Genetic Risk Factors for Stroke-Related Quantitative Traits and their Associated Ischaemic Stroke Subtypes

Lavinia Paternoster

Doctor of Philosophy

University of Edinburgh

March 2009

Abstract

Stroke is the 2nd leading cause of death in the UK and worldwide. 150,000 people have a stroke each year in the UK (ischaemic stroke being the most common) and a significant proportion of NHS resources go towards the treatment of these individuals (~£2.8 billion). Twin and family history studies have shown that having affected relatives makes you between 30 and 76% more likely to suffer a stroke, suggesting that there is a genetic component to the disease. So far, no genes have been convincingly associated with stroke. Intermediate traits may be useful tools for identifying genetic factors in complex disease. For stroke, two commonly used intermediate traits are carotid intima-media thickness (CIMT) and white matter hyperintensities (WMHs), which both show high heritabilities. These traits have both been studied widely for associations with many candidate gene polymorphisms.

In this thesis I systematically reviewed the literature for all genetic association studies of these two traits. Where particular associations have been studied in large numbers I meta-analysed the available data, developing novel methods for meta-analysis of genetic association data. I found there was substantial heterogeneity and small study bias in the literature and most polymorphisms have still been studied in too small numbers to make accurate conclusions. Apolipoprotein E (APOE) ε is the only polymorphism which shows a consistent association with CIMT, even when only the largest studies are analysed (MD 8μm (95% CI 6 to 11) between E4 and E3, and E3 and E2). No polymorphism has shown a convincing association with WMHs and interestingly APOE appears unlikely

to be associated with this trait. This is consistent with previous work that shows that APOE is associated with large artery but not small artery stroke.

Taking this hypothesis I attempted to investigate the association of APOE comparing patients who have had a large artery stroke with those who have had a small artery stroke in the Edinburgh Stroke Study cohort. However, genotyping of this polymorphism failed and I present investigatory analyses of problems from the genotyping laboratory.

Contents

List of Tables

List of Figures

Acknowledgements

The first thanks must go to Cathie Sudlow, who inspired me to take on this PhD in the first place and has always had time to listen to my ideas, discuss problems with me and worked relentlessly hard to try and ensure I had some 'real genotyping' data to analyse, even whilst away on maternity leave and despite the huge problems that we faced with this. Cathie always pushed me to publish everything we did, making me start writing my first paper in my first year, looking back now I can't thank her enough for this. It kept me writing, thinking about the final paper from the word go and made writing up so much easier. As well as guiding me through my PhD, Cathie has always been aware that I might leave some day, and has been instrumental in helping me forge my future academic career. I thank her unreservedly for this.

A huge thanks also goes to Steff Lewis, my own statistical guru. Steff never complained when I thrust a random paper in front of her and asked her to explain the statistics and helped me in developing the genetic meta-analysis methodology that I use in this thesis. Cathie and Steff read many many versions of the chapters that make up this thesis and their comments were always valuable and kept me enthusiastic about this thesis.

Many people contributed in some way to the work presented here. Everyone involved in the Edinburgh Stroke Study (notably Caroline Jackson and Aidan Hutchison for their help with the database and the WTCRF lab staff for answering my persistent questions) and Nahara Martinez-Gonzalez, Mabel Chung, Wanting Chen and Rebecca Charleton, who worked with me on the meta-analyses.

Other special people of note are Pippa Thomson, who led me through an MSc project, before helping me in my PhD with the interpretation of genotyping problems and David Porteous, who took time out of his busy schedule to help me remember my passion for genetics and advise me on my future career.

I also thank the Medical Research Council and the Newby Trust for supporting me financially.

Many friends and colleagues were sources of emotional support. I particularly want to thank my sailing girls for always being there for me, for providing some welcome distraction at times, and for showing me how to work hard and play hard. Special thanks must also go to Tom, for always feigning interest in my work and always reminding me that we should be striving for excellence.

My family's pride in me has been a massive source of motivation. I am so sorry that on occasion I 'brought my work home', taking over your dining room table, and sofa, and computer, and, Kyle, your room. Your work ethic and belief that I could achieve anything I wanted has made me who I am. Thank you.

Statement of Contribution

Brenda Thomas assisted in the design of the CIMT and APOE systematic review search strategy.

Several students, Nahara Martinez-Gonzalez, Wanting Chen, Mabel Chung and Rebecca Charleton assisted me with carrying out the search strategies, selecting papers for inclusion and data extraction for the meta-analyses presented in this thesis (often providing the necessary independent selection and extraction of data), either alongside or under the direction of myself. Steff Lewis and Cathie Sudlow helped to resolve difficulties/disagreements between observers.

The genetic meta-analysis method presented in chapter 2 was designed by myself, with advice from Steff Lewis and incorporates the gametan command produced (but not yet published) by Julian Higgins.

The Edinburgh Stroke Study (ESS), led by my supervisor, Cathie Sudlow and coordinated by Caroline Jackson, provided phenotypic data and DNA samples for my association study between APOE and stroke subtypes. The ESS data collection systems and methods, and the algorithm used to assign TOAST subtypes were developed by Caroline Jackson and Cathie Sudlow. The genotyping was carried out by the Wellcome Trust Clinical Research Facility of the Western General Hospital, Edinburgh. The ESS database has been maintained by Aidan Hutchison and Caroline Jackson, both of whom assisted me in the extraction of the relevant data from the database.

Pippa Thomson provided helpful advice following problems with genotyping the ESS samples. She suggested several lines of enquiry, which through discussion with the WTCRF lab we were able to pursue.

Declaration

I declare that this thesis has been composed by myself and that the work contained herein is my own (unless otherwise stated above). This work has not been submitted for any other degree or professional qualification.

Lavinia Paternoster
14 March 2009

Publications Arising from Thesis

Published:

Paternoster L, Martínez González NA, Lewis S, Sudlow C (2008). Association between apolipoprotein E genotype and carotid intima-media thickness may suggest a specific effect on large artery atherothrombotic stroke. *Stroke* **39(1)**:48-54.

(based on an earlier version of the apolipoprotein E and CIMT meta-analysis from chapter 3)

Paternoster L, Chen W, Sudlow CL (2009). Genetic determinants of white matter hyperintensities on brain scans: a systematic assessment of 19 candidate gene polymorphisms in 46 studies in 19,000 subjects. *Stroke* **40**:2020-2026.

(based on chapter 4)

In Press:

Paternoster L, Martinez-Gonzalez N, Charleton R, Chung M, Lewis S, Sudlow C. Genetic effects on carotid intima-media thickness (CIMT): Systematic assessment and meta-analyses of candidate gene polymorphisms studied in over 5000 subjects. *Circulation: Cardiovascular Genetics,* in press.

(based on chapter 3)

Abbreviations

ABI	Applied Biosystems Inc
ACE	angiotensin converting emzyme
ADD1	adducin 1
ADRB2	adrenergic beta-2 receptor
AF	atrial fibrilliation
AGT	Angiotensinogen
ALOX5AP	arachidonate 5-lipoxygenase-activating protein
ANOVA	analysis of variance
APOE	apolipoprotein E
ARWMC	age-related white matter changes
BIF	bifurcation
bp	basepair
CCA	common carotid artery
CI	confidence interval
CIMT	carotid intima-media thickness
CNV	copy number variation
CRP	c-reactive protein
CT	computed tomography
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DWMH	deep white matter hyperintensity
ECA	external carotid artery
EDTA	ethylenediaminetetraacetic acid
ESS	Edinburgh Stroke Study
FGG/FGA	fibrinogen gamma/fibrinogen alpha
FV	factor V
GP	General Practioner
GWAS	genomewide association study
HDL	high density lipoprotein
HR	hazard ratio
HuGENet	Human Genome Epidemiology Network
HWE	Hardy-Weinberg equilibrium
ICA	internal carotid artery
IGF1	insulin-like growth factor
IHD	ischaemic heart disease
IL6	interleukin 6
IPD	individual participant data
IQR	interquartile range
IS	ischaemic stroke
LACI	lacunar infarct
LAS	large artery ischaemic stroke
LD	linkage disequilibrium
LDL	low-density lipoprotein
LOA	limit of agreement
LOD	logarithm of odds
MAF	minor allele frequency
MD	mean difference
MeSH	Medical subject headings
MI	myocardial infarction

MRI	magnetic resonance imaging
MTHFR	methylenetetrahydrofolate reductase
NHS	National Health Service
NOS3	nitric oxide synthase 3
OCSP	Oxfordshire Community Stroke Project
OR	odds ratio
PACI	partial anterior circulation infarct
PCR	polymerase chain reaction
PDE4D	phosphodiesterase 4D
POCI	posterior circulation infarct
PON1	paraoxonase 1
PVH	periventricular hyperintensity
RA	renin angiotensin
Rn	reporter signal
SAS	small artery ischaemic stroke
SD	standard deviation
SE	standard error
SMD	standardised mean difference
SNP	single nucleotide polymorphism
TACI	total anterior circulation infarct
TIA	transient ischaemic attack
TOAST	Trial of Org 10172 in Acute Stroke Treatment
TPA	total plaque area
TPV	total plaque volume
UV	ultra violet
WMC	white matter changes
WMH	white matter hyperintensity
WML	white matter lesions
WTCRF	Wellcome Trust Clinical Research Facility

Preface

This thesis is organised into two main sections (A and B). Section A comprises two large-scale systematic reviews, one of the commonly studied genes for an association with carotid intima-media thickness, and the other of the commonly studied genes for an association with white matter hyperintensities. From section A I devised a hypothesis to test in the Edinburgh Stroke Study (ESS). I planned to genotype Apolipoprotein E and test for an association between this genotype and stroke subtypes.

The results and interpretation of the attempted APOE genotyping study in the ESS are presented in section B. Unfortunately, problems were encountered during the genotyping and so I could not carry out the planned association. Instead, I present a thorough investigation of the potential causes of genotyping problems and discuss future directions for genotyping in the ESS.

1 Introduction

In this chapter I introduce the topics that are combined in this thesis. I first define stroke and the various types and subtypes of the disease and discuss the public health impact, as well as the risk factors (including heritability). I then introduce the reader to methods used in genetic epidemiology to identify genes that influence disease (including the use of intermediate traits) and discuss the attempts to identify genes for stroke. I then present the history and methodology for systematic review and meta-analysis. Finally, I outline the aims of this thesis, to use systematic review and meta-analysis techniques to identify genetic polymorphisms that influence intermediate traits for stroke and then to attempt to test any hypotheses arising from this in a cohort of stroke patients collected in Edinburgh.

1.1 Stroke

1.1.1 Definition

A stroke is the sudden death of a portion of the brain due to lack of oxygen. This occurs when blood flow to the brain is interrupted, by blockage or rupture of an artery. The most common symptom is numbness and/or weakness of the face, arm or leg, normally on one side of the body. Other symptoms include difficulty speaking or swallowing, dizziness, confusion and - occasionally - unconsciousness. The symptoms vary according to the area of the brain that is affected and the severity of the symptoms tends to be associated with the size of the area of damaged brain tissue.

The World Health Organisation defines a stroke as:

> "a clinical syndrome characterized by rapidly developing clinical symptoms and/or signs of focal, and at times global (applied to patients in deep coma and those with subarachnoid haemorrhage), loss of cerebral function, with symptoms lasting more than 24h or leading to death, with no apparent cause other than that of vascular origin" [Hatano, 1976]

1.1.2 Main Pathological Types

Stroke is heterogeneous in its pathology. There are three main pathological types of stroke [Sudlow & Warlow, 1997]:

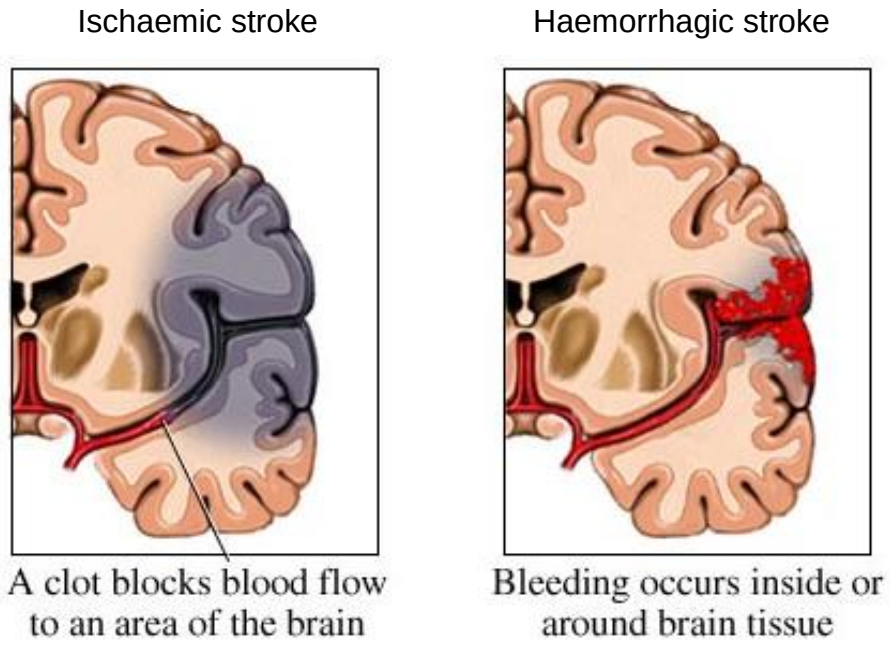


Figure 1.1 Diagram of ischaemic (occluded artery) and haemorrhagic (ruptured artery) stroke. Illustration by Nucleus Communications, Inc.

1.1.2.1 Ischaemic stroke

This accounts for 80% of all strokes and is caused by an occluded blood vessel.

1.1.2.2 Primary intracerebral haemorrhage

This accounts for 10% of all strokes and is caused by rupture of a blood vessel with leaking of blood into the brain tissue.

1.1.2.3 Subarachnoid haemorrhage

This accounts for 5% of all strokes and is caused by rupture of a blood vessel with leaking of blood into the subarachnoid space (within the skull, but outside the brain tissue).

1.1.2.4 Other

In community / population-based studies around 5% of strokes are of undetermined type, because of lack of appropriately timed brain scan or autopsy to distinguish reliably between the different pathologies.

In this thesis I will focus specifically on ischaemic stroke.

1.1.3 Ischaemic Stroke Subtypes

Ischaemic stroke can also be classified into subtypes. There are several methods of classifying and diagnosing ischaemic stroke subtypes. Two common methods are TOAST (Trial of Org 10172 in Acute Stroke Treatment) and OCSP (Oxfordshire Community Stroke Project). The OCSP uses clinical symptoms and signs to assign the patient to one of four categories that predicts the site, size and likely pathophysiological mechanism(s) of the ischaemic stroke: TACI – total anterior circulation infarct; PACI – partial anterior circulation infarct; POCI – posterior circulation infarct; LACI – lacunar infarct [Bamford *et al.,* 1991]. The TOAST classification is based directly on the presumed pathophysiological mechanisms and so is considered more suitable for investigating the relationship of risk factors to specific pathophysiological processes leading to ischaemic stroke [Adams, Jr. *et al.,* 1993]. One disadvantage is the requirement for a series of (often high-tech) investigations and so it is not suitable for a quick diagnosis or for use in less equipped clinics. The TOAST classification also leaves quite a large proportion of ischaemic strokes unclassified due to incomplete investigations or multiple possible mechanisms [Jackson & Sudlow, 2005]. It classifies ischaemic stroke into the following four subtypes:

1. Large-artery atherosclerosis

Clinical and imaging findings show evidence of stenosis or occlusion of a major brain artery. There is evidence of atherosclerosis and the infarct on imaging is more than 1.5cm in diameter. This diagnosis is made after excluding sources of cardioembolism. See figure 1.2.

2. Small-vessel occlusion

Dysfunction of the small perforating arteries results in a typical lacunar syndrome. Imaging shows a deep infarct of no more than 1.5cm. This diagnosis is made after sources of cardioembolism and >50% stenosis of an ipsilateral artery are excluded. See figure 1.3.

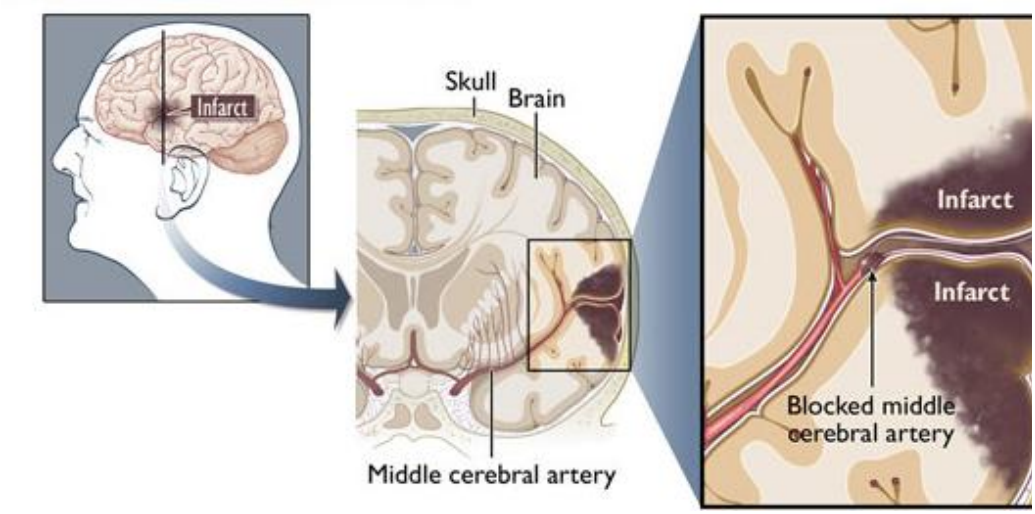


Figure 1.2 Diagram of large artery ischaemic stroke. Image taken from http://uwmedicine.washington.edu

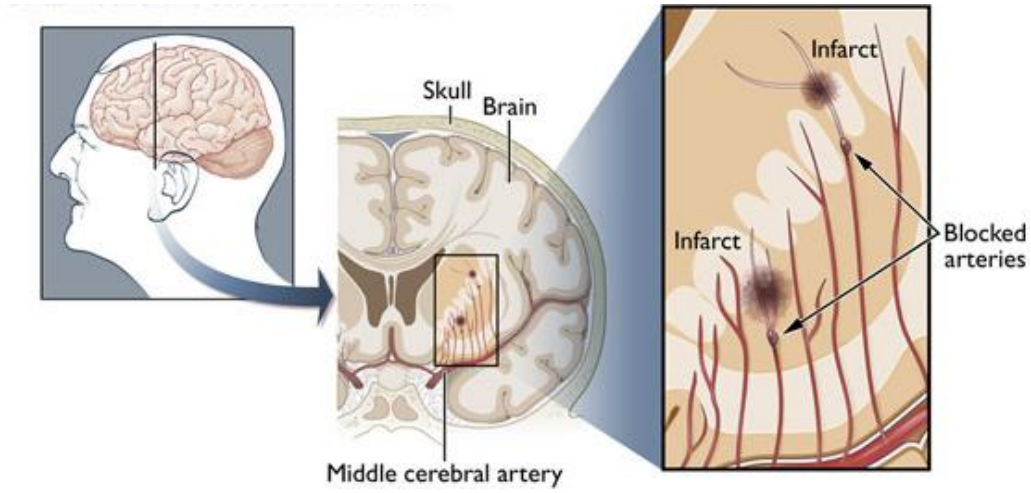


Figure 1.3 Diagram of small-vessel ischaemic stroke. Image taken from http://uwmedicine.washington.edu

3. Cardioembolism

Patients in this category have an occluded artery with a presumed cardiac source of embolism. Clinical and imaging findings are similar to that for large-artery atherosclerosis, but large artery atherosclerosis sources of thrombosis or embolism are excluded.

4. Other

This includes patients with rare causes of disease such as nonatherosclerotic vasculopathies, hypercoagulable states or haematological disorders. In addition to clinical and imaging findings, blood tests or arteriography help to diagnose these rarer causes.

1.1.4 Public Health Impact of Stroke

According to the World Health Organisation, cardiovascular disease (ischaemic heart disease and stroke combined) is the leading cause of death

and burden of disease worldwide, accounting for 32% of the deaths in women and 29% of the deaths in men in 2004 [World Health Organisation, 2004]. On its own stroke is second only to ischaemic heart disease, with 5.7 million deaths worldwide in 2004. In the UK stroke is also second only to ischaemic heart disease, with around 55,000 deaths caused by stroke in 2006 [Allender *et al.,* 2008]. In low-income countries stroke drops to the fifth leading cause of death [World Health Organisation, 2004]. There were 9 million new cases of stroke worldwide in 2004 (2 million in Europe). Clearly, stroke is of major public health importance. As well as accounting for a huge number of deaths, stroke is an important cause of disability worldwide, since a large number of people who have a stroke live with the disabling effects for many years. In 2004 there were an estimated 30 million stroke survivors in the world, 12 million of whom were described as having moderate or severe disability [World Health Organisation, 2004]. Stroke is described as the single biggest cause of major disability in the UK [Mackay & Mensah, 2004].

Stroke has a vast health care and economic burden. With stroke patients occupying 20% of all acute hospital beds and 25% of long-term beds, the direct cost of stroke to the NHS is thought to be around £2.8 billion per year [Department of Health, 2005].

It is thought that the burden of stroke will increase by the year 2030, mainly due to an increasingly ageing population [World Health Organisation, 2004].

1.1.5 Traditional Risk Factors

Traditional risk factors for stroke include non-modifiable factors such as age, sex, and ethnicity; and potentially modifiable factors such as hypertension, smoking, diabetes and atrial fibrillation [Goldstein *et al.*, 2006].

Age - Stroke risk has been found to double for each successive decade after age 55.

Sex - There is a higher incidence of stroke in men than in women (age-adjusted).

Ethnicity - There is a higher incidence of stroke in African Americans and East Asian individuals. This could possibly be due to the higher incidences of hypertension, obesity and diabetes in these populations

Hypertension - This is probably the most established modifiable risk factor for stroke. Individuals with higher blood pressure have increased risk of stroke. Antihypertensive treatment is associated with approximately 40% reduction in risk of stroke.

Smoking - Smoking doubles the risk of ischaemic stroke and is associated with a 3-fold increased risk of haemorrhagic stroke.

Diabetes - Diabetes is associated with an increased ischaemic stroke relative risk of between 1.8 and 6.

Atrial Fibrillation - AF is associated with a 3 to 4 fold increased risk of stroke. Stroke in patients with AF tend to be

larger and more disabling, and AF is associated with increased mortality.

Hyperlipidaemia - Increased cholesterol levels are associated with increased risk of ischaemic stroke.

Carotid stenosis - Ischaemic stroke is more frequent in patients with severe (>75%) carotid stenosis.

There are many other potential risk factors for stroke including diet, physical activity, hormone therapy, obesity, alcohol and drug abuse, and oral contraceptive use.

Some risk factors may be of particular importance to specific subtypes of stroke. For example, while diabetes has been shown to be a risk factor for ischaemic but not haemorrhagic stroke [Abbott *et al.*, 1987], hypertension has been associated with both haemorrhagic and ischaemic stroke [Sacco *et al.*, 1997]. Also, risk factor profiles may differ between specific subtypes of ischaemic stroke. However, as many risk factors are included in the definitions of specific subtypes, this is difficult to test without bias. In a systematic review, Jackson et al. [Jackson & Sudlow, 2005] found that both hypertension and diabetes were more common in lacunar than other subtypes of ischaemic stroke, but when risk factors were excluded from the stroke subtype definitions, this was only true for hypertension and the excess in lacunar ischaemic stroke was very small. They also found that atrial fibrillation and carotid stenosis were more common in non-lacunar stroke.

1.1.6 Heritability

Although stroke is not thought of as a genetic disease in the traditional sense, a family history of stroke has long been regarded as an important risk factor for the disease. Many studies have attempted to estimate how heritable stroke is using twin and family history studies. Flossmann et al. [2004] systematically reviewed the evidence for heritability of stroke. They identified 3 twin studies, 33 case control family history studies and 17 cohort family history studies, published between 1966 and 2003. They concluded that monozygotic twins had a 65% increased odds of being concordant for stroke compared with dizygotic twins. The case control studies showed that having a family history of stroke increased the odds of stroke by 76%. The cohort studies showed that a family history of stroke increased the odds by 30%. These estimates may be biased by various factors, including an unmeasured environmental contribution that may explain some of the supposed 'heritability'. In addition, many studies did not distinguish between haemorrhagic and ischaemic stroke. However, the twin studies (which are considered the most reliable and less influenced by confounding environmental factors) present a convincing case for at least a small genetic influence on the risk of stroke.

It has been shown that a family history of stroke is a stronger predictor of stroke when the affected relatives were younger [Flossmann *et al.*, 2004].

Few studies have assessed the influence family history has on odds of stroke stratified by stroke subtype, but those that have, have found that a family history of stroke is less frequent in cardioembolic stroke compared to large

and small artery stroke [Flossmann *et al.,* 2004; Schulz *et al.,* 2004], and is more frequent in large artery stroke than small artery stroke [Jerrard-Dunne *et al.,* 2003a]. This latter study also found that family history of MI was more common in patients with large artery stroke, than other subtypes.

Traditional risk factors for stroke (such as hypertension, diabetes, hyperlipidaemia) are known to have genetic components and could account for some of the heritability. Although adjusting for traditional risk factors diminished the association between family history and stroke, this still remained significant in a number of studies [Flossmann *et al.,* 2004], suggesting there are other genetic influences for stroke, beyond that expected for known risk factors.

1.2 Genetics

1.2.1 State of the Art

Studies that attempt to identify genetic variants that influence disease or phenotypic traits can be divided into two categories; linkage analysis studies, and association studies. Association studies can be candidate studies with a priori expectations, or genome-wide studies with no a priori expectations of the genes involved. Statistical methods and laboratory techniques have advanced to allow sophisticated analysis of genetic data.

1.2.1.1 Linkage studies

Linkage studies rely on the co-segregation of loci in pedigrees. Recombination between markers during meiosis occurs at a rate related to the distance between them. Therefore a disease/trait allele will be inherited in families along with a background section of the genome. By studying which genomic sections are commonly co-inherited with the disease/trait of interest in a family, the location of the variant of interest can be later refined [Dawn & Barrett, 2005]. Linkage analysis is generally 'genome-wide' or 'chromosome-wide' and only identifies large regions of linkage, not specific genes or mutations. This method is most useful for variants that have a large effect (which are often rare). Linkage studies also have their limitations for late-onset conditions such as stroke, since it is not necessarily appropriate to assign young people as unaffected, when they may go on to develop the disease in the future.

1.2.1.2 Association studies

Association studies, by contrast, are more useful for variants that are common, but have small effects [Risch & Merikangas, 1996]. This method looks for an association between the disease/trait and genetic variants in the population [Cordell & Clayton, 2005]. However, linkage disequilibrium (LD) between close markers means that the associated variant is not necessarily the causal variant.

Association studies can be either of candidate genes or genome-wide. Candidate gene studies require background knowledge to inform the choice of genes to be studied. This decision may be based on prior evidence of

association or linkage in the region, but are often selected with only tentative biological reasoning. Often little is known about the mechanistic pathways leading to a trait or disease and so selecting candidate genes this way can be difficult. Given the number of genes in the genome (~20,000), it is extremely unlikely a priori that disease risk genes will be selected for such studies, and so important genes are likely to be missed with this approach.

1.2.1.3 Genome-wide association studies

Genome-wide studies require no a priori expectation on which genes are associated with the disease or trait of interest. They usually involve genotyping of single nucleotide polymorphisms (SNPs) from across the entire genome. Associations with each SNP are then tested for. This can result in novel genes being identified as associated with the diseases/trait of interest. Genome-wide SNP chips have been developed that are either gene-centric; include large numbers of randomly selected SNPs from across the genome; or include 'tagging' SNPs that represent each LD block in the genome (thereby capturing as much of the variation as possible) [Li *et al.*, 2008]. SNP chips can now screen more than 1 million SNPs and the cost of genotyping has been rapidly decreasing, making genome-wide studies more affordable. However SNP chips do not capture all genomic variation and so this approach may miss some important genetic associations, demonstrating the continued need for candidate gene studies.

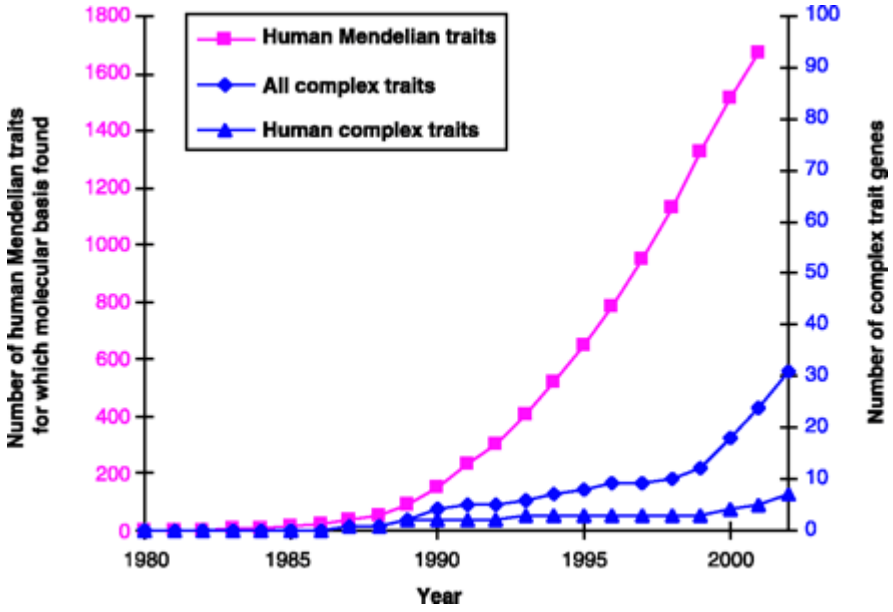


Figure 1.4 Cumulative numbers of identification of genes underlying human Mendelian traits and genetically complex traits in humans and other species. [Glazier *et al.*, 2002]

1.2.2 Complex Disease Genetics

1.2.2.1 Problems

Figure 1.4 shows the number of genes identified for Mendelian and complex traits up to 2002. The identification of the genes that cause Mendelian diseases has been straightforward and successful (see figure 1.4, pink squares). However, these diseases are relatively rare and hence of limited public health importance. Attentions have now turned to more common diseases that affect vast numbers of people and do not appear to be inherited in a Mendelian fashion, e.g. cancer, heart disease, schizophrenia, asthma and stroke etc. These are likely to be determined by a number of genetic and environmental factors. As most of these factors are likely to have modest effects, identifying them is difficult and, on the whole, attempts have been disappointing (see figure 1.4, blue triangles).

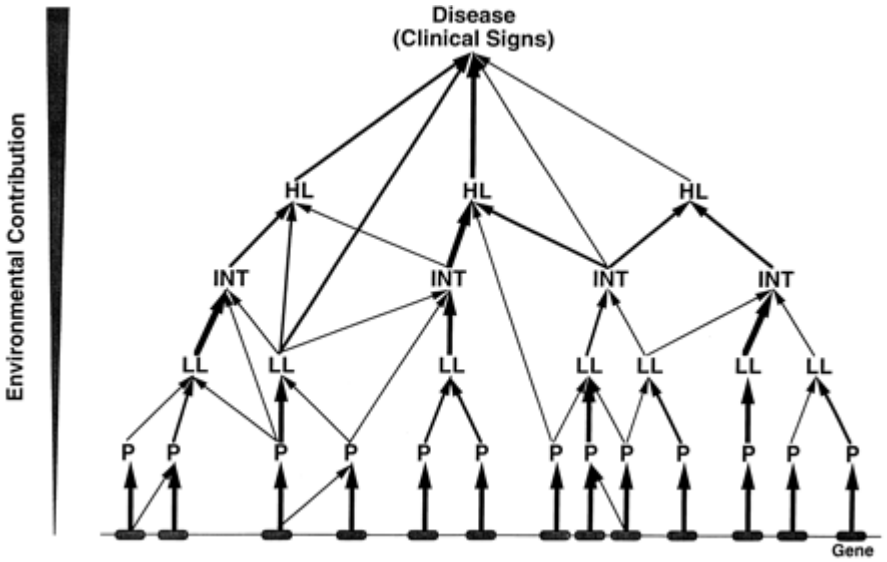


Figure 1.5 Schematic diagram of the relationship of genes and their products to intermediate phenotypes as well as the more overt clinical manifestations of a disease. The thickness of the arrows denotes the strength of the contribution of a lower-level factor to a higher-level factor. The inverted triangle on the left-hand side of the figure represents the (likely) diminishing effect of environmental conditions on factors integrated at lower and lower levels of a biochemical and physiological hierarchy. P=protein; LL=lower-level factor; INT=intermediate trait; HL=higher-level factor [Schork, 1997].

Figure 1.5 shows a schematic diagram of a possible pattern of causality for a complex disease. It is likely that there are many genetic factors that influence (to differing amounts) various protein levels and/or functions and intermediate phenotypes, which in turn, influence other intermediate phenotypes, ultimately resulting in the manifestation of disease. Alongside the genetic influences, there are also environmental contributions, which may have more of an effect at the higher (and later) levels (figure 1.5).

As the individual effect that a single variant will have on the occurrence of disease is likely to be very small, studies will require extremely large numbers of subjects to be statistically powered to detect them. Also, different genes will cause the same disease in different people (genetic heterogeneity) and not everyone with a particular 'causal' variant will

develop the disease (phenotypic heterogeneity), further complicating the identification of genes of importance. There may also be interactions between genes and/or environmental factors. All of these are likely reasons why the contribution of a gene to disease may be obscured and may explain why studies of genes influencing common complex disease have been conflicting [Schork, 1997].

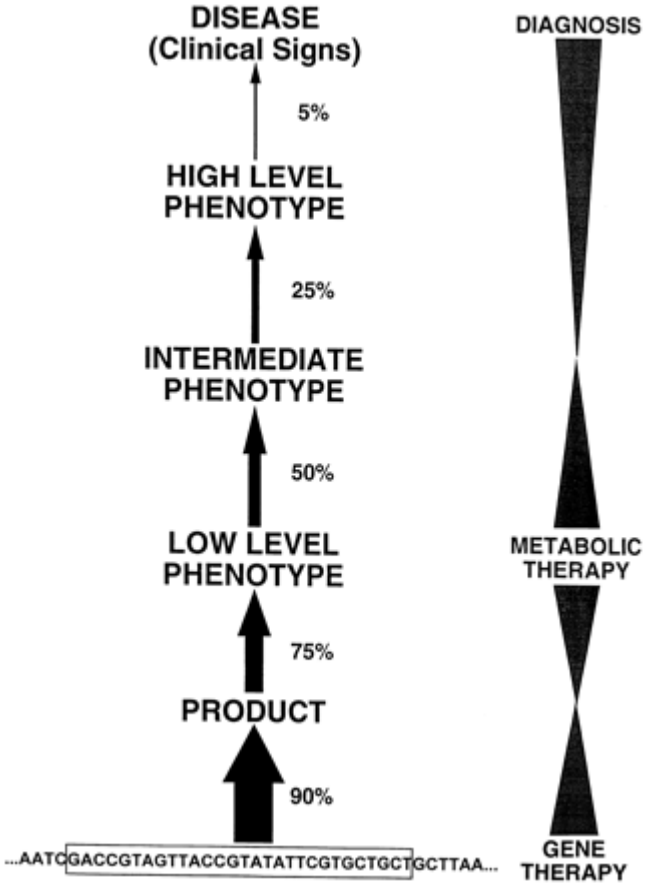


Figure 1.6 Schematic diagram of the relationship of a gene and its product to a single pathway that, when disrupted or dysfunctional, may contribute to disease. The thickness of the arrows between elements at different levels characterizes the strength of the contribution of a lower-level factor to a higher-level factor. The percentages next to an arrow give the hypothetical percentage of variation explained by the lower-level factor for the higher-level factor that they influence. The symbols on the right-hand side of the figure characterize the potential for diagnosis and therapeutic intervention at each level, with the size of the figures corresponding to the most realistic points for diagnosis or intervention [Schork, 1997].

1.2.2.2 Use of intermediate quantitative traits

One possible way to overcome the problems of identifying genetic factors for complex disease is to study the upstream intermediate traits [Majumder & Ghosh, 2005; Pan *et al.*, 2006]. Intermediate quantitative traits are often highly heritable and have a simpler genetic architecture than the disease end-point, as they are closer to the gene in the pathway (see figure 1.6). Therefore, it may be easier to identify the genetic polymorphisms that influence the intermediate trait.

Another advantage of intermediate traits is that they can be measured in the general population, usually with high accuracy. This is in contrast with the disease end-point, which is often late-onset and diagnosing somebody as a 'control', rather than a 'pre-case' may be difficult. Intermediate traits also tend to be quantitative (instead of binary), which increases the power of statistical analyses.

1.2.3 Identification of Genes that Cause Stroke

Many studies have assessed the genetic component of stroke. For example, HuGENavigator (http://hugenavigator.net), a database of human genetic epidemiology data which screens PubMed for relevant publications, lists 90 genes that have been studied for association with cerebral infarction. However, these attempts to identify genetic risk factors for stroke have been disappointing. Many studies have been conflicting and no single gene with a large effect has been identified. A meta-analysis of the 13 most commonly studied genes found only small associations for four genes (Factor V, MTHFR, prothrombin and ACE) [Casas *et al.*, 2004].

The deCODE group in Iceland has identified two potentially important genes for stroke using linkage analysis. A significant region of linkage was observed on chromosome 5q for ischaemic stroke [Gretarsdottir *et al.,* 2002]. A subsequent case control association study on a denser set of markers within this region identified PDE4D as the associated gene [Gretarsdottir *et al.,* 2003]. The association appeared to be specific to large artery and cardioembolic stroke, rather than small artery stroke. Another linkage study by the same group identified a significant region of linkage on chromosome 13q for stroke and MI. The subsequent case control association analysis identified ALOX5AP as the gene of interest [Helgadottir *et al.,* 2004]. This gene appeared to be associated with both haemorrhagic and ischaemic stroke. However, subsequent association studies of these two genes in other populations have been conflicting [Dichgans, 2007] and a meta-analysis of the association between genetic variants in the PDE4D gene and stroke reported no clear evidence of overall association [Bevan *et al.,* 2008].

Certain genes may predispose to all subtypes of stroke, whilst others may only predispose to specific subtypes [Dichgans & Markus, 2005]. Therefore studies which carefully classify stroke subtypes and analyse according to these, are important. Genes may also influence conventional risk factors (e.g. hypertension and diabetes) or may impact more directly on disease of the blood vessels or on neuronal susceptibility to an ischaemic or haemorrhagic insult. Other genes may not predispose to the disease itself but may influence the ability to recover from a stroke.

1.2.3.1 Quantitative traits for stroke

As described earlier, quantitative traits can provide greater statistical power for identifying genetic associations than the end-point of disease itself. For stroke two commonly used quantitative traits are carotid intima-media thickness (CIMT) and white matter hyperintensities on a brain scan (WMHs) [Dichgans & Markus, 2005]. These both have a strong genetic component. CIMT is a marker of subclinical atherosclerosis [Lorenz *et al.,* 2007], is a strong predictor of future myocardial infarction and stroke [Dijk *et al.,* 2006], and is associated more with large artery than small artery stroke. It is therefore a suitable quantitative trait for large artery stroke. WMHs are associated with a history of, and later progression to, small artery infarcts and clinical stroke [Leys *et al.,* 1999], and have been shown to be more prevalent in small artery compared to large artery stroke [Wiszniewska *et al.,* 2000]. They are therefore a suitable quantitative trait for small artery stroke.

Large numbers of studies have tested for associations between these traits and variation in many genes. However, these have produced conflicting results. Humphries & Morgan [2004] have published a narrative review on some of the genes which have been studied for an association with CIMT and report that the conflicting and under-powered studies make it difficult to determine the effects of these genes on CIMT. No review has been published on the association of genetic variants with WMH.

1.3 Systematic Review and Meta-Analysis

A meta-analysis is a statistical technique to calculate an overall summary outcome measure from the results from two or more studies. This may include some or all of the data available on a particular topic.

A systematic review is a review of all of the available evidence on a topic, which may or may not include a formal statistical meta-analysis.

A systematic review containing a meta-analysis provides a summary estimate calculated from all of the available data on a topic. This provides a (potentially) unbiased overall estimate with increased statistical power and so precision of the result.

1.3.1 History

The first documented combination of data from several studies was carried out by Karl Pearson in 1904 [Pearson, 1904]. He combined evidence from 11 datasets on the use of vaccines for typhoid. He estimated the overall correlation between typhoid inoculation and mortality, and concluded that this was weak. He also made observations analogous to what we now call 'heterogeneity' and 'statistical significance'.

Glass coined the term 'meta-analysis' in 1976 [Glass, 1976], and the technique became popular in the 1980s. In recent years, an explosion in the number of original studies published has led to a correspnding explosion in meta-

analyses, with a rise from approximately 250 meta-analyses published in 1990, to 2250 published in 2006 [Sutton & Higgins, 2008].

Systematic reviews and meta-analyses have been increasingly widely used to summarise the results of clinical trials, at least in part due to the establishment of the Cochrane Collaboration. This organization aims to:

> "Improve healthcare decision-making globally, through systematic reviews of the effects of healthcare interventions, published in The Cochrane Library."

As well as being a valuable resource for those searching for systematic reviews on a particular topic, the organisation also provides guidelines, assistance and software for investigators carrying out such reviews. Of particular relevance here is the Cochrane RevMan software[The Cochrane Collaboration, 2006], in which an entire review can be produced. It provides a step-by-step approach for preparing the text, tables, references, and for data input, and enables the user to perform meta-analysis and produce relevant graphs. Although the software has been designed for reviews of randomised trials of interventions, it can also be used for observational studies (albeit with limitations which I discuss in chapter 2).

The common use of systematic review and meta-analysis for observational epidemiological studies is more recent, and does not yet have the supportive infrastructure that the Cochrane Collaboration provides for randomised trials [Dickersin, 2002]. However, the HuGENet (Human Genome Epidemiology Network) collaboration has been recently formed to provide

similar guidelines and infrastructure for reviews of genetic epidemiology [Seminara *et al.*, 2007].

1.3.2 Systematic Review Methodology

A review is 'systematic' if it attempts to collate all evidence on a particular topic. To achieve this, firstly, the research question must be properly defined. For example 'is gene A associated with stroke?' is often insufficient. The reviewer must consider the types of studies to include (case-control, cohort, family studies), as well as the specific population of interest (e.g. early onset, elderly). Once a specific research question has been devised, a thorough search strategy can be built around this.

1.3.2.1 Database searching

Online databases such Medline and Embase aim to collate all health and medical journal articles and index them according to study details, including authors, title, journal and keywords, to allow these to be searched easily and the results downloaded into reference software such as Reference Manager.

The appropriate databases to search will depend on the field of study. For medical and health related journals, Medline and Embase provide good coverage. BIOSIS may be of more relevance for topics in general life sciences. Medline and Embase both index approximately 5000 journals, with approximately 3000 journals that are indexed in both (table 1.1). Depending on the subject, Embase or Medline may be more appropriate, but a comprehensive search of medical articles should generally include both.

Table 1.1 Numbers of citations indexed in Medline and Embase databases. Data from Ovid

Database	Dates	Number of journals indexed	Number of citations indexed	Number of citations added each year
Medline	1966 - present	~5250	~11.8 million	~520,000
Embase	1980 - present	~4550	~7.7 million	~500,000

Once the database/s have been selected, the next stage is to develop the search strategy. Different search strategies will be needed for each database, as the indexing terms they use may differ. As well as searching the citations for specific words or phrases, Medical Subject Headings (MeSH), which have been used to index the citations, can be used to identify relevant articles. Boolean terms, along with other syntax for specific queries can be incorporated to devise a sophisticated search strategy that matches the user's requirements. For a systematic review, the search strategy should be designed for maximum sensitivity whilst maintaining appropriate specificity. Table 1.2 shows an example of a multi-stage search strategy used to search the Medline database.

