *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005; 308: 419–21.
27 Giannakis E, Jokiranta TS, Male DA *et al.* A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *Eur J Immunol* 2003; 33: 962–9.
28 Butler JE, ed. *Immunochemistry of Solid-Phase Immunoassays*. Boca Raton, FL: CRC Press, Inc., 1991.
29 Tillett WS, Francis T. Serological reactions in pneumonia with a non-protein somatic fraction of pneumococcus. *J Exp Med* 1930; 52: 561–71.

30 Volanakis JE, Kaplan MH. Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc Soc Exp Biol Med* 1971; 136: 612–4.
31 Volanakis JE, Wirtz KWA. Interaction of C-reactive protein with artificial phosphatidylcholine bilayers. *Nature* 1979; 281: 155–7.
32 Gershov D, Kim S, Brot N, Elkon KB. C-reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an anti-inflammatory innate immune response: implications for systemic autoimmunity. *J Exp Med* 2000; 192: 1353–63.
33 zChang MK, Binder CJ, Torzewski M, Witztum JL. C-reactive protein binds to both oxidized LDL and apoptotic cells through recognition of a common ligand: phosphorylcholine of oxidized phospholipids. *Proc Natl Acad Sci USA* 2002; 99: 13043–8.
34 Hart SP, Alexander KM, MacCall SM, Dransfield I. C-reactive protein does not opsonize early apoptotic human neutrophils, but binds only membrane-permeable late apoptotic cells and has no effect on their phagocytosis by macrophages. *J Inflamm (Lond)* 2005; 2: 5.
35 Vigo C. Effect of C-reactive protein on platelet-activating factor-induced platelet aggregation and membrane stabilization. *J Biol Chem* 1985; 260: 3418–22.
36 Bhakdi S, Torzewski M, Paprotka K *et al.* Possible protective role for C-reactive protein in atherogenesis: complement activation by modified lipoproteins halts before detrimental terminal sequence. *Circulation* 2004; 109: 1870–6.
37 Rowe IF, Soutar AK, Trayner IM, Thompson GR, Pepys MB. Circulating human C-reactive protein binds very low density lipoproteins. *Clin Exp Immunol* 1984; 58: 237–44.
38 Du Clos TW. C-reactive protein reacts with the U1 small nuclear ribonucleoprotein. *J Immunol* 1989; 143: 2553–9.
39 Pepys MB, Booth SE, Tennent GA, Butler PJG, Williams DG. Binding of pentraxins to different nuclear structures: C-reactive protein binds to small nuclear ribonucleoprotein particles, serum amyloid P component binds to chromatin and nucleoli. *Clin Exp Immunol* 1994; 97: 152–7.
40 Kaplan MH, Volanakis JE. Interaction of C-reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C-reactive protein with pneumococcal C-polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J Immunol* 1974; 112: 2135–47.
41 Siegel J, Rent R, Gewurz H. Interactions of C-reactive protein with the complement system. I. Protamine-induced consumption of complement in acute phase sera. *J Exp Med* 1974; 140: 631–47.
42 Volanakis JE. Complement activation by C-reactive protein complexes. *Ann NY Acad Sci* 1982; 389: 235–50.
43 Pepys MB, Dash AC, Fletcher TC, Richardson N, Munn EA, Feinstein A. Analogues in other mammals and in fish of human plasma proteins C-reactive protein and amyloid P component. *Nature* 1978; 273: 168–70.