1.3.2.2 Selecting articles

Often a systematic search of the literature gives thousands of articles, many of which will be irrelevant. After searching, the investigators can then use their exact study criteria to select which citations are relevant, and which are not.

A systematic review may merely present all of the relevant papers, along with their results and stop there. However, if possible and appropriate it is likely that a review will go on to carry out a meta-analysis to summarise the data in a meaningful way.

Table 1.2 Demonstration search strategy for MEDLINE (Ovid format), for the topic 'Tamoxifen for breast cancer', taken from Cochrane Handbook.

1	randomized controlled trial.pt.
2	controlled clinical trial.pt.
3	randomized.ab.
4	placebo.ab.
5	drug therapy.fs.
6	randomly.ab.
7	trial.ab.
8	groups.ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	animals.sh. not (humans.sh. and animals.sh.)
11.	9 not 10
12.	exp Breast Neoplasms/
13.	(breast adj6 cancer$).mp.
14.	(breast adj6 neoplasm$).mp.
15.	(breast adj6 carcinoma$).mp.
16.	(breast adj6 tumour$).mp.
17.	(breast adj6 tumor$).mp.
18.	12 or 13 or 14 or 15 or 16 or 17
19.	exp Tamoxifen/
20.	tamoxifen.mp.
21.	19 or 20
22.	11 and 18 and 21

The 'adj6' operator indicates within six words; '$' indicates truncation; .mp. indicates a search of title, original title, abstract, name of substance word and subject heading word; .pt. indicates publication type; .ab. indicates a search of the abstract; .fs. indicates qualifiers for MeSH terms; .sh. indicates MeSH terms; exp …. / indicates a MeSH term and all of its subsidiary terms.

1.3.3 Meta-Analysis Methods

The method for combining the individual results to obtain an overall estimate of the results (summary statistic), depends on the research question and the type of data available. For dichotomous data common summary statistics are odds ratios (ORs), risk ratios and hazard ratios. For continuous data, mean differences (MDs) or standardised mean differences are commonly used as summary statistics. The first stage is to obtain study-

specific estimates and then a weighted average of these across all studies is calculated to give the summary statistic together with its p value and 95% confidence interval. Below I will describe two methods for study summary statistics; one for dichotomous data (ORs) and one for continuous data (MDs). I will then show the weighting methods used to obtain the pooled estimates.

1.3.3.1 Odds Ratio

Table 1.3 Notation used for odds ratio calculation. i refers to the ith study.

Risk Factor	Affected	Unaffected
Present	a_i	b_i
Absent	c_i	d_i

Table 1.3 shows the notation used in the following calculations:

$$OR_i = \frac{a_i\, d_i}{b_i\, c_i}$$

$$SE[\ln(OR_i)] = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}}$$

where OR = Odds ratio, SE=standard deviation

1.3.3.2 Mean difference

Table 1.4 Notation for mean difference calculation

Risk Factor	Mean of trait	Standard deviation	Sample size
Present	m_{1i}	SD_{1i}	n_{1i}
Absent	m_{2i}	SD_{2i}	n_{2i}

Table 1.4 shows the notation used in the following calculations:

$$MD_i = m_{1i} - m_{2i}$$

$$SE(MD_i) = \sqrt{\frac{SD_{1i}^2}{n_{1i}} + \frac{SD_{2i}^2}{n_{2i}}}$$

where MD= mean difference

1.3.3.3 Estimating pooled summary statistic

Whether OR or MD, the pooled summary estimate is calculated using the following formula:

$$\theta_{pooled} = \frac{\sum w_i \theta_i}{\sum w_i}$$

where θ denotes the summary statistic and w denotes the study weight.

The weights applied to each study depend on either a fixed- or a random-effects model. Next, I will describe these two models.

1.3.3.4 Fixed/Random effects model

A fixed effects model assumes that the true summary statistic from each study is the same. It will therefore weight the studies based on the standard error of each study. A random effects model assumes that studies estimate effects vary around a central estimate (following a normal distribution). The model incorporates an estimate of 'between study variation' into the

calculation of the pooled summary statistic. There are several methods for carrying out a fixed or random effects meta-analysis e.g.:

For fixed effects, dichotomous outcomes:

Mantel-Haenszel weight: $$w_i = \frac{b_i c_i}{N_i}$$

For fixed effects, continuous outcomes:

Inverse-Variance (IV) weight: $$w_i = \frac{1}{SE(\theta_i)^2}$$

For random effects, for both continuous and dichotomous outcomes:

DerSimonian & Laird (DL) weight: $$DL\ w_i = \frac{1}{SE(\theta_i)^2 + \tau^2}$$

where τ^2 is the variance between study estimates, and so takes into account the heterogeneity of studies.

These popular meta-analysis methods have been developed for the comparison of two groups. In studies where there are more than two groups, as is common in genetic association studies, these methods are limiting and so need to be expanded [Attia *et al.*, 2003; Salanti *et al.*, 2005]. This topic will be the focus of chapter 2 of this thesis.

1.4 Aims of Thesis

In this thesis I present the results from two systematic reviews and meta-analyses on the association between two quantitative stroke-related traits (CIMT and WMH) and commonly studied genes. I discuss the development of meta-analysis methods for genetic studies where there are generally three groups and a genetic model must be selected and use these methods where appropriate in my reviews. Following on from these reviews, I also report the results of a study attempting to test the association of the apolipoprotein E genotype with large artery compared with small artery stroke in a large hospital-based cohort of stroke patients, the Edinburgh Stroke Study.

SECTION A

2 Development of Genotype-Quantitative Trait Association Meta-Analysis Method

(Development of meta-analysis method)

In this chapter I discuss the limitations and problems with existing meta-analytical methods and I attempt to overcome these with a novel three-step approach. I devised and present three possible methods for determining the genetic model (step 2) and using test datasets discuss the merits and weaknesses of each. I finally present the chosen method to be used in chapter 3.

2.1 Introduction

Most traditional meta-analyses compare outcomes between two groups of patients (e.g. treatment and control arms of a randomized controlled trial) and so the most widely available statistical methods and software packages (e.g. RevMan [The Cochrane Collaboration, 2006]) have been designed to deal with this data structure. However, genetic association studies (along with some other observational studies) usually have more than two comparison groups. For the simplest genetic mutation (a to A) each individual will have one of 3 genotypes (aa, aA or AA). Therefore, traditional meta-analysis statistical methods and software packages that compare two groups are insufficient. As my PhD involves meta-analyses of intermediate phenotypes for stroke I will focus on the methods that allow meta-analysis of a continuous outcome where there are three genotypes (the 'trait increasing' allele denoted A).

2.2 Existing Methods

The problem of how to deal with genetic data when carrying out a meta-analysis has been addressed in the literature with a variety of methods. Most collapse the data into two groups and use the traditional methods and software. Methods particular to genetic meta-analysis have also been developed recently which analyse the data as three separate groups.

2.2.1 Methods Which Collapse Data into Two Groups

Method (i) Assume a dominant or recessive genetic model. Collapse the three genetic groups into two, based on an assumed genetic

model. The heterozygote Aa individuals are grouped with the AA or aa individuals depending on which model is adopted, e.g. [Juo *et al.,* 1999].

Dominant:	AA and Aa	compared to	aa
Recessive:	AA	compared to	Aa and aa

Method (ii) Compare the two extreme genotypes (AA and aa). This may be done when the underlying genetic model is thought to be co-dominant and so method (i) would be inappropriate, e.g. [Sayed-Tabatabaei *et al.,* 2003]

Method (iii) Often, if the genetic model is not known, multiple comparisons are made. Either the data are analysed according to multiple models (both recessive and dominant), or several comparisons between the individual genotypes are made (e.g. AA versus Aa and AA versus aa), e.g. [Rantala *et al.,* 2000]

A systematic review of all meta-analyses of genetic association studies up to August 2000 [Attia *et al.,* 2003] found that five out of the seven continuous outcome meta-analyses used multiple comparisons (method (iii)) to analyse the association and most failed to account for this multiple testing. The other two studies assumed a genetic model (method (i)). One gave explicit biological reasons for using the assumed model and the other gave no reason.

Adopting a certain genetic model is only appropriate if there is sufficient evidence to show that the correct one has been chosen. Using biological evidence is sensible if the evidence relates to the trait being studied. If there is evidence that a mutation works in a recessive way on one trait, this does not necessarily mean its influence on another trait is recessive. Often, when carrying out a meta-analysis one is constrained to using a particular genetic model, because many of the papers have presented their data according to this model and selecting this model generates the largest and most complete dataset possible. Providing the individual papers have chosen this genetic model sensibly, this model should be the most appropriate.

Comparing the two extremes (method (ii)) may show the largest difference, but removing the Aa genotype group (which is often much larger than the AA group) will reduce statistical power to detect an association.

2.2.2 Methods Which Analyse as Three Groups

Some methods have been developed that analyse the data as three groups. These either use a per-allele (co-dominant model), or attempt to analyse the data without assuming a particular genetic model. The methods are as follows:

Method (iv) Use a per-allele method, which assumes a co-dominant model. Ye *et al*. [2006] estimate a per-allele odds ratio using logistic regression. This method could be adapted to analyse continuous data and report the average mean difference between genotypes AA and Aa, and genotypes Aa and aa.

Method (v) A Bayesian model-free approach has been described for dichotomous outcomes [Minelli *et al.*, 2005]. This method is based around odds ratios but could be extended to analyse an association for a continuous outcome. It works on the basis that in a simple bi-allelic situation there are two odds ratios to be estimated; Aa compared with aa (OR_{Aa}) and AA compared with aa (OR_{AA}). The relationship between these two odds ratios is dependent on the genetic model. The method treats the log odds ratio of Aa versus aa ($\log OR_{Aa}$) as an unknown proportion (λ) of the log odds ratio of AA versus aa ($\log OR_{AA}$).

ie. $\lambda = \dfrac{\log OR_{Aa}}{\log OR_{AA}}$ and thus $OR_{Aa} = [OR_{AA}]^{\lambda}$

λ values of 0, 0.5 and 1 correspond to recessive, co-dominant and dominant respectively, but λ is allowed to take any value between 0 and 1. The study-specific $\log OR_{AA}$ is modelled as a normally distributed random effects parameter and λ is modelled as a fixed parameter. The study specific $\log OR_{Aa}$ is equal to the product of λ and the study-specific $\log OR_{AA}$. By estimating the log ORs and λ, this approach provides information of the genetic magnitude of the effect as well as the mode of inheritance. But, its results can be difficult to interpret as they depend on priors and the methods are inaccessible to those unfamiliar with Bayesian analysis.

Method (vi) A frequentist genetic model-free approach is described by Thakkinstian *et al*. [2005]. In this method, analysis of variance (ANOVA) is used to test for an overall association between a gene and a trait, but no estimation of λ or size of effect is made. The genetic-model-free ANOVA approach above is useful as it tests for an association without making any assumptions about the underlying genetic model. However, once an overall association has been found, it is still necessary to investigate the association further. The ANOVA result does not tell you which of the genotypes is associated with an increase or decrease in the trait or the size of any effect. After establishing an overall association a method is still required to determine which genetic model is appropriate, so that the data can then be analysed according to that model to determine the effect size.

2.3 Aim

I aimed to devise a new, easy to use method that deals with three comparison groups in a meta-analysis of the association with a continuous trait. By using the relationship between two mean differences to estimate λ (similar to that described in method (v)) the best genetic model (recessive, dominant or co-dominant) can be chosen. There are many ways λ can be estimated. I have devised and investigated three different methods of calculating λ and used real data to test these methods. However, as I will show, estimating λ is meaningless and often misleading if there is no

underlying association and so ANOVA provides a useful tool for establishing if there is any association to begin with.

First, I describe the ANOVA method, then I describe three different λ estimation methods, and then the meta-analysis mean difference method for the three genetic models. Finally I examine the three different λ estimation methods using a dataset that gives conflicting results between methods, discuss the methods and choose the most appropriate λ estimation method for my novel, simple, three-step approach for estimating the pooled association between a genetic polymorphism and a continuous trait.

To test these methods I used seven of the datasets collected for my carotid intima-media thickness (CIMT) meta-analyses (chapter 3). These datasets are shown in table 2.1. The number of studies in each dataset varies between 3 and 34. I carried out all analyses in Stata (version 7.0, [StataCorp., 2001]) and so provide instructions and code for this software package.

2.4 Methods Tested

2.4.1 Meta-ANOVA

Meta-ANOVA can be used to test for an overall association between genotype and a trait. By modelling 'genotype' and 'study' as independent categorical variables the between study differences can be accounted for and 'genotype' can be tested to see if it is a significant determinant of 'trait'. The ANOVA is weighted using (1/[standard error of the trait mean]^2), allowing larger and more precise studies to be weighted more heavily than small

Table 2.1 7 meta-analysis test datasets, studying the association between a gene and a trait. Each study reported 3 genotypes: sample size (n), mean value of the trait (mean) and standard deviation (SD). Data were taken from the CIMT datasets collected (chapter 3).

		genotype 1			genotype 2			genotype 3		
Gene	study	n	mean	SD	n	mean	SD	n	mean	SD
A	1	158	1.24	0.44	179	1.27	0.52	38	1.29	0.36
	2	262	0.90	0.23	215	0.92	0.24	42	0.91	0.18
	3	244	0.92	0.21	220	0.96	0.31	41	0.88	0.17
	4	43	0.37	0.07	57	0.35	0.08	18	0.45	0.13
	5	73	0.70	0.43	46	0.92	0.47	8	1.10	0.51
	6	89	0.53	0.38	72	0.55	0.34	6	0.60	0.42
	7	59	0.98	0.10	56	1.07	0.23	16	1.16	0.36
	8	1218	0.79	0.35	1013	0.79	0.32	217	0.81	0.29
B	1	23	0.75	0.19	88	0.68	0.19	76	0.74	0.17
	2	23	1.02	0.20	124	1.06	0.30	93	1.05	0.40
	3	18	1.14	0.40	47	0.94	0.28	36	0.81	0.28
	4	103	0.81	0.18	264	0.80	0.15	148	0.83	0.19
	5	16	0.7	0.08	70	0.76	0.08	46	0.78	0.07
	6	87	0.63	0.16	116	0.63	0.17	32	0.64	0.17
	7	147	0.99	0.36	149	1.06	0.54	60	1.20	0.59
	8	88	0.72	0.15	256	0.73	0.16	151	0.73	0.15
	9	31	0.62	0.15	55	0.63	0.18	62	0.63	0.13
	10	57	0.54	0.10	165	0.54	0.13	118	0.55	0.11
	11	33	0.52	0.04	80	0.54	0.06	37	0.53	0.04
	12	70	0.63	0.12	150	0.62	0.1	135	0.64	0.12
	13	42	1.04	0.23	100	1.08	0.33	77	1.01	0.19
	14	228	0.71	0.14	535	0.71	0.14	343	0.71	0.15
	15	7	0.71	0.05	22	0.76	0.09	22	0.8	0.10
	16	8	0.72	0.13	25	0.71	0.12	14	0.75	0.16
	17	65	0.74	0.27	69	0.76	0.307	41	0.88	0.35
	18	19	1.01	0.29	46	1.10	0.25	33	1.06	0.26
	19	1540	0.87	0.28	1640	0.87	0.29	477	0.87	0.26
	20	83	0.79	0.12	95	0.80	0.14	27	0.81	0.14
	21	116	0.56	0.16	180	0.59	0.23	84	0.58	0.17
	22	29	0.94	0.20	86	0.97	0.20	69	0.98	0.21
	23	29	0.73	0.54	62	0.77	0.55	36	0.91	0.48
	24	29	0.53	0.32	84	0.56	0.37	44	0.61	0.33
	25	30	0.58	0.12	27	0.67	0.11	45	0.79	0.18
	26	18	0.49	0.10	25	0.48	0.10	15	0.46	0.10
	27	35	0.55	0.10	57	0.53	0.10	28	0.57	0.11
	28	38	0.82	0.21	62	0.80	0.19	30	0.81	0.21
	29	1418	0.79	0.15	3264	0.80	0.16	1806	0.80	0.16
	30	10	0.42	0.06	21	0.43	0.07	17	0.40	0.08
	31	16	0.42	0.07	31	0.43	0.09	9	0.48	0.06
	32	8	1.32	0.29	28	1.29	0.33	28	1.29	0.30
	33	39	0.60	0.08	106	0.59	0.10	79	0.57	0.08
	34	25	0.64	0.06	34	0.72	0.05	29	0.78	0.06
C	1	130	0.74	0.23	36	0.67	0.18	7	0.74	0.08
	2	213	0.55	0.16	142	0.57	0.18	25	0.61	0.18
	3	3170	0.77	0.36	1668	0.78	0.44	245	0.76	0.37
D	1	33	0.85	0.23	155	0.97	0.25	66	1.00	0.24
	2	32	1.80	0.10	177	1.84	0.15	45	1.95	0.45
	3	26	0.99	0.50	261	1.04	0.49	62	1.10	0.48
	4	22	0.62	0.17	176	0.63	0.17	33	0.64	0.15
	5	24	0.52	0.05	90	0.54	0.05	30	0.52	0.04
	6	4	1.20	0.60	38	1.10	0.30	10	1.50	0.50

Gene	study	n	mean	SD	n	mean	SD	n	mean	SD
D cont.	7	33	0.59	0.13	200	0.63	0.38	28	0.75	0.35
	8	38	0.54	0.13	208	0.53	0.12	66	0.55	0.11
	9	10	0.72	0.25	77	0.70	0.25	25	0.80	0.25
	10	20	0.95	0.12	109	1.05	0.17	60	1.03	0.16
	11	750	0.75	0.14	3122	0.77	0.14	1392	0.77	0.18
	12	27	0.79	0.12	137	0.79	0.13	38	0.83	0.16
	13	10	0.61	0.15	65	0.64	0.14	20	0.76	0.17
	14	12	0.89	0.18	160	0.9	0.19	86	0.98	0.26
	15	13	0.78	0.15	150	0.93	0.21	90	0.93	0.23
	16	22	0.91	0.37	109	0.99	0.52	31	0.93	0.42
	17	146	0.69	0.13	650	0.72	0.15	283	0.70	0.13
	18	10	0.81	0.17	64	0.88	0.16	18	1.02	0.19
	19	34	0.62	0.12	161	0.68	0.17	58	0.89	0.15
	20	11	0.59	0.11	75	0.60	0.13	20	0.71	0.14
	21	373	0.73	0.16	1782	0.74	0.18	568	0.74	0.20
	22	18	0.76	0.17	158	0.81	0.21	49	0.83	0.23
	23	12	0.60	0.20	92	0.70	0.40	14	0.70	0.50
	24	17	0.86	0.23	89	0.78	0.16	21	0.88	0.30
	25	23	0.89	0.16	120	0.93	0.16	39	0.88	0.16
	26	717	0.71	0.11	3923	0.71	0.13	1124	0.72	0.13
	27	10	0.60	0.15	122	0.65	0.16	31	0.69	0.19
	28	634	0.72	0.13	1427	0.73	0.11	1126	0.75	0.13
	29	1459	0.72	0.12	5534	0.73	0.15	2311	0.74	0.14
	30	4	0.79	0.06	40	0.89	0.14	24	0.99	0.19
	31	4	0.90	0.00	22	0.96	0.17	7	0.87	0.08
	32	59	0.72	0.12	242	0.79	0.21	144	0.74	0.21
E	1	19	1.30	0.42	38	1.09	0.25	35	1.17	0.24
	2	47	0.85	0.17	19	0.78	0.21	5	0.72	0.15
	3	381	0.69	0.15	557	0.70	0.12	171	0.70	0.17
	4	320	0.77	0.14	500	0.77	0.14	180	0.79	0.16
	5	265	0.98	0.29	422	0.98	0.26	136	1.10	0.26
F	1	89	1.31	0.31	94	1.35	0.35	39	1.58	0.54
	2	35	0.83	0.19	38	0.86	0.19	22	0.88	0.20
	3	87	0.79	0.13	54	0.85	0.15	10	0.93	0.07
	4	28	0.98	0.21	72	1.03	0.17	20	1.23	0.18
	5	59	0.64	0.23	111	0.69	0.21	36	0.79	0.30
	6	47	0.86	0.29	55	0.93	0.22	22	0.89	0.22
	7	325	0.73	0.15	304	0.72	0.13	62	0.68	0.11
	8	1197	0.85	0.11	1542	0.86	0.10	508	0.86	0.11
	9	110	0.74	0.18	120	0.73	0.17	30	0.79	0.24
	10	171	0.67	0.13	84	0.67	0.12	20	0.67	0.15
	11	198	0.69	0.17	72	0.69	0.16	13	0.79	0.20
	12	312	1.02	0.16	378	1.02	0.17	125	1.03	0.16
	13	60	0.64	0.14	117	0.67	0.15	24	0.59	0.10
	14	346	0.76	0.20	316	0.75	0.18	52	0.74	0.18
	15	290	0.75	0.23	220	0.75	0.16	31	0.77	0.18
G	1	14	0.89	0.38	63	0.72	0.17	62	0.74	0.18
	2	88	0.76	0.15	91	0.76	0.13	17	0.78	0.18
	3	165	0.51	0.05	169	0.51	0.05	28	0.49	0.03
	4	140	1.14	0.22	111	1.13	0.21	35	1.12	0.23
	5	30	0.65	0.27	77	0.83	0.27	45	1.05	0.32
	6	262	0.88	0.19	198	0.88	0.18	36	0.90	0.17
	7	273	0.87	0.17	198	0.87	0.19	32	0.86	0.23
	8	110	0.88	0.19	146	0.85	0.16	29	0.90	0.17
	9	55	0.51	0.07	66	0.52	0.08	12	0.54	0.09

Table 2.2 Section of data entered into Stata for the meta-ANOVA between gene D and a continuous trait. The data have been rearranged from table 2.1 so that each genotype from each study represents a single observation in the meta-ANOVA and the standard error (SE) has been derived from the data.

study	genotype	n	mean	sd	se
1	1	33	0.85	0.23	0.04
1	2	155	0.97	0.25	0.02
1	3	66	1.00	0.24	0.03
2	1	32	1.80	0.10	0.02
2	2	177	1.84	0.15	0.01
2	3	45	1.95	0.45	0.07
3	1	26	0.99	0.50	0.10
3	2	261	1.04	0.49	0.03
3	3	62	1.10	0.48	0.06
.					
.					
.					
.					
32	3	144	0.74	0.21	0.02

studies and trait estimates with large variances (method derived from Thakkinstian *et al.* [2005]).

Table 2.2 shows a section of the example data for gene D, as entered into Stata. Study number is coded (1-32), there are 3 genotypes per study (coded as 1, 2 and 3) and each genotype has sample size (n), mean value of trait (mean) and standard deviation (SD) (from which the standard error (SE) was derived, from SD/√n).

Stata code:

```
xi: regress mean i.genotype i.study [aweight=1/se^2]
testparm _Igenotype*
```

Table 2.3 meta-ANOVA results for the seven example datasets. p-value for the test of 'genotype' as a significant variable in the model.

Gene	number of studies	p-value for association between genotype and trait
A	10	p=0.26
B	34	p=0.01
C	3	p=0.71
D	32	p<0.001
E	5	p=0.37
F	15	p=0.02
G	9	p=0.58

Table 2.3 shows the results obtained using the seven example datasets. Using this meta-ANOVA approach, three of the example datasets show a significant association (at $p<0.05$) between the trait and genotype (B, D and F). However, the results do not explain the nature of these associations. For these analyses to be biologically informative, we need to know which of the genotypes/alleles are causing an increase in the trait and by how much. So, carrying out a mean difference meta-analysis is still necessary, but we need a method for choosing the most appropriate genetic model.

Thakkinstian *et al*. [2005], who described this meta-ANOVA first stage approach, chose the most appropriate genetic model by calculating pooled mean differences between each pair of genotypes (between groups AA and aa [D_1], Aa and aa [D_2], and groups AA and Aa [D_3]). They then used the following rules to choose the most appropriate genetic model:

" (a) If $D_1 = D_3 \neq 0$ and $D_2 = 0$, then a recessive model is suggested.
(b) If $D_1 = D_2 \neq 0$ and $D_3 = 0$, then a dominant model is suggested.
(c) If $D_2 = -D_3 \neq 0$ and $D_1 = 0$, then a complete over-dominant model is suggested.
(d) If $D_1 > D_2 > 0$ and $D_1 > D_3 > 0$ (or $D_1 < D_2 < 0$ and $D_1 < D_3 < 0$, then a co-dominant model is suggested."

(NB. a complete over-dominant model occurs when the two homozygotes have equal values and the heterozygotes have a different value)

This method is limited as it is extremely unlikely that two mean differences would be exactly equal to each other, even if in the underlying model they were. There is no measurement of the error in this method of simply comparing mean differences and it requires a subjective judgement to be made on which model is most appropriate.

Typical post-hoc tests for ANOVA (e.g. Newman-Keuls, Tukey and Scheffe tests) may identify differences between the two extreme groups, but might not resolve what should be done with the intermediate group, and so are not useful when trying to deduce the genetic model. So, I investigated several alternative methods for choosing a genetic model.

2.4.2 Choosing a Genetic Model

For continuous traits, the genetic model can be described by the relationship:

$$\lambda = \frac{MD1}{MD2}$$

where

MD1 = the mean trait difference between Aa and aa, and

MD2 = the mean trait difference between AA and aa

i.e. MD1 is the effect of having one 'trait increasing' allele and MD2 is the effect of having two 'trait increasing' alleles. The ratio between these two relates to the genetic model.

λ takes theoretical values depending on the underlying genetic model: 0=recessive; 1=dominant; 0.5=co-dominant. All methods are based on this relationship but there are several different ways λ can be estimated. Here I describe and test three different methods.

2.4.2.1 Method 1

Calculate pooled MD1 and MD2 using traditional meta-analysis methods. I used both random and fixed effects analyses in STATA. The random effects analysis uses the DerSimonian & Laird method and the fixed effects analysis uses the Mantel-Haenszel method.

Stata code for random effects:

pooled MD1 **:**

```
metan nAa xAa sdAa naa xaa sdaa, random nostandard
```

pooled MD2:

```
metan nAA xAA sdAA naa xaa sdaa, random nostandard
```

Stata code for fixed effects:

pooled MD1:

```
metan nAa xAa sdAa naa xaa sdaa, fixed nostandard
```

pooled MD2:

```
metan nAA xAA sdAA naa xaa sdaa, fixed nostandard
```

An overall estimate of λ can then be calculated:

λ = (pooled MD1) / (pooled MD2)

Table 2.4 shows the estimation of both random effects and fixed effects pooled MD1 and MD2 for gene D. These pooled mean differences are then used to estimate λ:

'random effects' $\lambda = 0.025 / 0.050 = 0.50$

'fixed effects' $\lambda = 0.013 / 0.021 = 0.62$

which both suggest a co-dominant genetic model. The estimates of λ are both close to 0.5, but there is no estimation of the error.

Table 2.4 Results of MD1 and MD2 meta-analyses for example dataset D.

study	**MD1**	95% confidence interval (CI)	random weight (%)	fixed weight (%)	**MD2**	95% confidence interval (CI)	random weight (%)	fixed weight (%)
1	**0.120**	0.032 to 0.208	1.04	0.22	**0.150**	0.052 to 0.248	2.05	0.25
2	**0.040**	-0.001 to 0.081	3.63	0.98	**0.150**	0.014 to 0.286	1.21	0.13
3	**0.050**	-0.149 to 0.249	0.22	0.04	**0.103**	-0.122 to 0.328	0.49	0.05
4	**0.007**	-0.069 to 0.083	1.34	0.29	**0.016**	-0.072 to 0.104	2.37	0.30
5	**0.020**	-0.003 to 0.043	7.12	3.27	**0.000**	-0.025 to 0.025	6.58	3.87
6	**-0.100**	-0.696 to 0.496	0.02	0.00	**0.300**	-0.365 to 0.965	0.06	0.01
7	**0.040**	0.040 to -0.029	1.60	0.35	**0.160**	0.023 to 0.297	1.20	0.12
8	**-0.003**	-0.048 to 0.042	3.15	0.81	**0.009**	-0.041 to 0.059	4.51	0.95
9	**-0.020**	-0.185 to 0.145	0.31	0.06	**0.080**	-0.103 to 0.263	0.72	0.07
10	**0.100**	0.038 to 0.162	1.94	0.44	**0.080**	0.014 to 0.146	3.39	0.53
11	**0.020**	0.009 to 0.031	10.32	13.32	**0.020**	0.006 to 0.034	7.33	12.32
12	**0.000**	-0.050 to 0.050	2.70	0.66	**0.040**	-0.028 to 0.108	3.29	0.50
13	**0.030**	-0.069 to 0.129	0.83	0.17	**0.150**	0.031 to 0.269	1.51	0.16
14	**0.010**	-0.096 to 0.116	0.73	0.15	**0.090**	-0.026 to 0.206	1.58	0.17
15	**0.150**	0.062 to 0.238	1.03	0.21	**0.150**	0.056 to 0.244	2.15	0.26
16	**0.080**	-0.103 to 0.263	0.26	0.05	**0.020**	-0.194 to 0.234	0.54	0.05
17	**0.030**	0.006 to 0.054	6.73	2.87	**0.010**	-0.016 to 0.036	6.47	3.47
18	**0.070**	-0.042 to 0.182	0.66	0.13	**0.210**	0.073 to 0.347	1.19	0.12
19	**0.060**	0.012 to 0.108	2.88	0.72	**0.270**	0.214 to 0.326	4.05	0.75
20	**0.010**	-0.061 to 0.081	1.51	0.33	**0.120**	0.031 to 0.209	2.32	0.29
21	**0.010**	-0.008 to 0.028	8.28	4.97	**0.010**	-0.013 to 0.033	6.70	4.38
22	**0.050**	-0.035 to 0.135	1.10	0.23	**0.070**	-0.032 to 0.172	1.93	0.23
23	**0.100**	-0.040 to 0.240	0.43	0.09	**0.100**	-0.185 to 0.385	0.31	0.03
24	**-0.080**	-0.194 to 0.034	0.64	0.13	**0.020**	-0.149 to 0.189	0.83	0.08
25	**0.042**	-0.029 to 0.113	1.51	0.33	**-0.015**	-0.097 to 0.067	2.60	0.34
26	**-0.001**	-0.010 to 0.008	10.94	21.64	**0.009**	-0.002 to 0.020	7.48	19.78
27	**0.050**	-0.047 to 0.147	0.86	0.18	**0.090**	-0.025 to 0.205	1.60	0.18
28	**0.012**	0.001 to 0.023	10.25	12.70	**0.024**	0.011 to 0.307	7.40	14.85
29	**0.010**	0.003 to 0.017	11.31	33.02	**0.020**	0.012 to 0.028	7.59	33.75
30	**0.099**	0.024 to 0.174	1.39	0.30	**0.198**	0.100 to 0.296	2.04	0.24
31	**0.060**	-0.010 to 0.130	1.55	0.34	**-0.033**	-0.094 to 0.028	3.73	0.63
32	**0.070**	0.030 to 0.110	3.71	1.01	**0.020**	-0.026 to 0.066	4.79	1.11
pooled random	**0.025**	0.015 to 0.034			**0.050**	0.033 to 0.066		
pooled fixed	**0.013**	0.009 to 0.017			**0.021**	0.016 to 0.026		

2.4.2.2 Method 2

Calculate λ for each study (using MD1/MD2) and weight each study to obtain an overall estimate of λ across all studies.

The SEs for the two mean differences in a single study were averaged to obtain a study estimate of SE. Weighting of the studies was then done using inverse variance ($1/SE^2$).

Table 2.5 Results of λ estimation for example dataset, D. λ is calculated for each study and weighted by 1/(the mean standard error of the two mean differences)^2, to obtain the pooled estimate.

	MD1	MD2	λ (MD1/MD2)	weight	weight (%)
1	0.120	0.150	0.80	591	0.3
2	0.040	0.150	0.27	473	0.2
3	0.050	0.103	0.49	152	0.1
4	0.007	0.016	0.44	869	0.4
5	0.020	0.000	-	-	-
6	-0.100	0.300	-0.33	21	0.0
7	0.040	0.160	0.25	364	0.2
8	-0.003	0.009	-0.33	2560	1.2
9	-0.020	0.080	-0.25	210	0.1
10	0.100	0.080	1.25	1208	0.5
11	0.020	0.020	1.00	30416	13.8
12	0.000	0.040	0.00	1341	0.6
13	0.030	0.150	0.20	472	0.2
14	0.010	0.090	0.11	742	0.3
15	0.150	0.150	1.00	895	0.4
16	0.080	0.020	4.00	114	0.1
17	0.030	0.010	3.00	7899	3.6
18	0.070	0.210	0.33	353	0.2
19	0.060	0.270	0.22	1639	0.7
20	0.010	0.120	0.08	789	0.4
21	0.010	0.010	1.00	10914	4.9
22	0.050	0.070	0.71	654	0.3
23	0.100	0.100	1.00	90	0.0
24	-0.080	0.020	-4.00	255	0.1
25	0.042	-0.015	-2.80	881	0.4
26	-0.001	0.009	-0.11	50045	22.7
27	0.050	0.090	0.56	618	0.3
28	0.012	0.024	0.50	30390	13.8
29	0.010	0.020	0.50	72503	32.9
30	0.099	0.198	0.80	548	0.2
31	0.060	-0.033	0.27	601	0.3
32	0.070	0.020	0.49	2092	0.9
pooled			0.54		

Pooled λ = 0.54, suggesting a co-dominant genetic model. But, there is no estimation of the error.

2.4.2.3 Method 3

This final method aims to provide an estimate of λ and also provide a measurement of its precision. Weighted linear regression is used to estimate

λ with MD1 (from each study) as the dependent variable and MD2 (from each study) as the independent variable, restricted to pass through (0,0) and with each study weighted by $1/(SE)^2$ (where the study SE is estimated as in method 2 – the mean of the SEs from the two MDs).

The slope of the linear regression line represents λ and a 95% confidence interval (CI) of this estimate can be calculated.

As can be seen from figure 2.1 the theoretical λ values of 0, 0.5 and 1 correspond to three lines on the graph. So as well as estimating λ and obtaining a 95% CI, the graph provides a useful visual representation of the data. By plotting MD1 against MD2 and sizing the points according to the weight of each study, the fit of the data to the estimated model can be viewed.

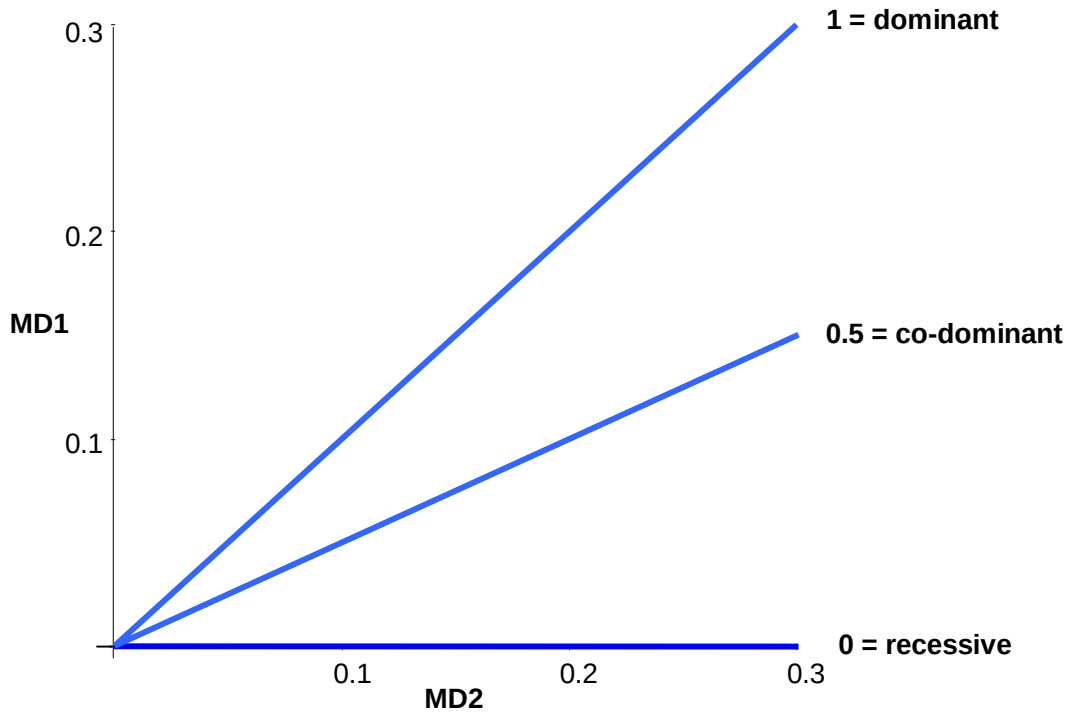


Figure 2.1 Plot of MD1 against MD2. A gradient of 0 represents a recessive genetic model, 0.5 represents a co-dominant model and 1 represents a dominant model.

After MD1 and MD2 for each study have been calculated the gradient of the regression line can be calculated.

Stata code:

```
regress MD1 MD2 [aweight=1/se^2], noconstant
```

Figure 2.2 shows the example of gene D. The estimated gradient of the regression line is 0.42 with a 95% CI of 0.27 to 0.57, suggesting that a co-dominant model is appropriate.

2.4.3 Mean Difference Meta-Analysis using Chosen Genetic Model

Once the most appropriate genetic model has been selected the corresponding pooled mean difference can be estimated. If a dominant or recessive model is selected then two of the three genotypes are combined and compared to the third genotype, using traditional meta-analysis methods for comparing two groups. If a co-dominant model is selected then the three genotypes are compared and the average per-allele mean difference is calculated.

Co-dominant: average of AA – Aa and Aa – aa
Dominant: AA,Aa – aa
Recessive: AA – Aa,aa

The analyses were carried out in Stata using the 'gametan' command designed by Julian Higgins [J Higgins, personal communication, Nov,2006].

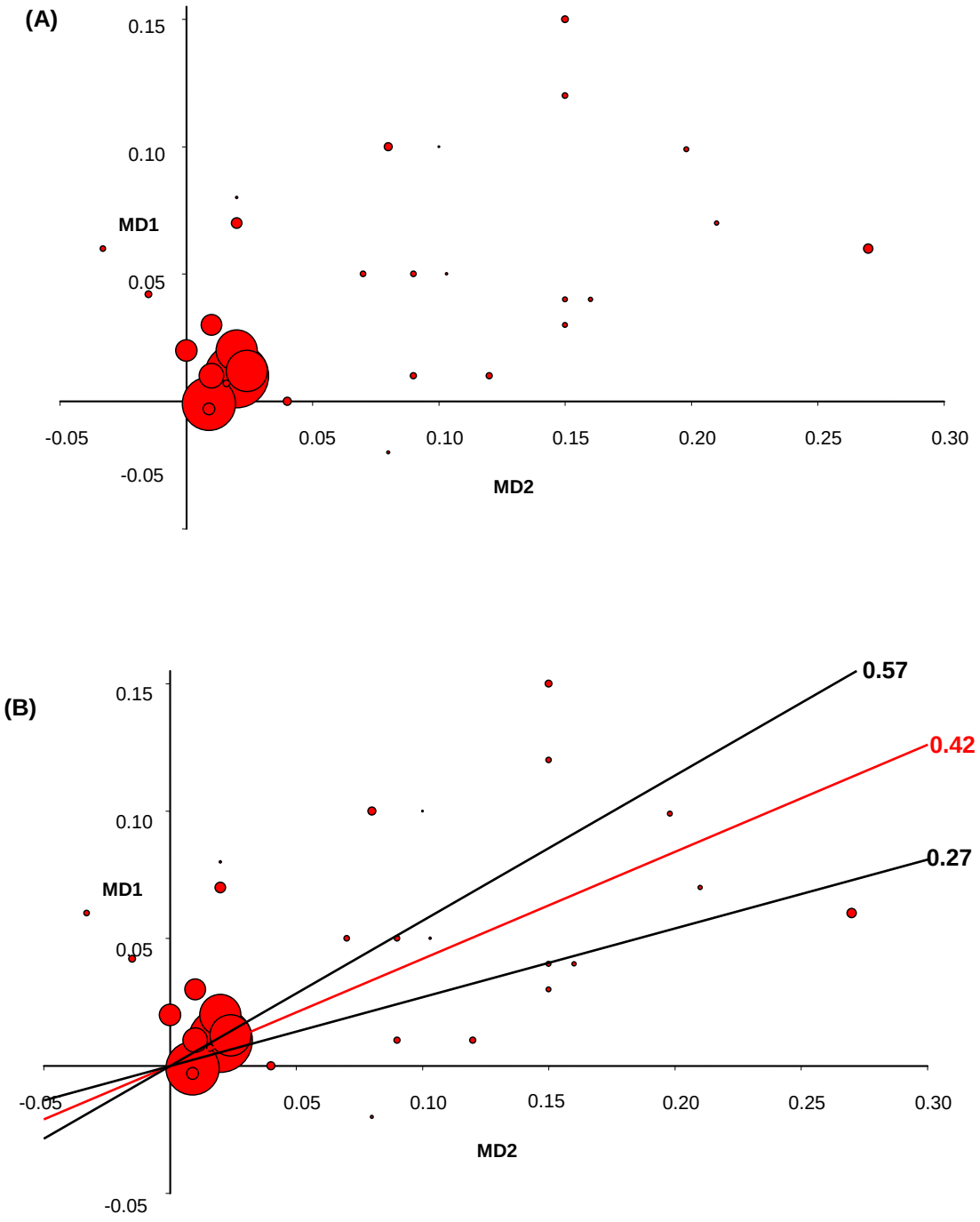


Figure 2.2 (A) Plot of MD1 against MD2 for example dataset D. The bubbles represent each study, with size proportional to weight. **(B)** Plot showing λ. The red line is the weighted regression line, with a gradient (λ) of 0.42. The black lines represent the 95% confidence limits for λ (0.27 to 0.57).

Stata code:

Co-dominant:

```
gametan AAn AAx AAsd Aan Aax Aasd aan aax aasd, codominant
```

Dominant:

```
gametan AAn AAx AAsd Aan Aax Aasd aan aax aasd, dominant
```

Recessive:

```
gametan AAn AAx AAsd Aan Aax Aasd aan aax aasd, recessive
```

The recessive and dominant analyses essentially collapse the data into two groups and carry out standard 'metan' analyses .

The co-dominant (per-allele) mean difference for gene D was 0.025 (95% CI 0.017 to 0.033) using random effects meta-analysis. Each step change (from aa to Aa, and from Aa to AA) corresponds to an increase in the trait of 0.025 units.

2.5 Results Using Test Data

Table 2.6 shows the results of all methods described in this chapter for all seven test datasets. Only genes B, D and F showed a significant association with the trait on ANOVA. The genes that did not show a significant association with the trait in the ANOVA analysis often had conflicting λ estimates using the three methods and a large confidence interval for method 3.

Table 2.6 Results of the meta-ANOVA method, three λ estimation methods and the mean difference method for the 7 example datasets. *denotes the ANOVA results that showed a significant association between genotype and trait (p<0.05), **gene is associated but λ estimation methods give conflicting genetic models.

	number of studies	meta-ANOVA (p-value)	λ method 1 (random)	λ method 1 (fixed)	λ method 2	λ method 3	genetic model chosen	Random effects pooled mean difference (95% CI)
A	8	0.26	0.79	0.31	-0.08	0.23 (-0.15 to 0.61)	no association	-
B	34	0.01*	0.46	0.67	0.70	0.52 (0.41 to 0.62)	co-dominant	0.014 (0.005 to 0.022)
C	3	0.71	0.08	0.75	-2.05	0.31 (-2.26 to 2.87)	no association	-
D	32	<0.001*	0.50	0.62	0.54	0.42 (0.27 to 0.57)	co-dominant	0.025 (0.017 to 0.033)
E	5	0.37	0.00	0.13	0.47	0.22 (-0.35 to 0.79)	no association	-
F	15	0.02*	0.10	0.63	0.79	0.23 (0.08 to 0.39)	**	-
G	9	0.58	0.03	0.00	0.01	0.38 (0.14 to 0.62)	no association	-

2.5.1 Genes B&D

All λ estimation methods suggest that co-dominant genetic models are the most appropriate for both B and D genes and the 95% CIs for method 3 only span one genetic model (0.5, co-dominant). Both genes showed a significant association with the trait in the meta-ANOVA analyses (p=0.01 and <0.001 respectively), suggesting that there is an association between genotype and trait for both of these genes. Using the most appropriate genetic model, co-dominant, the per-allele pooled mean difference was estimated to be 0.014 (95% CI, 0.005 to 0.022) for gene B and 0.025 (95% CI, 0.017 to 0.033) for gene D.