44 Baltz ML, de Beer FC, Feinstein A *et al.* Phylogenetic aspects of C-reactive protein and related proteins. *Ann NY Acad Sci* 1982; 389: 49–75.
45 Yasojima K, Schwab C, McGeer EG, McGeer PL. Generation of C-reactive protein and complement components in atherosclerotic plaques. *Am J Pathol* 2001; 158: 1039–51.
46 Sun H, Koike T, Ichikawa T *et al.* C-reactive protein in atherosclerotic lesions: its origin and pathophysiological significance. *Am J Pathol* 2005; 167: 1139–48.
47 Kushner I, Rzewnicki D, Samols D. What does minor elevation of C-reactive protein signify? *Am J Med* 2006; 119: 166.
48 Hutchinson WL, Noble GE, Hawkins PN, Pepys MB. The pentraxins, C-reactive protein and serum amyloid P component, are cleared and catabolized by hepatocytes *in vivo*. *J Clin Invest* 1994; 94: 1390–6.
49 Vigushin DM, Pepys MB, Hawkins PN. Metabolic and scintigraphic studies of radioiodinated human C-reactive protein in health and disease. *J Clin Invest* 1993; 91: 1351–7.
50 Shine B, de Beer FC, Pepys MB. Solid phase radioimmunoassays for human C-reactive protein. *Clin Chim Acta* 1981; 117: 13–23.
51 Woloshin S, Schwartz LM. Distribution of C-reactive protein values in the United States. *N Engl J Med* 2005; 352: 1611–3.
52 Hutchinson WL, Koenig W, Fröhlich M, Sund M, Lowe GDO, Pepys MB. Immunoradiometric assay of circulating C-reactive protein: age-related values in the adult general population. *Clin Chem* 2000; 46: 934–8.
53 Lakoski SG, Cushman M, Criqui M *et al.* Gender and C-reactive protein: data from the multiethnic study of atherosclerosis (MESA) cohort. *Am Heart J* 2006; 152: 593–8.
54 Kelley-Hedgepeth A, Lloyd-Jones DM, Colvin A *et al.* Ethnic differences in C-reactive protein concentrations. *Clin Chem* 2008; 54: 1027–37.
55 Yamada S, Gotoh T, Nakashima Y *et al.* Distribution of serum C-reactive protein and its association with atherosclerotic risk factors in a Japanese population: Jichi Medical School Cohort Study. *Am J Epidemiol* 2001; 153: 1183–90.
56 Greenfield JR, Samaras K, Jenkins AB *et al.* Obesity is an important determinant of baseline serum C-reactive protein concentration in monozygotic twins, independent of genetic influences. *Circulation* 2004; 109: 3022–8.
57 Yudkin JS, Stehouwer CDA, Emeis JJ, Coppack SW. C-reactive protein in healthy subjects: associations with obesity, insulin resistance, and endothelial dysfunction: a potential role for cytokines originating from adipose tissue? *Arterioscler Thromb Vasc Biol* 1999; 19: 972–8.
58 Calabro P, Chang DW, Willerson JT, Yeh ET. Release of C-reactive protein in response to inflammatory cytokines by human adipocytes: linking obesity to vascular inflammation. *J Am Coll Cardiol* 2005; 46: 1112–3.
59 Koenig W, Sund M, Fröhlich M, Löwel H, Hutchinson WL, Pepys MB. Refinement of the association of serum C-reactive protein concentration and coronary heart disease risk by correction for within-subject variation over time. The MONICA Augsburg Studies, 1984 and 1987. *Am J Epidemiol* 2003; 158: 357–64.
60 Fibrinogen Studies Collaboration, Danesh J, Lewington S *et al.* Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. *JAMA* 2005; 294: 1799–809.