2.5.2 Genes A, C, E & G

The three methods for genotype model estimation for genes A, C, E and G gave conflicting and often nonsensical results (e.g. -2.05 for gene C, method

2) and the 95% CIs for method 3 were very large, spanning multiple genetic models (e.g. -0.35 to 0.79 for gene E) . However, none of these genes showed a significant association with trait in the meta-ANOVA analyses (p=0.26, 0.71, 0.37, 0.58, respectively), and therefore estimating λ is essentially a meaningless task.

2.5.3 Gene F

Gene F was a significant variable in the meta-ANOVA analysis (p=0.02), suggesting an association between genotype and the trait. Despite this, the three λ estimation methods gave different results. Method 1 using random effects (λ=0.10) suggests a recessive genetic model, method 1 with fixed effects (λ=0.63) suggests a co-dominant genetic model, method 2 (λ=0.79) suggests a dominant model and the 95% CI for method 3 does not span any genetic model (0.08 to 0.39). In the next section I use this gene to compare the different λ estimation methods and explain why they give different results, as well as choosing which method to use in my later analyses (chapter 3).

2.6 Which Genetic Model Method to Use

As gene F gave conflicting results for the various genetic model estimation methods, I used this gene to compare methods and explain why they give different results.

Table 2.7 Results of MD1 and MD2 meta-analyses for example dataset F.

study	**MD1**	95% CI	random weight (%)	fixed weight (%)	**MD2**	95% CI	random weight (%)	fixed weight (%)
1	**0.040**	-0.056 to 0.136	0.81	0.45	**0.270**	0.089 to 0.451	2.52	0.26
2	**0.029**	-0.058 to 0.116	0.98	0.54	**0.042**	-0.061 to 0.145	5.19	0.81
3	**0.060**	0.012 to 0.108	3.04	1.75	**0.140**	0.089 to 0.191	8.33	3.25
4	**0.050**	-0.037 to 0.137	0.97	0.54	**0.250**	0.139 to 0.361	4.80	0.70
5	**0.050**	-0.021 to 0.121	1.47	0.82	**0.150**	0.036 to 0.264	4.64	0.66
6	**0.070**	-0.031 to 0.171	0.72	0.40	**0.030**	-0.094 to 0.154	4.23	0.56
7	**-0.012**	-0.034 to 0.010	11.93	8.14	**-0.050**	-0.082 to -0.018	9.47	8.12
8	**0.007**	0.005 to 0.009	39.03	62.01	**0.006**	0.003 to 0.009	10.30	66.68
9	**-0.017**	-0.064 to 0.029	3.30	1.90	**0.043**	-0.050 to 0.136	5.70	0.98
10	**-0.006**	0.038 to 0.026	6.54	4.01	**-0.008**	-0.079 to 0.062	7.08	1.73
11	**-0.004**	0.047 to 0.039	3.78	2.20	**0.091**	-0.022 to 0.203	4.73	0.68
12	**0.000**	0.025 to 0.025	10.22	6.73	**0.010**	-0.023 to 0.043	9.43	7.76
13	**0.024**	-0.021 to 0.069	3.50	2.03	**-0.055**	-0.108 to -0.002	8.19	2.99
14	**-0.010**	-0.039 to 0.019	7.79	4.89	**-0.020**	-0.073 to 0.033	8.20	3.01
15	**0.000**	-0.034 to 0.034	5.90	3.57	**0.020**	-0.049 to 0.089	7.19	1.81
Pooled random effects	**0.004**	-0.004 to 0.013			**0.039**	0.006 to 0.072		
Pooled fixed effects	**0.005**	-0.001 to 0.011			**0.008**	-0.001 to 0.017		

2.6.1 Results for gene F

2.6.1.1 Method 1

Table 2.7 shows the results for gene F.

random effects analysis $\lambda = 0.004 / 0.039 = 0.10$

fixed effects analysis $\lambda = 0.005 / 0.008 = 0.63$

This method is essentially the same as that by Thakkinstian *et al*. [2005] described in section 2.4.1, except here I go further than just calculating the MDs and observing similarities and differences. I actually calculate a ratio of these mean differences.

This method gives an estimate of λ but does not provide a measure of error for λ. As the MD1 and MD2 estimates both include aa individuals, but one also includes Aa individuals, while the other also includes AA individuals, typical properties of variance cannot be applied to estimate a 95% CI for λ. It is extremely unlikely that λ would be estimated to be exactly 0, 0.5, or 1 and so the nearest of these is the best guess, but with no measure of error we cannot tell how accurate the estimates are. However, from simply observing the very wide 95% CIs of MD1 and MD2 for both random and fixed analyses, it is clear that the error around the estimate of λ would be very large.

I used both random and fixed effects for this method as it is unclear which is the most appropriate. Although one would expect λ to be a 'fixed' parameter (i.e. the genetic model will be the same in all populations), the MDs may be either fixed or random (i.e. the effect size of the genotype on the trait may differ between populations). The random and fixed effects analyses gave very different results. The random effects analysis λ estimation is closest to a recessive genetic model (λ=0.10) and the fixed effects analysis λ estimation is closest to a co-dominant model (λ=0.63). This discrepancy arises because the two analyses weight the studies differently. The fixed effects analysis gives much larger weighting to study 8 (which has a much smaller MD2 compared to the other studies), due to the much smaller variance in this study, whilst the random effects analysis weights studies with smaller variances less, and so is more influenced by (larger) MD2 estimates from other studies.

There is a problem with both of these analyses, and this problem explains the discrepancy between the random and fixed effects analyses. When MD1 and

MD2 are pooled separately across studies, the within-study comparisons are broken, and imbalances in the data (such as differing sizes of MDs between studies) results in bias. This is similar to Simpson's paradox, which shows that something true of each subgroup, need not be true of the whole population [Altman & Deeks, 2002]. MD1 and MD2 being pooled separately results in studies being given different weights for the two MDs. For example, in study 8, in the random effects analysis, the weight for MD1 is 39% and the weight for MD2 is 10%. This study has particularly small MDs for both comparisons. So by weighting the MDs differently MD2 becomes falsely inflated compared to MD1, resulting in a low estimation of λ. This outcome is particularly enhanced in the random effects analysis compared to the fixed effects, as the fixed effects analysis appears to weight the studies more evenly. Even so, it is clear that calculating pooled MDs separately and then calculating λ is inherently flawed and other methods that maintain the within-study comparison would be more appropriate.

2.6.1.2 Method 2

Here I estimate λ for each study and weight across studies. This method was the simplest I could think of to maintain the within-study comparison of the data whilst estimating an overall estimate of λ.

Table 2.8 Results of λ estimation for example dataset, F. λ is calculated for each study and weighted by 1/(the average standard error of the two mean differences)^2, to obtain the pooled estimate.

	MD1	MD2	λ (MD1/MD2)	weight	weight (%)
1	0.040	0.270	0.15	249	0.34
2	0.029	0.042	0.69	443	0.60
3	0.060	0.140	0.43	1989	2.69
4	0.050	0.250	0.20	446	0.60
5	0.050	0.150	0.33	509	0.69
6	0.070	0.030	2.33	314	0.42
7	-0.012	-0.050	0.24	6349	8.57
8	0.007	0.006	1.17	44936	60.67
9	-0.017	0.043	-0.40	1110	1.50
10	-0.006	-0.008	0.74	3034	4.10
11	-0.004	0.091	-0.05	1933	2.61
12	0.000	0.010	0.00	5147	6.95
13	0.024	-0.055	-0.44	1694	2.29
14	-0.010	-0.020	0.50	3488	4.71
15	0.000	0.020	0.00	2427	3.28
pooled			0.79		

The results for gene F using this method are shown in table 2.8. This method suggests a dominant model (λ=0.79). The between study estimates vary quite considerably (from -0.44 to 2.33). The extreme estimates tend to come from small studies that have large standard errors of trait means and so are weighted less than larger studies.

This method weights study 8 much more than any other (weight = 61% for this study), resulting in the estimate of λ being close to 1.

Again, this method does not provide any measurement of precision for the estimate of λ.

2.6.1.3 Method 3

Figure 2.3 shows the example of gene F. The estimated gradient of the regression line (and hence the estimation of λ) is 0.23 with a 95% CI of 0.08 to 0.39. For this particular example the 95% CI does not include any of the three models, suggesting that none of recessive, dominant or co-dominant are appropriate. However, as the estimate is closer to 0 than 0.5 or 1, I chose to perform meta-analysis for this gene using a recessive genetic model.

Method 3 gives a different result to method 2 because a regression method will be more influenced by studies with a large MD1, MD2 or both, even though the two methods use the same weights. Method 2 is heavily influenced by study 8, which has very small MD1 and MD2. The MDs for study 8 are MD1=0.007 and MD2=0.006. Method 2 would have given the same result if the MDs were of a greater magnitude (e.g. 0.7 and 0.6 -more convincing evidence for a λ of 1.2), but for method 3 the study would have had more influence if the MDs had been 0.7 and 0.6. A study having more influence if it shows larger differences seems more appropriate. Observing figure 2.3A it seems that a linear regression line λ=0 is more sensible than λ=1. In addition, method 3 provides an estimate of the 95% CI surrounding the estimation of λ, which is useful in determining how precise it is.

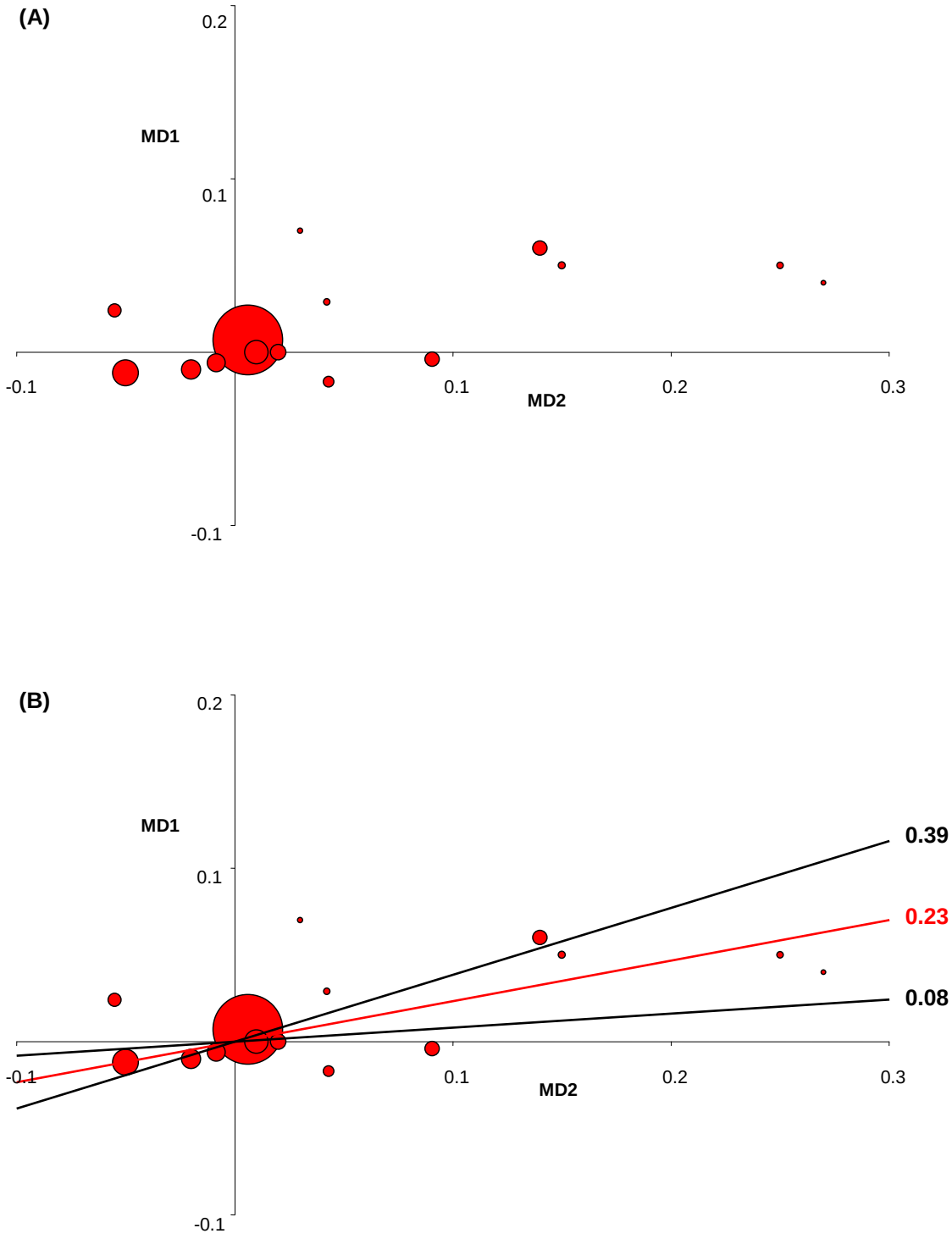


Figure 2.3 (A) Plot of MD1 against MD2 for example dataset F. The bubbles represent each study, with size proportional to weight. **(B)** Plot showing λ. The red line is the weighted regression line, with a gradient (λ) of 0.23. The black lines represent the 95% confidence limit for λ (0.08 to 0.39).

Carrying out a recessive meta-analysis, the mean difference was 0.031 (95% CI 0.000 to 0.061, p=0.051), which is (just) not significant, despite the meta-ANOVA showing a significant association (p=0.02). Carrying out the meta-analysis using other genetic models did not give a significant result either. The mean difference may not reach formal significance because the genetic model used is inaccurate (λ=0.23, not 0) and/or because the overall association was of marginal significance, and statistical significance is lost in the collapsing of two groups. If the two extreme genotypes are compared, there is a significant mean difference (MD=0.039, 95% CI = 0.006 to 0.072, p=0.02).

2.7 Discussion

Current methods used in meta-analysis of the association between continuous traits and genotypes are inadequate and often biased. Most studies make multiple comparisons and do not account for this multiple testing. I have investigated three methods of estimating which genetic model is the most appropriate, so that one meta-analysis can be carried out using this selected model.

Of the three methods that estimate λ, method 3, the linear regression method, seems to be the most appropriate and useful. Method 1 disrupts the within study comparison and hence introduces bias and method 2 heavily weights studies with small SE, regardless of the magnitude of the MDs, which may give counter-intuitive results. Methods 1 and 2 provide no measurement of error of the estimate of λ. Method 3 provides not only a 95%

CI, but also a graphical representation of how well the genetic model fits the data.

Before the genetic model is estimated, meta-ANOVA is required to test for an overall association. This is necessary as trying to choose the best genetic model when there is no association is meaningless and gives spurious and imprecise results as I have shown (genes A, C, E and G). The 95% CI of λ in some cases spanned all three genetic models.

Estimating λ using a linear regression method is more sophisticated than simply carrying out two or three meta-analyses and then comparing the results to choose the best model, as the estimate of λ along with its CI shows how strong the evidence is for choosing a particular genetic model. It also means there is no issue of multiple testing.

The method I describe here also allows the data to be analysed as using a co-dominant (per-allele) model, should that be appropriate. Many software packages for meta-analysis (such as RevMan [The Cochrane Collaboration, 2006]) cannot do this.

For genes B and D, significant p-values in the meta-ANOVA suggest an association and the linear regression estimates λ to be close to 0.5, suggesting a co-dominant model. Using a co-dominant model to analyse these associations shows the mean differences are small but significant. For gene F, although a significant p-value was obtained in the meta-ANOVA there

was no association when analysing the data using the chosen genetic model. λ was estimated to be 0.23 (95% CI 0.08 to 0.39) which does not include any of the three assumed models, so the closest model was chosen: recessive. As theoretically a genetic model where λ is 0.23 is possible, this highlights that genetic studies that assume a model to be recessive, dominant or co-dominant have limitations. Comparing AA with aa for this gene there is a significant difference and so in specific cases of marginal significance, a meta-ANOVA may show association, whilst the mean difference meta-analysis does not.

Eliminating genes from the mean difference meta-analysis that showed no overall association on meta-ANOVA does mean that results for genes which have not quite reached significance in the first stage are not plotted out on a forest plot. It is sometimes of interest to view a forest plot of the mean differences even if there is no statistically significant association. It may be that the association between the gene and trait has not reached significance because it has not been studied in large enough numbers yet or that a significant association is only found in a subgroup of the studies (e.g. in those individuals of a particular ethnicity). However, as a forest plot compares two groups and there is no unbiased way of selecting which two groups to compare for these genes, a forest plot is inappropriate. If it is still desirable to view a display of this data the three means and SDs for each study could be plotted.

The method I devised is simple and quick to use and is an improvement on most current methods used to analyse genetic meta-analyses.

2.8 Method Used in Future Chapters

For the meta-analyses in chapter (3) of this thesis, I use a three-step meta-analysis approach:

i. Use the meta-ANOVA method (described in section 2.4.1) to test for an overall association between a polymorphism and a trait,

ii. For those that show a significant association, investigate the genetic model using a novel linear regression method (as described in section 2.4.2.3)

iii. Use the most appropriate genetic model from ii) to carry out a traditional two group comparison or per-allele mean difference meta-analysis (as described in section 2.4.3).

The Stata (version 7.0, [StataCorp., 2001]) step by step code I devised is shown in appendix 1.

3 CIMT Systematic Review and Meta-Analysis

This chapter comprises a systematic review and meta-analyses of the most commonly studied genetic polymorphisms in association with carotid intima-media thickness.

3.1 Introduction

3.1.1 Carotid Intima-Media Thickness

Carotid intima-media thickness (CIMT) is an atherosclerotic trait, measured non-invasively by B-mode ultrasonography (figure 3.1). The carotid artery is the main artery supplying oxygenated blood to the head. There is one carotid artery on each side of the neck. CIMT has been commonly investigated and typical mean CIMT values reported in population-based studies were between 0.63 and 0.80mm [Lorenz *et al.,* 2007]. The standard deviations of the means from these studies were consistently 0.15 or 0.16.

CIMT is a marker of atherosclerosis and a surrogate of vascular disease [Greenland *et al.,* 2000] and is a strong predictor of future myocardial infarction and stroke [Lorenz *et al.,* 2007]. CIMT has been shown to be

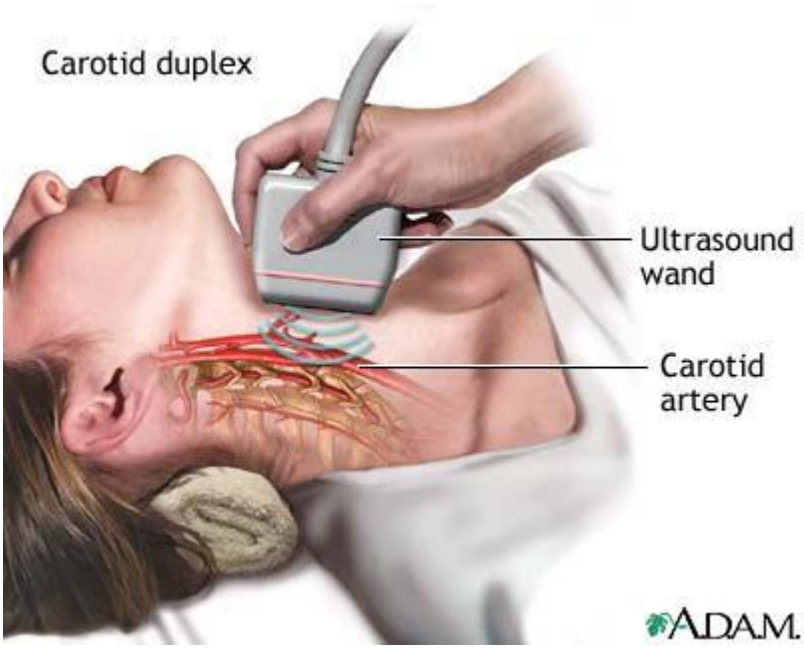


Figure 3.1 Carotid artery ultrasound scan procedure. Illustration from A.D.A.M. Inc.

greater in patients with large artery, compared to those with small artery ischaemic stroke, with a mean difference between the two groups of 0.16mm (95% CI 0.09 to 0.23) [Pruissen *et al.,* 2007].

CIMT is a commonly used intermediate phenotype for early atherosclerosis and large artery stroke [Dichgans & Markus, 2005]. Studying CIMT may be a powerful way to determine which genes influence risk of large artery stroke.

3.1.2 Measurement Methods

CIMT can be measured using B-mode real-time imaging, with a transducer being placed against the neck close to the carotid artery. An image is normally returned to a screen and image-analysing software can give estimates of CIMT at various positions. There are many different sites of the carotid artery that can be measured and many different ways in which the thickness can be reported.

There are different sections to the carotid artery: the common carotid artery (CCA) – the first part of this arterial group that branches from the brachiocephalic artery on the right and the aortic arch on the left side; the internal carotid artery (ICA) – one of the two branches from the CCA which supplies blood to the brain; the external carotid artery (ECA) – the other branch from the CCA, supplying blood to the anterior parts of the neck and the face; and the place where these three arteries join - the bifurcation (BIF). Measurements can be made in any of these segments. The ECA thickness is not important as a predictor of stroke, while ICA is harder to measure than

the CCA, which is more reproducible [O'Leary *et al.*, 1991]. Figure 3.2, taken from Lorenz [2007] shows the different definitions that several large clinical studies have used for the sections of CIMT measurement. Studies use very different definitions for the CCA, some of which do not even overlap, illustrating the considerable heterogeneity in carotid measurement methods across studies.

Some studies measure both the near and far walls. Van Bortel [2005] suggests that measuring only the far wall may be more precise, because near wall measurements are performed at the trailing edge of the ultrasound pulse and variability is higher for this wall [Wendelhag *et al.*, 1991]; [Wikstrand & Wiklund, 1992]. Some studies report a maximum CIMT, whilst others report a mean across several measurements.

The inconsistencies in CIMT measurement have been well documented and

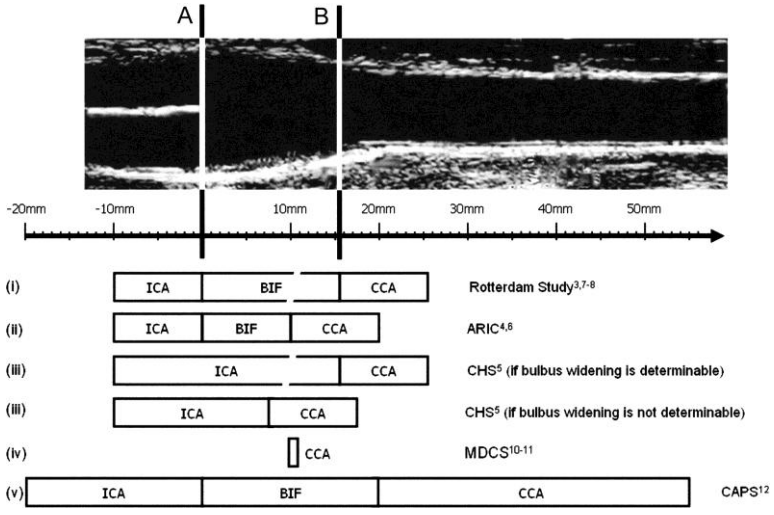


Figure 3.2 Definitions of the carotid segments in several large scale clinical studies. Illustration from [Lorenz *et al.*, 2007].

have led to the Mannheim Intima-Media Thickness consensus [Touboul *et al.,* 2004], which aims to persuade future studies to use standardised definitions and measurement methods, to enable more meaningful comparisons of results across studies. They propose that standard measurements should be of a plaque free region of the far wall of the CCA, ICA or BIF.

A study has shown that inter- and intra-observer variability of CIMT measurements is small [O'Leary *et al.,* 1991].

3.1.3 Heritability

The first estimate of the heritability of CIMT was extremely high – 0.92 [Duggirala *et al.,* 1996]. This study used 46 sibships from Mexico City, and probably overestimated the heritability, as it was very small and did not account for shared environmental factors. Subsequent studies have produced more moderate heritability estimates. The Northern Manhattan Family Study report an age- and sex-adjusted CCA CIMT heritability of 0.39 using 440 subjects from 77 community based families [Juo *et al.,* 2004]. The Framingham Heart Study, which studied data from 1886 subjects from 586 families reported an age- and sex-adjusted CCA CIMT heritability of 0.44 [Fox *et al.,* 2003]. A further study among 565 subjects from 154 families with a parent affected with carotid artery atherosclerosis found a higher heritability of 0.61, adjusted for age, sex hypertension, diabetes mellitus and lipoprotein (a) [Moskau *et al.,* 2005]. A study of 252 diabetic subjects estimated the age- sex- and race-adjusted heritability as 0.32 [Lange *et al.,* 2002]. Despite the wide-ranging estimates, it seems that CIMT has at least a moderate heritability.

3.1.4 Genetic Associations

Hundreds of studies have attempted to identify the genes that are responsible for this genetic influence on CIMT. More than 90 candidate genes have been studied for an association with CIMT [Manolio *et al.*, 2004; Pollex & Hegele, 2006]. However, these have been conflicting and generally only included small numbers of subjects, preventing firm conclusions from being made. Although some reviews have attempted to provide an overview of the area of genetics of CIMT, none has aimed to do this in a systematic and quantitative way.

3.1.5 Aims

I aimed to identify all studies that have analysed the association between CIMT and any gene. For the most commonly studied genes I systematically sought every relevant paper and carried out meta-analyses to provide a summary estimate of the association using all available data. I also aimed to identify sources of heterogeneity between studies

3.2 Methods

3.2.1 Initial Search Strategy

I sought all papers describing the association between any gene and CIMT, using comprehensive, electronic search strategies in Medline (1966 to end 2007) and Embase (1980 to end 2007). I combined MeSH terms and textwords to ensure a highly sensitive search strategy. Table 3.1 shows the search strategy for Medline; a similar search was used in Embase.

Table 3.1 Medline search strategy for all genetic CIMT studies.

Stage 1 Medline search strategy*	
1	exp carotid artery diseases/ge
2	exp carotid arteries/
3	(carotid adj8 (atherosclero$ or stenos$ or plaque$ or imt or cimt or arteriosclero$ or intima media$ or intimal media$ or ultrasound or sclero$ or atheroma$ or wall or thick$)).tw.
4	2 or 3
5	exp genetics/ or exp genotype/ or exp inheritance patterns/ or exp "linkage (genetics)"/ or exp phenotype/ or exp "variation (genetics)"/ or chromosomes/ or exp genes/ or exp genome/
6	(polymorphi$ or genotyp$ or gene or genes or genetic$ or allel$ or mutat$).tw.
7	5 or 6
8	4 and 7
9	1 or 8
10	limit 9 to humans

* I used a similar, appropriately adapted strategy for Embase

3.2.2 Genes Selected for Meta-Analysis

I read the titles of all studies identified from the search and excluded any papers that were obviously not relevant. I then read the abstracts (or full papers where no abstract was available) of all remaining studies and retained all potentially relevant studies (any original study of the association between any gene and CIMT). I listed all genes that had been studied in association with CIMT and calculated the approximate number of studies and subjects for each gene.

Table 3.2 Medline search strategy for MTHFR CIMT studies.

Stage 2 Medline search strategy for MTHFR*	
1	exp carotid artery diseases/ge
2	exp carotid arteries/
3	(carotid adj8 (atherosclero$ or stenos$ or plaque$ or imt or cimt or arteriosclero$ or intima media$ or intimal media$ or ultrasound or sclero$ or atheroma$ or wall or thick$)).tw.
4	1 or 2 or 3
5	exp "Methylenetetrahydrofolate Reductase (NADPH2)"/ge [Genetics]
6	(MTHFR or methylenetetrahydrofolate or c677t or nadph2).tw.
7	methylene tetrahydrofolate.tw.
8	5 or 6 or 7
9	4 and 8

* I used a similar, appropriately adapted strategy for Embase. The specific terms used in the other gene-specific searches are shown in appendix 2.

I selected, for my systematic review and meta-analysis, any gene that had been studied in an estimated total of >7000 subjects. I also selected any gene studied in an estimated total of >3000 subjects if the largest study had >3000 subjects. These cut-offs were chosen to restrict the detailed analysis to those polymorphisms for which results were likely to be the most precise and reliable, avoiding meta-analyses of multiple small studies of less extensively studied polymorphisms which would be likely to yield largely uninformative - or even potentially misleading - results. The precise cut-off chosen was based on feasibility.

3.2.3 Gene Specific Searches and Study Selection

To ensure that all potentially relevant papers had been identified, I carried out a series of supplementary searches for the selected genes in Medline and Embase, replacing the general genetics terms with gene-specific terms (see table 3.2 and appendix 2). Again, I read all the titles or abstracts and retained all relevant studies. A second person (one of: Nahara Martinez-Gonzalez, Rebecca Charleton, Mabel Chung) independently read all the titles or abstracts and selected the papers they felt to be relevant. Comparing these, I compiled a final list of relevant studies.

I obtained the full articles of these potentially relevant studies. Studies in all languages were included, and I obtained translations where necessary. I checked the reference lists of all relevant articles for further relevant studies that may have been missed by the electronic searches.

Studies were included if they had assessed the association between variation in one of the selected genes and a measure of the thickness of the intima-media of the carotid artery. I excluded studies of IMT of other arteries, studies of frank atheroma and plaque, and studies that had only measured a change or rate of change in CIMT. To avoid double counting, where two or more studies used overlapping subjects, I used only the largest available published dataset and excluded the other study/ies.

3.2.4 Data Extraction

I extracted the following information from the papers and entered it into pre-designed spreadsheets: first author and year of publication; total number of subjects studied; country in which the study was conducted; ethnicity of the subjects; types of subjects studied (e.g. healthy volunteers, general population sample, subjects with hypertension, subjects with diabetes); mean age and gender distribution of subjects; candidate gene(s) and polymorphism(s) studied; number of subjects with each genotype; whether the genotypes of subjects conformed to Hardy-Weinberg equilibrium; method of CIMT measurement; mean CIMT (and standard deviation) of subjects with each genotype.

Where studies had presented data separately for subjects defined by different criteria (such as ethnicity, or presence/absence of specific medical condition), I extracted data for each group separately, and analysed these as separate sub-studies (e.g. i, ii, iii).

A second person (one of Nahara Martinez-Gonzalez, Rebecca Charleton or Mabel Chung) independently reviewed study eligibility and extracted the information and data from each study. We resolved differences by discussion and mutual consensus, and if necessary discussed with Steff Lewis or Cathie Sudlow, to reach consensus.

3.2.5 Data Manipulation

Where papers did not present data in the required format I had to carry out transformations of the data. The common transformations are shown below, with examples presented in appendix 3. Other transformations that were particular to certain studies are also presented in appendix 3.

- **Combining groups of subjects within a paper**. For some studies I had to combine means from two or more groups to obtain the data of interest, for example: to combine the mean (and SD) CIMT for men and women, where they were presented separately; to determine the CIMT results per APOE group (E2,E3,E4) where results for all genotypes were presented separately; and in many cases to estimate the mean age across all subjects within a study.

  I used the following formula to obtain the overall mean when combining two groups:

$$\text{mean}_{\text{total}} = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}$$

  where n_i represents the sample size of the ith group, and μ_i represents the mean of the ith group.

  I used the following formula to obtain the overall variance when combining two groups:

$$\text{variance}_{\text{total}} = \frac{n_1(\sigma_1^2 + \mu_1^2) + n_2(\sigma_2^2 + \mu_2^2)}{n_1 + n_2} - \mu_{\text{total}}^2$$

where σ^2_i represents the variance of the ith group

I expanded these formulae to include more than two groups where necessary.

An example of this data transformation is presented in appendix 3.1 and an example of combining measurements within individuals (which is different) is presented in appendix 3.2.

- **Converting standard errors to standard deviations.** Where studies had only reported the former I used the following formula:

$$\text{standard deviation} = \text{standard error} \times \sqrt{\text{n}}$$

where n is the size of the sample for which the standard error refers.

- **Converting confidence intervals to standard deviations.** Where studies had only reported the former I used the following formula:

$$\text{standard deviation} = \frac{\text{upper - lower}}{3.92} \times \sqrt{\text{n}}$$

where 'upper' and 'lower' are the limits of the 95% CI.

- **Estimating numbers of subjects.** In some papers, only genotype proportions were reported, not actual numbers for each genotype. In these cases, I used the total number of subjects and the genotype frequencies to estimate the number of subjects per genotype, but often several actual values were possible and so I had to make a best guess.

  An example of this data transformation is presented in appendix 3.3.

- **Other transformations**

  There was a small number of papers that required specific transformations, such as estimating numbers from a graph, or transforming the CIMT data from sums to means. These specific cases are presented in appendix 3.4 to 3.10.

3.2.6 Attempts to Acquire Missing Data

Where papers did not present the required data, and it could not be calculated, I contacted the corresponding authors of the papers. I designed a standardised data collection form and emailed this along with a letter to each author (see appendix 4).

3.2.7 Statistical Analysis

I used the three-step meta-analysis method described in chapter 2 to investigate the association with CIMT of each genetic polymorphism.

The method is briefly as follows:

1. Determine whether there is an overall association between each genotype and CIMT, by carrying out a meta-analysis of variance (meta-ANOVA) of CIMT, with study and genotype as categorical variables, weighting studies by the inverse of the square of the standard error of the mean CIMT.
2. Where I found a statistically significant overall association ($p<0.05$) in step 1, I went on to determine the most appropriate genetic model (λ), using a linear regression method to estimate λ (where 0=recessive, 0.5=co-dominant, 1=dominant).
3. Using the most appropriate genetic model from step 2, I calculated pooled mean CIMT differences between genotype groups (combining two genotype groups for recessive and dominant models and calculating a per-allele mean difference for co-dominant models).

Most polymorphisms are single mutations, resulting in two alleles and therefore three genotypes. Apolipoprotein E (APOE) has three alleles (ε2, ε3, ε4), making six genotypes. Conventionally the rare ε2ε4 genotype is commonly disregarded and the remaining genotypes are grouped into three groups: E2 (ε2ε2, ε2ε3); E3 (ε3ε3); E4 (ε3ε4, ε4ε4). I analysed APOE using these three groups and so the genetic models refer to these groupings and not the individual genotypes. For example the 'co-dominant model' does not represent a per-allele difference, but the (equal) difference between E4 and E3 genotypes , and E3 and E2 genotypes.

Where there was only one study for a particular polymorphism I carried out ANOVA (instead of meta-ANOVA), and I based the genetic model selection on the single study (λ = MD1/MD2). Where MD1 is the mean CIMT difference Aa and aa genotypes, and MD2 is the mean CIMT difference between AA and aa genotypes.

Where full genotype data needed for meta-ANOVA and genetic model selection were unavailable but the relevant studies had consistently reported and analysed data according to a particular genetic model, I used that model for the meta-analysis of that polymorphism.

I assessed the extent of heterogeneity between studies using the I^2 statistic. I^2 is an estimate of the percentage of variation between studies that cannot be attributed to chance [Higgins *et al.,* 2003].

Before carrying out the meta-analyses, I pre-specified several subgroup analyses. These were: study size (splitting into large and small, where large studies are those larger than the mean number of subjects across all eligible studies or sub-studies for that polymorphism); ethnicity (White, East Asian, South Asian, Black); vascular risk status (high – subjects with a history of vascular disease or with vascular risk factors such as diabetes or hypertension, low – healthy subjects or from a general population). I carried out these sub-group analyses for all polymorphisms that showed a significant overall effect and had been studied in sufficient number of studies to allow this analysis. The within subgroup I^2 statistics are reported and I

tested for significant heterogeneity between subgroups using the Q-test, as suggested by Deeks *et al.* [2001].

I carried out all analyses in Stata (version 7.0 [StataCorp., 2001]) – code for the three-step meta-analysis method is shown in appendix 1. Random effects mean differences were calculated in the primary analyses. Fixed effects mean differences were also calculated in secondary analyses.

I could not include in the formal meta-analyses any study for which the necessary data were unavailable (even after contacting authors). I quantified the proportion of these unavailable data and in an attempt to minimise the impact of bias due to missing data, I extracted qualitative statements on the presence or absence of an association from the papers (where available) and informally assessed how the inclusion of this data may have affected the conclusions.

3.3 Results

3.3.1 Genes Commonly Studied

The first stage of the search strategy yielded 2319 papers, 384 of which appeared to be potentially relevant from reading the titles and abstracts. Appendix 5 shows the full table of the estimated numbers of all genes. The top part of the table is shown in table 3.3. I carried out formal systematic

Table 3.3 Most studied for CIMT genes (top section of full table – appendix 5). Grey shaded genes are those selected for systematic review and meta-analysis.

Gene	Total Subjects	Number of publications	Largest Study
Apolipoprotein E	37493	47	12491
Angiotensin converting enzyme	23935	51	5321
Methylenetetrahydrofolate reductase	14205	33	3247
Nitric oxide synthase 3	9434	19	2448
Paraoxonase 1	8921	27	1786
Adducin 1	8535	5	6471
Angiotensinogen	7515	19	737
Interleukin 6	7190	10	2421
C-reactive protein	6603	3	4641
CD14 molecule	5943	7	1110
Factor V	5828	5	3750
Toll-like receptor 4	5638	6	2955
Apolipoprotein A1/C3	5363	8	2265
Hemachromatosis	5288	4	2932
Adrenergic beta-2 receptor	5249	1	5249
Angiotensin II receptor, type 1	5117	14	737
Cholesteryl ester transfer protein	4387	7	2632
Fibrinogen gamma/alpha	4274	1	4274
Insulin-like growth factor 1	4239	2	3769
Lipoprotein lipase	4178	10	2445
Adiponectin	4035	4	1745
Apolipoprotein B	3386	7	326
Hepatic lipase	3181	4	2268
Toll-like receptor 2	3000	2	2955
peroxisome proliferator-activated receptor alpha	2991	2	2301
peroxisome proliferator-activated receptor gamma	2963	2	1379
Tumor necrosis factor 1	2737	1	2737
Maxtix metallopeptidase 3	2531	5	1111

Table 3.4 Function and estimated and final numbers of relevant studies and subjects for the 13 selected polymorphisms.

Gene (polymorphism)	Function of protein product	Estimated number of studies (subjects)	Final number of studies fulfilling inclusion criteria (subjects)
APOE (ε2, ε3, ε4)	Lipid metabolism	47 (37493)	30 (32995)
ACE (I/D)	Renin-angiotensin system (BP/fluid balance)	51 (23935)	39 (20105)
MTHFR (677 C/T)	Homocysteine metabolism	33 (14205)	20 (10487)
NOS3 (Glu298Asp)	Vascular smooth muscle + endothelial function	19 (9434)	12 (7475)
ADD1 (Gly460Trp)	Endoskeletal protein involved in BP regulation	5 (8535)	4 (6056)
PON1 (Gln192Arg)	LDL modification	27 (8921)	14 (4651)
IL6 (-174 G/C)	Cytokine involved in acute phase response	10 (7190)	7 (4595)
IGF1 (192bp allele)	Interacts with insulin to control carbohydrate metabolism	2 (4239)	1 (5132)
ADRB2 (Gln27Glu)	Intracellular signal transduction	1 (5249)	1 (5173)
CRP (5 SNPs -790 A/T, 1919 A/T, 2667 G/C, 3872 G/A, 5237 A/G)	Inflammation	3 (6603)	1 (4641)
FGG/FGA (7 SNP haplotype)	Coagulation factor	1 (4274)	1 (4274)
AGT (Met235Thr)	Renin-angiotensin system (BP/fluid balance)	19 (7515)	11 (3528)
FV (Leiden)	Activation of thrombin	5 (5828)	3 (3525)

reviews for all genes with an estimated total of >7000 subjects (APOE, ACE, MTHFR, PON1, NOS3, ADD1, AGT, IL6) and any additional gene which had been studied in a total of >3000 subjects where the largest individual study included >3000 subjects (ADRB2, FGG/FGA, CRP, IGF1, FV).

3.3.2 Study Selection for Meta-Analyses

I identified 122 studies (103,804 individual subjects – 112,713 when 'multi-counting' subjects for whom multiple genes were analysed in individual studies) that had analysed the association between CIMT and one of the 13 genes of interest.

The final numbers of relevant papers and subjects for each genetic polymorphism after carrying out supplementary searches and excluding any overlapping or irrelevant papers are shown in table 3.4. The final numbers of relevant studies for each gene were often substantially smaller than the estimated numbers from stage 1. This was because the gene-specific searches added only a few papers, but papers were excluded where they did not fulfill the inclusion criteria after careful consideration, or contained overlapping or identical groups of subjects.

3.3.3 Collection of Missing Data

Of the 122 studies of interest, 38 (including 21,794 subjects) did not have the necessary full data available in the published papers. 19% of data was initially missing. I contacted authors of these studies and 13 authors responded with the necessary missing data [Altamura *et al.,* 2007; Asakimori

et al., 2003; de Maat *et al.,* 2003; Fortunato *et al.,* 2003; Junyent *et al.,* 2006; Karvonen *et al.,* 2002b; Karvonen *et al.,* 2004; Kelemen *et al.,* 2004; Lembo *et al.,* 2001; Mayosi *et al.,* 2005; McDonald *et al.,* 2005; Varda *et al.,* 2005; Visvikis *et al.,* 2000]. I therefore managed to retrieve 26% (5596/21794 subjects) of the "missing" data and reduced the overall proportion of "missing" data to 14%. Table 3.5 shows, for each genetic polymorphism, the number of papers with (i) full data available from the published paper, (ii) full data made available from the authors, (iii) full data not available, despite contacting authors. The papers that still had 'missing data' after this data collection stage are shaded grey in the study characteristics table (table 3.6).

3.3.4 Study Characteristics

Table 3.6 shows the summary characteristics extracted from all relevant sub-studies for the 13 genes. Sample sizes ranged from 47 to 9304 (mean=708). White subjects from Europe, Australia and the US made up the majority of the subjects. Several studies were carried out in Eastern Asian subjects from China, Japan and Taiwan (one recruited subjects from Canada with Eastern Asian heritage). One study recruited subjects from Canada with Southern Asian heritage. Four studies were carried out in Black Americans.

Subjects were mostly middle-aged to elderly. Most were from general population samples or healthy volunteers, but some were of selected subjects at high vascular risk. Genotypes were mostly in Hardy-Weinberg equilibrium (81% of studies), and where they were not (6%), the subjects were generally selected patient groups, for whom Hardy-Weinberg equilibrium would not necessarily be expected.

Table 3.5 Number of studies (and subjects) with and without sufficient data for meta-analysis, both from the publications and from correspondence with the authors for the 13 genes.