61 Fibrinogen Studies Collaboration, Wood AM, White I, Thompson SG, Lewington S, Danesh J. Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27,247 adults in 15 prospective studies. *Int J Epidemiol* 2006; 35: 1570–8.

62 Danesh J, Wheeler JG, Hirschfield GM *et al.* C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med* 2004; 350: 1387–97.

63 Kuller LH, Tracy RP, Shaten J, Meilahn EN, for the MRFIT Research Group. Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study. *Am J Epidemiol* 1996; 144: 537–47.

64 Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N Engl J Med* 1997; 336: 973–9.

65 Tracy RP, Lemaitre RN, Psaty BM *et al.* Relationship of C-reactive protein to risk of cardiovascular disease in the elderly. Results from the Cardiovascular Health Study and the Rural Health Promotion Project. *Arterioscler Thromb Vasc Biol* 1997; 17: 1121–7.

66 Agewall S, Wikstrand J, Fagerberg B. Prothrombin fragment 1+2 is a risk factor for myocardial infarction in treated hypertensive men. *J Hypertens* 1998; 16: 537–41.

67 Ridker PM, Buring JE, Shih J, Matias M, Hennekens CH. Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women. *Circulation* 1998; 98: 731–3.

68 Harris TB, Ferrucci L, Tracy RP *et al.* Associations of elevated interleukin-6 and C-reactive protein levels with mortality in the elderly. *Am J Med* 1999; 106: 506–12.

69 Jager A, van Hinsbergh VWM, Kostense PJ *et al.* von Willebrand factor, C-reactive protein, and 5-year mortality in diabetic and nondiabetic subjects. *Arterioscler Thromb Vasc Biol* 1999; 19: 3071–8.

70 Koenig W, Sund M, Fröhlich M *et al.* C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (monitoring trends and determinants in cardiovascular disease) Augsburg Cohort Study, 1984 to 1992. *Circulation* 1999; 99: 237–42.

71 Danesh J, Whincup P, Walker M *et al.* Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ* 2000; 321: 199–204.

72 Mendall MA, Strachan DP, Butland BK *et al.* C-reactive protein: relation to total mortality, cardiovascular mortality and cardiovascular risk factors in men. *Eur Heart J* 2000; 21: 1584–90.

73 Packard CJ, O'Reilly DSJ, Caslake MJ *et al.* Lipoprotein-associated phospholipase A_2 as an independent predictor of coronary heart disease. West of Scotland Coronary Prevention Study Group. *N Engl J Med* 2000; 343: 1148–55.

74 Rouvanen M, Vitk-Kajander M, Palosuo T *et al.* Infections, inflammation, and the risk of coronary heart disease. *Circulation* 2000; 101: 252–7.

75 Strandberg TE, Tilvis RS. C-reactive protein, cardiovascular risk factors, and mortality in a prospective study in the elderly. *Arterioscler Thromb Vasc Biol* 2000; 20: 1057–60.

76 Lowe GDO, Yarnell JWG, Rumley A, Bainton D, Sweetnam PM. C-reactive protein, fibrin D-dimer, and incident ischemic heart disease in the Speedwell study: are inflammation and fibrin turnover linked in pathogenesis? *Arterioscler Thromb Vasc Biol* 2001; 21: 603–10.

77 Pirro M, Bergeron J, Dagenais GR *et al.* Age and duration of follow-up as modulators of the risk for ischemic heart disease associated with high plasma C-reactive protein levels in men. *Arch Intern Med* 2001; 161: 2474–80.

78 Folsom AR, Aleksic N, Catellier D, Juneja HS, Wu KK. C-reactive protein and incident coronary heart disease in the atherosclerosis risk in communities (ARIC) study. *Am Heart J* 2002; 144: 233–8.

79 Pradhan AD, Manson JE, Rossouw JE *et al.* Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative Observational Study. *JAMA* 2002; 288: 980–7.

80 Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med* 2002; 347: 1557–65.

81 Sakkinen P, Abbott RD, Curb JD, Rodriguez BL, Yano K, Tracy RP. C-reactive protein and myocardial infarction. *J Clin Epidemiol* 2002; 55: 445–51.

82 Luc G, Bard JM, Juhan-Vague I *et al.* C-reactive protein, interleukin-6, and fibrinogen as predictors of coronary heart disease: the PRIME Study. *Arterioscler Thromb Vasc Biol* 2003; 23: 1255–61.

83 van der Meer IM, de Maat MPM, Kiliaan AJ, van der Kuip DA, Hofman A, Witteman JC. The value of C-reactive protein in cardiovascular risk prediction: the Rotterdam Study. *Arch Intern Med* 2003; 163: 1323–8.

84 Koenig W, Löwel H, Baumert J, Meisinger C. C-reactive protein modulates risk prediction based on the Framingham Score: implications for future risk assessment: results from a large cohort study in southern Germany. *Circulation* 2004; 109: 1349–53.

85 Lowe GDO, Sweetnam PM, Yarnell JW *et al.* C-reactive protein, fibrin D-dimer, and risk of ischemic heart disease: the Caerphilly and Speedwell studies. *Arterioscler Thromb Vasc Biol* 2004; 24: 1957–62.

86 Pai JK, Pischon T, Ma J *et al.* Inflammatory markers and the risk of coronary heart disease in men and women. *N Engl J Med* 2004; 351: 2599–610.

87 Cushman M, Arnold AM, Psaty BM *et al.* C-reactive protein and the 10-year incidence of coronary heart disease in older men and women: the cardiovascular health study. *Circulation* 2005; 112: 25–31.

88 Lawlor DA, Davey Smith G, Rumley A, Lowe GDO, Ebrahim S. Associations of fibrinogen and C-reactive protein with prevalent and incident coronary heart disease are attenuated by adjustment for confounding factors. British Women's Heart and Health Study. *Thromb Haemost* 2005; 93: 955–63.

89 Ridker PM, Rifai N, Cook NR, Bradwin G, Buring JE. Non-HDL cholesterol, apolipoproteins A-I and B100, standard lipid measures, lipid ratios, and CRP as risk factors for cardiovascular disease in women. *JAMA* 2005; 294: 326–33.

90 St-Pierre AC, Cantin B, Bergeron J *et al.* Inflammatory markers and long-term risk of ischemic heart disease in men: a 13-year follow-up of the Quebec Cardiovascular Study. *Atherosclerosis* 2005; 182: 315–21.

91 Wilson PWF, Nam B-H, Pencina M, D'Agostino RB Sr, Benjamin EJ, O'Donnell CJ. C-reactive protein and risk of cardiovascular disease in men and women from the Framingham Heart Study. *Arch Intern Med* 2005; 165: 2473–8.

92 Boekholdt SM, Hack CE, Sandhu MS *et al.* C-reactive protein levels and coronary artery disease incidence and mortality in apparently healthy men and women: The EPIC-Norfolk prospective population study 1993–2003. *Atherosclerosis* 2006; 187: 415–22.

93 Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med* 2006; 145: 21–9.

94 Folsom AR, Chambless LE, Ballantyne CM *et al.* An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the atherosclerosis risk in communities study. *Arch Intern Med* 2006; 166: 1368–73.

95 Koenig W, Khuseyinova N, Baumert J *et al.* Increased concentrations of C-reactive protein and IL-6 but not IL-18 are independently associated with incident coronary events in middle-aged men and women: results from the MONICA/KORA Augsburg case-cohort study, 1984–2002. *Arterioscler Thromb Vasc Biol* 2006; 26: 2745–51.

96 May M, Lawlor DA, Brindle P, Patel R, Ebrahim S. Cardiovascular disease risk assessment in older women: can we improve on Framingham? British Women's Heart and Health prospective cohort study. *Heart* 2006; 92: 1396–401.

97 Stork S, Feelders RA, van den Beld AW *et al.* Prediction of mortality risk in the elderly. *Am J Med* 2006; 119: 519–25.

98 Cao JJ, Arnold AM, Manolio TA *et al.* Association of carotid artery intima-media thickness, plaques, and C-reactive protein with future cardiovascular disease and all-cause mortality: the Cardiovascular Health Study. *Circulation* 2007; 116: 32–8.

99 Sattar N, Murray HM, McConnachie A *et al.* C-reactive protein and prediction of coronary heart disease and global vascular events in the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER). *Circulation* 2007; 115: 981–9.

100 Woodward M, Rumley A, Welsh P, MacMahon S, Lowe G. A comparison of the associations between seven hemostatic or inflammatory variables and coronary heart disease. *J Thromb Haemost* 2007; 5: 1795–800.

101 Erbel R, Möhlenkamp S, Lehmann N *et al.* Sex related cardiovascular risk stratification based on quantification of atherosclerosis and inflammation. *Atherosclerosis* 2008; 197: 662–72.