Gene	Number of studies with sufficient data for analysis in publication	Number of studies where authors provided me with necessary data	Papers which provided me with necessary data	Number of studies with unavailable data.
APOE	23 (31316)	4 (937)	Karvonen 2002; Asakimori 2003;Junyent 2006; Altamura 2007	3 (742)
ACE	29 (16934)	1 (104)	Varda 2005	9 (3067)
MTHFR	10 (6235)	3 (1710)	DeMaat 2003; Kelemen 2004; McDonald 2005	7 (2542)
NOS3	5 (4015)	1 (375)	Lembo 2001	6 (3085)
ADD1	3 (5636)	-		1 (420)
PON1	5 (905)	3 (1647)	Visvikis 2000; Fortunato 2003; Karvonen 2004	6 (2099)
IL6	4 (2272)	1 (823)	Mayosi 2005	2 (1500)
IGF1	1 (5132)	-		-
ADRB2	1 (1573)†	-		-
CRP	1 (4641)	-		-
FGG/FGA	1 (4274)†	-		-
AGT	6 (1255)	-		5 (2273)
FV	2 (3055) †	-		1 (470)

Table 3.6 Characteristics of studies included for each of the 13 selected genes

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
APOE												
[Terry *et al.,* 1996]	254	US	White	Coronary angiography referrals	59±9	50	High	✓	CCA	Both	Both	Mean of max of 4 sites
[Cattin *et al.,* 1997]	254	Italy	White	Population sample	53±7	46	Low	✓	CCA	Both	Both	Mean of right and left
[Kogawa *et al.,* 1997]i	349	Japan	E.Asian	NIDDM patients	60±11	58	High	✓	CCA/BIF	*	*	Mean of 3 sites
[Kogawa *et al.,* 1997]ii	231	Japan	E.Asian	Non-diabetic subjects	51±11	37	Low	✓	CCA/BIF	*	*	Mean of 3 sites
[Olmer *et al.,* 1997]†	66	France	White	Haemodialysis patients	50±15	50	High	✓	CCA	Far	Both	Mean of 3 each side
[Vauhkonen *et al.,* 1997]i‡	83	Finland	White	NIDDM patients	56±7	52	High	✓	CCA/BIF	Far	Both	Mean of max of 4 sites
[Vauhkonen *et al.,* 1997]ii‡	123	Finland	White	Population sample	54±5	46	Low	✓	CCA/BIF	Far	Both	Mean of max of 4 sites
[Sass *et al.,* 1998]	144	France	White	Population sample	41±4	52	Low	*	CCA	*	Both	Mean of 2 each side
[Zhang *et al.,* 1998]	52	China	E.Asian	CHD patients	57±8	100	High	✓	CCA/BIF/ICA	Far	Both	Mean of 8 sites
[Guz *et al.,* 2000]	261	Turkey	White	Haemodialysis patients	46±15	57	High	✓	CCA	*	Both	Mean of 3 each side
[Hanon *et al.,* 2000]	312	France	White	Patients with vascular risk factors/disease	49±12	53	High	*	CCA	Far	Right	One measurement
[Horejsi *et al.,* 2000]	112	Czech Republic	White	Lipoprotein disorder patients	53±*	45	High	*	CCA	Far	*	Mean of max of 3 sites
[Ilveskoski *et al.,* 2000]	189	Finland	White	Population sample	54±3	100	Low	✓	CCA	Both	Both	Max of 4 sites
[Slooter *et al.,* 2001]	5264	Netherlands	White	Population sample	69±9	41	Low	✓	CCA	Far	Both	Mean of left and right
[Tabara *et al.,* 2001]	202	Japan	E.Asian	Population sample	70±9	32	Low	✓	CCA	Far	Right	Mean of 3 sites
[Haraki *et al.,* 2002]	95	Japan	E.Asian	Healthy subjects	50±8	100	Low	✓	CCA	Far	Right	Mean of 9 sites
[Karvonen *et al.,* 2002b]i	258	Finland	White	Hypertensive patients	51±6	100	High	✓	CCA/BIF/ICA	Far	Both	Mean of 20 sites
[Karvonen *et al.,* 2002b]ii	253	Finland	White	Population sample	51±6	100	Low	✓	CCA/BIF/ICA	Far	Both	Mean of 10 sites
[Asakimori *et al.,* 2003]	162	Japan	E.Asian	Haemodialysis patients	55±11	52	High	✓	CCA	Far	Both	Maximum
[Beilby *et al.,* 2003]	1079	Australia	White	Population sample	53±13	50	Low	✓	CCA	Far	Both	Mean of 3 each side
[Li *et al.,* 2003]	92	China	E.Asian	Hypertensive patients	64±11	55	High	✓	CCA	*	*	*
[Xiang *et al.,* 2003]i	253	China	E.Asian	NIDDM patients	*	*	High	✓	CCA/BIF/ICA	Far	Both	Mean of 3 each side
[Xiang *et al.,* 2003]ii	106	China	E.Asian	Healthy controls	*	*	Low	✓	CCA/BIF/ICA	Far	Both	Mean of 3 each side
[Elosua *et al.,* 2004]	2723	US	White	Population sample	59±10	48	Low	✓	CCA	Both	Both	Mean of max each side
[Fernandez *et al.,* 2004]	225	Spain	White	CHD patients	61±8	85	High	*	CCA	Far	Both	Mean of 3 each side
[Kahraman *et al.,* 2004]	118	Turkey	White	Renal transplant recipients	40±8	68	High	*	CCA	*	Both	Mean of left and right
[Bednarska *et al.,* 2005]	127	Poland	White	Alcoholics	49±6	100	High	✓	CCA	Far	Both	Mean of 3 each side
[Bleil *et al.,* 2006]	182	US	White	Hypertensive patients	56±9	100	High	✓	CCA/BIF/ICA	Both	Both	Mean of all sites
[Brenner *et al.,* 2006]†	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left
[Debette *et al.,* 2006]	5764	France	White	Population sample	74±5	40	Low	✓	CCA	Far	Both	Mean of right and left

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
[Junyent *et al.,* 2006]	163	Spain	White	Familial hypercholesterolaemia patients	47±*	*	High	✓	CCA	Far	Both	Mean of right and left
[Volcik *et al.,* 2006]i	3187	US	Black	Population sample	range 45-64**	*	Low	✓	CCA/BIF/ICA	*	Both	Mean of 6 sites
[Volcik *et al.,* 2006]ii	9304	US	White	Population sample	range 45-64**	*	Low	✓	CCA/BIF/ICA	*	Both	Mean of 6 sites
[Altamura *et al.,* 2007]i	68	Italy	White	Alzheimer disease patients	75±8	31	High	*	CCA	*	Both	Mean of right and left
[Altamura *et al.,* 2007]ii	33	Italy	White	Vascular dementia patients	77±8	51	High	*	CCA	*	Both	Mean of right and left
[Wohlin *et al.,* 2007]	437	Sweden	White	Population sample	all 75	100	Low	✓	CCA	Far	Both	Mean of 3 each side
ACE												
[Castellano *et al.,* 1995]	187	Italy	White	Population sample	58±3	52	Low	✓	CCA/BIF/ICA		Both	Mean of all sites
[Dessi-Fulgheri *et al.,* 1995]	240	Italy	White	Outpatients without vascular risk factors	53±7	57	Low	✓	CCA/BIF/ICA	Both	Both	Mean
[Markus *et al.,* 1995]	101	UK	White	Ischaemic CVD patients	65±9	68	High	✓	CCA	Far	*	Maximum
[Kauma *et al.,* 1996]	515	Finland	White	Hypertensive patients	51±6	49	High	✓	CCA	Far	Both	Mean of max at each site
[Pujia *et al.,* 1996]	132	Italy	White	NIDDM patients	50±10	100	High	✓	CCA	Far	Both	Mean of 6 sites
[Kogawa *et al.,* 1997]i	356	Japan	E.Asian	NIDDM patients	60±11	58	High	✓	CCA/BIF	*	*	Mean of 3 sites
[Kogawa *et al.,* 1997]ii	235	Japan	E.Asian	Non-diabetic subjects	51±11	37	Low	✓	CCA/BIF	*	*	Mean of 3 sites
[Watanabe *et al.,* 1997][‡]	169	Japan	E.Asian	Healthy volunteers	59±6	51	Low	✓	CCA/BIF/ICA	Both	Both	Mean
[Arnett *et al.,* 1998]	495	US	White	Population sample	59±6	42	Low	✓	CCA/BIF/ICA	Far	Both	Mean of 6 sites
[Frost *et al.,* 1998]	148	Germany	White	IDDM patients	30±7	38	High	✓	CCA	Far	Both	Maximum
[Girerd *et al.,* 1998]	340	France	White	Patients with vascular risk factors/disease	49±12	53	High	✓	CCA	Far	Right	One measurement
[Sass *et al.,* 1998]	150	France	White	Population sample	41±4	52	Low	✓	CCA	Far	Both	Mean of all
[Ferrieres *et al.,* 1999]	355	France	White	Population sample	54±7	100	Low	✓	CCA	Far	Both	Mean of 12 sites
[Huang *et al.,* 1999]	219	Finland	White	Population sample	54±3	100	Low	✓	CCA	Far	Both	Maximum
[Hung *et al.,* 1999]	1106	Australia	White	Population sample	53±12	50	Low	✓	CCA	Far	Both	Mean of 6 sites
[Nergizoglu *et al.,* 1999]	51	Turkey	White	Hemodialysis patients	36±9	69	High	✓	CCA	Far	Both	Mean of 6 sites
[Pit'ha *et al.,* 1999]	47	Czech Republic	White	Hypertensive patients	62±3	100	High	✓	CCA	Far	Both	Mean of 10 sites
[Jeng, 2000]	175	China	E.Asian	Hypertensive patients	57±10	52	High	X	CCA	Far	Both	Mean of right and left
[Pontremoli *et al.,* 2000][†]	215	Italy	White	Hypertensive patients	48±9	62	High	✓	CCA	Far	Both	Mean of 3 sites
[Taute *et al.,* 2000]	98	Germany	White	PAD patients	61±9	79	High	✓	CCA	Far	Both	Maximum
[Mannami *et al.,* 2001]	3657	Japan	E.Asian	Population sample	60±12	46	Low	✓	CCA	Both	Both	Mean of 4 sites
[Markus *et al.,* 2001][†]	287	UK	White	Population sample	61±8	100	Low	✓	CCA	Far	Both	Mean
[Tabara *et al.,* 2001]	205	Japan	E.Asian	Healthy population sample	70±9	32	Low	✓	CCA	Far	Right	Mean
[Balkestein *et al.,* 2002]	380	Belgium	White	Population sample	40±16	50	Low	✓	CCA	Far	Right	Mean of 3 sites
[Diamantopoulos *et al.,* 2002]	184	Greece	White	NIDDM patients	62±8	41	High	✓	CCA	Far	Both	Max of mean from each side
[Kawamoto *et al.,* 2002][‡]	184	Japan	E.Asian	In-patients being evaluated for possible atherosclerosis	67±14	47	High	✓	CCA	Far	Both	Mean of right and left

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
[Piao *et al.*, 2002][‡]	262	Japan	E.Asian	NIDDM patients	58±10	66	High	*	CCA/BIF/ICA	*	Both	Mean of 6 sites
[Czarnecka *et al.*, 2004]i	127	Poland	White	Population sample – parents	51±5	40	Low	✓	CCA	Both	Both	*
[Czarnecka *et al.*, 2004]ii	157	Poland	White	Population sample – offspring	24±5	50	Low	✓	CCA	Both	Both	*
[Li *et al.*, 2004]	102	China	E.Asian	Hypertensive patients	54±9	*	High	X	CCA	Both	Both	Mean of 12 sites
[Pall *et al.*, 2004]i	120	Hungary	White	Hypertensive students	16±1	53	High	✓	CCA	*	*	Mean of 3 sites
[Pall *et al.*, 2004]ii	58	Hungary	White	Non-hypertensive students	*	*	Low	✓	CCA	*	*	Mean of 3 sites
[Bednarska *et al.*, 2005]	130	Poland	White	Alcoholics	48±6	100	High	✓	CCA	Far	Both	Mean of each side
[Sleegers *et al.*, 2005]	6488	Netherlands	White	Population sample	69±9	41	Low	✓	CCA	Both	Both	*
[Varda *et al.*, 2005]i	56	Slovenia	White	Offspring of CVD patients	18±6	52	High	✓	CCA/ICA	*	Both	Mean of 4 sites
[Varda *et al.*, 2005]ii	48	Slovenia	White	Subjects without parental history of CVD	18±6	52	Low	✓	CCA/ICA	*	Both	Mean of 4 sites
[Bilici *et al.*, 2006]	64	Turkey	White	Memory impaired patients	57±13	83	High	✓	CCA	Far	Both	Mean of right and left
[Brenner *et al.*, 2006][†]	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left
[Burdon *et al.*, 2006][†]	737	US	White	NIDDM patients & their siblings	61±10	43	High	✓	CCA	Both	Both	Mean of 20 sites
[Islam *et al.*, 2006]	224	Finland	White	Population sample	34±2	54	Low	✓	CCA	Far	Left	Mean of 4 sites
[Yamasaki *et al.*, 2006][†]	690	Japan	E.Asian	NIDDM patients	63±7	52	High	*	CCA/BIF/ICA	*	Both	Mean of max
[Bartoli *et al.*, 2007][‡]	53	Italy	White	Systemic sclerosis patients	60±11	11	High	✓	CCA	Far	Both	Mean of right and left
[Tanriverdi *et al.*, 2007]	88	Japan	E.Asian	Coronary angiography patients	55±11	55	High	X	CCA	*	Both	Mean of 8 sites
MTHFR												
[Arai *et al.*, 1997]	222	Japan	E.Asian	NIDDM patients	60±8	73	High	✓	BIF	Both	Both	Maximum
[Demuth *et al.*, 1998][†]	144	France	White	Patients with vascular risk factors/disease	48±13	46	High	✓	CCA	Far	Right	*
[Mazza *et al.*, 1999]	95	Italy	White	NIDDM patients	53±10	35	High	✓	CCA	Far	Both	Mean of 6 sites
[McQuillan *et al.*, 1999][†]	1111	Australia	White	Population sample	53±13	50	Low	✓	CCA	Far	Both	Mean of 6 sites
[Kawamoto *et al.*, 2001][†]	136	Japan	E.Asian	Patients with vascular risk factors	74±12	45	High	✓	CCA	Far	Both	Mean
[Lim *et al.*, 2001]	151	Taiwan	E.Asian	End stage renal disease patients	55±14	42	High	✓	CCA	Both	*	Mean
[Markus *et al.*, 2001][†]	279	UK	White	Population sample	61±8	100	Low	✓	CCA	Far	Both	Mean
[Pallaud *et al.*, 2001][‡]	121	France	White	Population sample	43±5	64	Low	✓	CCA	Far	Both	Mean
[Passaro *et al.*, 2001]	120	Italy	White	Healthy post-menopausal women	62±4	0	Low	✓	CCA	Both	Both	Mean of max
[Ravera *et al.*, 2001]	206	Italy	White	Hypertensive patients	48±9	*	High	✓	CCA	Far	Both	Mean of 3 sites
[Scaglione *et al.*, 2002]	124	Italy	White	NIDDM patients	65±8	76	High	✓	CCA	Far	Both	Mean of 6 sites
[de Maat *et al.*, 2003]	691	Denmark	White	Population sample	All 60	47	Low	✓	CCA/BIF/ICA	Both	Right	Mean of 3 sites
[Inamoto *et al.*, 2003]	3247	Japan	E.Asian	Population sample	59±13	48	Low	✓	CCA	Both	Both	Mean
[Kelemen *et al.*, 2004]i	260	Canada	White	Population sample	49±*	49	Low	✓	CCA/BIF/ICA	Both	Both	Mean of max
[Kelemen *et al.*, 2004]ii	275	Canada	E.Asian	Population sample	47±*	53	Low	✓	CCA/BIF/ICA	Both	Both	Mean of max

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
[Kelemen *et al.,* 2004]iii	283	Canada	S.Asian	Population sample	48±*	54	Low	✓	CCA/BIF/ICA	Both	Both	Mean of max
[Durga *et al.,* 2005]	815	Netherlands	White	Patients with high homocysteine	60±6	72	High	✓	CCA	Both	Both	Mean of max
[McDonald *et al.,* 2005]	201	Australia	White	Population sample	37±*	44	Low	X	CCA	Both	*	Mean of 6 sites
[Linnebank *et al.,* 2006]	714	Germany	White	Vascular event patients	64±9	49	High	✓	CCA	Far	*	Mean
[Yamasaki *et al.,* 2006]†	690	Japan	E.Asian	NIDDM patients	63±7	52	High	*	CCA/BIF/ICA	*	Both	Mean of max
[Fernandez *et al.,* 2007]‡	61	Spain	White	Patients with coronary disease	68±7	82	High	*	CCA	Far	Both	Mean of 6 sites
[Liu *et al.*, 2007]	541	Taiwan	E.Asian	Healthy volunteers	53±15	50	Low	✓	CCA	Far	Either	Mean of 4 sites
NOS3												
[Lembo *et al.,* 2001]	375	Italy	White	Hypertensive patients	54±*	55	High	✓	CCA/BIF/ICA	Both	Both	Maximum
[Karvonen *et al.,* 2002a]i	505	Finland	White	Hypertensive patients	51±6	49	High	✓	CCA	Far	Both	Mean of 10 sites
[Karvonen *et al.,* 2002a]ii	519	Finland	White	Population sample	51±7	50	Low	✓	CCA	Far	Both	Mean of 10 sites
[Asakimori *et al.,* 2003]†	163	Japan	E.Asian	Haemodialysis patients	55±11	52	High	✓	CCA	Far	Both	Maximum
[Schmoelzer *et al.,* 2003]‡	932	Italy	White	Population sample	53±6	55	Low	✓	CCA/BIF/ICA	Both	Both	Mean of 12 sites
[Paradossi *et al.,* 2004]	118	Italy	White	Population sample	30±5	39	Low	✓	CCA	*	Both	Mean of max
[Czarnecka *et al.,* 2005]i	127	Poland	White	Population sample – parents	51±5	40	Low	✓	CCA	Both	Both	*
[Czarnecka *et al.,* 2005]ii	167	Poland	White	Population sample – offspring	24±5	50	Low	✓	CCA	Both	Both	*
[Spoto *et al.,* 2005]	131	Italy	White	Haemodialysis patients	61±13	60	High	✓	CCA/BIF/ICA	Far	Both	Mean of 12 sites
[Wolff *et al.,* 2005]	2448	Germany	White	Population sample	62±10	51	Low	✓	CCA	Far	Both	Mean of 20 sites
[Brenner *et al.,* 2006]†	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left
[Burdon *et al.,* 2006]†	737	US	White	NIDDM patients & their siblings	61±10	43	High	✓	CCA	Both	Both	Mean of 20 sites
[Lekakis *et al.,* 2006]‡	122	Greece	White	Coronary angiography patients	61±10	84	High	*	CCA/BIF/ICA	Far	Both	Mean of max of 6 sites
[Bhuiyan *et al.,* 2007]‡	661	US	White	Population sample	37±4	40	Low	✓	CCA	Far	Both	Mean of max of 6 sites
ADD1												
[Castellano *et al.,* 1997]	173	Italy	White	Population sample	57±5	50	Low	✓	CCA	Far	Both	Mean
[Balkestein *et al.,* 2002]	380	Belgium	White	Population sample	40±16	49	Low	✓	CCA	Far	Right	Mean of 3 sites
[Sarzani *et al.,* 2006]‡	420	Italy	White	Medical student volunteers	23±2	52	Low	X	CCA/BIF	Both	Both	Mean of max of 8 sites
[Yazdanpanah *et al.,* 2006]	5083	Netherlands	White	Population sample	69±9	40	Low	✓	CCA	Both	Both	Mean of 6 sites
PON1												
[Cao *et al.,* 1998]‡	170	France	White	NIDDM patients	55±8	78	High	✓	CCA	*	Both	Mean of 32 sites
[Sakai *et al.,* 1998]	139	Japan	E.Asian	NIDDM patients	62±14	47	High	✓	CCA	Far	Both	Mean of max
[Dessi *et al.,* 1999]	196	Italy	White	Population sample	55±12	61	Low	✓	CCA	Far	Both	Mean of 6 sites
[Visvikis *et al.,* 2000]	362	France	White	Population sample	*	48	Low	✓	CCA	Both	Both	Mean of 4 sites
[Markus *et al.,* 2001]†	288	UK	White	Population sample	61±8	100	Low	✓	CCA	Far	Both	Mean

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
[Fortunato *et al.*, 2003]	286	Italy	White	Population sample	55±8	0	Low	✓	CCA	Both	Both	Mean of 6 sites
[Hu *et al.*, 2003]	152	China	E.Asian	NIDDM patients	59±12	63	High	✓	CCA	*	Both	Mean of 6 sites
[Campo *et al.*, 2004][†]	208	Italy	White	Hypercholesterolemia patients	57±10	48	High	✓	CCA/BIF/ICA	*	Both	Mean
[Karvonen *et al.*, 2004]i	496	Finland	White	Hypertensive patients	*	*	High	✓	CCA/BIF/ICA	Far	Both	Mean of 6 sites
[Karvonen *et al.*, 2004]ii	503	Finland	White	Population sample	*	*	Low	✓	CCA/BIF/ICA	Far	Both	Mean of 6 sites
[Srinivasan *et al.*, 2004]i[‡]	307	US	White	Population sample	33±7**	44**	Low	✓	CCA/BIF/ICA	Far	Both	Mean of maximum
[Srinivasan *et al.*, 2004]ii[‡]	129	US	Black	Population sample	33±7**	44**	Low	✓	CCA/BIF/ICA	Far	Both	Mean of maximum
[Burdon *et al.*, 2005][†]	527	US	White	NIDDM patients & their siblings	62±10	44	High	X	CCA	Both	Both	Mean of 20 sites
[Van Himbergen *et al.*, 2004]	285	Netherlands	White	Familial hypercholesterolaemia patients	48±*	40	High	✓	CCA	Both	Both	Mean
[Brenner *et al.*, 2006][†]	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left
[Roest *et al.*, 2006]	133	Netherlands	White	Paediatric lipid clinic patients	*	*	High	✓	CCA	Far	Both	Mean
IL6												
[Rauramaa *et al.*, 2000]	92	Finland	White	Population sample	55±3	100	Low	✓	BIF	Far	Both	Mean of maximum
[Rundek *et al.*, 2002]	71	US	Black& White	Population sample	70±12	45	Low	✓	CCABIF/ICA	Both	Both	Mean of 12 sites
[Chapman *et al.*, 2003]	1109	Australia	White	Population sample	53±13	50	Low	✓	CCA	Far	Both	Mean of 6 sites
[Jerrard-Dunne *et al.*, 2003b]	1000	UK	White	Population sample	range 50-65	*	Low	✓	CCA	Far	Both	*
[Mayosi *et al.*, 2005]	823	UK	White	Hypertensive patients & their relatives	54±*	48	High	✓	CCA	Far	Both	Maximum
[Markus *et al.*, 2006][†]	810	Italy	White	Population sample	58±11	50	Low	✓	CCA	Far	Both	Mean of max
[Yamasaki *et al.*, 2006][†]	690	Japan	E.Asian	NIDDM patients	63±7	52	High	*	CCA/BIF/ICA	*	Both	Mean of max
IGF1												
[Schut *et al.*, 2003]	5132	Netherlands	White	Population sample	65±6	43	Low	✓	CCA	Both	Both	Mean
ADRB2												
[Hindorff *et al.*, 2005][‡]	5249	US	Black& White	Population sample	73±*	44	Low	✓	CCA	Both	Both	Mean of max
CRP												
[Lange *et al.*, 2006]	4641	US	White	Population sample	73±6	40	Low	✓	CCA/ICA	*	Both	Mean of maximum
FGG/FGA												
[Kardys *et al.*, 2007][†]	4274	Netherlands	White	Population sample	70±9	41	Low	✓	CCA	Both	Both	Mean
AGT												
[Barley *et al.*, 1995]	100	UK	White	Patients with TIA or stroke	65±9	66	High	✓	CCA	Far	*	Maximum
[Arnett *et al.*, 1998]	475	US	White	Population sample	59±6	42	Low	X	CCA/BIF/ICA	Far	Both	Mean of 6 sites
[Jeng, 1999]	175	Taiwan	E.Asian	Hypertensive patients	57±9	52	High	X	CCA	Far	Both	Mean of right and left

Study (first author & publication year)	No. of subjects	Country	Ethnicity of subjects	Type of subjects	Mean age ± SD	% male	Vascular risk status	HWE	CIMT measurement method			
									Carotid segment	Near/far carotid wall	Right/left carotid	Mean/max
[Pontremoli *et al.,* 2000]†	215	Italy	White	Hypertensive patients	48±9	62	High	✓	CCA	Far	Both	Mean of 3 sites
[Pallaud *et al.,* 2001]†	161	France	White	Population sample	43±5	48	Low	✓	CCA	Far	Both	Mean
[Tabara *et al.,* 2001]	205	Japan	E.Asian	Population sample	70±9	32	Low	✓	CCA	Far	Right	Mean of 3 sites
[Bozec *et al.,* 2003]	98	France	White	Hypertensive patients	51±8	62	High	X	CCA	Far	Right	Mean of 3 sites
[Brenner *et al.,* 2006]†	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left
[Burdon *et al.,* 2006]†	737	US	White	NIDDM patients & their siblings	61±10	43	High	✓	CCA	Both	Both	Mean of 20 sites
[Islam *et al.,* 2006]	202	Finland	White	Population sample	34±3	54	Low	✓	CCA	Far	Left	Maximum of 4 sites
[Yamasaki *et al.,* 2006]†	690	Japan	E.Asian	NIDDM patients	63±7	52	High	*	CCA/BIF/ICA	*	Both	Mean of maximum
FV												
[Garg *et al.,* 1998]‡	1292	US	Black& White	CHD patients & their siblings	56±11	47	High	*	CCA/BIF/ICA	Far	Both	Mean of 6 sites
[Fox *et al.,* 2004b]‡	1763	US	White	CHD patients' offspring	57±10	49	High	✓	CCA	Both	Both	Mean of maximum
[Brenner *et al.,* 2006]†	470	France	White	Ischaemic stroke patients	range 18-85	*	High	*	CCA	Far	Both	Mean of right and left

Grey shaded studies are those which were not included in the analyses because complete data were unavailable.

* information not available from publication, **data only available for whole study so estimated to be equal for each sub-study, †studies with all result data unavailable from the publication, ‡ studies with result data only relating to a particular genetic model available in the publication.

APOE: apolipoprotein E; ACE: angiotensin I converting enzyme (peptidyl-dipeptidase A) 1; MTHFR: 5,10-methylenetetrahydrofolate reductase (NADPH); NOS3: nitric oxide synthase 3 (endothelial cell); ADD1: adducin 1 (alpha); PON1: paraoxonase 1; IL6: interleukin 6 (interferon, beta 2); IGF1: insulin-like growth factor 1 (somatomedin C); ADRB2: adrenergic, beta- 2-, receptor, surface; CRP: C-reactive protein, pentraxin-related; FGG/FGA: fibrinogen gamma chain/alpha chain; AGT: angiotensinogen (serpin peptidase inhibitor, clade A, member 8); FV: coagulation factor V (proaccelerin, labile factor); HWE: Hardy Weinberg equilibrium; CCA: common carotid artery; BIF: bifurcation; ICA: internal carotid artery; NIDDM: non-insulin-dependent diabetes mellitus; CHD: coronary heart disease; CVD: cerebrovascular disease; IDDM: insulin-dependent diabetes mellitus; PAD: peripheral artery disease.

Method of CIMT measurement varied quite considerably between studies. Where possible I had selected the common carotid artery measurement, and so most studies' results are from this segment only (115 studies, using the authors' definitions). In a smaller number of studies (43), the authors had only presented data combining multiple segments (e.g. overall mean of CCA, ICA & BIF). In two studies, only the bifurcation was measured.

Most studies measured the far wall only (92), a smaller number measured both walls (40), none measured the near wall only, and 27 studies did not report which wall was measured.

The majority of studies (133) combined measurements from both the right and left carotid arteries to produce a value for each patient, fewer studies measured only the right (11) or the left (2), and one study reported measuring 'either'. Thirteen studies did not report which side they had measured.

By far the most variable part of the measurement method was the number of sites measured and whether means or maximums were recorded. Most (110) recorded the mean of all sites measured, with the number of sites measured varying enormously from three to 32. 12 recorded the maximum measurement from all sites (often not reporting how many sites were measured, so the sonographer may have looked for the thickest portion from all scans). 27 studies recorded the 'mean of maximum', where the maximum from each site or each side was averaged, again with varying numbers of

sites measured. A few other studies used other methods: one study reported 'maximum of mean', measuring the mean from each side and the reporting the maximum of these; two studies reported that they only made one measurement for each subject. Eight studies did not report how they combined measurements for the overall value.

Often the CIMT measurement methods used were poorly reported and it was difficult to tell exactly how many measurements were taken and how these were combined to create the overall value. If we ignore the number of measurements taken, the most common method was to measure the mean of the far wall of both common carotid arteries (in 48 studies).

Table 3.7 shows the relevant CIMT data per genotype extracted from each study.

3.3.5 Overall Results

Table 3.8 shows the overall results from the three steps of the analysis.

Of the 13 genes reviewed:

- Eight genes (APOE, ACE, MTHFR, NOS3, ADD1, PON1, IL6 & AGT) had been studied in more than one study and could be analysed using the 3-step meta-analysis approach.

Table 3.7 CIMT data for each study with full data available. Sample size, CIMT mean and CIMT SD per genotype are shown.

Study	N	mean	SD	N	mean	SD	N	mean	SD
APOE		E4			E3			E2	
Terry 1996	66	1	0.24	155	0.97	0.25	33	0.85	0.23
Cattin 1997	45	1.95	0.45	177	1.84	0.15	32	1.80	0.10
Kogawa 1997i	62	1.098	0.482	261	1.043	0.485	26	0.993	0.495
Kogawa 1997ii	33	0.640	0.149	176	0.631	0.172	22	0.624	0.172
Sass 1998	30	0.52	0.04	90	0.54	0.05	24	0.52	0.05
Zhang 1998	10	1.5	0.5	38	1.1	0.3	4	1.2	0.6
Guz 2000	28	0.75	0.35	200	0.63	0.38	33	0.59	0.13
Hanon 2000	66	0.545	0.105	208	0.533	0.115	38	0.536	0.134
Horejsi 2000	25	0.8	0.25	77	0.7	0.25	10	0.72	0.25
Ilveskoski 2000	60	1.03	0.16	109	1.05	0.17	20	0.95	0.12
Slooter 2001	1392	0.77	0.18	3122	0.77	0.14	750	0.75	0.14
Tabara 2001	38	0.83	0.16	137	0.79	0.13	27	0.79	0.12
Haraki 2002	20	0.76	0.17	65	0.64	0.14	10	0.61	0.15
Karvonen 2002i	86	0.98	0.26	160	0.90	0.19	12	0.89	0.18
Karvonen 2002ii	90	0.93	0.23	150	0.93	0.21	13	0.78	0.15
Asakimori 2003	31	0.93	0.42	109	0.99	0.52	22	0.91	0.37
Beilby 2003	283	0.70	0.13	650	0.72	0.15	146	0.69	0.13
Li 2003	18	1.02	0.19	64	0.88	0.16	10	0.81	0.17
Xiang 2003i	58	0.89	0.15	161	0.68	0.17	34	0.62	0.12
Xiang 2003ii	20	0.71	0.14	75	0.60	0.13	11	0.59	0.11
Elosua 2004	568	0.74	0.20	1782	0.74	0.18	373	0.73	0.16
Fernandez 2004	49	0.83	0.23	158	0.81	0.21	18	0.76	0.17
Kahraman 2004	14	0.7	0.4	92	0.7	0.4	12	0.6	0.2
Bednarska 2005	21	0.88	0.30	89	0.78	0.16	17	0.86	0.23
Bleil 2006	39	0.875	0.16	120	0.932	0.16	23	0.89	0.16
Debette 2006	1124	0.722	0.129	3923	0.712	0.125	717	0.713	0.107
Junyent 2006	31	0.69	0.19	122	0.65	0.16	10	0.60	0.15
Volcik 2006i	1126	0.746	0.134	1427	0.734	0.113	634	0.722	0.126
Volcik 2006ii	2311	0.743	0.144	5534	0.733	0.149	1459	0.723	0.115
Altamura 2007i	24	0.986	0.190	40	0.887	0.136	4	0.788	0.063
Altamura 2007ii	7	0.867	0.082	22	0.960	0.168	4	0.900	0
Wohlin 2007	144	0.74	0.21	242	0.79	0.21	59	0.72	0.12

Study	N	mean	SD	N	mean	SD	N	mean	SD
ACE		DD			ID			II	
Castellano 1995	76	0.74	0.17	88	0.68	0.19	23	0.75	0.19
Dessi 1995	93	1.05	0.4	124	1.06	0.3	23	1.02	0.2
Markus 1994	36	0.811	0.276	47	0.939	0.279	18	1.135	0.395
Kauma 1996	148	0.83	0.19	264	0.80	0.15	103	0.81	0.18
Pujia 1996	46	0.778	0.07	70	0.759	0.08	16	0.700	0.08
Kogawa 1997i	60	1.200	0.586	149	1.062	0.541	147	0.990	0.364
Kogawa 1997ii	32	0.640	0.173	116	0.631	0.171	87	0.629	0.162
Arnett 1998	151	0.731	0.15	256	0.730	0.16	88	0.720	0.15
Frost 1998	62	0.63	0.13	55	0.63	0.18	31	0.62	0.15
Girerd 1998	118	0.547	0.111	165	0.538	0.129	57	0.536	0.103
Sass 1998	37	0.53	0.04	80	0.54	0.06	33	0.52	0.04
Ferrieres 1999	135	0.64	0.12	150	0.62	0.10	70	0.63	0.12
Huang 1999	77	1.01	0.19	100	1.08	0.33	42	1.04	0.23
Hung 1999	343	0.71	0.15	535	0.71	0.14	228	0.71	0.14
Nergizoglu 1999	22	0.80	0.10	22	0.76	0.09	7	0.71	0.05
Pit'ha 1999	14	0.747	0.16	25	0.713	0.12	8	0.723	0.13
Jeng 2000	41	0.877	0.354	69	0.756	0.307	65	0.737	0.273
Taute 2000	33	1.06	0.26	46	1.10	0.25	19	1.01	0.29
Mannami 2001	477	0.87	0.26	1640	0.87	0.29	1540	0.87	0.28
Tabara 2001	27	0.81	0.14	95	0.80	0.14	83	0.79	0.12
Balkestein 2002	84	0.582	0.17	180	0.585	0.23	116	0.555	0.16
Diamantopoulos 2002	69	0.98	0.21	86	0.97	0.20	29	0.94	0.20
Czarnecka 2004i	36	0.91	0.48	62	0.77	0.55	29	0.73	0.54
Czarnecka 2004ii	44	0.61	0.33	84	0.56	0.37	29	0.53	0.32
Li 2004	45	0.79	0.18	27	0.67	0.11	30	0.58	0.12
Pall 2004i	28	0.57	0.11	57	0.53	0.10	35	0.55	0.10
Pall 2004ii	15	0.46	0.10	25	0.48	0.10	18	0.49	0.10
Bednarska 2005	30	0.81	0.21	62	0.80	0.19	38	0.82	0.21
Sleegers 2005	1806	0.80	0.16	3264	0.80	0.16	1418	0.79	0.15
Varda 2005i	9	0.48	0.06	31	0.43	0.09	16	0.42	0.07
Varda 2005ii	17	0.40	0.08	21	0.43	0.07	10	0.42	0.06
Bilici 2006	28	1.29	0.30	28	1.29	0.33	8	1.32	0.29
Islam 2006	79	0.57	0.08	106	0.59	0.10	39	0.60	0.08
Tanriverdi 2007	29	0.78	0.06	34	0.72	0.05	25	0.64	0.06

Study	N	mean	SD	N	mean	SD	N	mean	SD
MTHFR		TT			TC			CC	
Arai 1997	39	1.58	0.54	94	1.35	0.35	89	1.31	0.31
Mazza 1999	22	0.875	0.197	38	0.862	0.193	35	0.833	0.186
Lim 2001	10	0.93	0.07	54	0.85	0.15	87	0.79	0.13
Passaro 2001	20	1.23	0.18	72	1.03	0.17	28	0.98	0.21
Ravera 2001	36	0.79	0.30	111	0.69	0.21	59	0.64	0.23
Scaglione 2002	22	0.89	0.22	55	0.93	0.22	47	0.86	0.29
DeMaat 2003	62	0.682	0.112	304	0.720	0.134	325	0.732	0.153
Inamoto 2003	508	0.860	0.107	1542	0.861	0.099	1197	0.854	0.114
Kelemen 2004i	30	0.7874	0.2421	120	0.7272	0.1734	110	0.7443	0.1847
Kelemen 2004ii	20	0.6653	0.1540	84	0.6675	0.1179	171	0.6737	0.1311
Kelemen 2004iii	13	0.7852	0.2020	72	0.6906	0.1563	198	0.6947	0.1693
Durga 2005	125	1.03	0.16	378	1.02	0.17	312	1.02	0.16
McDonald 2005	24	0.586	0.099	117	0.665	0.149	60	0.641	0.142
Linnebank 2006	52	0.74	0.18	316	0.75	0.18	346	0.76	0.2
Liu 2007	31	0.77	0.18	220	0.75	0.16	290	0.75	0.23
NOS3		Glu/Glu			Glu/Asp			Asp/Asp	
Lembo 2001	158	1.24	0.44	179	1.27	0.52	38	1.29	0.36
Karvonen 2002i	244	0.915	0.214	220	0.959	0.311	41	0.878	0.173
Karvonen 2002ii	262	0.896	0.225	215	0.922	0.243	42	0.911	0.183
Paradossi 2004	43	0.37	0.07	57	0.35	0.08	18	0.45	0.13
Czarnecka 2005i	73	0.70	0.43	46	0.92	0.47	8	1.10	0.51
Czarnecka 2005ii	89	0.53	0.38	72	0.55	0.34	6	0.60	0.42
Spoto 2005	59	0.98	0.10	56	1.07	0.23	16	1.16	0.36
Wolff 2005	1218	0.79	0.35	1013	0.79	0.32	217	0.81	0.29
ADD1		Gly/Gly			Gly/Trp			Trp/Trp	
Castellano 1997	130	0.74	0.23	36	0.67	0.18	7	0.74	0.08
Balkestein 2002	213	0.551	0.156	142	0.572	0.182	25	0.611	0.177
Yazdanpanah 2006	3170	0.77	0.36	1668	0.78	0.44	245	0.76	0.37
PON1		QQ			QR			RR	
Sakai 1998	14	0.89	0.38	63	0.72	0.17	62	0.74	0.18
Dessi 1999	88	0.755	0.151	91	0.758	0.130	17	0.779	0.184
Visvikis 2000	165	0.51	0.05	169	0.51	0.05	28	0.49	0.03
Fortunato 2003	140	1.14	0.22	111	1.13	0.21	35	1.12	0.23

Study	N	mean	SD	N	mean	SD	N	mean	SD
Hu 2003	30	0.65	0.27	77	0.83	0.27	45	1.05	0.32
Karvonen 2004i	262	0.88	0.19	198	0.88	0.18	36	0.90	0.17
Karvonen 2004ii	273	0.87	0.17	198	0.87	0.19	32	0.86	0.23
van Himbergen 2005	110	0.88	0.19	146	0.85	0.16	29	0.9	0.17
Roest 2006	55	0.51	0.07	66	0.52	0.08	12	0.54	0.09
IL6		GG			GC			CC	
Rauramaa 2000	19	1.30	0.42	38	1.09	0.25	35	1.17	0.24
Rundek 2002	47	0.85	0.17	19	0.78	0.21	5	0.72	0.15
Chapman 2003	381	0.69	0.15	557	0.70	0.12	171	0.70	0.17
Jerrard-Dunne 2003	317	0.77	0.14	495	0.77	0.14	188	0.79	0.16
Mayosi 2005	265	0.98	0.29	422	0.98	0.26	136	1.10	0.26
IGF1		noncarriers			1 192bp allele			2 192bp alleles	
Schut 2003	617	0.78	0.14	2275	0.77	0.15	2240	0.76	0.14
ADRB2		Gln/Gln			Gln/Glu & Glu/Glu				
Hindorff 2005	1952	1.49	0.71	3221		1.50		0.68	
CRP*	1919A/T		2667G/C		3872G/A		5237A/G		790A/T
Lange 2006	P=0.48	P=0.88	P=0.33	P=0.29	P=0.46	P=0.33	P=0.13	P=0.65	P=0.12
FGG/FGA	Haplotype 1	Haplotype 2	Haplotype 3	Haplotype 4	Haplotype 5	Haplotype 6	Haplotype 7		
Kardys 2007	P=0.50	P=0.30	P=0.03	P=0.39	P=0.57	P=0.93	P=0.14		
AGT		MM			MT			TT	
Barley 1995	44	0.879	0.230	44	0.948	0.343	12	0.873	0.175
Arnett 1998	123	0.730	0.152	213	0.726	0.156	139	0.728	0.144
Jeng 1999	32	0.781	0.330	37	0.818	0.445	106	0.762	0.241
Tabara 2001	10	0.77	0.15	69	0.80	0.13	126	0.80	0.13
Bozec 2003	42	0.550	0.129	35	0.583	0.156	21	0.596	0.101
Islam 2006	76	0.59	0.08	86	0.60	0.09	40	0.57	0.09
FV		Wildtype factor V			Containing leiden mutation				
Garg 1998	1209	0.810	0.283	83		0.782		0.283	
Fox 2004	1692	0.6	0.17	71		0.58		0.08	

* there are 2 p-values for each CRP SNP because the association was tested in both white and black participants, except 790A/T which was only studied in black participants.

Of the others:

- One genes (IGF1) had been studied in only one study, and was analysed using ANOVA rather than meta-ANOVA.
- Two genes (ADRB2 & FV) had only been studied using a dominant genetic model and so I had to analyse them using this model.
- One gene (FGG/FGA), from only one study did not have SNP-specific data presented as only a haplotype analysis had been done. I simply reported the results from this haplotype analysis.
- One gene (CRP) did not have the necessary result data presented in the paper, but I could extract the association p-values.

Meta-ANOVA/ANOVA found no overall association between genotype and CIMT for six of the ten genetic polymorphisms for which this analysis was possible: NOS3, ADD1, PON1, IL6, AGT, and so these genes were not studied further. Overall significant associations ($p<0.05$) between genotype and CIMT were found for four of the ten genetic polymorphisms: APOE, ACE, MTHFR & IGF1. These genes went onto stage 2, where the most appropriate genetic model was estimated, and then in stage 3 the mean differences were estimated for the selected genetic models for each polymorphism. The estimated sizes of the effects of the polymorphisms are reported under the gene specific headings below.

As ADRB2 and FV had only been analysed in a dominant fashion, these genes went straight onto step 3 of determining the mean differences. The

Table 3.8 Results of the 3-step meta-analysis of the association between CIMT and polymorphisms in 13 selected genes.

Gene	Number of studies (subjects) in analyses	Step 1: meta-ANOVA	Step 2: λ (95% CI)	Step 2: Selected genetic model	Step 3: Random effects pooled mean CIMT difference between genotypes with selected model, μm (95% CI)
APOE (ε2,ε3,ε4)	32 (32253)	p<0.001	0.4 (0.3 to 0.6)	co-dominant	25 (17 to 33)
ACE (I/D)	34 (17038)	p=0.005	0.5 (0.4 to 0.6)	co-dominant	14 (5 to 22)
MTHFR (677 C/T)	15 (7945)	p=0.02	0.2 (0.1 to 0.4)	recessive/none	31 (0 to 61)
NOS3 (Glu298Asp)	8 (4390)	p=0.3	-	-	-
ADD1 (Gly460Trp)	3 (5636)	p=0.7	-	-	-
PON1 (Gln192Arg)	9 (2552)	p=0.6	-	-	-
IL6 (-174 G/C)	5 (3095)	p=0.4	-	-	-
AGT (Met235Thr)	6 (1255)	p=0.5			
IGF1 (192 bp allele)	1 (5132)	p=0.004	0.50 (no CI)	co-dominant	10 (4 to 16)
CRP (5 SNPs)	1 (4641)	No SNPs associated - p-values range from 0.12 to 0.88			
ADRB2 (Gln27Glu)	1 (5173)	study used dominant model		dominant	10 (-29 to 49)
FV (Leiden)	2 (3055)	both studies used dominant model		dominant	-20 (-29 to -12)
FGG/FGA (7 SNP haplotype)	1 (4274)	haplotype analyses: no significant association			-

mean difference for ADRB2 was not significant, but for FV a significant mean difference was found (further details below). The FGG/FGA study had only reported haplotype analyses, and found that there were no significant associations, so I did not study this gene further.