102 Danesh J, Kaptoge S, Mann AG *et al.* Long-term interleukin-6 levels and subsequent risk of coronary heart disease: two new prospective studies and a systematic review. *PLoS Med* 2008; 5: e78.

103 Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease. *JAMA* 1998; 279: 1477–82.

104 Sarwar N, Danesh J, Eiriksdottir G *et al.* Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation* 2007; 115: 450–8.

105 Bennet A, Di Angelantonio E, Erqou S *et al.* Lipoprotein (a) levels and risk of future coronary heart disease: large-scale prospective data. *Arch Intern Med* 2008; 168: 598–608.

106 Sattar N, McConnachie A, O'Reilly D *et al.* Inverse association between birth weight and C-reactive protein concentrations in the MIDSPAN Family Study. *Arterioscler Thromb Vasc Biol* 2004; 24: 583–7.

107 Lakoski SG, Cushman M, Palmas W, Blumenthal R, D'Agostino RB Jr, Herrington DM. The relationship between blood pressure and C-reactive protein in the Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Coll Cardiol* 2005; 46: 1869–74.

108 Timpson NJ, Lawlor DA, Harbord RM *et al.* C-reactive protein and its role in metabolic syndrome: Mendelian randomization study. *Lancet* 2005; 366: 1954–9.

109 Schindler TH, Cardenas J, Prior JO *et al.* Relationship between increasing body weight, insulin resistance, inflammation, adipocytokine leptin, and coronary circulatory function. *J Am Coll Cardiol* 2006; 47: 1188–95.

110 Boekholdt SM, Meuwese MC, Day NE *et al.* Plasma concentrations of ascorbic acid and C-reactive protein, and risk of future coronary artery disease, in apparently healthy men and women: the EPIC-Norfolk prospective population study. *Br J Nutr* 2006; 96: 516–22.

111 Miller M, Zhan M, Havas S. High attributable risk of elevated C-reactive protein level to conventional coronary heart disease risk factors: the Third National Health and Nutrition Examination Survey. *Arch Intern Med* 2005; 165: 2063–8.

112 Clapp BR, Hirschfield GM, Storry C *et al.* Inflammation and endothelial function: direct vascular effects of human C-reactive protein on nitric oxide bioavailability. *Circulation* 2005; 111: 1530–6.

113 Lafuente N, Azcutia V, Matesanz N *et al.* Evidence for sodium azide as an artifact mediating the modulation of inducible nitric oxide synthase by C-reactive protein. *J Cardiovasc Pharmacol* 2005; 45: 193–6.

114 Nagoshi Y, Kuwasako K, Cao YN, Kitamura K, Eto T. Effects of C-reactive protein on atherogenic mediators and adrenomedullin in human coronary artery endothelial and smooth muscle cells. *Biochem Biophys Res Commun* 2004; 314: 1057–63.

115 Pepys MB, Hawkins PN, Kahan MC *et al.* Proinflammatory effects of bacterial recombinant human C-reactive protein are caused by contamination with bacterial products, not by C-reactive protein itself. *Circ Res* 2005; 97: e97–103.

116 Swafford AN Jr, Bratz IN, Knudson JD *et al.* C-reactive protein does not relax vascular smooth muscle: effects mediated

by sodium azide in commercially available preparations. *Am J Physiol Heart Circ Physiol* 2005; 288: H1786–95.

117 Taylor KE, Giddings JC, van den Berg CW. C-reactive protein-induced *in vitro* endothelial cell activation is an artefact caused by azide and lipopolysaccharide. *Arterioscler Thromb Vasc Biol* 2005; 25: 1225–30.

118 van den Berg CW, Taylor KE, Lang D. C-reactive protein-induced in vitro vasorelaxation is an artefact caused by the presence of sodium azide in commercial preparations. *Arterioscler Thromb Vasc Biol* 2004; 24: e168–71.

119 van den Berg CW, Taylor KE. Letter regarding article by Li *et al.* "C-reactive protein upregulates complement-inhibitory factors in endothelial cells". *Circulation* 2004; 110: e542.

120 Pepys MB, Hirschfield GM, Tennent GA *et al.* Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature* 2006; 440: 1217–21.

121 Paul A, Ko KWS, Li L *et al.* C-reactive protein accelerates the progression of atherosclerosis in apolipoprotein E-deficient mice. *Circulation* 2004; 109: 647–55.

122 Hirschfield G. Effects of fluvastatin and bezafibrate combination on plasma fibrinogen, t-plasminogen activator inhibitor and C-reactive protein levels in coronary artery disease patients with mixed hyperlipidaemia (FACT Study) – a rebuttal. *Thromb Haemost* 2001; 85: 189–90.

123 Reifenberg K, Lehr H-A, Baskal D *et al.* Role of C-reactive protein in atherogenesis: can the apolipoprotein E knockout mouse provide the answer? *Arterioscler Thromb Vasc Biol* 2005; 25: 1641–6.

124 Trion A, de Maat MPM, Jukema JW *et al.* No effect of C-reactive protein on early atherosclerosis development in apolipoprotein E*3-Leiden/human C-reactive protein transgenic mice. *Arterioscler Thromb Vasc Biol* 2005; 25: 1635–40.

125 Tennent GA, Hutchinson WL, Kahan MC *et al.* Transgenic human CRP is not pro-atherogenic, pro-atherothrombotic or pro-inflammatory in apoE[-/-] mice. *Atherosclerosis* 2008; 196: 248–55.

126 Kovacs A, Tornvall P, Nilsson R, Tegnér J, Hamsten A, Björkegren J. Human C-reactive protein slows atherosclerosis development in a mouse model with human-like hypercholesterolemia. *Proc Natl Acad Sci USA* 2007; 104: 13768–73.

127 Pepys MB. Isolation of serum amyloid P component (protein SAP) in the mouse. *Immunology* 1979; 37: 637–41.

128 Pepys MB, Baltz M, Gomer K, Davies AJS, Doenhoff M. Serum amyloid P-component is an acute-phase reactant in the mouse. *Nature* 1979; 278: 259–61.

129 Kerr SW, Jiang H, Sellati R *et al.* Generation and characterization of a C-reactive protein knockout mouse. *Arterioscler Thromb Vasc Biol* 2007; 27: e69.

130 Ridker PM, on behalf of the JUPITER Study Group. Rosuvastatin in the primary prevention of cardiovascular disease among patients with low levels of low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein: rationale and design of the JUPITER trial. *Circulation* 2003; 108: 2292–7.

131 Mora S, Ridker PM. Justification for the use of statins in primary prevention: an intervention trial evaluating rosuvastatin (JUPITER) – can C-reactive protein be used to target statin therapy in primary prevention? *Am J Cardiol* 2006; 97: 33A–41A.

132 Baigent C, Keech A, Kearney PM *et al.* Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomized trials of statins. *Lancet* 2005; 366: 1267–78.

133 Kinlay S. Low-density lipoprotein-dependent and -independent effects of cholesterol-lowering therapies on C-reactive protein: a meta-analysis. *J Am Coll Cardiol* 2007; 49: 2003–9.

134 Ridker PM, Rifai N, Pfeffer MA, Sacks F, Braunwald E. Long-term effects of pravastatin on plasma concentration of C-reactive protein. *Circulation* 1999; 100: 230–5.

135 Albert MA, Danielson E, Rifai N, Ridker PM. Effect of statin therapy on C-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. *JAMA* 2001; 286: 64–70.

136 Pearson T, Ballantyne C, Sisk C, Shah A, Veltri E, Maccubbin D. Comparison of effects of ezetimibe/simvastatin versus simvastatin versus atorvastatin in reducing C-reactive protein and low-density lipoprotein cholesterol levels. *Am J Cardiol* 2007; 99: 1706–13.

137 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomized placebo-controlled trial. *Lancet* 2002; 360: 7–22.

138 MRC/BHF Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol-lowering therapy and of antioxidant vitamin supplementation in a wide range of patients at increased risk of coronary heart disease death: early safety and efficacy experience. *Eur Heart J* 1999; 20: 725–41.