3.3.6 Apolipoprotein E Results

I found 30 relevant studies (36 sub-studies, 32,995 subjects) for the association between the APOE ε polymorphism and CIMT. After contacting authors, full data were still unavailable for three studies (4 sub-studies, 742 subjects), resulting in 2% missing data.

The meta-ANOVA of 32 sub-studies (32,253 subjects) found an overall association between APOE and CIMT, with a p-value of <0.001. The linear

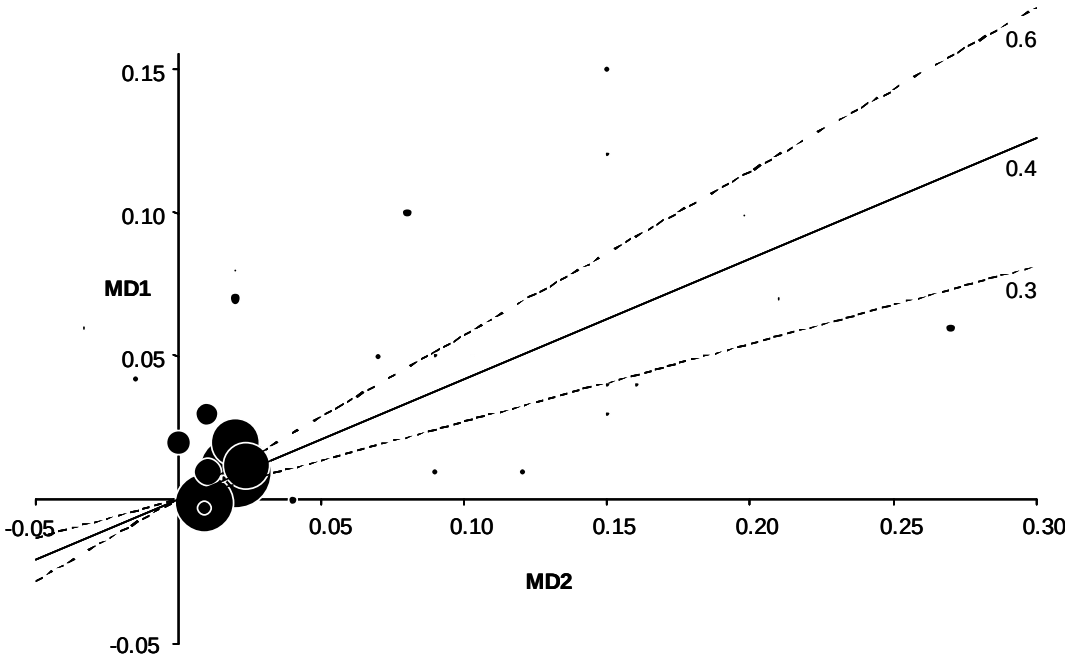


Figure 3.3. λ estimation for APOE using weighted linear regression, λ=0.4 (95% CI, 0.3 to 0.6).

regression to estimate λ is shown in figure 3.3. λ was estimated to be 0.4 (95% CI, 0.3 to 0.6), which is close to and has a CI including 0.5, suggesting that a co-dominant model is the most appropriate.

Figure 3.4 shows the forest plot of the co-dominant mean difference analysis. 'Co-dominant' in the case of APOE implies an equal difference between E4

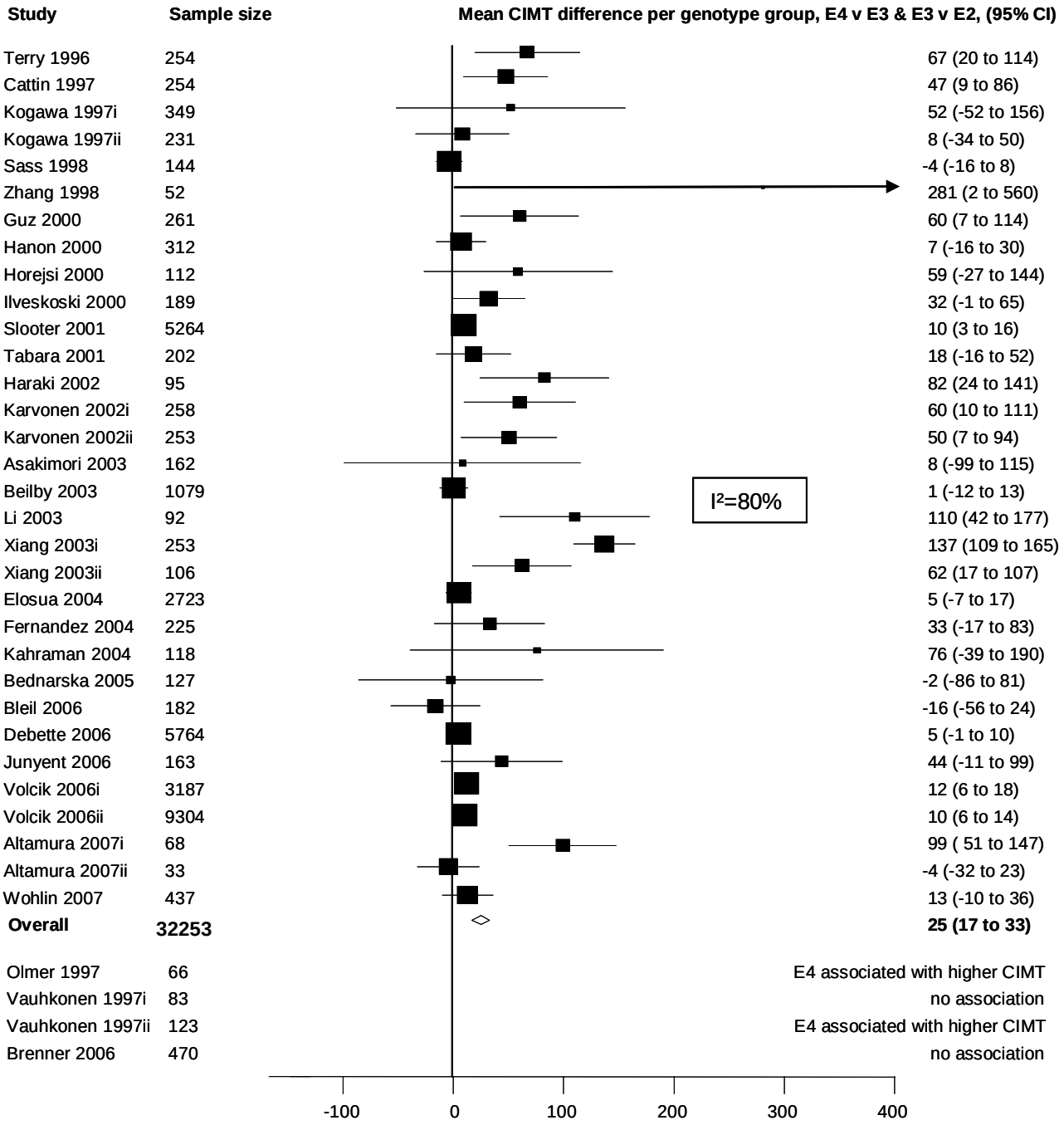


Figure 3.4 Study and pooled mean difference in CIMT between APOE genotype groups E4 and E3 and between E3 and E2, using a random effects method.

and E3, and between E3 and E2 genotypes, and the 'mean difference' estimates the size of this equal step-wise difference. The random effects pooled mean difference was 25μm (95% CI 17 to 33, p<0.001). When carrying out the same analysis using a fixed effects method the pooled mean difference was 10μm (95% CI 8 to 13, p<0.001) (not shown).

The I^2 estimate of heterogeneity amongst the studies was 80%, showing substantial heterogeneity beyond that expected by chance. The subgroup analyses shown in figure 3.5 go some way to explaining the sources of this heterogeneity. There was a substantially larger pooled mean CIMT difference amongst subjects of vascular high risk, compared with low risk subjects and in Eastern Asian subjects compared with White or Black subjects, for both comparisons the heterogeneity was statistically significant between subgroups (Q-test p<0.001 for both). There was significant heterogeneity between the small and large subgroups (Q-test, p<0.001),

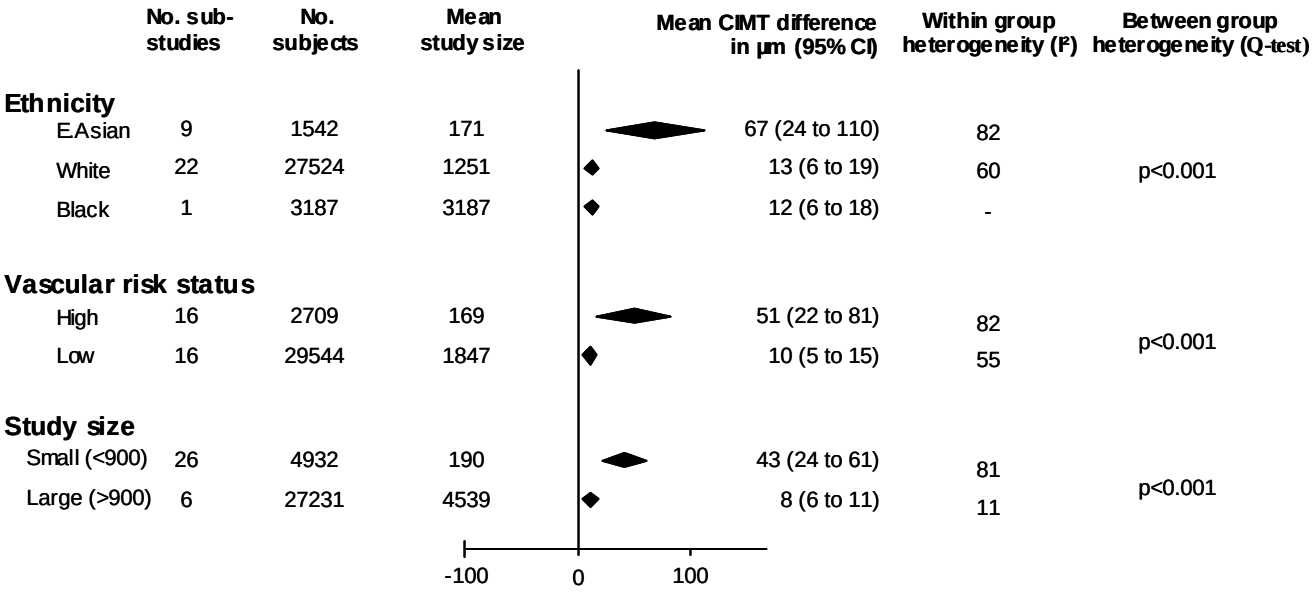


Figure 3.5 Subgroup sensitivity meta-analysis for APOE.

suggesting the presence of small study bias, with the smaller studies showing a more pronounced effect. Focusing on just the larger (and presumably more reliable) studies, a significant association between APOE and CIMT remains, and the excess heterogeneity (I^2) is reduced to 11%. The mean difference estimate is however smaller than the overall estimate, with a mean difference of 8μm (95% CI 6 to 11) per step from E2 to E3 to E4 genotype groups, suggesting that the biases in the literature lead to over-estimation of the effect.

Qualitative statements of the results from studies that could not be included in the meta-analysis are shown at the bottom of figure 3.4. Of the three studies with unavailable data for APOE, one found an association between E4 genotypes and higher CIMT [Olmer *et al.*, 1997], another found a similar association, but only in the non-diabetic subgroup [Vauhkonen *et al.*, 1997], and the other found no association [Brenner *et al.*, 2006]. All three of these studies were comparatively small (66, 206 and 470 subjects) and would not have contributed to the analysis including only larger subjects.

3.3.7 Angiotensin Converting Enzyme Results

I found 39 relevant studies (43 sub-studies, 20,105 subjects) for the association between the ACE I/D polymorphism and CIMT. After contacting authors, full data were still unavailable for nine studies (3067 subjects), resulting in 15% missing data.

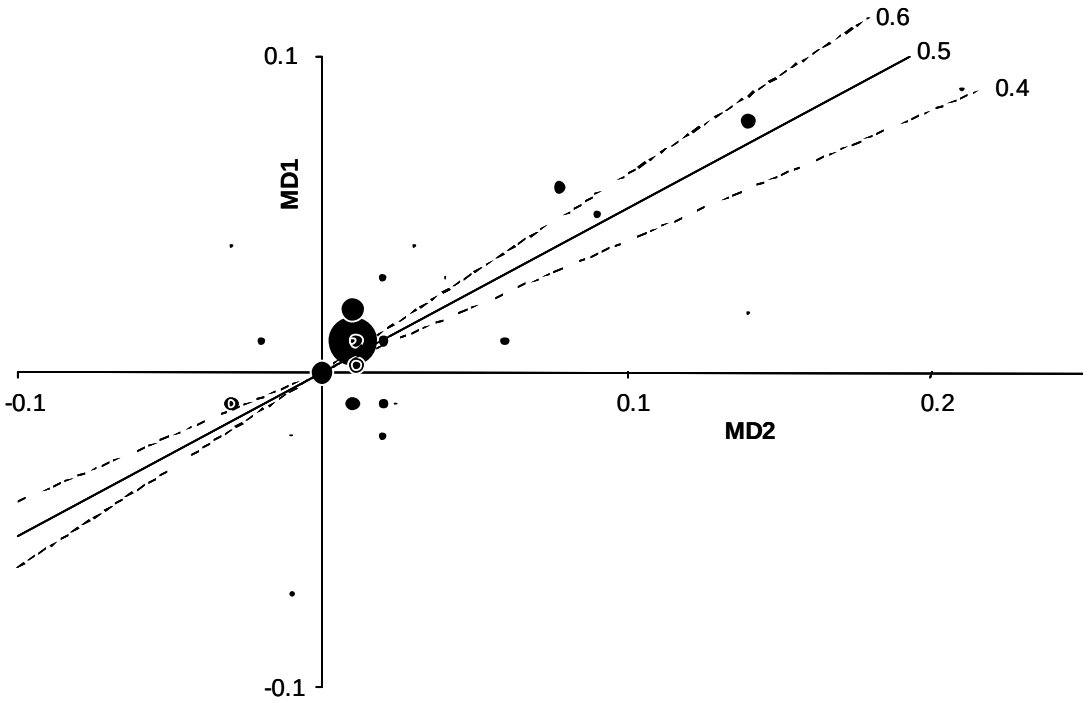


Figure 3.6 λ estimation for ACE using weighted linear regression, λ=0.5 (95% CI, 0.4 to 0.6)

The meta-ANOVA of 34 sub-studies (17,038 subjects) found an overall association between ACE and CIMT, with a p-value of 0.005. The linear regression to estimate λ is shown in figure 3.6 λ was estimated to be 0.5 (95% CI, 0.4 to 0.6), suggesting that a co-dominant model is the most appropriate.

Figure 3.7 shows the forest plot of the co-dominant mean difference analysis. The random effects pooled per-allele mean difference was 14μm (95% CI, 5 to 22, p=0.002). When carrying out the same analysis using a fixed effects method the pooled mean difference was 9μm (95% CI 6 to 12, p<0.001) (not shown).

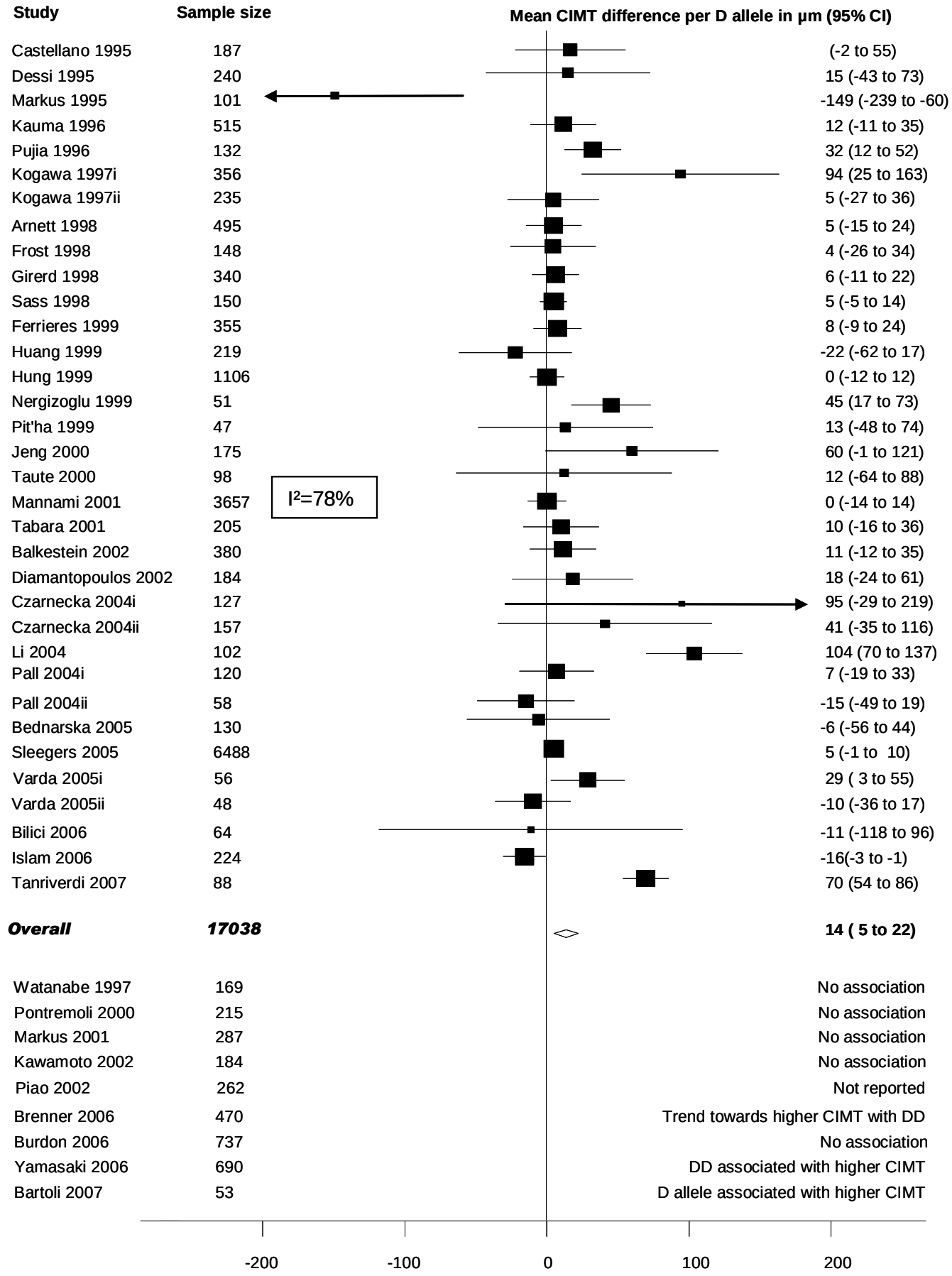


Figure 3.7 Study and pooled mean difference in CIMT per D allele of the ACE I/D polymorphism, using a random effects method.

The I^2 estimate of heterogeneity amongst the studies was 78%, showing substantial heterogeneity beyond that expected by chance. In a similar way to APOE, the subgroup analyses shown in figure 3.8 go some way to explaining the sources of this heterogeneity. There was a trend towards larger pooled mean CIMT difference amongst Eastern Asian subjects compared with White subjects, but the heterogeneity between subgroups was not statistically significant (Q-test, p=0.2). The pooled mean CIMT was larger in high vascular risk subjects compared to low risk subjects and the heterogeneity was statistically significant between subgroups (Q-test p<0.001).

There was significant heterogeneity between the small and large subgroups (Q-test, p=0.002), suggesting the presence of small study bias, with the smaller studies showing a more pronounced effect. On pooling just the larger (more reliable) studies, the association between ACE and CIMT becomes non-significant (p=0.08).

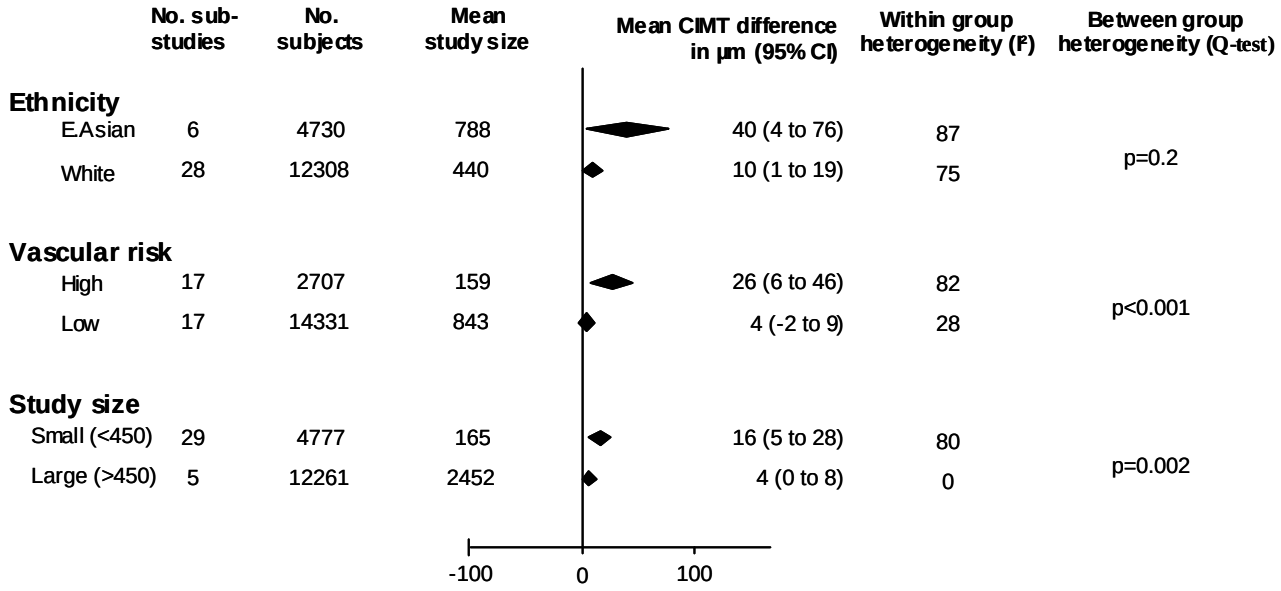


Figure 3.8 Subgroup sensitivity meta-analysis for ACE.

Qualitative statements of the results from studies that could not be included in the meta-analysis are shown at the bottom of figure 3.7. Most of the nine studies reported no association. Three would have contributed to the 'larger studies' analysis (> the mean of 450) [Brenner *et al.,* 2006; Burdon *et al.,* 2006; Yamasaki *et al.,* 2006]. Of these, two [Brenner *et al.,* 2006; Yamasaki *et al.,* 2006] found that the D allele was associated with increased CIMT, so their inclusion could potentially have strengthened the association between ACE and CIMT.

3.3.8 Methylenetetrahydrofolate Reductase Results

I found 20 relevant studies (22 sub-studies, 10,487 subjects) for the association between the MTHFR C677T polymorphism and CIMT. After contacting authors, full data were still unavailable for seven studies (2,542 subjects), resulting in 24% missing data.

The meta-ANOVA of 15 sub-studies (7,945) found an overall association between MTHFR and CIMT, with a p-value of 0.02. The linear regression to estimate λ is shown in figure 3.9. λ was estimated to be 0.2 (95% CI, 0.1 to 0.4). The CI range does not include any of the genetic model values of λ. It is however, closest to 0 and so I chose to carry out the mean difference analysis using a recessive model.

Figure 3.10 shows the forest plot of the recessive mean difference analysis. The random effects pooled mean difference was 31μm (95% CI 0 to 61, p=0.051). This is not quite statistically significant, despite there being a

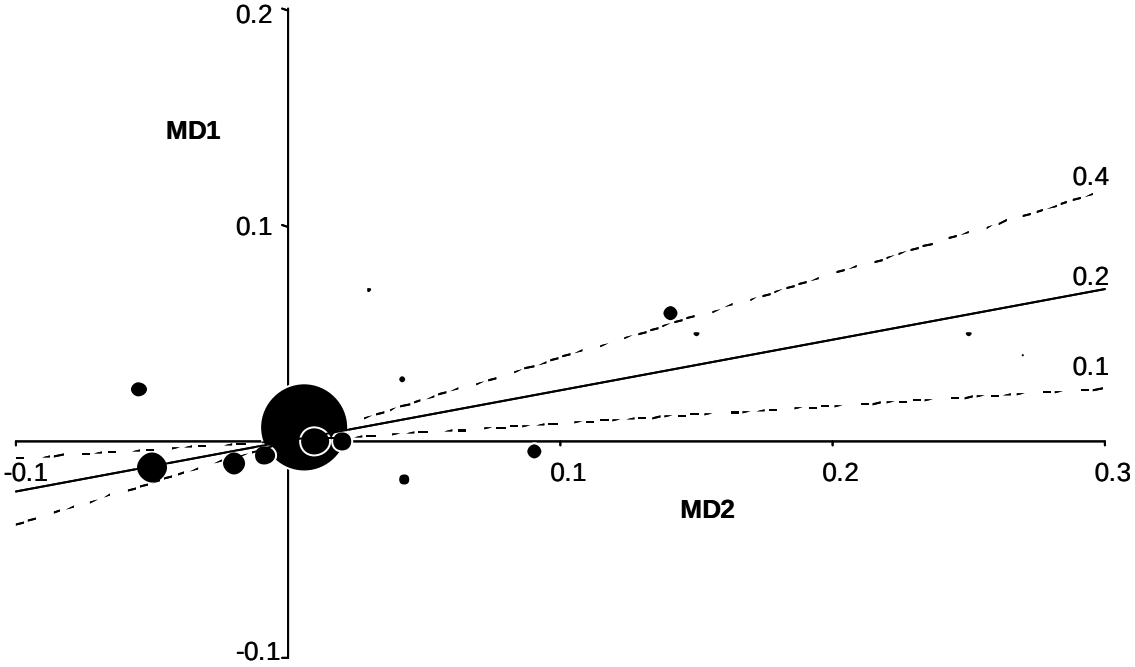


Figure 3.9. λ estimation for MTHFR using weighted linear regression, λ=0.2 (95% CI, 0.1 to 0.4)

significant overall association. When carrying out the same analysis using a fixed effects method the pooled mean difference was 2μm (95% CI -6 to 10, p=0.646) (not shown).

The I^2 estimate of heterogeneity amongst the studies was 83%, showing substantial heterogeneity beyond that expected by chance. In a similar way to APOE and ACE, the subgroup analyses shown in figure 3.10 go some way to explaining the sources of this heterogeneity. The trends seen are similar to those of APOE and ACE. Studies of Eastern and Southern Asian subjects showed a trend towards higher mean CIMT difference than those of white subjects, although the heterogeneity between subgroups was not statistically significant (Q-test, p=0.1). Studies of subjects at high vascular risk showed a higher mean CIMT difference than those in subjects of low risk; the

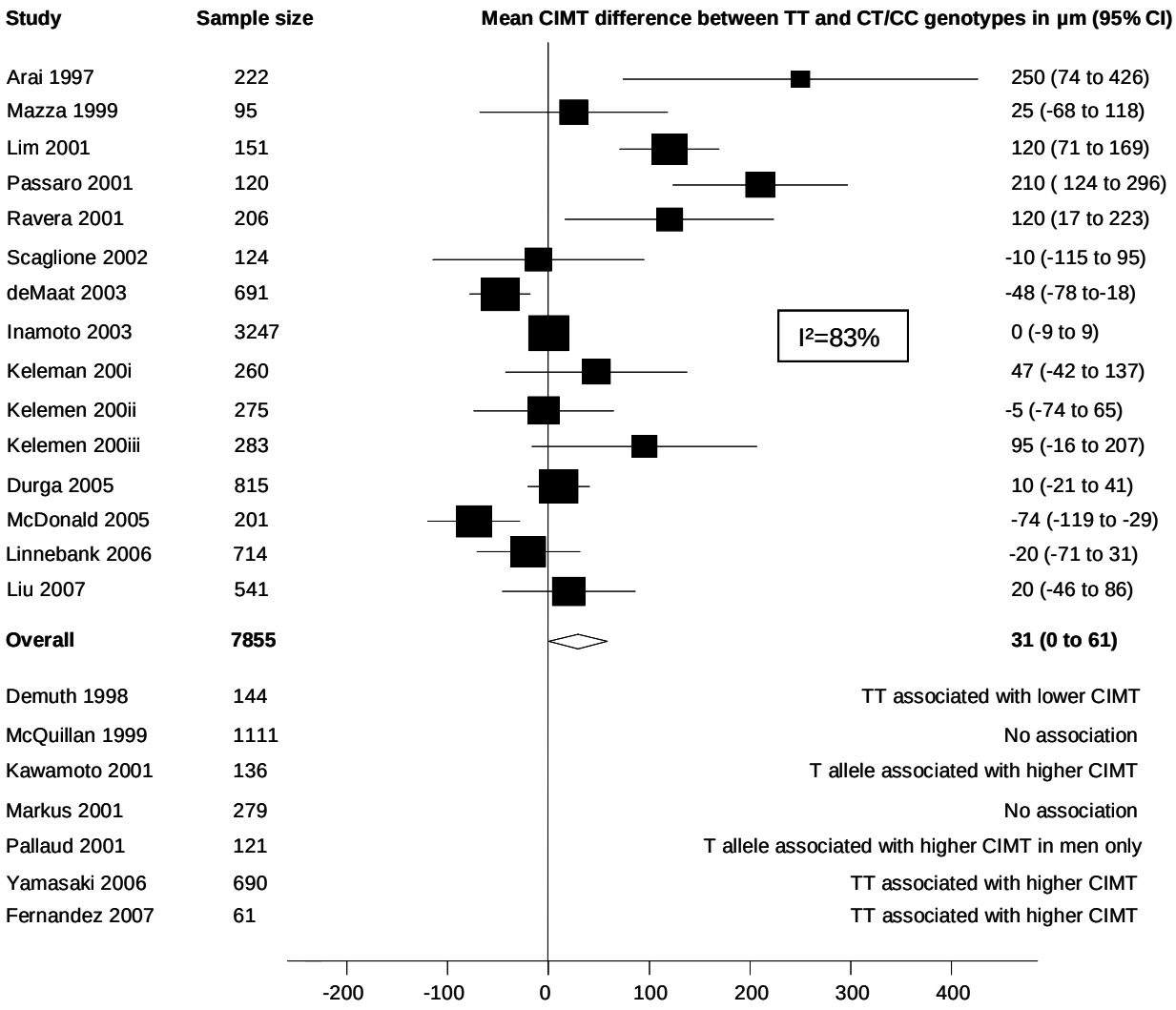


Figure 3.10 Study and pooled mean difference in CIMT between TT and CT/CC genotypes of the MTHFR C677T polymorphism, using a random effects method**.**

heterogeneity between subgroups was significant (Q-test, p=0.002). There was significant heterogeneity between the small and large subgroups (Q-test, p<0.001), suggesting the presence of small study bias, with the smaller studies showing a much more pronounced effect. On pooling just the larger (more reliable) studies, there was no significant association between MTHFR and CIMT and the trend was in the opposite direction to the overall result.

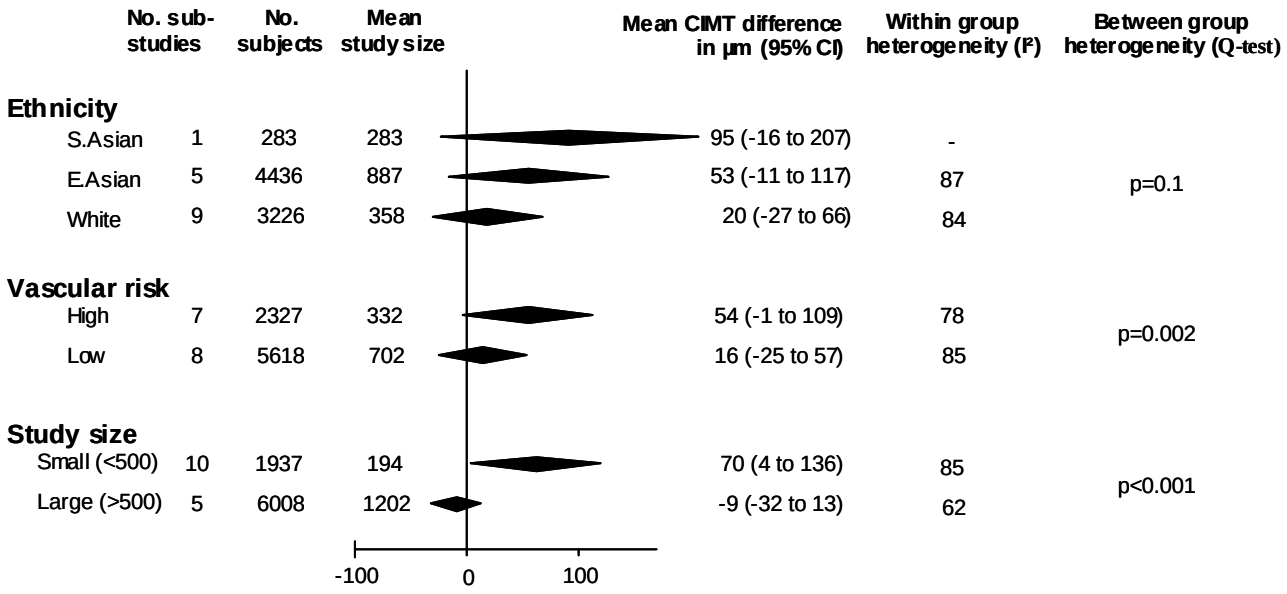


Figure 3.11 Subgroup sensitivity meta-analysis for MTHFR.

Qualitative statements of the results from studies that could not be included in the meta-analysis are shown at the bottom of figure 3.10. Of these seven studies, two would have contributed to the 'larger studies' analysis (> the mean of 500). The larger [McQuillan *et al.*, 1999] reported no association between MTHFR and CIMT while the other [Yamasaki *et al.*, 2006] did find an association, so overall their inclusion would be unlikely to greatly affect the overall conclusions.

3.3.9 Nitric Oxide Synthase 3 Results

I found 12 relevant studies (14 sub-studies, 7,475 subjects) for the association between the NOS3 Glu298Asp polymorphism and CIMT. After contacting authors, full data were still unavailable for six studies (3,085 subjects), resulting in 41% missing data.

The meta-ANOVA found no overall association between NOS3 and CIMT, with a p-value of 0.3, and so this gene was not analysed further.

One of the studies with missing data reported a marginal association between the T allele and CIMT [Bhuiyan *et al.*, 2007]. This paper had a sample size of 661, so its inclusion would be unlikely to greatly affect the overall conclusion. The other five 'missing' studies reported no association.

3.3.10 Adducin 1 Results

I found four relevant studies (6,056 subjects) for the association between the ADD1 Gly460Trp polymorphism and CIMT. After contacting authors, full data were still unavailable for one studies (420 subjects), resulting in 7% missing data.

The meta-ANOVA of three studies (5636 subjects) found no overall association between ADD1 and CIMT, with a p-value of 0.7.

The study with missing data [Sarzani *et al.*, 2006] had analysed the association according to a dominant model and found that the Trp allele was associated with an increased CIMT, but only in male subjects. This could be a spurious result or there could be a real interaction. Sex-specific data was not available from the other studies and so this could not be tested, and so this gene was not analysed further.

3.3.11 Paraoxonase 1 Results

I found 14 relevant studies (16 sub-studies, 4,651 subjects) for the association between PON1 Gln192Arg polymorphism and CIMT. After contacting authors, full data were still unavailable for six studies (seven sub-studies, 2,099 subjects), resulting in 45% missing data.

The meta-ANOVA of eight studies (nine sub-studies, 2,552 subjects) found no overall association between PON1 and CIMT, with a p-value of 0.6.

A large proportion of the relevant data were unavailable for this gene. Only one of the studies with missing data reported a significant association, but only in females [Srinivasan *et al.*, 2004]. Again, this could be due to an interaction effect or just spurious and there were insufficient data to test this in the whole dataset, so this gene was not analysed further.

3.3.12 Interleukin 6 results

I found seven relevant studies (4,595 subjects) for the association between the IL6 -174G/C polymorphism and CIMT. After contacting authors, full data were still unavailable for two studies (1,500 subjects), resulting in 33% missing data.

The meta-ANOVA of five studies (3,095 subjects) found no overall association between IL6 and CIMT, with a p-value of 0.4.

Of the two studies with missing data, one had only studied haplotypes (which included this gene and two other inflammatory genes – IL1 and CD14) and reported a statistically significant association between the gene-variant score and CIMT [Markus *et al.*, 2006] and the other found no association between IL6 and CIMT [Yamasaki *et al.*, 2006], and so this gene was not analysed further.

3.3.13 Angiotensinogen Results

I found eleven relevant studies (3,528 subjects) for the association between the AGT Met235Thr polymorphism and CIMT. After contacting authors, full data were still unavailable for five studies (2,273 subjects), resulting in 64% missing data.

The meta-ANOVA of six studies (1,255 subjects) found no overall association between AGT and CIMT, with a p-value of 0.5, and so this gene was not analysed further.

A large proportion of the relevant data were unavailable for this gene. However, none of these studies reported a significant association between AGT and CIMT and so they would seem unlikely to have altered the result shown here.

3.3.14 Insulin-like Growth Factor 1 Results

Only 1 study (5132 subjects) had analysed the association between the IGF1 192bp allele and CIMT [Schut *et al.*, 2003]. ANOVA found an overall

association between IGF1 and CIMT, with a p-value of 0.004. MD1/MD2 for this study gave a λ of 0.5, suggesting that the polymorphism is co-dominant. The co-dominant random effects per-allele mean difference was 10μm (95% CI, 4 to 16, p=0.001). Although this is a statistically significant association, it relies only on one study.

3.3.15 C-Reactive Protein Results

Only one study (4641 subjects) had analysed the association between CRP and CIMT [Lange *et al.,* 2006]. Five SNPs were studied in this gene. The results data were not available and could not be back-calculated, to enable ANOVA analysis for this paper. However, they reported that there was no association between any of these SNPs and CIMT (p-values ranging from 0.12 to 088).

3.3.16 Adrenergic Beta-2 Receptor Results

Only one study (1573 subjects) had analysed the association between the ADRB2 Gln27Glu polymorphism and CIMT [Hindorff *et al.,* 2005]. This study analysed the polymorphism in a dominant model, and only presented data for the two groups. I therefore calculated the mean difference between these two groups for this study and found that the dominant model mean difference was not significant, 10μm (95% CI -29 to 49, p=0.618).

3.3.17 Factor V Results

I found three relevant studies (3,525 subjects) for the association between the FV Leiden mutation and CIMT. After contacting authors, full data were still unavailable for one study (470 subjects), resulting in 13% missing data.

The two studies with available data both analysed the data using a dominant model, and only presented data for the corresponding two groups. I therefore calculated the random effects pooled mean difference between FV Leiden-positive and -negative subjects. There was a significant association, -20μm (95% CI, -29 to -12, $p<0.001$), suggesting the FV Leiden mutation may be associated with a decrease in CIMT, despite neither study reporting an independently significant result. The study with unavailable data [Brenner *et al.*, 2006] found no association between the FV Leiden mutation and CIMT. Since it was much smaller than the two included studies, its inclusion would be unlikely to have greatly altered the overall result.

3.3.18 Fibrinogen Gamma/Fibrinogen Alpha Results

Only one study (4,274 subjects) had analysed the association between the FGG and FGA gene polymorphisms [Kardys *et al.*, 2007]. However, this study only carried out a haplotype analysis of both genes (including 7 SNPs) and did not report individual SNP distributions in relation to CIMT phenotypes. Therefore, it was not possible to carry out a straightforward association analysis on these data. However, the paper reported that no haplotypes were significantly associated.

3.3.19 Comparison of Sub-Group Analyses for APOE, ACE & MTHFR

Figure 3.12 shows the subgroup meta-analyses for APOE, ACE and MTHFR, grouped by analysis. This shows that all three polymorphisms show very similar subgroup patterns. The studies of Asian subjects consistently had larger pooled mean differences than White or Black American populations. The difference between subgroups was highly significant for APOE ($p<0.001$), but not significant for the other two polymorphisms (ACE $p=0.2$; MTHFR $p=0.1$). Studies of subjects at high vascular risk consistently showed larger pooled mean differences. The difference between subgroups was significant for all polymorphisms (APOE $p<0.001$; ACE $p<0.001$; MTHFR $p=0.002$). Smaller studies, also consistently showed larger pooled mean differences. The difference between subgroups was significant for all polymorphisms (APOE $p<0.001$; ACE $p=0.002$; MTHFR $p<0.001$). The larger pooled mean differences amongst smaller studies are suggestive of small study bias. For each polymorphism, there was less heterogeneity between the larger studies (I^2 values are between 80 and 85 for small studies and between 0 and 62 for large studies). The results from the ethnicity and vascular risk status subgroup analyses may suggest that there is an interaction effect with these factors. However, the high risk studies that show significantly larger pooled differences were smaller than the low risk studies for all polymorphisms (mean sample sizes for APOE, ACE and MTHFR were 169, 159 and 332 respectively for the high risk studies and 1847, 843 and 702 for the low risk studies) and for APOE the studies of East Asian subjects had a smaller mean sample size (171) than the studies of White subjects (1251). Therefore, study size may explain the apparent differences seen for ethnicity and vascular risk status.

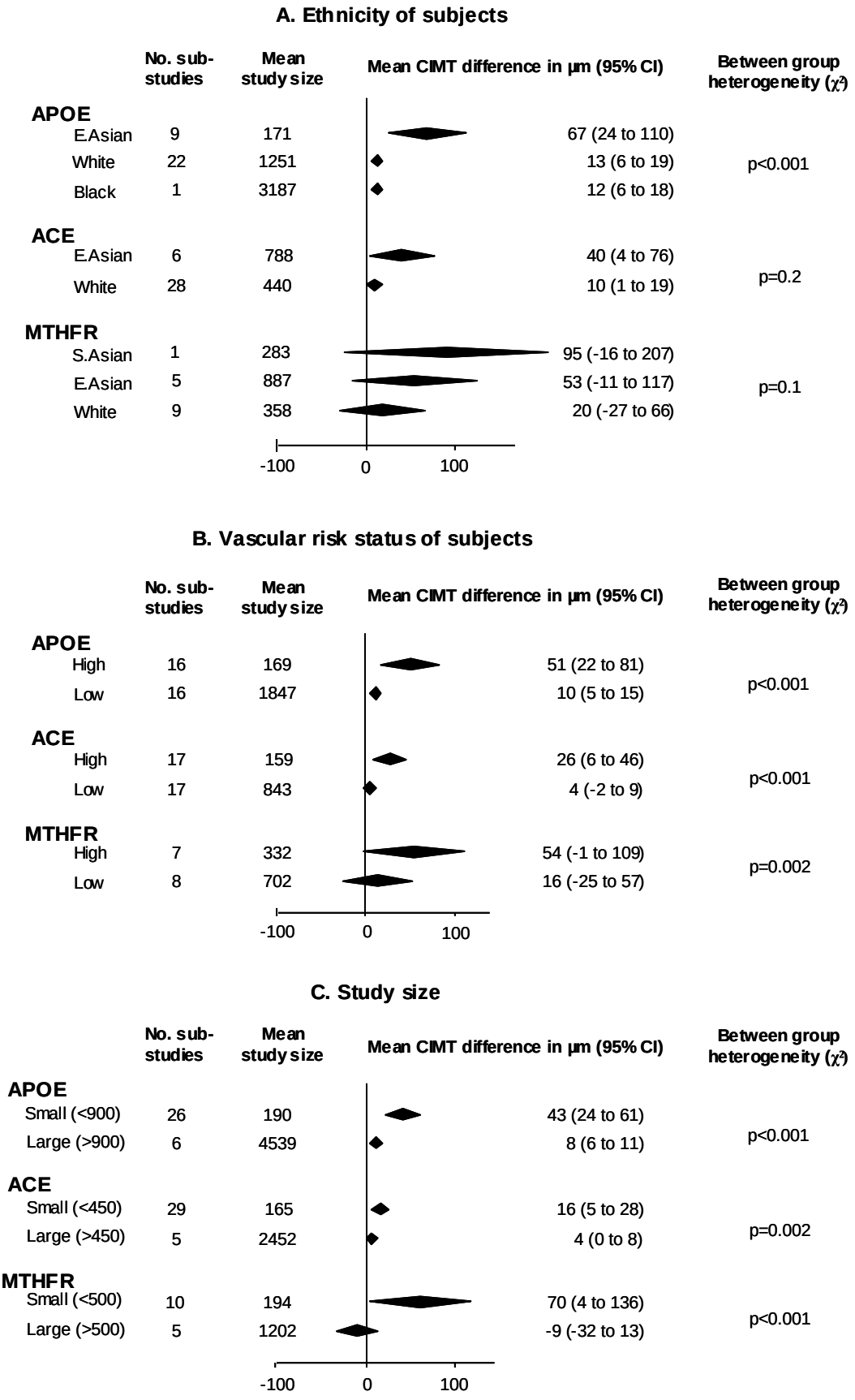


Figure 3.12 Subgroup meta-analyses for APOE, ACE and MTHFR. A. shows the ethnicity subgroup analysis. B. shows the vascular risk status subgroup analysis. C. shows the study size subgroup analysis where for each gene the mean study size was used to split into small and large. For each gene the appropriate genetic model was used. Co-dominant for APOE and ACE and recessive for MTHFR**.**

Table 3.9 % of missing data, meta-ANOVA p-values and mean differences before and after the acquisition of extra data by contacting authors.