139 Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from Mendelian randomization. *Lancet* 2005; 365: 224–32.

140 Davey Smith G, Harbord R, Milton J, Ebrahim S, Sterne JA. Does elevated plasma fibrinogen increase the risk of coronary heart disease? Evidence from a meta-analysis of genetic association studies. *Arterioscler Thromb Vasc Biol* 2005; 25: 2228–33.

141 Keavney B, Danesh J, Parish S *et al.* Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *Int J Epidemiol* 2006; 35: 935–43.

142 Mendel G. *Versuche über Pflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn., Bd. IV für das Jahr 1865*, Abhandlungen: 1866; 3–47.

143 Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002; 31: 730–6.

144 Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.

145 Pankow JS, Folsom AR, Cushman M *et al.* Familial and genetic determinants of systemic markers of inflammation: the NHLBI Family Heart Study. *Atherosclerosis* 2001; 154: 681–9.

146 MacGregor AJ, Gallimore JR, Spector TD, Pepys MB. Genetic effects on baseline values of C-reactive protein and serum amyloid A protein: a comparison of monozygotic and dizygotic twins. *Clin Chem* 2004; 50: 130–4.

147 Wörns MA, Victor A, Galle PR, Höhler T. Genetic and environmental contributions to plasma C-reactive protein and interleukin-6 levels – a study in twins. *Genes Immun* 2006; 7: 600–5.

148 Reiner AP, Barber MJ, Guan Y *et al.* Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet* 2008; 82: 1193–201.

149 Verzilli C, Shah T, Casas JP *et al.* Bayesian meta-analysis of genetic association studies with different sets of markers. *Am J Hum Genet* 2008; 82: 859–72.

150 Kovacs A, Green F, Hansson LO *et al.* A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 2005; 178: 193–8.

151 Miller DT, Zee RY, Suk Danik J *et al.* Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann Hum Genet* 2005; 69: 623–38.

152 Kathiresan S, Larson MG, Vasan RS *et al.* Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level. *Circulation* 2006; 113: 1415–23.

153 Kardys I, de Maat MP, Uitterlinden AG, Hofman A, Witteman JC. C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. *Eur Heart J* 2006; 27: 1331–7.

154 Casas JP, Shah T, Cooper J *et al.* Insight into the nature of the CRP-coronary event association using Mendelian randomization. *Int J Epidemiol* 2006; 35: 922–31.

155 Lange LA, Carlson CS, Hindorff LA *et al.* Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA* 2006; 296: 2703–11.

156 CRP CHD Genetics Collaboration. Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomization. *Eur J Epidemiol* 2008; 23: 531–40.

157 Antman EM, Cohen M, Bernink PJ *et al.* The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA* 2000; 284: 835–42.

158 Morrow DA, Rifai N, Antman EM *et al.* C-reactive protein is a potent predictor of mortality independently of and in combination with troponin T in acute coronary syndromes: a TIMI 11A substudy. Thrombolysis in myocardial infarction. *J Am Coll Cardiol* 1998; 31: 1460–5.

159 Biasucci LM, Liuzzo G, Grillo RL *et al.* Elevated levels of C-reactive protein at discharge in patients with unstable angina predict recurrent instability. *Circulation* 1999; 99: 855–60.

160 Toussas SG, Zairis MN, Lyras AG *et al.* Early prognostic usefulness of C-reactive protein added to the thrombolysis in myocardial infarction risk score in acute coronary syndromes. *Am J Cardiol* 2005; 96: 533–7.

161 Correia LCL, Lima JC, Rocha MS, D'Oliveira Junior A, Pericles Esteves J. Does high-sensitivity C-reactive protein add prognostic value to the TIMI-Risk Score in individuals with non-ST elevation acute coronary syndromes? *Clin Chim Acta* 2007; 375: 124–8.

162 Tello-Montoliu A, Marin F, Roldan V *et al.* A multimarker risk stratification approach to non-ST elevation acute coronary syndrome: implications of troponin T, CRP, NT pro-BNP and fibrin D-dimer levels. *J Intern Med* 2007; 262: 651–8.