Gene	% of data missing		meta-ANOVA p-value		Meta-analysis mean difference	
	before	after	before	after	before	after
APOE	5	2	<0.001	<0.001	22 (14 to 30)	25 (17 to 33)
ACE	16	15	0.01	0.01	14 (5 to 23)	14 (5 to 22)
MTHFR	41	24	0.01	0.02	54 (16 to 91)	31 (0 to 61)
NOS3	45	40	0.3	0.3		
PON1	64	45	0.3	0.6		
IL6	51	33	0.7	0.4		

3.3.20 Minimising Bias by Obtaining Unpublished Data

I determined the impact of attempting to collect missing data by comparing the before and after data collection percentages of missing data, and by carrying out the same analyses without the extra acquired data to see if different conclusions would have been drawn had I not included these papers. Results are shown in table 3.9. The percentage of missing data for some polymorphisms reduced hugely after acquisition of extra data. However, in most cases this did not have a significant effect on the results. The only polymorphism for which the overall conclusions would have been different is MTHFR, which showed a stronger association with CIMT before the extra data were collected, and had a mean difference that was much larger and a confidence interval that did not include 0 (54μm, 95% CI 16 to 91).

3.3.21 Other Potential Genes of Interest

There were 132 other genes studied in smaller numbers, which did not make it through to the meta-analysis stage of this review. Many of these showed preliminary evidence for an association, but these findings would need to be replicated in much larger samples before the results could be considered reliable.

3.4 Discussion

Through carrying out a large systematic search, I identified more than 140 genes that had been studied in association with CIMT. I then reviewed in detail 122 studies (112,713 subjects) that had analysed the association between CIMT and the 13 most commonly studied genes. Most of these did not show convincing evidence for an association. APOE, ACE and MTHFR all showed a significant association with CIMT in the meta-ANOVA analysis. Of these, APOE ε was the only polymorphism that still showed an association when the analysis was restricted to larger studies only. The results suggest that although there is an association between APOE and CIMT, the size of this association is over-estimated in the literature, due to small study bias. I also found significant associations with CIMT of IGF1 and FV Leiden, but as they were only analysed in a few studies, these findings are still preliminary and warrant further investigation.

3.4.1 Meaning of Effect Size

The pooled estimate for the per-group mean difference between E4 and E3, and E3 and E2 groups across all studies was 25μm (95% CI 17 to 33). When restricted to only the larger studies this estimate dropped to 8μm (95% CI 6

to 11). Whilst still significant this is a rather small difference (equivalent to approximately 0.05 of one standard deviation)

3.4.2 Effect of Sample Size

Despite large numbers of studies (with hundreds of thousands of subjects) assessing the associations between candidate genes and CIMT, very few firm conclusions can be made. However, this may not be surprising due to the nature of complex traits. CIMT is a complex trait with many possible environmental and genetic risk factors, and it is likely that any genes that are associated will show only modest effects. This may explain why APOE was the only gene with robust enough data to still show a significant effect in the large studies. If the effect size is very small, then the number of subjects required for the test of association to be powerful enough is extremely large and perhaps APOE was the only gene with sufficient numbers. The meta-analysis of APOE was the largest, comprising of 32,253 subjects. The overall number of subjects in the 'large studies' APOE subgroup was 27,231, much larger than the 12,261 for ACE or 6,008 for MTHFR. This is important for informing new studies. Probably, many tens of thousands of subjects in very well carried out studies are required before effects as small as that seen for CIMT can be properly identified.

3.4.3 Subgroup Analyses

The subgroup analyses for all three polymorphisms (APOE, ACE & MTHFR) show similar trends. Asian subjects and subjects at high vascular risk tended to have greater mean differences than white subjects and subjects at low vascular risk, across all polymorphisms. However, Asian ethnicity and

vascular high-risk status correlate with sample size, i.e. high-risk subjects and Asian subjects tend to be studied in smaller studies than population volunteers and white subjects. Therefore, when there is heterogeneity between studies it is difficult to determine the cause. It may be the case that there is an interaction effect with vascular risk or ethnicity, or perhaps the heterogeneity observed is caused by small study bias and it just so happens that Eastern Asian and high vascular risk subjects tend to be studied in smaller numbers. As this phenomenon is observed for all polymorphisms it is more likely that it is caused by small study bias rather than real interactions. In a previous meta-analysis of ACE and CIMT [Sayed-Tabatabaei *et al.,* 2003] (which is updated in the present study), the increased size of the association in high risk subjects was attributed to an interaction with smoking after further investigation [Sayed-Tabatabaei *et al.,* 2004]. Whilst this is possible, their investigations do not rule out the possibility that it is just due to small study bias.

3.4.4 Genetic Model Selection

For each polymorphism with an overall association, I chose the best genetic model using a linear regression method. For APOE and ACE, the analysis suggested a co-dominant model should be used. These are likely to be correct as APOE has been shown to follow a linear genetic model for lipid levels and coronary risk [Matsuoka *et al.,* 2000] and serum levels for ACE follow a co-dominant genetic model [Rigat *et al.,* 1990]. For MTHFR the estimated λ was 0.2 (95% CI ranging from 0.1 to 0.4). This does not include any of the assumed genetic models (recessive=0, co-dominant=0.5, dominant=1), suggesting that none of these models are appropriate for this

polymorphism. It is biologically feasible that λ could be equal to 0.2 in a (not completely) co-dominant model and this result demonstrates how genetic studies are limited by assuming certain genetic models. For two polymorphisms (ADRB2 and FV) the data available from publications only allowed me to analyse the associations using dominant models. However, it is likely that this is the most appropriate model as it was chosen by the paper's authors because it is the accepted genetic model for these polymorphisms.

3.4.5 Missing Data

I attempted to obtain all relevant data by contacting authors when important data were unavailable from the publications. Despite this, I was unable to collect full data for a large number of studies and for some polymorphisms the majority of studies. The large proportion of unavailable data highlights the impact of not collecting all data. Many systematic reviews include in their selection criteria only papers with available data. I have shown this can miss a large proportion of the relevant data. This probably introduces bias (known as reporting bias), as papers which do not fully report the data may not have found an association and so any estimates from a meta-analysis may over-estimate the association. I aimed to minimise this bias by consistently reporting the overall qualitative results from these studies alongside the meta-analyses. For most polymorphisms, I found that these missing studies would be unlikely to change the results significantly. However, two large studies for ACE with missing data found significant associations and so may have strengthened the association between ACE and CIMT.

The comparison from before and after the additional data were acquired from some authors showed similar results for most polymorphisms. The only polymorphism for which different conclusions would have been drawn before the extra data collection was MTHFR. The meta-ANOVA p-value was smaller (p=0.01, compared to 0.02 after) and the recessive mean difference was 54μm (95% CI 16 to 91) before and 31μm (95% CI 0 to 61) after. Therefore, without the inclusion of the extra data one would have concluded that there was a clear association between MTHFR and CIMT. However, after inclusion of the extra data the influence of MTHFR on CIMT is much less clear. Even for the polymorphisms for which the results did not significantly change, the inclusion of the extra data was important as it allows more of the data of interest to be assessed and so removes some of the potential bias that only including published results can cause.

3.4.6 Linkage Studies

Two linkage studies have identified quantitative trait loci for CIMT. Although none have replicated any of the candidate gene findings, they have identified some potential novel genes for CIMT. One reported a maximum log odds (LOD) score of 4.1 at 161cM on chromosome 12, and subsequently found evidence of association with an atherosclerosis candidate gene (SCARB1, a high density lipoprotein receptor, cell-surface glycoprotein) from the region of linkage [Fox *et al.*, 2004a]. The other identified 2q33-35 as a region with significant linkage (LOD=3.08), including the NOSTRIN, IGFBP2 and IGFBP5 genes, none of which have yet been independently tested for an association with CIMT [Wang *et al.*, 2005].

3.4.7 Missing Heritability

CIMT seemed to be a promising candidate as an intermediate trait for studying the genetics of stroke. It seems to correlate with and predict stroke risk and is highly heritable, so why has studying it been so far fruitless in identifying associated genes? Perhaps the initial estimates of heritability are over-estimated, there are many genes that influence CIMT and each have a very small undetectable effect, and/or the 'wrong' genes or polymorphisms could have been studied as candidates to date.

There is evidence that APOE may be specifically associated with large artery stroke and so one would perhaps expect an association with CIMT. Despite a significant overall association being found in the meta-analysis (even after restricting to only the larger studies), it is smaller than expected if APOE acts through CIMT to have an influence on susceptibility to stroke. It may be the case that there are other stroke pathways in addition to the CIMT pathway that APOE affects. If that was the case then instead of analysing CIMT increasing the power to detect an association with APOE compared with stroke, it may decrease the power.

CIMT may follow the trend of other genetic studies for complex traits. Recent extremely large genome wide association studies have identified genetic association with important genes for complex traits such as diabetes, despite the candidate gene studies for these traits being relatively fruitless [The Wellcome Trust Case Control Consortium, 2007]. The 'big players' may

emerge from these huge genome wide association studies simply because they have not been previously studied as candidates.

4 WMH Systematic Review and Meta-analysis

This chapter comprises a systematic review and meta-analyses of the most commonly studied genetic polymorphisms in association with white matter hyperintensities.

4.1 Introduction

4.1.1 White Matter Hyperintensities

White matter hyperintensities (WMH) are changes of the white matter in the brain which show up as hyperintensities (increased signal intensities) on MRI (magnetic resonance imaging) or hypointensities on CT (computed tomography) [Fazekas *et al.,* 2002]. Small amounts of WMH are thought to be the consequence of normal ageing [Awad *et al.,* 1986]. These changes in the brain are often asymptomatic, but it has been shown that the presence of WMH is associated with a history of, and later progression to small-vessel

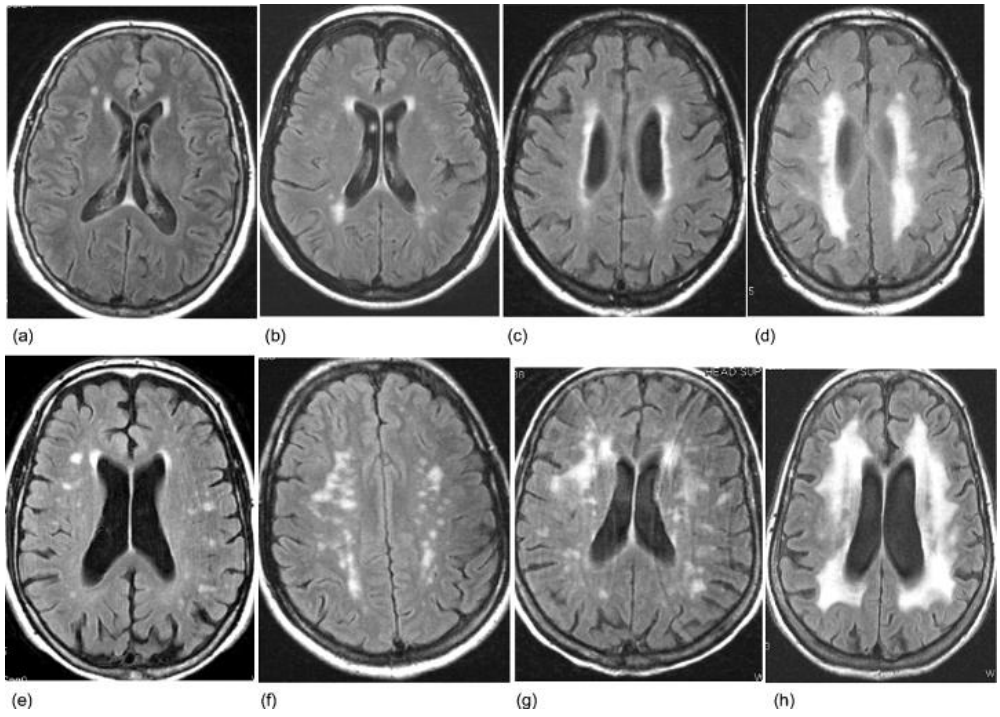


Figure 4.1 Taken from Bronge et al. [1999], showing various types/stages of white matter hyperintensities (WMH). (a) small 'caps' adjacent to the frontal horns. (b) pronounced caps next to the frontal and posterior horns. (c) periventricular bands. (d) pronounced periventricular bands extending into deep WM. (e) punctuate deep WMH. (f) and (g) punctuate and patchy deep WMH. (h) confluent deep WMH.

disease (lacunar infarcts) and clinical small artery stroke [Leys *et al.,* 1999].

WMH are more prevalent in patients with lacunar ischemic stroke than in those with other stroke subtypes [Wiszniewska *et al.,* 2000] and so are considered a useful quantitative trait for studying small artery (lacunar) ischemic stroke [Dichgans & Markus, 2005]. It has been suggested that lacunar infarction associated with WMH may reflect one subtype of small vessel disease pathology, with isolated lacunar infarction being the other subtype and having a different underlying vascular pathology [Markus, 2008]. WMH are commonly seen in normal ageing, with prevalence estimates between 10% and 100% in different elderly populations [Bronge *et al.,* 1999]. WMH are more prevalent and severe in patients with cardiovascular disease and cardiovascular disease risk factors [Meyer *et al.,* 1992]. Figure 4.1 shows the various stages and types of WMH.

4.1.2 Definitions

The term 'WMH' describes the phenomenon of a hyperintensity on an MRI scan in an area of white matter within the brain. Several other terms used in the literature also describe WMH. 'Leukoaraiosis', 'white matter changes' ('WMC') and 'white matter lesions' ('WML') are all broad terms that describe disease of the white matter, whether benign changes seen with normal ageing or changes associated with stroke, dementia or other diseases. Often individual papers will use these terms with their own definition to describe a particular type of WMH. 'Age-related white matter changes' ('ARWMC')

specifically refers to the benign changes seen with normal ageing. For consistency I use the term WMH throughout this thesis.

4.1.3 Measurement Methods

As well as heterogeneity of terminology used to describe WMH, there is also heterogeneity in how these changes are measured. WMH can be graded according to scales of hyperintensity severity or the volume of the 'lesion' can be estimated.

- Methods which aim to estimate the volume of WMH are appealing because they provide an objective quantitative measure of the changes seen in the brain, which may in turn increase statistical power when testing for associations with WMH. However, the measurement stage is time consuming and requires expensive equipment and high quality MRI protocols [Fazekas *et al.*, 2002]. In addition, the WMH volume may not fully represent the clinical impact, since distribution and location are also of importance.
- Grading scales have been developed to categorize the severity of WMH semi-quantitatively. There are many scales in use; some for use with CT, some for use with MRI and some for use with either [Scheltens *et al.*, 1998]. All scales rate WMH according to extent and severity, some rate periventricular hyperintensities (PVH) and deep white matter hyperintensities (DWMH) separately and some rate different parts of the brain separately. Three commonly used rating scales are shown in table 4.1. Many more grading scales also exist,

Table 4.1. Three commonly used grading scales for white matter hyperintensity (WMH)

Name/Author	Scan	PVH/DWMH	Grades (numbers and definitions)	Areas scored	Total score
Fazekas [Fazekas *et al.,* 1987]	MRI	PVH	0= absence 1= 'caps' or pencil-thin lining 2= smooth 'halo' 3= irregular PVH extending into the DWM	PVH as whole	0-3
		DWMH	0 = absence 1= punctuate foci 2= beginning confluence of foci 3= large confluent areas	DWMH as whole	0-3
Scheltens [Scheltens *et al.,* 1993]	MRI	PVH	0= absent 1= ≤ 5mm 2= > 5mm and <10mm	scored separately & summed: Occipital 'caps' Frontal 'caps' Lateral ventricle 'bands'	0-6
		DWMH	0= absent 1= < 3mm, n ≤5 2= <3mm, n>6 3= 4-10mm, n≤5 4= 4mm, n≤5 5= >11mm, n>1 6= confluent	scored separately & summed: Frontal Parietal Occipital Temporal	0-24
ARWMC [Wahlund *et al.,* 2001]	Both	Both	0= no lesions 1= focal lesions 2= beginning confluence of lesions 3= diffuse involvement of the entire region	one score for both PVH & DWMH	0-3

including simply dichotomising patients into those with and without WMH (often with very different cut-offs). Almost every study uses a subtly different scale, having imposed their own slight modification on the popularly reported scales. Scheltens *et. al.* [1998] provides an overview of most of these scales. They conclude that many scales lack reproducibility and that 'the ideal rating scale does not yet exist'.

4.1.4 Heritability

The NHLBI twin study, amongst 74 monozygotic and 71 dizygotic World War II veteran twins, showed that 0.71 of the variability of WMH volume was due to additive genetic influences (after correcting for age and head size [Carmelli *et al.*, 1998]. This is surprisingly high for a late developing condition, but twin studies tend to over-estimate genetic effects and this study was small and limited to an older population at probable high cerebrovascular risk. However, in 2004, two studies confirmed the high heritability of WMH volume. The Genetic Epidemiology Network Of Arteriopathy (GENOA) carried out a study in 483 subjects that were part of 434 hypertensive sibling pairs. They reported a heritability of 0.80 amongst this sample [Turner *et al.*, 2004]. Also in 2004, the Framingham Study estimated that 0.55 of the variability of WMH volume was due to additive genetic effects (after adjusting for sex, age and cranial volume) in a large population based study (n=1330), spanning a broad age range [Atwood *et al.*, 2004].

4.1.5 Genetic Associations

As heritability estimates for WMH have been consistently moderate to high it is possible that there are individual polymorphisms which have a reasonably large effect on this trait. Many studies have analysed the association between candidate gene polymorphisms and WMH. Most of these have been small with apparently conflicting results, and are difficult to interpret in isolation.

4.1.6 Aims

I aimed to bring together all studies of the association between any genetic polymorphism and WMH and to perform detailed methodological assessments and meta-analyses of studies of polymorphisms studied in sufficiently large numbers of subjects to make this appropriate. My intention was to provide an up-to-date summary of what is so far reliably known about genetics of WMH and which genes have been studied. By pooling studies, the power to detect associations can be increased and potential reasons for heterogeneity of study results can be explored.

4.2 Methods

4.2.1 Initial Search Strategy

I sought all papers describing studies of the association between any gene and WMH using a comprehensive search strategy in Medline (1966 to end 2007) and Embase (1980 to end 2007), using all MeSH terms and textwords associated with WMH and all MeSH terms and textwords associated with genetics (see table 4.2)

Table 4.2 Electronic literature search strategies.

Medline (1966 to end 2007)

	Search Terms
1	exp genetics/ or exp genotype/ or exp inheritance patterns/ or exp "linkage (genetics)"/ or exp phenotype/ or exp "variation (genetics)"/ or chromosomes/ or exp genes/ or exp genome/
2	(polymorphi$ or genotyp$ or gene or genes or genetic$ or allel$ or mutat$).tw.
3	1 or 2
4	exp Leukoaraiosis/
5	(leukoaraiosis or leucoaraiosis or leukoariosis or leucoariosis or MARCD or microangiopathy related cerebral damage or microangiopathy-related cerebral damage).tw.
6	(white matter lesion$ or WML or white matter hyperintensit$ or WMH or white matter change$ or small vessel disease or small-vessel disease or microangiopath$).tw.
7	4 or 5 or 6
8	7 and 3
9	limit 8 to humans

Embase (1980 to end 2007)

	Search terms
1	exp genetics/ or exp heredity/ or exp genetic disorder/ or genetic epidemiology/ or exp genetic analysis/ or exp population genetic parameters/ or quantitative trait/ or exp molecular genetics/ or exp genetic parameters/ or exp gene mapping/
2	(polymorphi$ or genotyp$ or gene or genes or genetic$ or allel$ or mutat$).tw.
3	1 or 2
4	exp LEUKOARAIOSIS/
5	(leukoariosis or leucoariosis or leukoaraiosis or leucoaraiosis or white matter lesion$ or WML or white matter hyperintensit$ or WMH or MARCD or microangiopath$ or white matter change$ or small vessel disease or small-vessel disease).tw.
6	4 or 5
7	3 and 6
8	limit 7 to human

I read the titles of all studies identified from the search and excluded any papers that were obviously not relevant, then read the abstracts of all remaining studies and retained all potentially relevant studies (any original study of the association between any gene and WMH). A second person (Wanting Chen) independently identified relevant papers from these searches. I compared our two lists and compiled a final list of relevant studies. Disagreements were resolved by discussion and where necessary by consultation with a third person (Cathie Sudlow). I listed all genes that have been studied in association with WMH and calculated the number of studies and subjects for each gene.

4.2.2 Genes/Studies Selected for Meta-Analysis

Any gene for which the initial search identified more than 2000 subjects studied was carried forward to formal meta-analysis. To ensure that all relevant papers were identified I carried out supplementary gene specific searches in Medline and Embase (replacing the general genetics terms with gene name terms) (see appendix 6). Again, a second person (Wanting Chen) independently identified the relevant papers from these searches and disagreements were resolved by consultation with Cathie Sudlow. I also checked the reference lists of the identified papers for further studies. All studies that had measured the volume or grade of WMH in any area of the brain were included. Studies with all types of subjects, including those with prior stroke were included. Papers in all languages were sought. Where studies appeared to use overlapping subject samples only the largest (with data available) were included in the analyses.

4.2.3 Data Extraction

I designed a data extraction form (appendix 7), which two independent observers (myself and Wanting Chen) used to extract the following data from each study identified as potentially relevant:

1. First author and year of publication
2. Study name or research group name (if applicable)
3. Number of subjects
4. Gene and polymorphism studied
5. Definition of WMH and the measurement method used
6. Country in which the study was conducted
7. Genotyping method
8. Whether genotyping was done blind to WMH assessment and vice versa
9. Subject demographics; age, sex, ethnicity, whether from a particular patient group (e.g. patients with hypertension) or population sample/healthy volunteers.
10. Concordance of genotypes with Hardy-Weinberg equilibrium (and I calculated this directly where possible)
11. Results (see below for alternative forms)

Results were presented in three alternative forms in the papers:

- WMH volume measured – mean & SD reported
- WMH graded – mean & SD reported
- WMH graded – numbers of subjects in each grade reported

I analysed these three types of data separately. For the studies where WMH was graded and numbers of subjects counted, the studies either reported the number of subjects in several grades or chose a particular grade to be the cut-off and analysed the number of subjects above and below this cut-off for each genotype. Where more than two groups were reported I chose as close to the following as possible for the cut-off: DWMH that were early confluent or confluent (Fazekas scale 2 or 3, or equivalent) were included in the upper group and only PVH that were classed as irregular (Fazekas scale 3 or equivalent) were included in the upper group. This cut-off is the most commonly used and so reduced heterogeneity between studies.

Where results were presented separately for several different brain locations I selected data from the deep white matter sub-scale only; this allowed the most consistent comparison across studies. Where possible, the studies that included different groups of subjects were treated as separate sub-studies, for example those with and without dementia, hypertension or cerebral infarcts on brain imaging were separated for the purpose of the meta-analyses.

4.2.4 Data Manipulation

Where papers did not present data in the required format I had to carry out transformations of the data. Some of these were similar to those for CIMT and I refer the reader to section 3.2.5. Some specific manipulations of the data such as estimating values from graphs and back-calculating from odds ratios are presented in appendix 8.

4.2.5 Statistical Analysis

Most studies presented the data in either a dominant or recessive model, so for each polymorphism I analysed the data according to the most widely used genetic model from amongst the included studies. The most commonly used model is generally the most biologically appropriate. Furthermore, this approach allows the maximum number of relevant studies to be included in the meta-analyses.

I carried out meta-analyses in Cochrane RevMan software (version 4.3[The Cochrane Collaboration, 2006]). For dichotomous data studies I calculated study specific and pooled odds ratios (OR). For continuous data studies I calculated study specific and pooled standardised mean differences (SMD), which measure the difference in units of standard deviation. I used random effects in the primary analyses and also carried out the analyses using fixed effects.

I used the I^2 statistic to assess heterogeneity between studies, where I^2 estimates the percentage of variation between studies that cannot be attributed to chance [Higgins *et al.,* 2003].

4.3 Results

4.3.1 Studies Identified in Initial Search

Using the search strategies in table 4.2, 831 papers were found in Medline and 1239 papers were found in Embase. After duplicates were removed there were 1398 individual papers. After reading the titles and abstracts of these, 45 studies were found to be potentially relevant for this review (after removing overlapping studies) [Amar *et al.,* 1998; Bachmann *et al.,* 1996; Barber *et al.,* 1999; Bartres-Faz *et al.,* 2001; Bigler *et al.,* 2003; Bornebroek *et al.,* 1997b; Bornebroek *et al.,* 1997a; Bracco *et al.,* 2005; Bronge *et al.,* 1999; de Leeuw *et al.,* 2004; Decarli *et al.,* 1999; Doody *et al.,* 2000; Fornage *et al.,* 2007; Gormley *et al.,* 2007; Gurol *et al.,* 2006; Hadjigeorgiou *et al.,* 2007; Han *et al.,* 2005; Hassan *et al.,* 2002; Hassan *et al.,* 2004a; Hassan *et al.,* 2004b; Henskens *et al.,* 2005; Hirono *et al.,* 2000; Hogh *et al.,* 2007; Khan *et al.,* 2007; Kohara *et al.,* 2003; Lunetta *et al.,* 2007; Maia *et al.,* 2006; Nebes *et al.,* 2001; Purandare *et al.,* 2006; Reitz *et al.,* 2007; Sawada *et al.,* 2000; Schmidt *et al.,* 1996; Schmidt *et al.,* 2000; Schmidt *et al.,* 2001; Seifert *et al.,* 2006; Sierra *et al.,* 2002; Skoog *et al.,* 1998; Sleegers *et al.,* 2005; Steffens *et al.,* 2003; Szolnoki *et al.,* 2004; Szolnoki, 2007; van Rijn *et al.,* 2006; van Rijn *et al.,* 2007; Verpillat *et al.,* 2001; Wen *et al.,* 2006]. After carrying out gene specific searches for the most commonly studied genes (APOE, ACE, MTHFR and AGT), only one further study was identified [Kuller *et al.,* 1998] and searching the reference lists of included papers found no further studies. This gave a total of 46 independent studies

that had analysed the association between a particular genetic polymorphism and WMH. Table 4.3 shows the numbers of studies (and participants) for each of the genes studied for an association with WMH. 19 genes had been studied in a total of approximately 19,000 subjects (ranging between 40 and 8546 for a particular gene). Most of these genes are involved in lipid metabolism, vascular tone or blood pressure regulation.

4.3.2 Study Selection for Meta-Analyses

Four genetic polymorphisms (APOE (ε), ACE (I/D), MTHFR (C677T) and AGT (Met235Thr)) had been studied in more than 2000 subjects and so were included in the meta-analyses. APOE was studied in 24 studies (8546 subjects), MTHFR was studied in 3 studies (2796 subjects), ACE was studied in 9 studies (2319 subjects) and AGT was studied in 6 studies (2702 subjects).

For APOE, ACE and AGT several relevant studies did not present the required data in their publications and so they could not contribute quantitatively to the meta-analyses. However, I considered them qualitatively in the results. All the relevant studies, along with details of the studied subjects are presented in table 4.4. Studies were conducted in Europe, Japan, Hong Kong and USA. Many of the studies recruited hospital patients while some recruited subjects from the general population. Study participants were generally middle aged to elderly (mean age ranged from 52 to 85). Methodological details of the included studies are shown in table 4.5. Most studies reported that brain scan operators were blinded to genotype and that genotypes were in HWE. The method of WMH quantification

Table 4.3 Number of studies (and subjects) published by the end of 2007 assessing the association between any gene and WMH.

Gene	Polymorphism*	Function†	Number of studies (subjects)	Studies
Apolipoprotein E	ε2, ε3, ε4	Lipid metabolism	24 (8546)	[Amar *et al.,* 1998; Barber *et al.,* 1999; Bartres-Faz *et al.,* 2001; Bigler *et al.,* 2003; Bornebroek *et al.,* 1997b; Bracco *et al.,* 2005; Bronge *et al.,* 1999; de Leeuw *et al.,* 2004; Decarli *et al.,* 1999; Doody *et al.,* 2000; Gurol *et al.,* 2006; Hirono *et al.,* 2000; Hogh *et al.,* 2007; Kuller *et al.,* 1998; Lunetta *et al.,* 2007; Maia *et al.,* 2006; Nebes *et al.,* 2001; Sawada *et al.,* 2000; Schmidt *et al.,* 1996; Seifert *et al.,* 2006; Skoog *et al.,* 1998; Steffens *et al.,* 2003; Szolnoki *et al.,* 2004; Wen *et al.,* 2006]
Methylenetetrahydrofolate reductase	677 C/T	Homocysteine metabolism	3 (2796)	[Hassan *et al.,* 2004b; Kohara *et al.,* 2003; Szolnoki *et al.,* 2004]
Angiotensin converting enzyme	I/D	Renin-angiotensin system (BP/fluid balance)	9 (2319)	[Amar *et al.,* 1998; Bartres-Faz *et al.,* 2001; Gormley *et al.,* 2007; Hassan *et al.,* 2002; Henskens *et al.,* 2005; Purandare *et al.,* 2006; Sierra *et al.,* 2002; Sleegers *et al.,* 2005; Szolnoki *et al.,* 2004]
Angiotensinogen	Met235Thr	Renin-angiotensin system (BP/fluid balance)	6 (2702)	[Gormley *et al.,* 2007; Henskens *et al.,* 2005; Schmidt *et al.,* 2001; Sierra *et al.,* 2002; van Rijn *et al.,* 2007; Verpillat *et al.,* 2001]
Matrix metalloproteinase -3 and -9	Haplotype tagging SNPs	Breakdown of extracellular matrix	1 (1427)	[Fornage *et al.,* 2007]
C reactive protein	Haplotype tagging SNPs	Inflammation	1 (1323)	[Reitz *et al.,* 2007]

Endothelial nitric oxide synthase	Glu298Asp	Regulates vascular smooth muscle and endothelial function	3 (1222)	[Hassan *et al.,* 2004a; Henskens *et al.,* 2005; Verpillat *et al.,* 2001]
Angiotensin II receptor 1	A1166C	Renin-angiotensin system (BP/fluid balance)	3 (1160)	[Henskens *et al.,* 2005; Sierra *et al.,* 2002; van Rijn *et al.,* 2007]
Adducin 1	Gly460Trp	Encodes cytoskeletal protein involved in blood pressure regulation	1 (1014)	[van Rijn *et al.,* 2006]
Endothelial 1	Not reported	Vasoconstriction	1 (829)	[Verpillat *et al.,* 2001]
Aldosterone synthase	-344 C/T	Blood pressure regulation	1 (758)	[Verpillat *et al.,* 2001]
Kinesin light chain 1	185 A/C & 406C/T	Organelle transport	1 (493)	[Szolnoki, 2007]
Paraoxonase 1	Met55Leu & Arg192Gln	LDL modification	2 (343)	[Hadjigeorgiou *et al.,* 2007; Schmidt *et al.,* 2000]
Cytochrome B	242 C/T, 640 A/G & 930 A/G	Phagocyte oxidase system	1 (316)	[Khan *et al.,* 2007]
Intercellular adhesion molecule 1	Lys649Glu	Inflammatory response (leukocyte-endothelial adhesion) and endothelial barrier function	1 (220)	[Han *et al.,* 2005]
Presenilin 1	Not reported	Catalyzes deposits of amyloid-beta	1 (65)	[Bornebroek *et al.,* 1997a]
Apolipoprotein C	Not reported	Lipid metabolism	1 (58)	[Bartres-Faz *et al.,* 2001]
Dystrophia myotonica-protein kinase	CTG repeat	Modulation of cardiac contractility	1 (40)	[Bachmann *et al.,* 1996]

* polymorphisms defined using their common name: 677 C/T notation refers to the DNA base change; Glu298Asp notation refers to the amino acid change.

† functions obtained from UniProtKB/Swiss-Prot (http://www.ebi.ac.uk/swissprot)

Table 4.4 Subject characteristics of studies included in the meta-analyses of associations between WMH and APOE, ACE, MTHFR and AGT.

Study	N	Country	Subjects	Male %	Mean age	HWE
APOE						
[Schmidt *et al.,* 1996]	214	Austria	Population sample	50	61	yes
[Bornebroek *et al.,* 1997b]	25	Netherlands	Patients with hereditary cerebral haemorrhage with amyloidosis - Dutch type	48	52	yes
[Amar *et al.,* 1998](i)	29	UK	Patients at memory disorder clinic with infarcts on CT/MRI brain scan	NR	72	yes
[Amar *et al.,* 1998](ii)	149	UK	Patients at memory disorder clinic without infarcts	NR	72	yes
[Kuller *et al.,* 1998]*	3480	USA	Elderly population sample	NR	>70	NR
[Skoog *et al.,* 1998](i)	72	Sweden	Population sample with dementia (DSM-III-R criteria)	NR	all 85	NR
[Skoog *et al.,* 1998](ii)	117	Sweden	Population sample without dementia	NR	all 85	NR
[Barber *et al.,* 1999]*	72	UK	Patients with dementia	52	77	NR
[Bronge *et al.,* 1999]*	60	Sweden	Patients with Alzheimer's disease	38	64	yes
[Decarli *et al.,* 1999]	396	USA	Twins recruited from register of Armed Forces veterans	100	72	NR
[Doody *et al.,* 2000]	104	USA	Patients with Alzheimer's disease	24	74	NR
[Hirono *et al.,* 2000]	131	Japan	Patients with dementia	23	74	yes
[Sawada *et al.,* 2000]	55	Japan	Patients with Alzheimer's disease	36	76	NR
[Bartres-Faz *et al.,* 2001]	58	Spain	Patients with age associated memory impairment	NR	67	yes
[Nebes *et al.,* 2001]	92	USA	Population sample	NR	74	NR
[Bigler *et al.,* 2003]*	215	USA	Population sample	NR	>65	NR
[Steffens *et al.,* 2003]*	245	USA	Patients with major depression	33	70	NR
[de Leeuw *et al.,* 2004](i)	427	Netherlands	Population - with hypertension	49	72	yes
[de Leeuw *et al.,* 2004](ii)	402	Netherlands	Population - without hypertension	49	72	yes
[Szolnoki *et al.,* 2004]	944	Hungary	Patients with cognitive complaints or headaches	54	62	yes
[Bracco *et al.,* 2005]	82	Italy	Patients with Alzheimer's disease	32	72	NR
[Gurol *et al.,* 2006]*	96	USA	Patients with Alzheimer's disease, mild cognitive impairment or cerebral amyloid angiopathy	50	75	NR
[Maia *et al.,* 2006]*	23	Portugal	Patients with primary intracerebral haemorrhage	50	72	yes
[Seifert *et al.,* 2006]	101	Austria	Patients with nontraumatic intracerebral haemorrhage	NR	69	yes

Study	N	Country	Subjects	Male %	Mean age	HWE
[Wen *et al.,* 2006]	67	Hong Kong	Patients with lacunar infarct	46	71	NR
[Hogh *et al.,* 2007]*	75	Denmark	Population sample	NR	82	NR
[Lunetta *et al.,* 2007]*	815	MIRAGE‡	Patients with Alzheimer's disease and siblings	41	73	NR
ACE						
[Amar *et al.,* 1998](i)	29	UK	Patients at memory disorder clinic with infarcts on brain scan	NR	72	yes
[Amar *et al.,* 1998](ii)	146	UK	Patients at memory disorder clinic without infarcts	NR	72	yes
[Bartres-Faz *et al.,* 2001]*	58	Spain	Patients with age associated memory impairment	NR	67	yes
[Hassan *et al.,* 2002]	84	UK	Patients with lacunar syndrome + compatible lesion	52	70	yes
[Sierra *et al.,* 2002]	60	Spain	Patients with hypertension	60	54	no
[Sleegers *et al.,* 2005]	494	Netherlands	Population sample	52	69	yes
[Szolnoki *et al.,* 2004]	961	Hungary	Patients with cognitive complaints or headaches	54	62	yes
[Henskens *et al.,* 2005]*	93	Netherlands	Patients with hypertension	60	55	yes
[Purandare *et al.,* 2006]	97	UK	Patients with dementia	53	75	yes
[Gormley *et al.,* 2007]	294	UK	Patients with small vessel disease (& infarct on scan)	66	67	no
MTHFR						
[Kohara *et al.,* 2003]	1721	Japan	Population sample	51	59	yes
[Hassan *et al.,* 2004b]	114	UK	Patients with lacunar syndrome + compatible lesion on brain scan	59	67	yes
[Szolnoki *et al.,* 2004]	961	Hungary	Patients with cognitive complaints or headaches	54	62	yes
AGT						
[Schmidt *et al.,* 2001]	396	Austria	Population sample	48	60	yes
[Verpillat *et al.,* 2001]*	829	France	Population sample	42	69	yes
[Sierra *et al.,* 2002]	60	Spain	Patients with hypertension	60	54	yes
[Henskens *et al.,* 2005]*	93	Netherlands	Patients with hypertension	60	55	yes
[Gormley *et al.,* 2007]	280	UK	Patients with small vessel disease (& infarct on scan)	66	67	yes
[van Rijn *et al.,* 2007]	1044	Netherlands	Population sample	41	70	yes

* studies with result data unavailable, † MIRAGE sample includes subjects from USA, Canada, Greece and Germany.

HWE= Hardy Weinberg equilibrium, NR= information not available

Table 4.5 Methods of genotyping and phenotyping for studies included in the meta-analyses of associations between WMH and APOE, ACE, MTHFR and AGT.

Study	Genotyping method	scanner blind to genotype?	location	grade/ volume	WMH rating method	dichotomous/ continuous	cut-off[†]
APOE							
[Schmidt *et al.,* 1996]	PCR + CfoI	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
[Bornebroek *et al.,* 1997b]	PCR + HhaI	yes	DW	grade	Scheltens	continuous	-
[Amar *et al.,* 1998](i)	PCR + CfoI	yes	PV	grade	+/-	dichotomous	+/-
[Amar *et al.,* 1998](ii)	PCR + CfoI	yes	PV	grade	+/-	dichotomous	+/-
[Kuller *et al.,* 1998]*	PCR + HhaI	?	PV+DW	grade	0-9	-	-
[Skoog *et al.,* 1998](i)	IEF	yes	PV+DW	grade	+/-	dichotomous	+/-
[Skoog *et al.,* 1998](ii)	IEF	yes	PV+DW	grade	+/-	dichotomous	+/-
[Barber *et al.,* 1999]*	PCR + CfoI	yes	DW	grade	Scheltens	continuous	-
[Bronge *et al.,* 1999]*	micro-sequencing	yes	DW	grade	Scheltens	continuous	-
[Decarli *et al.,* 1999]	PCR + HhaI	yes	?	volume	-	continuous	-
[Doody *et al.,* 2000]	PCR + HhaI	yes	DW	grade	Scheltens	continuous	-
[Hirono *et al.,* 2000]	PCR + HhaI	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
[Sawada *et al.,* 2000]	PCR + HhaI	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
[Bartres-Faz *et al.,* 2001]	PCR + HhaI	yes	DW	grade	Scheltens	continuous	-
[Nebes *et al.,* 2001]	?	?	DW	grade	0-9	dichotomous	4
[Bigler *et al.,* 2003]*	PCR +HhaI	?	DW	grade	4 point scale	-	-
[Steffens *et al.,* 2003]*	PCR + HhaI	?	DW	volume	segmentation	continuous	-
[de Leeuw *et al.,* 2004](i)	PCR + CfoI	yes	DW	volume	3 sizes, count	continuous	-
[de Leeuw *et al.,* 2004](ii)	PCR + CfoI	yes	DW	volume	3 sizes, count	continuous	-
[Szolnoki *et al.,* 2004]	PCR + CfoI	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
[Bracco *et al.,* 2005]	PCR + HhaI	?	PV+DW	grade	ARWMC	dichotomous	5
[Gurol *et al.,* 2006]*	PCR + HhaI	yes	?	volume	segmentation	continuous	-
[Maia *et al.,* 2006]*	PCR + CfoI	yes	DW	grade	ARWMC	-	-
[Seifert *et al.,* 2006]	PCR + HhaI	?	?	grade	Fazekas	dichotomous	2
[Wen *et al.,* 2006]	PCR + HhaI	yes	?	volume	segmentation	continuous	-
[Hogh *et al.,* 2007]*	PCR + CfoI	?	PV+DW	grade	Scheltens	continuous	-
[Lunetta *et al.,* 2007]*	?	yes	?	grade	100 point scale	continuous	-

ACE							
[Amar *et al.,* 1998](i)	PCR	yes	PV	grade	+/-	dichotomous	+/-
[Amar *et al.,* 1998](ii)	PCR	yes	PV	grade	+/-	dichotomous	+/-
[Bartres-Faz *et al.,* 2001]*	PCR	yes	DW	grade	Scheltens	continuous	-
[Hassan *et al.,* 2002]	PCR	yes	PV	grade	0-4 grade	dichotomous	2
[Sierra *et al.,* 2002]	PCR	yes	PV+DW	grade	Fazekas	dichotomous	2
[Sleegers *et al.,* 2005]	PCR	yes	DW	volume	3 sizes, count	continuous	-
[Szolnoki *et al.,* 2004]	I specific probe	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
[Henskens *et al.,* 2005]*	PCR	yes	DW	volume	3 sizes, count	continuous	-
[Purandare *et al.,* 2006]	PCR	yes	DW	grade	Scheltens	continuous	-
[Gormley *et al.,* 2007]	RFLP	yes	?	grade	Fazekas	dichotomous	3
MTHFR							
[Kohara *et al.,* 2003]	PCR	yes	PV+DW	grade	Fazekas	dichotomous	3
[Hassan *et al.,* 2004b]	?	yes	PV+DW	grade	Fazekas	dichotomous	3
[Szolnoki *et al.,* 2004]	PCR	yes	PV+DW	grade	Fazekas	dichotomous	PV3 DW2
AGT							
[Schmidt *et al.,* 2001]	PCR+ Asp I	yes	PV+DW	grade	Fazekas	dichotomous	2
[Verpillat *et al.,* 2001]*	?	?	PV+DW	grade	Scheltens	dichotomous	severe
[Sierra *et al.,* 2002]	PCR+SfaNI	yes	PV+DW	grade	Fazekas	dichotomous	2
[Henskens *et al.,* 2005]*	multilocus assay	yes	DW	volume	3 sizes, count	continuous	-
[Gormley *et al.,* 2007]	RFLP	yes	?	grade	Fazekas	dichotomous	3
[van Rijn *et al.,* 2007]	Taqman	?	DW	volume	mean volume	continuous	-

* studies with result data unavailable, ?=information not available, † numbers denote the grade which was considered to be in the upper group, +/- denotes where WMH was just defined as present or absent.

PV= periventricular, DW= deep white matter.

varied between studies, but most studies used a grading scale and most only studied the deep WM. Where two or more distinct populations were studied within one study, for the purposes of the meta-analyses, these were split into sub-studies. These included subjects with and without dementia, infarcts and hypertension (substudies are denoted (i) and (ii) in tables and figures).

4.3.3 Apolipoprotein E Results

24 studies (8546 subjects) had assessed the association between WMH and the APOE ε polymorphism [Amar *et al.,* 1998; Barber *et al.,* 1999; Bartres-Faz *et al.,* 2001; Bigler *et al.,* 2003; Bornebroek *et al.,* 1997b; Bracco *et al.,* 2005; Bronge *et al.,* 1999; de Leeuw *et al.,* 2004; Decarli *et al.,* 1999; Doody *et al.,* 2000; Gurol *et al.,* 2006; Hirono *et al.,* 2000; Hogh *et al.,* 2007; Kuller *et al.,* 1998; Lunetta *et al.,* 2007; Maia *et al.,* 2006; Nebes *et al.,* 2001; Sawada *et al.,* 2000; Schmidt *et al.,* 1996; Seifert *et al.,* 2006; Skoog *et al.,* 1998; Steffens *et al.,* 2003; Szolnoki *et al.,* 2004; Wen *et al.,* 2006]. From nine of these (5081 subjects, i.e. 59% of the total number of subjects from relevant studies) data were missing from the papers which prevented them from being included quantitatively in the meta-analyses [Barber *et al.,* 1999; Bigler *et al.,* 2003; Bronge *et al.,* 1999; Gurol *et al.,* 2006; Hogh *et al.,* 2007; Kuller *et al.,* 1998; Lunetta *et al.,* 2007; Maia *et al.,* 2006; Steffens *et al.,* 2003]. However, I have shown qualitative results for these studies, which allows an informal assessment of their potential impact on the results. Most studies presented data allowing analysis of the association between APOE and WMH with regards to the presence or absence of the ε4 allele in the genotype, and so this was the genetic model used in the meta-analysis (ε4+ versus ε4 -).

Depending on what data the papers presented they were included in one of three meta-analyses:

- Eleven studies/sub-studies contributed to the comparison of numbers of subjects in lower and upper WMH grades between genotype groups (figure 4.2a). None of the individual studies showed a significant association between ε4+/ε4- genotype and WMH, and overall there was no significant association with the random effects model (OR 0.97, 95% confidence interval (CI) 0.78 to 1.21), and no detectable heterogeneity between the contributing studies (I^2=0%). Analysing the data using a fixed effects model gave a similar result (OR 0.96, 95% CI 0.77 to 1.20).
- Three studies contributed to the analysis of standardised mean difference in grade between genotype groups (figure 4.2b). Again, none of these studies showed a significant difference in WMH between ε4+/ε4- genotypes (pooled random effects SMD= 0.30, 95% CI -0.02 to 0.62), and there was no detectable heterogeneity between studies (I^2=0%). Analysing the data using a fixed effects model gave exactly the same result.
- Four studies/sub-studies contributed towards the analysis of standardised mean difference in volume between genotype groups (figure 4.2c). Although one of these studies found that ε4+ genotypes had a significantly larger standardised mean than ε4 - genotypes [de Leeuw *et al.*, 2004], the pooled random effects SMD was not statistically significant (SMD=0.15, 95% CI -0.04 to 0.33). There was, however, substantial heterogeneity between studies (I^2=51%). Analysing the data using a fixed effects model gave a

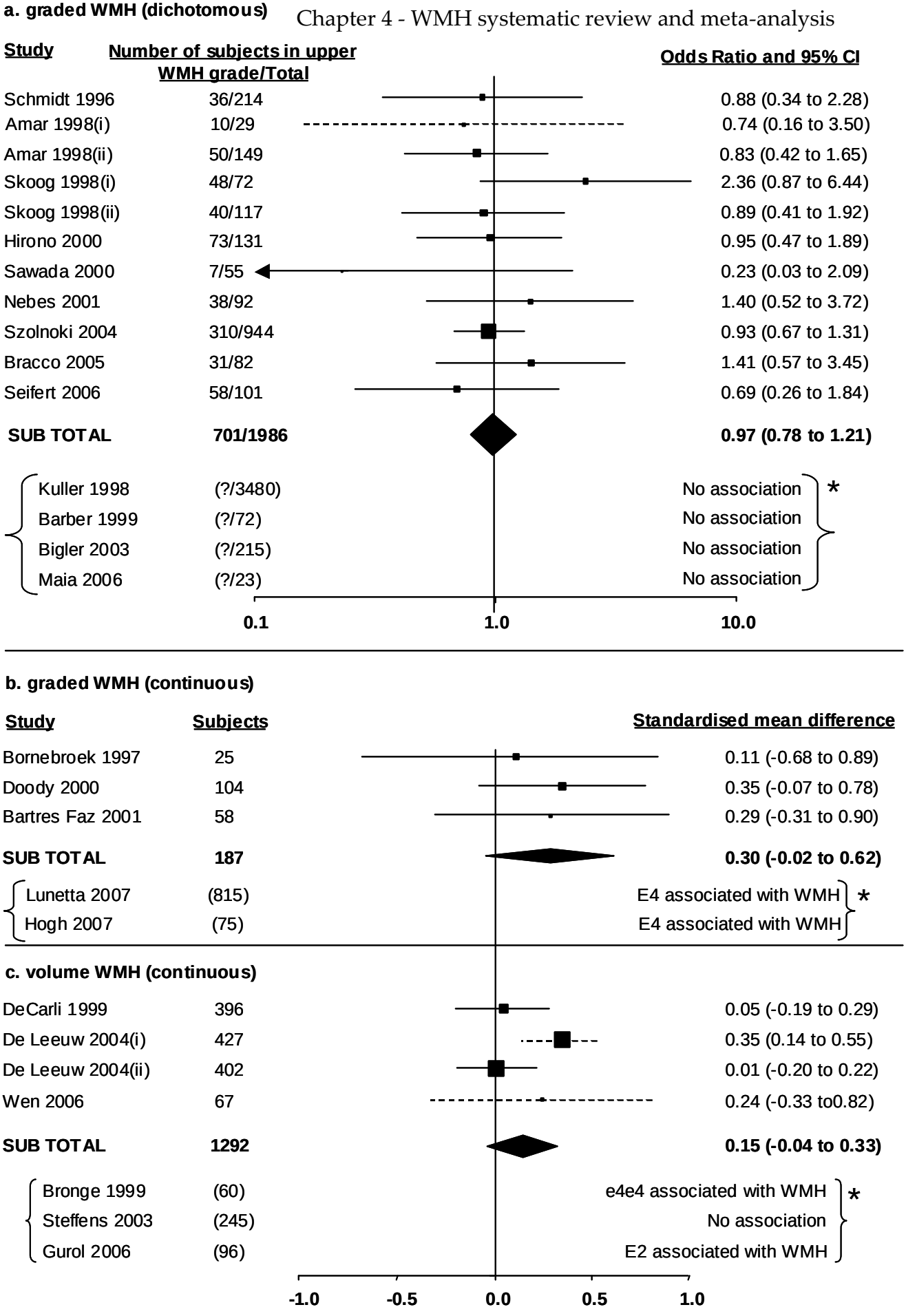


Figure 4.2 Study and pooled effects of the association between WMH and APOE genotype (ε4+ versus ε4-),using random effects. a. odds ratio between upper and lower WMH grade. b. standardised mean difference in WMH grade. c. standardised mean difference in WMH volume. Dashed lines – subjects with infarcts or hypertension.

marginally significant result (SMD=0.15, 95% CI 0.03 to 0.27).

Of the nine studies from which data for meta-analysis could not be extracted, six had measured grade of WMH (figure 4.2). Two of these reported a significant association, one was very small (n=75) [Hogh *et al.,* 2007] and the other was relatively large (n=815) [Lunetta *et al.,* 2007]. Of the four that reported no association, three were small [Barber *et al.,* 1999; Bigler *et al.,* 2003; Maia *et al.,* 2006], but one was conducted among 3480 subjects and so was larger than the total number of subjects from all studies contributing to the current meta-analysis [Kuller *et al.,* 1998]. Three studies with missing data had measured WMH volume. The largest of these (n=245) found no overall difference in WMH volume between genotypes [Steffens *et al.,* 2003]. The other two reported apparent, but different, associations between APOE and WMH (one found that ε4ε4 homozygotes had a significantly increased WMH volume compared to other genotypes [Bronge *et al.,* 1999], and the other that ε2+ genotypes increased WMH volume [Gurol *et al.,* 2006]). These were both very small studies with <100 subjects.

4.3.4 Angiotensin Converting Enzyme Results

Nine studies (2316 subjects) had assessed the association between ACE (I/D) and WMH [Amar *et al.,* 1998; Bartres-Faz *et al.,* 2001; Gormley *et al.,* 2007; Hassan *et al.,* 2002; Henskens *et al.,* 2005; Purandare *et al.,* 2006; Sierra *et al.,* 2002; Sleegers *et al.,* 2005; Szolnoki *et al.,* 2004]. From two of these (151 subjects, i.e. 7% of the total number of subjects from relevant studies) data for meta-analyses were not available in the publications, but qualitative results could be extracted and are shown [Bartres-Faz *et al.,* 2001; Henskens *et al.,*

2005]. The largest included study analysed the data according to a recessive model (DD v ID/II) [Szolnoki *et al.,* 2004]. Furthermore, this model has been used in previous large analyses of the association between ACE and both MI and ischaemic stroke [Agerholm-Larsen *et al.,* 2000], and so was the model used in the meta-analyses.

- Six studies/sub-studies measured grade of WMH and contributed to the comparison of numbers of subjects with upper and lower WMH grades between genotype groups (figure 4.3 a). The pooled random effects estimate suggests a significant association between ACE (I/D) and WMH (OR 1.95, 95%CI 1.09 to 3.48). The fixed effects model was also significant (OR 1.36, 95% 1.08 to 1.72). However, there was substantial heterogeneity between study results (I^2=71%). Although three individual studies/sub-studies found a significant association, all were small. However, it is of possible interest that these three studies were conducted among subjects with lacunar syndrome [Hassan *et al.,* 2002], infarcts on scan [Amar *et al.,* 1998] or hypertension [Sierra *et al.,* 2002] and so at high risk of developing small vessel disease.
- One study with data available analysed the WMH grade as a continuous variable, but this study was small [Purandare *et al.,* 2006]. Although the result is in the same direction as for the dichotomous analysis, the SMD was not significant (figure 4.3b).
- One study with data available measured WMH volume [Sleegers *et al.,* 2005] and found no association with ACE genotype (figure 4.3c).

The required data for meta-analysis could not be extracted from two studies. One had measured WMH volume and analysed data under a recessive

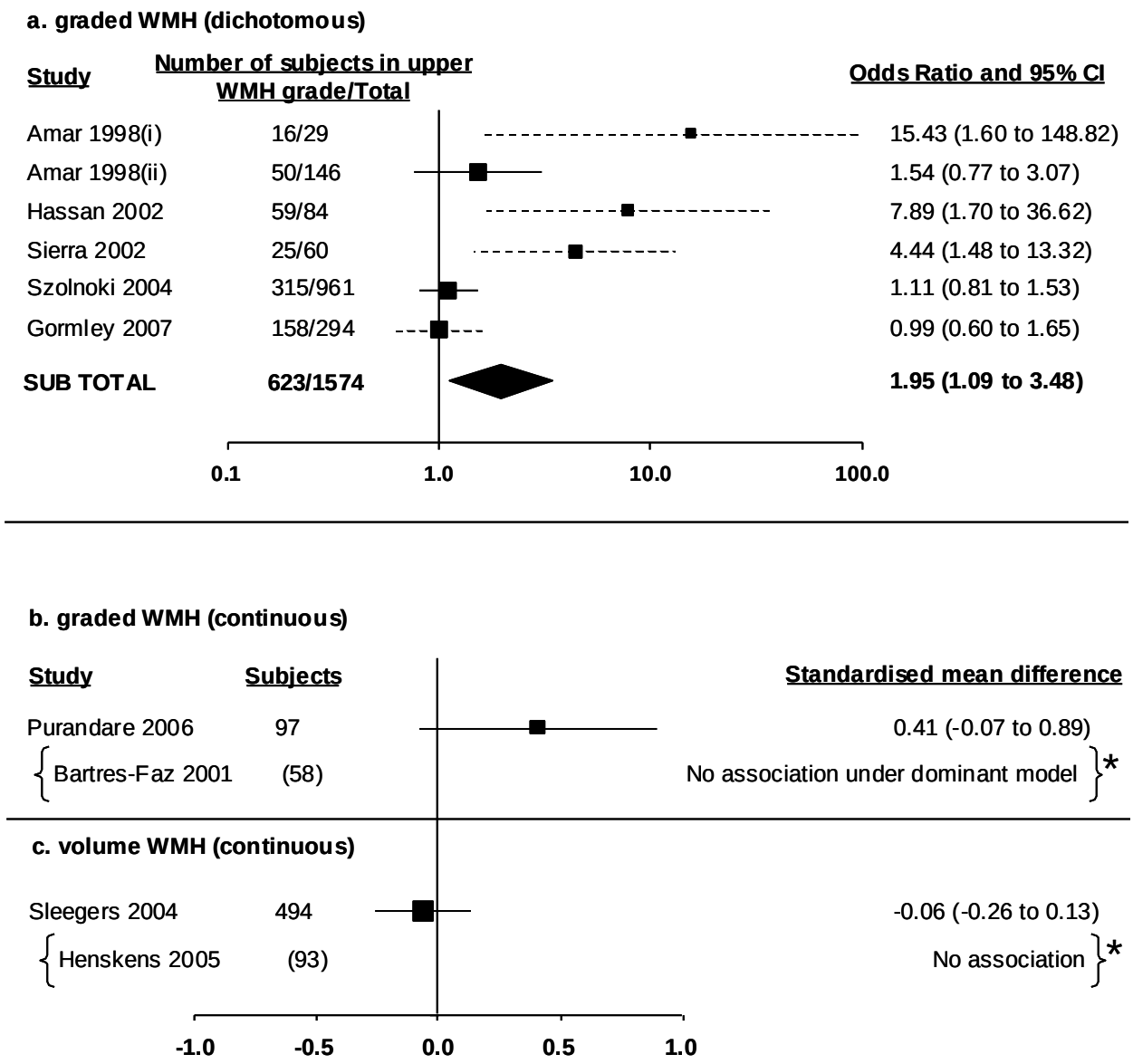


Figure 4.3 Study and pooled effects of the association between WMH and ACE genotype (DD versus ID / II) using random effects. a. odds ratio between upper and lower WMH grade. b. standardised mean difference in WMH grade. c. standardised mean difference in WMH volume. Dashed lines – subjects with infarcts or hypertension.

model (as used above) [Henskens *et al.,* 2005], while the other had measured WMH grade and analysed these data according to a dominant model [Bartres-Faz *et al.,* 2001]. Neither study reported an association between ACE (I/D) and WMH.

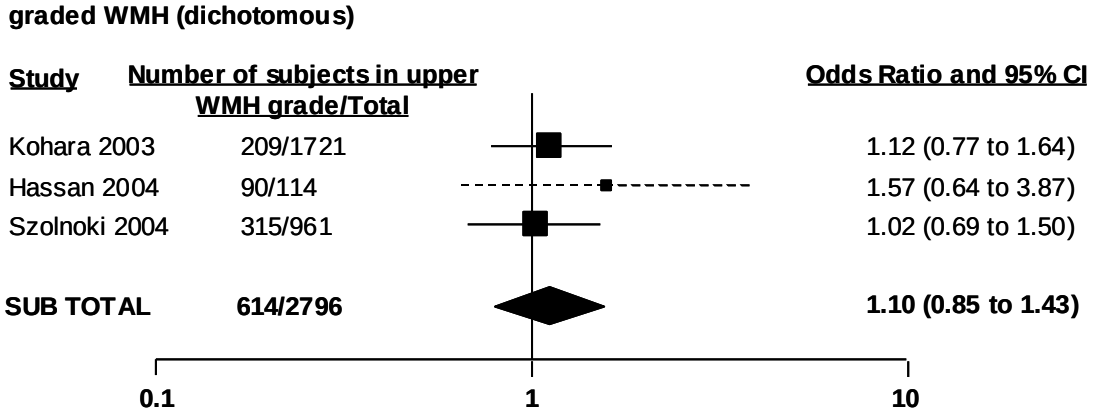


Figure 4.4 Study and pooled odds ratios of the association between upper and lower WMH and MTHFR genotype (TT versus TC / CC) using random effects. Dashed lines – subjects with infarcts.

4.3.5 Methylenetetrahydrofolate Reductase Results

Three studies (2796 subjects) had assessed the association between MTHFR (C677T) and WMH [Hassan *et al.,* 2004b; Kohara *et al.,* 2003; Szolnoki *et al.,* 2004]. The most common genetic model for analysing the data was the recessive model (TT v CT/CC), and so this was used in the meta-analysis. All studies had measured WMH grade and none lacked data for inclusion in our meta-analysis.

- None of the studies individually showed an association between the MTHFR polymorphism and WMH, and overall there was no significant association with the random effects model (OR 1.10, 95% CI 0.85 to 1.43) (figure 4.4) or the fixed effects model (OR 1.11, 95% CI 0.85 to 1.43). There was no excess heterogeneity (I^2=0%).

4.3.6 Angiotensinogen Results

Six studies (2702 subjects) had assessed the association between AGT (Met235Thr) and WMH [Gormley *et al.,* 2007; Henskens *et al.,* 2005; Schmidt

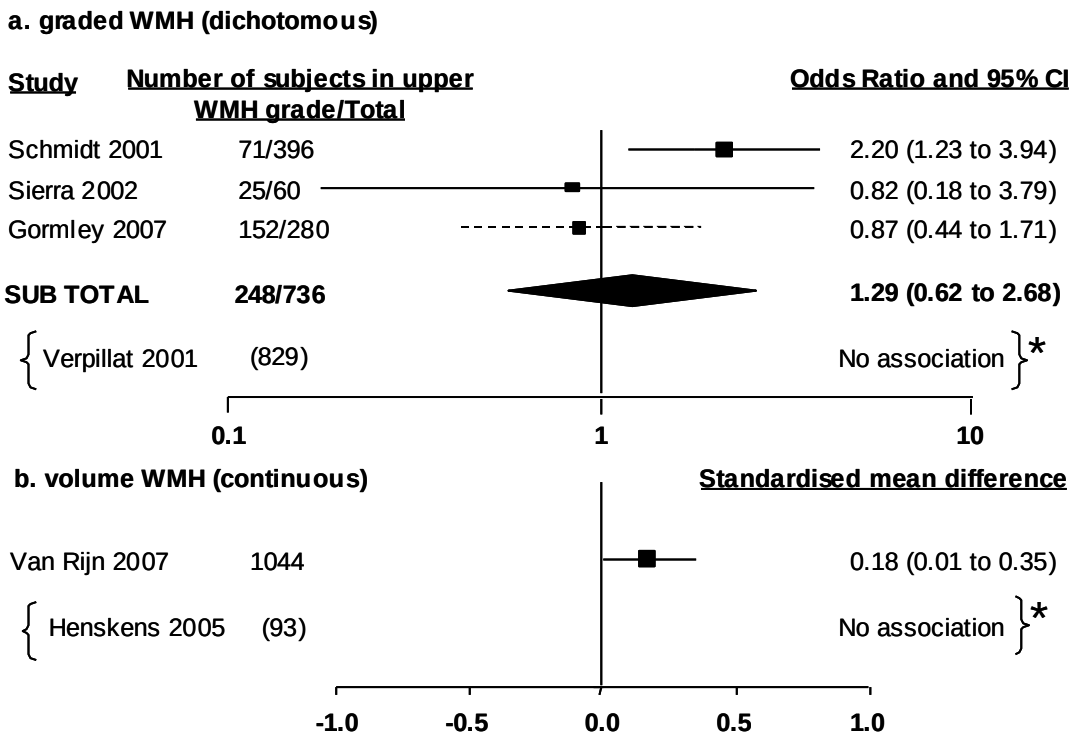


Figure 4.5 Study and pooled effects of the association between WMH and AGT genotype (TT versus TM / MM) using random effects. a. odds ratio between upper and lower WMH grade. b. standardised mean difference in WMH volume. Dashed lines – subjects with infarcts or hypertension.

et al., 2001; Sierra *et al.,* 2002; van Rijn *et al.,* 2007; Verpillat *et al.,* 2001]. From two of these (922 subjects, i.e. 34% of the total number of subjects from relevant studies) data for meta-analyses were not available in the publications, but qualitative results could be extracted and are shown [Henskens *et al.,* 2005; Verpillat *et al.,* 2001]. The most common genetic model for analysing the data was the recessive model (TT v MT/MM), and so this was used in the meta-analysis.

- Three studies measured grade of WMH and contributed to the comparison of numbers of subjects with upper and lower WMH grades between genotype groups (figure 4.5a). The pooled random effects estimate suggests no association between AGT and WMH (OR

1.29, 95% CI 0.62 to 2.68) and the fixed effects model gave a similar result (OR 1.38, 95% CI 0.90 to 2.11).

- One study with data available measured WMH volume. They found a small but significant association between TT and WMH (figure 4.5b) [van Rijn *et al.*, 2007].

One large study (n=829) with data unavailable had measured WMH grade [Verpillat *et al.*, 2001]. This study reported no significant association. One study with data unavailable had measured WMH volume [Henskens *et al.*, 2005]. This study was small (n=93) and reported no significant association.

4.3.7 Other Potential Genes of Interest

Table 4.3 includes all the other genes that have been studied for their association with WMH, generally just in one or a few studies and small numbers (<1500 subjects). Many of these genes showed preliminary evidence for an association with WMH (e.g. CYP11B2, protein kinase on chromosome 19, and ICAM), but these need to be replicated in much larger samples before any conclusions can be drawn.

4.4 Discussion

Through carrying out a large systematic search I identified 19 genes that had been studied for an association with WMH. I then reviewed in detail and carried out meta-analyses for those genes which had been studied in more than 2000 subjects: APOE (ε), ACE (I/D), MTHFR (C677T) and AGT

(Met235Thr). None of these showed a convincing association with WMH and ACE I/D was the only polymorphism for which the evidence suggests a possible association.

4.4.1 Lack of Evidence

Despite the potential promise of WMH as a quantitative intermediate phenotype for study of genetic influences on small vessel disease, and the large number of studies of many genes (representing mainly lipid metabolism, vascular tone and blood pressure regulation pathways), these studies are generally individually small and by the end of 2007 only four genetic polymorphisms had been studied in a total of more than 2000 subjects.

Reliable conclusions cannot be drawn when the number of subjects studied is small because of lack of precision of results. Thus, meta-analyses were only conducted where the total number of subjects was in excess of 2000. Even with this approach there is much potential for small study (mainly publication) and other sources of bias.

4.4.2 No Association Found with APOE, MTHFR or AGT

No convincing association was found between WMH and APOE, MTHFR or AGT. Although there was a substantially large proportion of missing data in the APOE and AGT analyses it is unlikely that the inclusion of any these missing studies would have led to the identification of an association between APOE or AGT and WMH – indeed their inclusion would almost

certainly have strengthened the conclusion of no association, based on qualitative assessment.

For APOE, 95% CIs of the meta-analyses include the possibility of a small or moderate association, but studies not included (because of missing data) mainly showed no association and so anything other than an extremely modest association seems unlikely. This is consistent with results of previous work by our group. In a previous meta-analysis of the association between APOE and stroke, it was found that in the few studies that studied the association of APOE with subtypes of ischaemic stroke, there appeared to be an association with large artery stroke, but not small artery stroke [Sudlow *et al.*, 2006]. As WMH are associated with - and reflect the vascular pathology underlying of - small artery stroke, this is consistent with the notion that APOE is less important in the disease pathway of small artery stroke.

For MTHFR and AGT, the wide 95% CI includes the possibility of an association, but so far no association has been found with WMH, and much larger studies will be needed to detect a small to moderate association.

4.4.3 ACE May be Associated With WMH

ACE I/D was the only polymorphism to show an overall association with WMH (measured and analysed as a grade). However, none of the four studies which had missing data (and so did not contribute to our graded meta-analysis) found an association and so the apparent association from our

meta-analysis may well be prone to bias. Only 7% of the data were missing, but often unreported results are negative (reporting bias) and so would probably decrease the association if included.

4.4.4 Infarct and Hypertension Samples

For all polymorphisms studied in this meta-analysis, the only studies which gave individually positive results are those carried out in subjects that have had a clinical stroke, have infarcts on scan or are hypertensive (dotted lines in the figures). It could be that there is actually an interaction effect and the genotype is more influential on WMH in those subjects already with infarcts or hypertension. This result could also represent bias as these studies use hospital subjects, which tend to be small and so may be prone to small study bias. The three studies that were significant in the ACE analysis all had infarcts or hypertension and were very small studies (number of subjects ranging from 29 to 84). Further work with large numbers of patient subjects is required to ascertain whether the association is a result of bias or a real interaction effect. This correlation between high-risk samples and small samples has been observed in other meta-analyses [Sayed-Tabatabaei *et al.*, 2003] and one needs to be careful to interpret the meaning of this.

4.4.5 Limitations

As always the comprehensiveness of a meta-analysis depends on the data available from the individual studies of which it is made up of. Several restrictions on the data from these studies prevented me from being as thorough in my analyses as I would have liked.

4.4.5.1 WMH measurement method

Several methods for measuring WMH have been implemented and so for each polymorphism the data could not always be combined into one analysis. It may be that there is only enough power to observe an association when an accurate measurement of WMH volume is taken and that grading simply is not powerful enough, or that the real difference is observed when the subjects are dichotomised into 'normal' WMH variation and 'abnomal' WMH variation. Volume methods have the advantage of theoretically increased statistical power but require more sophisticated and costly imaging equipment and software and also lack information on the WMH location and so may lack information on the clinical impact. Some grading scales grade various parts of the brain separately and so may provide more clinically relevant information. However, since these separate grades are then often combined into one global estimate of WMH, the information on severity in different regions is lost. A further problem with dichotomous analyses is that they may be data derived and so prone to bias. I also question whether data using the grading scale should be analysed as a continuous trait. Grading WMH into several categories is unlikely to produce a normal distribution and assumptions made in the analysis of a continuous trait break down.

Within the grade and volume methods many different techniques were used. I attempted to make these as consistent as possible across studies when selecting which data to extract. For example, despite several grades being used across studies it was fairly easy to consistently pick a cut-off that

represented the same amount of WMH irrespective of the scale used. Most studies analysed here use graded measures of WMH. Volume studies (if carried out carefully) may produce more consistent results.

4.4.5.2 Genotype model

The analysis was also limited by the genotype models used in individual studies. For all four polymorphisms analysed here I chose the most commonly used genotypic model to allow the largest analysis of the data. These models were usually chosen by the individual authors because they were backed up with plausible biological explanations but it is possible that the model is incorrect with regards to WMH. For example one excluded study for APOE [Gurol *et al.,* 2006] showed an association between ε2+ genotypes and WMH but as almost all the other papers had analysed the data with respect to the ε4 allele, a meta-analysis of the impact of ε2 was not possible.

4.4.5.3 Missing Data

Missing data obviously adds to the limitations of the analyses. However, I attempted to minimise this by qualitatively assessing these studies and in most cases it seemed that their inclusion would be unlikely to affect the results. Most meta-analysis studies in stroke have used 'availability of relevant data from publication' as a study inclusion criterion. They are therefore not as comprehensive as perhaps a meta-analysis should be, especially as reporting bias is likely to mean that the unavailable data are quite different to the available data. And so, one strength of this WMH

meta-analysis is that I did attempt to assess qualitatively the impact of the missing data. I did not attempt to obtain missing data from the individual authors as the value of this would have been minimal.

4.4.6 Comparing Results to Genome-Wide Linkage Scans

A genome-wide linkage scan for WMH volume has been carried out in the Framingham study [DeStefano *et al.,* 2006]. A 10cM density microsatellite genome linkage scan was performed on 747 subjects in 237 families. A significant log odds (LOD) score of 3.69 was observed at 4cM on chromosome 4; a region in which no candidate gene has so far been studied in WMH. A suggestive LOD score of 1.78 was also observed at 95cM on chromosome 17; this region is within 10cM of the ACE gene and so this could tie in with the results observed in this meta-analysis. It could be that the ACE I/D polymorphism is the causal allele and linkage disequilibrium in the region caused a suggestive linkage peak at 95cM, or perhaps the causal allele is at a different location in the region and linkage disequilibrium explains the association with ACE, or even that there are multiple alleles of interest in this region. It is also possible that this suggestive peak is just a spurious artefact and there could be no true linkage between a gene in this region and WMH.

Another genome-wide scan of 366 microsatellites carried out in the GENOA (Genetic Epidemiology Network of Arteriopathy) study using 488 subjects from 223 sibships found only tentative evidence of linkage (maximum LOD scores of 1.30 to 1.99) for WMH volume in several novel regions [Turner *et al.,* 2005].

Regions containing APOE, MTHFR and AGT have not yet been highlighted as possible regions of interest in genome-wide linkage scans. This could be because these genes are not associated with WMH at all, that the genome-wide linkage scans have not been powerful enough to pick up modest signals or that the regions which these three genes lie in may not be well covered by the genotyping chips used in these studies. More recently designed SNP (single nucleotide polymorphism) chips have much greater coverage (newest generation of chips in excess of one million SNPs).

4.4.7 Missing Heritablility

As quoted in the introduction to this chapter, the estimates for the heritability of WMH have been consistently high, ranging from 55 to 71% [Atwood *et al.,* 2004; Carmelli *et al.,* 1998; Turner *et al.,* 2004]. Of all the genes studied in candidate gene association studies, no large genetic influences have yet been found. There are several explanations for this. The initial estimates of heritability may have been false or misleading. Of particular note is that the heritability studies were of WMH volume and the association studies were predominantly of WMH grade. Therefore, volume may be a more heritable trait than WMH grade. This may suggest that using WMH volume may be more appropriate in the association studies. However this would need to be investigated further. Other possible explanations are the polymorphisms analysed here are associated but that methodological issues have prevented the detection of this, or that other novel genetic polymorphisms yet to be studied have important effects on WMH. All of these reasons probably contribute somewhat to the discrepancy between

high heritability of WMH and the lack of convincing genetic associations identified so far.

4.5 Conclusion

No genetic polymorphism has been convincingly associated with WMH and ACE I/D is the only polymorphism for which the evidence suggests a possible association with WMH. This meta-analysis shows that APOE (ε) is unlikely to be associated with WMH consistent with previous work showing APOE to be associated with large artery but not small artery stroke. The genetics of WMH is a promising area of study, but like many other areas of complex disease genetics it requires much larger studies and internationally agreed measurement methods to allow comparability of study results and to improve opportunities for pooling data and meta-analyses. The ideal WMH measure for future studies would need to be heritable, representative of WMH severity, reliable and repeatable, feasible for use in large samples and statistically powerful.

5 Systematic Review Discussion

In this chapter I discuss the results from the two systematic reviews from chapters 3 and 4 and I put the results in the context of the wider stroke literature. I also discuss the strengths and limitations of the meta-analyses. Finally, I use the results of the meta-analyses to design a hypothesis to test in the Edinburgh Stroke Study

5.1 Findings

Despite carotid intima-media thickness (CIMT) and white matter hyperintensities on brain scans (WMH) having been studied in association with many genes in hundreds of studies using thousands of subjects, few firm conclusions can yet be made about which genes are associated with these traits. The study methodologies for both traits were somewhat heterogeneous, particularly for WMH, where the different ways of reporting extent of WMH meant that not all studies could be combined to obtain a single pooled estimate. No gene has shown a convincing association with WMH, with ACE showing only a possible association. APOE is the only gene showing a convincing association with CIMT, with a meta-analysis restricted to the large studies showing an overall per genotype group mean difference of 8μm (95% CI, 6 to 11), with E4 greater than E3, and E3 greater than E2.

5.1.1 Sample Size

The fact that the meta-analysis between APOE and CIMT was both the largest (>32,000 subjects) and the only one showing significant overall association may indicate that the others are still under-powered. Or, this pattern may just be because a promising genetic association with many positive studies is more likely to be replicated in different populations and by different groups, whilst an association that has shown little promise in early studies is less likely to be studied further. Regardless, it appears that very large numbers of subjects are required reliably to detect what appear to be only small associations with intermediate traits for stroke.

The sample size calculations for CIMT associations are shown in table 5.1. Most genes in the CIMT meta-analyses had minor allele frequencies (MAFs) of (or the equivalent of, for APOE) approximately 0.45 (APOE, ACE, IL6, AGT), 0.30 (MTHFR, NOS3, PON1, IGF1) or 0.20 (ADD1). The table shows that the required number of individuals to detect a 10μm per genotype difference for a co-dominant model with a MAF of 0.45, to achieve 80% power at $p < 0.05$ is >4000 (assuming a mean and SD CIMT of 700 ± 160μm [Lorenz *et al.*, 2007]). The mean sample size of the 'large' studies for APOE was 4539 and so these were powered adequately to detect this effect size. The mean sample size of the 'large' studies for ACE was 2452 and so these were only powered to detect effect sizes in the region of 20μm. The CI for the mean CIMT difference between the ACE genotypes was 0 to 8μm and so studies have not been adequately powered to detect any small effect that ACE might have. The sample size calculations suggest that if there is a small effect for ACE (~5μm) then more than 16,000 subjects will be needed to detect a significant association ($p<0.05$). The combined total of the 'large' studies do not reach this total (n=12,260). Table 5.1 shows that compared to the co-dominant model, recessive and dominant models require much larger sample sizes to detect the same effect size. For MTHFR (MAF = 0.30, recessive) to detect a 10μm difference requires more than 24,000 subjects.

Table 5.2 shows the sample size calculations for dichotomous WMH graded associations. If there is a true association between WMH and APOE or MTHFR then the odds ratio (OR) is likely to only be in the region of 1.2. The table shows that overall no polymorphisms were studied in enough subjects (pooled) to be adequately powered to detect an odds ratio (OR) of 1.2 (>4650 for APOE and ACE - MAF 0.45 dominant (according to the minor allele

Table 5.1 Sample sizes required to achieve 80% power to detect a p<0.05 significant mean difference (ranging from 5 to 50µm), with minor allele frequencies of 0.2, 0.3 and 0.45. The shaded columns show the values for APOE and ACE (MAF 0.45, co-dominant) and MTHFR (MAF 0.30, recessive), assuming a mean and SD CIMT of 700 ± 160µm. Calculated using Quanto version 1.2.3 [Gauderman WJ & Morrison JM, 2006].

	Minor allele frequency								
	0.20			0.30			0.45		
Effect size	Recessive (r)	Co-dominant (c)	Dominant (d)	r	c	d	r	c	d
5µm	209,300	25,112	34,880	98,131	19,132	32,158	49,764	16,233	38,088
10µm	52,322	6,275	8,717	24,560	4,780	8,037	12,438	4,055	9,519
20µm	13,078	1,566	2,176	6,130	1,192	2,006	3,107	1,001	2,377
50µm	2,089	247	345	977	187	318	494	158	377

Table 5.2 Sample sizes required to achieve 80% power to detect a p<0.05 significant OR (ranging from 1.1 to 2), with minor allele frequencies of 0.2, 0.3 and 0.45. The shaded columns show the values for APOE and ACE (MAF 0.45, dominant) and MTHFR (MAF 0.30, recessive), assume a case:control ratio of 1:1 (with and without WMH). Calculated using Quanto version 1.2.3 [Gauderman WJ & Morrison JM, 2006].

	Minor allele frequency								
	0.20			0.30			0.45		
Odds ratio	Recessive (r)	Co-dominant (c)	Dominant (d)	r	c	d	r	c	d
1.1	86154	10504	14814	40594	8082	13854	20814	6956	16704
1.2	22628	2802	4008	10714	2176	3798	5552	1898	4652
1.5	4138	536	794	1986	426	780	1060	386	992
2.0	1248	174	268	610	142	276	340	136	368

frequency), >10,700 for MTHFR – MAF 0.30 recessive). Most individual studies were not powered to detect an OR of less than 2 (>368 subjects for APOE and ACE, >610 subjects for MTHFR). Calculations assume a case:control ratio of 1:1. All sample size calculations were carried out in Quanto (version 1.2.3 [Gauderman WJ & Morrison JM, 2006]).

Generally very few studies have been large enough to detect the small effect sizes that probably exist, and if there are interacting factors and/or phenotypic heterogeneity within studies that increase the complexity of the association, then even larger studies will be needed to achieve the same power. Even after pooling studies in meta-analyses in this thesis, the sample sizes achieved are not sufficient to detect the most likely effect sizes with strong power in most cases. The samples required to achieve appropriate statistical power are often well outside the scope of what single studies can feasibly achieve. This highlights the need for large and consistent studies that can be pooled successfully in the future.

5.1.2 The Effects of Risk Factors, Ethnicity, and Study Size

For both the CIMT and WMH meta-analyses the more extreme estimates of effect (mean differences or odds ratios) were seen in studies of subjects considered to be at high risk of vascular disease. This included those with a history of vascular disease or with vascular risk factors such as hypertension. However, these studies tended to be based on small hospital samples and so are prone to small study bias. CIMT subgroup analyses also showed that studies of Eastern Asian subjects appeared to have more extreme estimates than studies of White subjects. But, again the Eastern Asian studies tended

to be smaller and so may be prone to small study bias. An ethnicity sensitivity analysis was not carried out for WMH as there were very few non-white studies and no evidence of heterogeneity between studies.

The same subgroup patterns appear for all genes (higher effect estimates in Eastern Asian and high vascular risk subjects). In a previous study investigators have noticed a tendency for high risk subjects to show more extreme estimates of association and have concluded that there must be important interactions [Sayed-Tabatabaei *et al.,* 2003]. However, the consistent pattern seen in the CIMT and WMH results show that these differences may be explained by study size bias and so results of this type need to be interpreted with caution.

5.2 Potentially Important Genes and Gene Pathways for Stroke

The genes that have been studied for an association with CIMT, WMH or even stroke have been chosen by the investigators of these studies because they make good candidates. This is often because they are already known to play a major role in a pathway that is considered important for the trait or disease endpoint. This explains why the same genes have been investigated for WMH, CIMT, ischaemic stroke (IS) and ischaemic heart disease (IHD). Many of the genes studied in my meta-analyses have also been included in a recent meta-analysis of commonly studied genes for IHD [Kitsios & Zintzaras, 2007]. It also explains why many of the genes studied are related. When a pathway is considered important, it is common for several of the key players in this pathway to be investigated. For each of the genes reviewed in

Table 5.3 Pathways and genes included in my meta-analyses of the association of commonly studied genes with CIMT and WMH.

Pathway	Genes in CIMT meta-analysis	Genes in WMH meta-analysis
Lipid metabolism	APOE PON1	APOE
Vascular homeostasis	ACE AGT NOS3 IGF1 ADRB2	ACE AGT
Metabolic factors	MTHFR	MTHFR
Haemostasis	Factor V FGG/FGA	
Inflammation	CRP IL6	
Blood pressure regulation	ADD1	ADD1

this thesis, I will now discuss the pathways that they are involved in, showing why various candidates seem attractive choices for influencing stroke and its intermediate traits. Table 5.3 shows the pathways that genes from the meta-analyses (chapter 3 and 4) are involved in.

5.2.1 Lipid Metabolism

Since cholesterol levels are an important risk factor for stroke and other vascular diseases, genes from the lipid metabolism pathway are ideal candidates for stroke and have been studied extensively for both IHD and IS. APOE (which I studied in both the CIMT and WMH meta-analyses) and PON1 (in the CIMT meta-analysis) are both key candidates from this pathway.

5.2.1.1 Apolipoprotein E (APOE)

The apolipoprotein E protein (ApoE) is an LDL receptor ligand encoded by the APOE gene, located on chromosome 19q13.2. Three common alleles exist: ε2, ε3, and ε4, resulting from missense mutations at two locations. These produce three isoforms: E2, E3 and E4. The E3 isoform has a cysteine amino acid at position 112 and an arginine amino acid at position 158, E2 has cysteine at both positions and E4 has arginine at both. The ε3 allele (considered the 'normal' allele) is the most frequent, accounting for between 50 and 90% in different populations; the ε4 allele is the next most frequent (5-35%); ε2 the least frequent (1-15%) [Mahley & Rall, Jr., 2000].

Apolipoprotein levels vary according to genotype, with ε2 associated with increased plasma levels and ε4 with decreased plasma levels [Davignon *et al.,* 1988]. Apolipoproteins bind with free cholesterol, phospholipids, cholesterol esters and some triacylglycerols to form lipoproteins. ApoE helps to stabilize and solubolize lipoproteins as they circulate in the blood and interacts with specific lipoprotein receptors to alter the circulating levels of cholesterol [Eichner *et al.,* 2002].

The association between APOE and cholesterol levels is well documented. ε2 is associated with lower- low-density lipoprotein (LDL) cholesterol levels, and ε4 with higher levels [Cattin *et al.,* 1997]. LDL cholesterol molecules contribute to the development and progression of atherosclerosis. It has been shown that ε2 lowers cholesterol levels by ~14 mg/dl and ε4 raises them by ~8 mg/dl [Hallman *et al.,* 1991]. As much as 10% of the total variation in

cholesterol levels in the population is accounted for by the APOE gene locus, more than for any other gene identified so far [Mahley & Rall, Jr., 2000].

As the association between high cholesterol (specifically LDL cholesterol) and IHD risk is well established, it is unsurprising that APOE has been considered an important genetic risk factor candidate for IHD. APOE knockout mice develop spontaneous atherosclerosis, suggesting that the presence of the APOE gene and its protein product apoE is protective [Zhang *et al.*, 1992]. A meta-analysis of large human studies (only including studies with more than 500 IHD cases) reported that there was an approximately linear relationship between APOE ε genotype and IHD risk (ordered ε2ε2, ε2ε3, ε2ε4, ε3ε3, ε3ε4, ε4ε4). The OR for ε2 carriers compared to ε3ε3 subjects was 0.80 (95% CI, 0.70 to 0.90) and for ε4 carriers compared to ε3ε3 subjects was 1.06 (95% CI, 0.99 to 1.13) [Bennet *et al.*, 2007].

A systematic review and meta-analysis of APOE ε genotypes and IS found no clear evidence for an overall association between ε4 carriers and IS when restricting the analysis to only the larger (more than 200 cases) studies (OR= 0.99, 95% CI, 0.88 to 1.11). There was some evidence that there may be a specific association between ε4 carriers and the large artery subtype of stroke (OR= 1.33, 95% CI, 0.99 to 1.78) [Sudlow *et al.*, 2006]. However, this is based on a small proportion of the studies, so may be susceptible to reporting bias and so warrants further investigation.

Taking all this evidence together with the results of my meta-analysis of the association between APOE and CIMT, it seems likely that the APOE ε4 is

associated with increased CIMT and confers a risk of atherosclerosis, which may lead to large artery IS or IHD. The apparent lack of association with WMH (a phenotype related to small vessel disease) and other ischaemic stroke subtypes including small vessel disease stroke, suggests that small vessel disease is distinct from atherosclerosis and large vessel disease and is not influenced by APOE.

Other apolipoproteins (apoA-I/C-III/A-IV and apoB), lipoprotein receptors and key enzymes with functional roles in homeostasis and lipid metabolism have been suggested as possible genetic sources of risk for lipid levels and so for cardiovascular and cerebrovascular disease, but so far little is known about the influence of these genes on atherosclerosis and vascular disease [Nieminen, 2006].

5.2.1.2 Paraoxonase 1

Paraoxonase 1 is a calcium–dependent glycoprotein synthesised in the liver. It binds to HDL and prevents oxidation of LDL [Mackness *et al.,* 1998]. Oxidised LDL is important in the atherosclerotic pathway [Mertens & Holvoet, 2001].