163 Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97: 1837–47.

164 Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 2002; 105: 310–5.

165 Conroy RM, Pyorala K, Fitzgerald AP *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003; 24: 987–1003.

166 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007; 335: 136.

167 Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* 2007; 297: 611–9.

168 Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; 247: 2543–6.

169 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15: 361–87.

170 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004; 23: 723–48.

171 Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157–72.

172 Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928–35.

173 Kroop IG, Shackman NH. The C-reactive protein determination as an index of myocardial necrosis in coronary artery disease. *Am J Med* 1957; 22: 90–8.

174 Kushner I, Broder ML, Karp D. Control of the acute phase response. Serum C-reactive protein kinetics after acute myocardial infarction. *J Clin Invest* 1978; 61: 235–42.

175 de Beer FC, Hind CRK, Fox KM, Allan R, Maseri A, Pepys MB. Measurement of serum C-reactive protein concentration in myocardial ischaemia and infarction. *Br Heart J* 1982; 47: 239–43.

176 Kinjo K, Sato H, Ohnishi Y *et al.* Impact of high-sensitivity C-reactive protein on predicting long-term mortality of acute myocardial infarction. *Am J Cardiol* 2003; 91: 931–5.

177 Suleiman M, Aronson D, Reisner SA *et al.* Admission C-reactive protein levels and 30-day mortality in patients with acute myocardial infarction. *Am J Med* 2003; 115: 695–701.

178 Pietilä KO, Harmoinen AP, Jokiniitty J, Pasternack AI. Serum C-reactive protein concentration in acute myocardial infarction and its relationship to mortality during 24 months of follow-up in patients under thrombolytic treatment. *Eur Heart J* 1996; 17: 1345–9.

179 Suleiman M, Khatib R, Agmon Y *et al.* Early inflammation and risk of long-term development of heart failure and mortality in survivors of acute myocardial infarction predictive role of C-reactive protein. *J Am Coll Cardiol* 2006; 47: 962–8.

180 Bursi F, Weston SA, Killian JM, Gabriel SE, Jacobsen SJ, Roger VL. C-reactive protein and heart failure after myocardial infarction in the community. *Am J Med* 2007; 120: 616–22.

181 Kavsak PA, MacRae AR, Newman AM *et al.* Elevated C-reactive protein in acute coronary syndrome presentation is an independent predictor of long-term mortality and heart failure. *Clin Biochem* 2007; 40: 326–9.

182 Maroko PR, Carpenter CB, Chiariello M *et al.* Reduction by cobra venom factor of myocardial necrosis after coronary artery occlusion. *J Clin Invest* 1978; 61: 661–70.

183 Pinckard RN, O'Rourke RA, Crawford MH *et al.* Complement localization and mediation of ischemic injury in baboon myocardium. *J Clin Invest* 1980; 66: 1050–6.

184 Nijmeijer R, Lagrand WK, Lubbers YTP *et al.* C-reactive protein activates complement in infarcted human myocardium. *Am J Pathol* 2003; 163: 269–75.

185 de Beer FC, Baltz ML, Munn EA *et al.* Isolation and characterization of C-reactive protein and serum amyloid P component in the rat. *Immunology* 1982; 45: 55–70.

186 Gill R, Kemp JA, Sabin C, Pepys MB. Human C-reactive protein increases cerebral infarct size after middle cerebral artery occlusion in adult rats. *J Cereb Blood Flow Metab* 2004; 24: 1214–8.

187 Thompson D, Pepys MB, Wood SP. The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* 1999; 7: 169–77.

Correspondence: Professor MB Pepys, FRS, FMedSci, Head, Division of Medicine, Royal Free Campus and Director, Centre for Amyloidosis & Acute Phase Proteins, University College London Medical School, Rowland Hill Street, London NW3 2PF, UK.
(fax: +44 20 7433 2803; e-mail: m.pepys@medsch.ucl.ac.uk).