The PON1 gene is on 7q21.3. There are two commonly studied missense mutations: Q192R and L55M. The L55M polymorphism affects serum concentration of PON1 [Garin *et al.,* 1997] and Q192R affects efficiency of the enzyme [Humbert *et al.,* 1993]. PON1 knockout mice have high levels of

oxidised LDL and are more prone to atherosclerosis than wild type mice when fed a high fat diet [Shih *et al.*, 1998].

A meta-analysis of the association between PON1 polymorphisms and IHD found a significant overall association between the 192 R allele and IHD (RR 1.12, 95% CI 1.07 to 1.16). However, sub-analysis of only the larger studies suggest this result could be prone to small study bias (RR 1.05, 95% CI 0.98 to 1.13) [Wheeler *et al.*, 2004]. A narrative review reports that depleted PON1 serum concentration and activity may be better predictors of IHD than any polymorphism studied so far, as studies have found statistically significant associations with activity and concentration, but not with genetic polymorphisms [Mackness & Mackness, 2004]. This may indicate that other polymorphisms of the gene yet to be investigated could be important or that there are other factors regulating PON1 including other genetic polymorphisms or environmental factors. This finding could also suggest that PON1 activity and concentration is associated with IHD by reverse causation, where the onset of IHD is the cause of the decreased activity and not a consequence of it.

A meta-analysis of the association between the Q192R PON1 polymorphism and stroke found an overall significant association (OR 1.64, 95% CI, 1.39 to 1.94). However, this only included four studies and 460 stroke patients and so may be prone to publication bias [Ranade *et al.*, 2005].

I included the Q192R polymorphism in my meta-analyses of genetic influences on CIMT, but found no overall statistically significant association.

Studies of PON1 associations with WMH were only assessed in two small studies and so were not included in my meta-analyses. However, one study reported no association between L55M or Q192R and WMH [Schmidt *et al.*, 2000] , while the other reported that the QQ genotype at the Q192R locus was associated with WMH [Hadjigeorgiou *et al.*, 2007] (the opposite direction of the association proposed for stroke and IHD). However this was of marginal statistical significance (p=0.02) and the study was very small (n=79) and so this is probably due to chance.

A review has suggested that the discrepancies seen for PON1 association for IHD and CIMT may be due to an interaction with smoking, based on recent findings in a small study (of less than 200 Finnish men) that non-smokers with LL at residue 55 had a higher mean CIMT than M carriers, whereas smokers who were M carriers had a higher mean CIMT than LL subjects [Humphries & Morgan, 2004]. This requires further investigation.

This pathway is a strong candidate for influencing atherosclerosis and large artery ischaemic stroke. Other genes of the lipid metabolism pathway, yet to be studied in large numbers, may prove to be important. Many of these are likely to have very small effects and so only very large studies (and – perhaps – meta-analyses of these) which consider interactions and focus on specific disease or trait definitions, will help to tease apart the associations.

5.2.2 Vascular Homeostasis

Pathways controlling vascular architecture and function are an obvious place to search for genes that predispose to atherosclerosis, stroke and CVD. The renin angiotensin (RA) system and the nitric oxide synthase (NOS) system both have important roles in function of the vessels. The RA system has been extensively studied (ACE and AGT genes featured in both the CIMT and WMH meta-analyses) and NOS3 from the NOS system was studied in the CIMT meta-analysis. Other genes involved in vascular homeostasis were also included in the CIMT meta-analysis (IGF1 and ADRB2).

5.2.2.1 Renin angiotensin system

The RA system is a hormone system involved in blood pressure regulation. Angiotensin converting enyme (ACE) converts inactive angiotensinogen (AGT) to the vasconstrictor angiotensin II (which is mediated by angiotensin II receptor type 1 (AGTR1)) and also inactivates the vasodilator bradykinin, hence regulating vascular tone, vascular smooth muscle proliferation, and endothelial function [Carluccio *et al.,* 2001].

Angiotensin converting enzyme gene (ACE)

ACE is the most extensively studied gene in the RA system. The presence (insertion, I) or absence (deletion, D) of a 287 base-pair alu (a short interspersed nuclear element) repeat sequence in reverse orientation in intron 16 17q23 of this gene has been shown to be associated with substantially different levels of plasma ACE [Rigat *et al.,* 1990]. This polymorphism accounts for 47% of the variation in ACE plasma level with the DD genotype being associated with the highest levels.

Two meta-analyses of the association between the ACE I/D polymorphism and MI or IHD only found an overall significant association in the smaller studies (IHD (whites) OR 1.29, 95% CI 1.15 to 1.43; MI (whites) OR 1.47, 95% CI 1.30 to 1.66; MI (all ethnicities) RR 1.57, 95% CI 1.38 to 1.78); meta-analyses of the larger studies showed no association (IHD OR 1.07, 95% CI 0.97 to 1.17; MI (whites) OR 0.99, 95% CI 0.88 to 1.12; MI (all ethnicities) RR 0.99, 95% CI 0.90 to 1.08) [Agerholm-Larsen *et al.,* 2000; Keavney *et al.,* 2000].

A meta-analysis of the association between ACE I/D and IS in persons of European descent found an overall significant association between the DD genotype and ischaemic stroke risk (OR 1.21, 95% CI 1.08 to 1.35) [Casas *et al.,* 2004]. A meta-analysis of this association in persons of non-European descent also found an overall significant association in the Chinese individuals (OR 1.90, 95% CI 1.23 to 2.93), but not the Japanese (OR 1.74, 95% CI 0.88 to 3.42) [Ariyaratnam *et al.,* 2007]. However, there was evidence in this second meta-analysis of small study bias and significant heterogeneity. Some studies have suggested that the association with ischaemic stroke is specific to lacunar stroke, whilst others have not found an association between lacunar stroke and the ACE DD genotype, and a meta-analysis of these studies shows that this relationship is still unclear [Gormley *et al.,* 2007].

The ACE I/D polymorphism featured in both my CIMT and WMH meta-analyses. I found an overall significant association between the D allele and increased CIMT, but no significant pooled mean CIMT difference when only

the large studies where analysed. Some large studies that I could not include in my analysis because data were unavailable, did show an association between the D allele and increased CIMT and so their inclusion may have strengthened the evidence for an association. Therefore, I cannot rule out a modest association between ACE and CIMT.

It has been suggested that MTHFR and APOE genotypes, smoking and alcohol consumption may be important interacting factors for the association between ACE I/D and ischaemic stroke [Szolnoki & Melegh, 2006]. However, this was based on the findings of one study with 1341 subjects. These interactions could not be tested in my meta-analysis, but may warrant further investigation. The ACE – CIMT meta-analysis I carried out in chapter 3 is an update of that by [Sayed-Tabatabaei et al. 2003]. They too found that the association was more pronounced in the high risk subjects, but they did not investigate the effects of study size as a potential confounder of this finding. They attributed this finding to gene-environment interaction and went on to investigate potential interacting risk factors in their own cohort study, discovering a significant association between ACE I/D and CIMT only in the presence of smoking [Sayed-Tabatabaei *et al.*, 2004]. However, other studies assessing this interaction have been inconsistent [Sass *et al.*, 1998]. It is quite likely that if ACE I/D is associated with CIMT that there are important interactions with traditional risk and environmental factors. However, the correlation between high risk individuals and study size means that the positive studies may be biased and there is no convincing evidence for any association between ACE I/D and CIMT at present. It will be important to consider potential interacting factors in future studies of ACE.

Angiotensinogen (AGT)

AGT has also been studied for its involvement in IHD and stroke. A methionine to threonine substitution in exon 2 of the AGT gene on 1q42-43 (M235T) has been shown to associated with AGT concentration [Bloem *et al.*, 1997].

Many studies have assessed the association between AGT M235T and IHD, but these have been conflicting and a meta-analysis reported no overall association [Sethi *et al.*, 2003]. Studies assessing the association between AGT and stroke have been conflicting [Bersano *et al.*, 2008]. But, some studies have found specific associations with lacunar stroke [Nakase *et al.*, 2007; Takami *et al.*, 2000].

AGT M235T was included in my meta-analyses for both CIMT and WMH. I found no overall association in either meta-analysis. However, the largest WMH study (n = 1044) did report a significant association and this polymorphism has been studied in very relatively small numbers so far, so there may be an association yet to be confirmed in larger studies.

Other renin angiotensin system genes

Other genes of this important pathway may be associated with CIMT and WMH and/or confer a risk of ischaemic stroke. Angiotensin II receptor, type 1 (AGTR1) is another gene which has been studied for associations with IHD, IS, CIMT and WMH. It was not studied in enough subjects to be included in

either of my meta-analyses. But has been implicated in individual studies with MI (in an interation with ACE I/D) [Tiret *et al.,* 1994] and IS (in a small study) [Rubattu *et al.,* 2004], and possibly particularly with lacunar stroke (althought this was a very small study) [Takami *et al.,* 2000].

Genes from the RA pathway are likely to be the focus of many candidate gene associations in the future. Studying very large numbers of individuals and carefully defining disease and phenotypes, will enable the role of these genes in vascular disease to be determined.

5.2.2.2 Nitric oxide synthase

The nitric oxide synthase (NOS) system is important for endothelial function, including regulation of tone, integrity and growth. NOS is an enzyme which acts on L-arginine to produce nitric oxide (NO), a vasodilator [Andrew & Mayer, 1999]. Endothelial NOS (NOS3) is presumed to be responsible for most of the endothelial and vascular effects of NO.

A G to T mutation at nucleotide position 894 of the NOS3 gene results in a glutamic acid to aspartic acid substitution at amino acid 298, which reduces NOS3 activity [Tesauro *et al.,* 2000]. This is the only common non-synonymous variant, but other potentially important polymorphisms include a 27-base pair repeat polymorphism in intron 4 and a T to C mutation 786 base pairs upstream of the NOS3 gene (T-786C). NOS3 knockout mice are highly sensitive to focal cerebral ischemia [Samdani *et al.,* 1997].

A meta-analysis of three NOS3 polymorphisms found that the intron 4 polymorphism was significantly associated with IHD (per-allele OR 1.12, 95% CI 1.01 to 1.24). Significant associations were also found for the Glu298Asp (per-allele OR 1.17, 95% CI 1.07 to 1.28) and T-786C (per-allele OR 1.17, 95% CI 1.07 to 1.28) polymorphisms. However, these were prone to small study bias [Casas *et al.*, 2006]. A meta-analysis of Glu298Asp NOS3 polymorphism and ischaemic stroke reported no overall association (recessive OR 0.98, 95% CI 0.76 to 1.26) [Casas *et al.*, 2004].

The NOS3 Glu298Asp polymorphism was included in my meta-analysis of CIMT. I found no overall association. NOS3 polymorphisms were only studied in three WMH studies (total n=1222) and so were not included in my WMH meta-analysis, but the studies showed mixed results.

It has been reported that there may be interactions between smoking, ACE genotype, MTHFR genotype and NOS3 genotype which together associate with ischaemic stroke [Szolnoki & Melegh, 2006]. These interactions could not be tested in my meta-analysis, as the information on these factors was not available from the individual studies. Unlike other genes, there is no definitive functional gene variant and studies of the association between NOS3 and cardiovascular events have not consistently studied the same polymorphisms [Napoli & Ignarro, 2007], making reviews of this literature difficult. As the IHD meta-analysis referred to above suggests, perhaps the less-studied intron 4 polymorphism is more important than Glu298Asp [Casas *et al.*, 2006]. The authors of this study suggest that future work for this gene and CVD should include gene-wide tagging polymorphisms to

capture variation across the whole gene, in a large-scale genetic association study.

5.2.2.3 Insulin-like growth factor 1 (IGF1)

IGF1 is a mitogenic peptide hormone with an established role in growth and differentiation. More recent work suggest that it also acts as a vascular protective factor by stimulating NO production, resulting in decreased vascular smooth muscle proliferation and vasodilation [Walsh *et al.,* 1996].

IGF1 levels have been shown to be decreased in atherosclerotic plaques [Okura *et al.,* 2001] and circulating levels are decreased in patients with cardiovascular disease [Ezzat *et al.,* 2008]. IGF1 levels have been shown to be ~20% lower in individuals without the 192-base pair wildtype allele of the gene [Vaessen *et al.,* 2001].

This polymorphism featured in my meta-analysis for CIMT. Only one study had assessed the association between the 192-base pair polymorphism of IGF1 and CIMT, but this study was of substantial size (5132 subjects) and found a significant association [Schut *et al.,* 2003]. I found the per-allele mean difference (for the non-192-base pair allele) to be 10μm, suggesting that, as might be expected, this mutation of the IGF1 gene confers a risk of atherosclerosis. The association reported was particularly pronounced in a subset of hypertensive subjects, suggesting the polymorphism may modulate the risk in these individuals more than in non-hypertensive subjects. As the only evidence for this association comes from one study, this will need to be

repeated in other studies before any firm conclusions can be made, but this gene looks like a promising candidate for atherosclerosis, CIMT, and so for IHD and IS.

5.2.2.4 Adrenergic beta 2 receptor (ADRB2)

Adrenergic beta 2 receptors are G protein-coupled receptors that mediate a cardiovascular response when stimulated by adrenaline [Guimaraes & Moura, 2001]. Specifically, they regulate dilation of arteries, resulting in the increased perfusion of organs needed for the fight-or-flight response.

The gene ADRB2, which encodes the β2 adrenergic receptor, is located on 5q31-32. Several polymorphisms have been identified [Johnson & Terra, 2002], the most commonly studied being Gln27Glu and Arg16Gly, which are in almost complete linkage dysequilibrium with each other. These two polymorphisms have been shown to be related to down regulation of the receptor [Green *et al.,* 1994]. Beta adrenergic receptor agonists (β-blockers) are used to treat and prevent further coronary events and so polymorphisms of the ADRB2 are potential candidates for cardiovascular disease.

In a large single prospective study of more than 5000 subjects (702 of which had a coronary event and 438 of which had a stroke in the 10 years of follow-up) Glu27 carriers were found to have a lower risk of coronary events than Gln27 homozygotes (RR 0.82, 95% CI 0.70 to 0.95). However, there was no such association found for stroke (RR 0.94, 95% CI 0.77 to 1.15) or

cardiovascular events as a whole (RR 0.93, 95% CI 0.82 to 1.05) [Heckbert *et al.*, 2003; Johnson & Terra, 2002].

The ADRB2 Gln27Glu polymorphism was included in my meta-analysis for CIMT. Only one large study was identified and this found no association (mean CIMT difference between Glu carriers and Gln homozygotes was 10μm, 95% CI –29 to 49). This finding along with the lack of association seen for stroke could imply that although ADRB2 may be an important risk factor and drug target for coronary events, it is not associated with other cardiovascular disease. Although this is based on only a few studies and ADRB2 could still be found to have a wider role in cardiovascular disease.

Other polymorphisms and complex promoter region haplotypes of this gene have been shown to alter receptor expression [Drysdale *et al.*, 2000] and could also be studied for an association with CIMT and/or stroke.

5.2.3 Metabolic Factors

5.2.3.1 MTHFR

Methylenetetrahydrofolate reductase (MTHFR) is an enzyme which reduces 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, which acts as a carbon donor in the remethylation of homocysteine to methionine. Elevated homocysteine levels have been shown to promote atherosclerosis, potentially through several mechanisms including endothelial dysfunction [Welch & Loscalzo, 1998].

The C677T polymorphism of the MTHFR gene produces an alanine to valine substitution, which increases the thermolability of the enzyme, hence reducing activity (especially in folate-deficient individuals) [Frost *et al.,* 1998]. Homocysteine levels have been found to be ~25% higher in TT individuals compared with CC individuals [Brattstrom *et al.,* 1998].

A meta-analysis of 32 studies (14870 subjects) of the association between C677T and stroke found that overall there was a significant association between the T allele and increased stroke risk (OR 1.18, 95% CI 1.09 to 1.29) [Cronin *et al.,* 2005], largely mediated through homocysteine levels [Casas *et al.,* 2005]. However, there were insufficient data in the individual studies to detect stroke subtype differences or an interaction with folate levels. MTHFR and homocysteine seem to be important in conferring risk of stroke, but so far the efficacy of homocysteine lowering treatment (namely folic acid supplementation) in reducing the risk of stroke is unclear [Hankey, 2006].

The case for an association between MTHFR and IHD has been less convincing. A meta-analysis of 80 studies (>57,000 subjects) found a small overall association between TT genotype and increased risk of IHD (OR 1.14, 95% CI 1.05 to 1.24), but this was prone to bias and when stratified geographically the association was only significant in the Middle Eastern subjects (OR 2.61, 95% CI 1.81 to 3.75) [Lewis *et al.,* 2005]. An earlier meta-analysis reported that the TT genotype was associated with IHD, particularly under low folate conditions. It also suggested that heterogeneity between geographic populations may result from varying levels of folate in the diets

of these populations, particularly in the use of vitamin supplements and folate fortification of breakfast cereal in North America [Klerk *et al.,* 2002].

MTHFR was included in my meta-analyses for both CIMT and WMH. I found no overall association between MTHFR C677T polymorphism and WMH. I found a significant association for CIMT, but this was very small and prone to study bias. The larger studies showed no association between C677T and CIMT.

As mentioned above, MTHFR activity is particularly reduced when, in addition to the mutation at 677, there is also folate deficiency, suggesting that folate may be an important interacting factor for CIMT association and risk of stroke. It has been suggested that folate intake may only be associated with CIMT when it is at a critically low level [Durga *et al.,* 2005] and therefore the association between CIMT and MTHFR C677T may only be detected in individuals below this critical folate level. Information on folate status of individuals was lacking in most studies and so assessment of this potential interacting risk factor was not possible in my meta-analysis but would be of interest for future work in this area.

5.2.4 Haemostasis

The formation of a clot, and its potential blockage of an artery can partially or completely occlude blood flow to vital organs. Lack of blood supply to the heart or brain leads to ischemic damage, resulting in an MI or stroke and so genes with a role in haemostasis are ideal candidates for stroke genetics.

5.2.4.1 Factor V

The Factor V Leiden variant of the gene has a missense mutation (G1691A), which abolishes the activated protein C (APC) cleavage site, resulting in factor V being resistant to APC inactivation. Resistance to APC inactivation has been found to be a risk factor for venous thrombosis [Svensson & Dahlback, 1994], but its role in arterial thrombosis is still debated.

A meta-analysis of 20 studies (>40,000 subjects) found that there was a significant overall association between factor V Leiden and IHD (per-allele RR 1.17, 95% CI 1.08 to 1.28), but this result is prone to small study bias and the seven studies with >500 subjects showed no significant association [Ye *et al.*, 2006]. A meta-analysis of genetic associations for stroke found a significant overall association between factor V Leiden and IS (dominant OR 1.33, 95% CI 1.12 to 1.58), with no evidence of publication bias[Casas *et al.*, 2004].

The factor V Leiden polymorphism was included in my meta-analysis for CIMT, but has so far not been studied for WMH. I found a statistically significant overall association between factor V Leiden and CIMT, but in the

opposite direction to that expected from the disease meta-analyses described above (the factor V Leiden mutation appears to decrease CIMT, whereas it increases risk of IHD and IS). However, only two studies were included in this meta-analysis and both were in families of IHD patients, so this could a chance result or due to bias. It would be interesting to see if the association can be replicated in a community based study.

5.2.4.2 Fibrinogen

Fibrinogen is an important coagulation factor. It is cleaved by thrombin to form fibrin after vascular injury [Herrick *et al.,* 1999]. Fibrinogen has three polypeptide chains: α, β and γ, encode by three genes clustered on chromosome 4q: FGA, FGB and FGG, respectively.

FGG/FGA haplotypes have been shown to be associated with the structure of the fibrin network [Mannila *et al.,* 2006] and fibrin structure has been shown to be associated with IHD [Fatah *et al.,* 1992]. FGG/FGA haplotypes have been shown to be associated with ischaemic stroke (OR 1.36, 95% CI 1.09 to 1.69) and MI (OR 1.51, 95% CI 1.18 to 1.93) in single studies, but these have been relatively small (less than 400 subjects) [Cheung *et al.,* 2008; Mannila *et al.,* 2005]. In my systematic review, only one study had assessed the association between FGG/FGA haplotypes and CIMT. This study found no association. No study had assessed the association between WMH and any fibrinogen polymorphism.

There is also evidence for association between FGB mutations and fibrinogen plasma levels, as well as stroke, but only in small numbers [Hassan & Markus, 2000]. In addition, a large individual participant data meta-analysis (154211 subjects) has shown a significant association between plasma fibrinogen levels and IHD (HR 2.42, 95% CI 2.24 to 2.60, per 1g/L increase in fibrinogen level) and plasma fibrinogen levels and stroke (HR 2.06, 95% CI 1.83 to 2.33, per 1g/L increase in fibrinogen level) [Danesh *et al.,* 2005].

There is some preliminary evidence and background reasoning for an association between polymorphisms of the fibrinogen genes and stroke (and its intermediate traits). However, much more work is needed on these polymorphisms before conclusions can be made with any confidence.

5.2.4.3 Other haemostatic genes

Other haemostatic genes (e.g. factor VII, prothrombin, factor XIII, platelet glycoprotein receptor, HPA2, von Willebrand factor, plasminogen activator inhibitor-I) have been studied for an association with stroke and MI, but these have mostly been relatively small and with conflicting results [Hassan & Markus, 2000]. A meta-analysis of seven haemostatic gene polymorphisms and IHD found that both plasminogen activator inhibitor 1 (RR 1.06, 95% CI 1.02 to 1.10) and prothrombin (G20210A) (RR 1.31, 95% CI 1.12 to 1.52) were significantly associated with MI [Ye *et al.,* 2006] and in a meta-analysis of genetic polymorphisms and stroke, significant associations were found for prothrombin (G20210A) (OR 1.44, 95% CI 1.11 to 1.86), plasminogen activator inhibitor 1 (OR 1.47, 95% CI 1.13 to 1.92) and glycoprotein Ib-α (OR 1.88, 95% CI 1.28 to 2.76) polymorphisms [Casas *et al.,* 2004]. Therefore, it seems likely

that genes of the haemostatic system play a role in stroke susceptibility, but which genes are the key players and how exactly they are associated with stroke and its intermediate phenotypes is still relatively unknown.

5.2.5 Inflammation

Inflammation is the process by which the body responds to injury. Therefore atherogenesis will elicit an inflammatory response. The nature of this inflammatory response may play a key role in the extent and outcome of the disease process. Considerable evidence now exists for inflammation as a key process in pathogenesis of atherosclerosis and stroke [Libby *et al.,* 2002].

5.2.5.1 C-reactive protein (CRP)

C-reactive protein is as an inflammatory marker, present in atherosclerotic plaques [Torzewski *et al.,* 2000]. CRP levels can rise rapidly in response to cytokines (such as IL6). Several polymorphisms of the CRP gene are associated with plasma levels of the protein [Lange *et al.,* 2006]. However, the association of CRP plasma levels with IHD, stroke and CIMT remain controversial

CRP levels have been consistently associated with increased risk of IHD and several polymorphisms in the gene encoding CRP have been consistently associated with CRP levels. However, it is not clear whether the association between CRP and IHD is causal or just reflects CRP being a marker of disease. Studies attempting to test for associations between CRP genetic polymorphisms and IHD have so far been found no significant association.

However, they may have been massively underpowered (collectively including only a few thousand cases), as a sample size calculation suggests that 15000 IHD cases may be required for such a study to be sufficiently powered [Casas *et al.,* 2008]. A meta-analysis found that high concentrations of plasma CRP are associated with stroke [Kuo *et al.,* 2005]. However, whether this relationship is causal is still to be determined.

The association between CIMT and polymorphisms in the gene encoding CRP was only studied in one study in my meta-analysis. This study found no association between CIMT and any of the five polymorphisms assessed, but included 4641 subjects and may have been under-powered and perhaps much larger studies need to be carried out.

5.2.5.2 Interleukin 6 (IL6)

IL6 is a cytokine that regulates C-reactive protein production during the inflammatory response [Heinrich *et al.,* 2003]. There are several common polymorphisms in the IL6 gene, G-174C being the most widely studied. This polymorphism has been shown to be associated with plasma IL6 [Terry *et al.,* 2000] and CRP levels [Vickers *et al.,* 2002]. As mentioned above there does appear to be an association between CRP levels and IHD, but it is unclear whether this is causal or a result of reverse causation [Casas *et al.,* 2008].

The G-174C polymorphism has been shown to be associated with IHD and stroke. However, the direction of the association differs between studies [Georges *et al.,* 2001; Humphries *et al.,* 2001; Kelberman *et al.,* 2004; Pola *et al.,*

2003; Um *et al.*, 2005], with some concluding that GG is the risk genotype, and others, CC.

I included IL6 in my meta-analysis for CIMT, and found no overall association. However, a study for which separate G-174C data were unavailable, showed that when studying three inflammatory genes (IL-6, Il-1 receptor antagonist and endotoxin receptor) in combination, the gene variant score was associated with CIMT [Markus *et al.*, 2006]. This study was relatively small (810 subjects), but suggests that other inflammatory genes may be of importance and that interactions may exist between these genes.

Other inflammatory genes have been studied for association with IHD, but so far none of these have been found to have a significant association [Kitsios & Zintzaras, 2007].

5.2.6 Blood Pressure Regulation

Hypertension is an established risk factor for stroke and so genes that may regulate blood pressure are ideal candidates for stroke.

5.2.6.1 Adducin 1 (ADD1)

Adducin is a heterodymeric cytoskeletal protein consisting of three subunits (α,β and γ) encoded by three genes: ADD1, ADD2 and ADD3 [Matsuoka *et al.*, 2000]. The α subunit, encoded by ADD1 on chromosome 14p16.3, regulates the activity of transmembrane ion pumps.

Studies using Milan hypertensive and normotensive strains of rat have shown that the Phe316Tyr polymorphism of the ADD1 gene accounts for a large proportion of the blood pressure difference between these strains [Bianchi *et al.,* 1994]. The Gly460Trp polymorphism and other polymorphisms in the ADD1 region have also been implicated in human hypertension, although results have not been consistent between studies [Bianchi *et al.,* 2005]. It has been suggested that these inconsistencies are due to interactions with other variables including the ACE I/D polymorphism and sodium levels (which have not always been taken into account). Even more conflicting are the results of studies assessing an association between ADD1 polymorphisms and cardiovascular disease and stroke.

ADD1 was a commonly studied gene in my meta-analysis of CIMT. I found no overall significant association between ADD1 and CIMT. However, most studies suggested there could be important interacting factors that need to be studied further; sex [Sarzani *et al.,* 2006]; diabetes [Yazdanpanah *et al.,* 2006]; the ACE I/D polymorphism [Balkestein *et al.,* 2002]. The association between ADD1 and WMH had only been assessed in one study (n=1014) and so did not reach the cut-off of 2000 set for 'commonly studied' genes for my WMH meta-analysis. However, this study found no overall association.

Although ADD1 appears to play an important role in hypertension, its association with stroke and intermediate traits remains uncertain. As hypertension is an important risk factor for stroke, ADD1 seems likely to have some role in the susceptibility to stroke. However, if the ADD1

influence on hypertension is small and involves interactions, the association with a trait or disease further down the causal pathway will be even smaller and more complex. In addition it is unclear whether hypertension influences all stroke subtypes similarly. In a very large meta-analysis of 61 prospective studies (>958,000 subjects) it was reported that at age 40-69 years, each difference of 20mm Hg usual systolic blood pressure is associated with a twofold difference in stroke death rate (but IS was not subdivided into subtypes). History of hypertension is used in the TOAST classification for diagnosis of lacunar, but not large artery stroke [Adams, Jr. *et al.*, 1993]. The available evidence from systematic reviews and meta-analyses of observational epidemiological studies suggests that there is no difference between pathological types and subtypes of stroke in the influence of hypertension on stroke risk [Jackson & Sudlow, 2005].

5.3 Limitations of the CIMT and WMH meta-analyses

5.3.1 Novel Genetic Meta-Analysis Method for Continuous

I devised a novel meta-analysis method that allows the pooled overall association to be assessed and the nature of this association to be estimated without the issue of multiple testing or relying on previous knowledge of the genetic model. In my method I first tested the data for an overall association. If I detected an overall association, I went on to determine the most appropriate genetic model using linear regression, and then to estimate the pooled effect size (in this case the pooled mean difference in CIMT) corresponding to this genetic model.

This proved to be a useful method for my CIMT dataset. Most traditional meta-analyses compare two groups. However, deciding which two-way comparison to be made of the three genotypes was difficult and not really meaningful for the genes showing no overall association. Only studying in detail those genes that first show an overall association seems sensible.

A previous method for choosing the genetic model by comparing multiple, pooled mean differences [Thakkinstian *et al.*, 2005], is flawed as it disregards the structure of the data (the connection between mean differences within an individual study). The linear regression method I developed provides an obejective measure of the error around the chosen genetic model. As can be seen for APOE and ACE, the evidence for a co-dominant model is strong (tight 95% CIs around a λ of 0.5).

With my method, no forest plots are constructed for polymorphisms showing no overall association. However, I believe that for my data showing plots of these 'associations' would not have been very meaningful. If it is felt appropriate to view a forest plot of a mean difference for a polymorphism that showed no overall association, this can still be done, but should be reported as an exploratory exercise.

My method cannot be carried out in the easy-to-use Cochrane RevMan software [The Cochrane Collaboration, 2006] and must be carried out in a more flexible statistical package and this does require some statistical knowledge and programming ability. I carried out all stages of the method in STATA version 7.0 [StataCorp., 2001] and the code that I developed for

this can be applied universally (if data are supplied in a specific format), and so run by anyone with the STATA software.

My three-step meta-analysis method was not used in the WMH analyses, as for these datasets the individual studies had often selected a genetic model and had only reported the data according to that model. I therefore used the most common genetic model for my analysis and did not first test for an overall association. This meant I could include a larger amount of data. However, I cannot exclude the possibility that other genetic models may have found associations where I found none. For example, despite ACE I/D being analysed according to a recessive model, with a just significant overall effect on WMH, this does not necessarily mean that ACE I/D acts on WMH in a recessive fashion.

5.3.2 Missing Data

In both my meta-analyses there was a large amount of 'unavailable data' (from studies that had not reported in sufficient detail to allow the necessary data extraction). I contacted authors and attempted to collect the unavailable data for CIMT. However, I did not attempt to retrieve data from authors of the WMH studies, as it seemed unlikely that the inclusion of these data would have influenced the results at all.

A particular strength of the meta-analyses presented here is that, unlike other meta-analyses of stroke [Casas *et al.,* 2004], I chose to assess qualitatively the impact that studies with unavailable data would have made, had I been able

to include them. In most cases I was able to conclude that inclusion of these data would not have impacted substantially on the overall results, but it was only by considering ALL relevant published studies, whether or not they had data available for meta-analyses, that I could draw these conclusions.

The large proportion of data that were unavailable from individual publications from my systematic reviews highlights the need to be thorough when publishing results and that results for all analyses performed should be presented, even results that are not statistically significant. This avoids the reporting bias that otherwise occurs. It is common for significant results to be published in full, whilst non-significant results are briefly mentioned without the appropriate data reported, or even more problematic, not published at all. Scientists feel under pressure to selectively highlight only positive results, either because it fits with their prior beliefs or because they believe it is easier to impress journal editors and reviewers, and more interesting for readers. Journals have increasingly stringent word limits and this puts further pressure on authors to not present all of their methods and findings in full. This may be overcome in some journals by publishing supplementary material online. However, not all journals offer this facility.

5.3.3 Limitations of Meta-Analysis

Despite meta-analysis being an extremely useful tool for systematically summarizing all available data on a topic, there are limitations. Pooling large amounts of data should increase the power to detect associations, but when there is heterogeneity in study design and individuals participating, extra noise is introduced and true associations may be hard to detect. One

criticism of meta-analyses is that they can only be as good as the individual studies that they comprise. However even in cases of heterogeneous or methodologically poor individual studies, thoughtful meta-analyses which seek to explore reasons for heterogeneity by performing detailed assessments of study methodology, characteristics and sources of bias can be very informative.

I attempted to limit the heterogeneity between studies in my meta-analyses by being consistent when selecting which specific trait to use (i.e. CIMT measured at a specific location, or choosing as consistent a definition of WMH as possible). However, this was only possible so far as the individual studies allowed. Some investigators have called for phenotypes and diseases to be more consistently reported in the literature to allow for more sensible meta-analyses to take place, and hopefully this will have a positive influence on meta-analyses in the future (e.g. the Mannheim Intima-Media Thickness Consensus [Touboul *et al.,* 2004]).

If studies use different populations and different definitions of disease, this may lead to real differences in their results (although large qualitative differences in the genetic effects under study seem unlikely – e.g. men and women may have their CIMT affected to a different degree by some genetic polymorphisms but the existence of or direction of effect is unlikely to differ). As seen in my meta-analyses the characteristics of studies often co-vary making it difficult to tell whether (for example) differences between study estimates of an association are due to the ethnicity, hypertensive status of the subjects, or the size of the study.

Covariates are important for almost all phenotypes. However, unless all studies have measured and reported on the same set of covariates, one cannot assess these within a meta-analysis and so any association may be masked by other variables. Interactions may also be important, whether with other genes or with environmental factors, and these too cannot be systematically assessed when individual studies have tested for different interactions. Without thorough assessment of covariates it is unclear whether any significant association with a polymorphism is acting via an existing risk factor, or if it is associated independently of known risk factors. A further problem with studying covariates and interactions in a systematic review or meta-analysis is that it is probable that the studies may have tested many factors and only reported the significant results. One important difference between the WMH and CIMT meta-analyses is that the WMH studies tended to use older subjects. This difference could explain why APOE appears to be associated with CIMT, but not with WMH.

Individual participant data (IPD) meta-analysis is a more thorough and powerful way of summarizing data from multiple studies, and has been described as the 'yardstick' for meta-analysis [Clarke & Stewart, 2001]. This method will still suffer from problems of heterogeneity if different definitions of the trait or different populations are studied, but at least with all of the raw individual data available it is more likely that these factors can be investigated. IPD meta-analysis relies heavily on co-operation of study investigators and is best adopted within a consortium of the relevant

investigators. It involves more work than meta-analysis of summary data, and is often not possible as many of the datasets cannot be obtained.

However, meta-analyses of published data (with additional summary data from investigators if necessary and where feasible) are useful ways of summarizing all of the available data to date and identifying hypotheses that can be tested in future studies, as well as identifying pit-falls and informing on sample size and phenotypic definition issues for future studies.

5.3.4 Are WMH and CIMT Useful Intermediate Traits?

WMH and CIMT, being highly heritable and strongly related to stroke subtypes, appear to be powerful intermediate traits for the genetic study of stroke. However, so far, when the available data are carefully scrutinised, they have not lived up to expectations and there are few convincing associations. But this is true of many complex traits.

The lack of success with these intermediate traits could be due to several factors:

- **The initial estimates of heritability may have been over-estimated**

Many studies have found heritability to be high for CIMT and WMH. However, twin, sib-ship and family history studies are all thought to over-estimate the heritability, due to shared environmental factors along with the difficulty of completely controlling for their confounding effects [Guo, 2001]. It is probably the case that these traits are heritable but not to the level of the

more extreme estimates (0.92 for CIMT; 0.80 for WMH) [Duggirala *et al.,* 1996; Turner *et al.,* 2004], instead being rather closer to the more modest estimates (0.32 for CIMT; 0.55 for WMH) [Atwood *et al.,* 2004; Lange *et al.,* 2002].

- **The genes studied so far are associated with the traits, but studies have so far failed to detect this**

This may be due to insufficient sample sizes, heterogeneity, or interactions. If the 'common disease common variant' hypothesis is correct, there are likely to be many genes influencing CIMT and WMH, each with a small effect. Identification of these may require much larger study sizes than those used so far. There is obvious heterogeneity between studies. Although I have attempted to limit this in my meta-analyses, it may still be present enough to cloud the underlying genetic associations. For other complex traits, genetic interactions between multiple loci have been shown to produce larger effect sizes than the sum of the effects of the single gene variants (e.g. [Williams *et al.,* 2000] and this may also be the case for WMH and CIMT.

- **Other polymorphisms, yet to be studied, are more important.**

Perhaps work so far has focused on the wrong genes and pathways. Although these genes all have strong biological evidence to justify their association with stroke and its intermediate traits, there are ~20,000 genes [Clamp *et al.,* 2007] and selecting which are expected to be associated with a trait is akin to finding 'a needle in a haystack'. Novel polymorphisms may be identified by genome-wide association studies (GWAS). We are now in an era of large scale GWAS, with promising novel polymorphisms being

identified for common complex disease [Cambien, 2007]. So far there has only been one small preliminary GWAS for stroke and none for CIMT or WMH, but this is likely to change in the near future. The GWAS for stroke only included 249 cases, found no SNPs with genomewide significant association with stroke, but identified many SNPs for follow-up ($p<1 \times 10^{-5}$), none of which are in genes that have been considered strong candidates so far [Matarin *et al.*, 2007]. In the next few years large scale GWAS and follow-up studies will likely identify some important stroke genes that have yet to be studied.

It could also be the case that the right genes, but the wrong polymorphisms have been studied. On the whole the polymorphisms have been selected due to their impact on gene expression or function, and so would be expected to be of importance, but there may still be other more important polymorphisms. Also of interest is the increasing evidence that copy number variation (CNV) may be as, or more, important than single nucleotide polymorphsisms (SNPs). My meta-analyses have not just focused on SNPs, but have included small-scale insertions, deletions, and repeats, that have been shown to have functional effects. However, the larger-scale variations in copy number (CNVs) have been more difficult to identify. There is a lot of focus in the genetic community now to characterise CNVs within the genome and test these for association with traits and disease and this may be fruitful.

Other traits may prove to be useful in identifying genetic factors for stroke. For example, total plaque area (TPA) and total plaque volume (TPV) have

been discussed as potential traits for large artery stroke [Pollex & Hegele, 2006].

5.3.5 Are Small Associations Clinically Relevant?

So far there is no clear evidence for any large genetic effects for CIMT, WMH or stroke itself. Small associations have been identified (although, these are by no mean confirmed beyond doubt), for example the association I report here between CIMT and APOE. This association is likely to be in the region of a 20μm difference between E2 and E4 individuals. The question then, is how clinically relevant is such a small difference?

A meta-analysis on the association between CIMT and stroke reported that a difference of 100μm confers a relative risk of 1.18 (95%CI 1.16 to 1.21) [Lorenz *et al.,* 2007]. Therefore, if APOE acts only through CIMT, the risk that this gene confers on stroke is very small (E4 individuals will have a 3.5% increased risk compared to E2 individuals). Identifying individuals' APOE genotype status is therefore unlikely to be important alone in predicting stroke risk for individuals. However, it has been suggested that these polymorphisms with small effects could be more important as drug targets. As the effects of inherited variants are limited, the effect of drug treatments are not. An example is the gene for 3-hydroxy-3-methyl glutaryl-coenzyme A reductase (HMGCR), for which the SNPs in this gene have only a ~5% effect on LDL levels, whereas drugs targeting the encoded protein of this gene decrease LDL levels by ~30% [Altshuler *et al.,* 2008].

5.3.6 Lessons Learnt

Future studies of genetic association for stroke and its intermediate phenotypes will need to address two concepts:

- Heterogeneity of phenotypes
- Interactions between risk factors (genetic and environmental)

To do this successfully studies will need to:

- Be much larger than previous studies.
- Include careful consistent definitions of the disease/trait being studied.
- Include extensive phenotyping, so heterogeneity and interactions can be considered.
- Include extensive genotyping (using high-throughput genotyping techniques) to identify novel polymorphisms and study gene-gene interactions.
- Perhaps use other intermediate traits in addition to CIMT and WMH to study small- and large- artery stroke and other stroke subtypes.
- Attempt to carry out replications in similar populations.
- Involve sophisticated methods of analysis to deal with the vast amounts of data and the identification of interactions and sources of heterogeneity.

5.4 Hypothesis for Further Investigation

The strongest conclusions that can be made from the meta-analyses are for APOE. APOE has been studied in by far the largest numbers for both CIMT and WMH. APOE was the only gene that showed a significant association with CIMT overall and when restricting the analysis to only the larger (and probably more reliable) studies. Conversely, there appeared to be no association between APOE and WMH in my meta-analysis. This is in keeping with previous studies of the association between APOE and stroke that suggest an association with large artery stroke, but no association with small artery stroke. However, few studies have tested explicitly for a difference in association of APOE with large versus small artery ischaemic stroke. I decided to test this hypothesis in a large cohort of stroke patients. This study is described in section B of this thesis.

SECTION B

6 Association Between Ischaemic Stroke Subtype and APOE Genotype in a Hospital-Based Stroke Cohort

(ESS APOE Genotyping Study)

In this chapter I describe the methodology and analysis plan for the genotyping study I undertook, within the Edinburgh Stroke Study, in which I planned to test the association between Apolipoprotein E genotype and ischaemic stroke subtype. Unfortunately, problems with the genotyping meant I could not carry out the planned association analysis, and so instead I present investigatory analyses to determine the causes of the genotyping problems.

6.1 Introduction

The Edinburgh Stroke Study (ESS) recruited from May 2002 to May 2005 and aimed to register all stroke and TIA (transient ischaemic attack) patients seen at the Western General Hospital, and to follow-up all those who presented with a stroke for recurrent stroke, myocardial infarction, death and disability. The aim was to study causes and consequences of stroke and TIA. Data collected included medical background and family history information, clinical details of the presenting stroke or TIA, investigation results (including imaging), and follow-up. Follow-up of stroke patients was for between one and four years. So far the ESS has produced several publications, including an assessment of the impact of the requirement for consent and comparisons between ischaemic stroke subtypes of the prevalence of traditional risk factors and of the prognosis for recurrent vascular events [Jackson *et al.,* in press; Jackson *et al.,* 2008; Jackson *et al.,* in press]. Blood samples for extraction and storage of DNA, as well as for storage of plasma were taken from most of the patients included in the cohort, with a view to future genetic and other biomarker studies.

I aimed to use this cohort to test the association between APOE epsilon genotype and ischaemic stroke subtypes, specifically comparing the distribution of APOE genotypes between patients who have had a large artery ischaemic stroke (LAS), with those that have had a small artery ischemic stroke (SAS). This follows from the hypothesis generated from my systematic review work which suggests that the APOE epsilon genotype is associated with CIMT (and LAS), but not with WMH (and SAS).

6.2 Methods

6.2.1 Subject Recruitment

The aim was to recruit into the ESS any patient with a definite or probable stroke or TIA who was admitted as an inpatient to – or seen in an outpatient clinic at - the Western General Hospital, Edinburgh between May 2002 and May 2005. All those recruited gave informed consent. Patients could consent to any or all of the following:

- use of their data for research
- contact with their GP and access to their medical records
- future follow-up
- storage of blood samples for future biological and genetic analyses

2160 patients were recruited, around 1500 of whom had a confirmed stroke. In a comparison of the ESS participants and a contemporaneous stroke audit from the same hospital, with the same target population but no requirement for consent, it was found that, during an 18 month period (October 2002 to March 2004 inclusive), the ESS recruited 88% of eligible participants [Jackson *et al.*, 2008]. The need for consent may have introduced selection bias relative to the target population. Participants were more likely than non-participants to have had a milder stroke, but there were some very mild stroke patients who were missed due to a shorter stay in hospital, meaning that consent could not always be obtained prior to discharge.

6.2.2 Data & Sample Collection

Baseline data were collected on standardized forms by stroke specialist doctors who assessed the patients (see appendix 9). Clinical diagnosis of stroke and the subtype was confirmed by clinical assessment and imaging investigations. Stroke subtype was classified according to both the OCSP (Oxford Community Stroke Project) and a modified TOAST (Trial of Org 10172 in Acute Stroke Treatment) classification (see section 1.1 for discussion of classification methods and appendix 10 for the modified TOAST algorithm used in the ESS). 2 x 2ml EDTA blood samples were collected by either doctors assessing each patient or trained research nurses from the Wellcome Trust Clinical Research Facility (WTCRF) within the Western General Hospital. Both samples were transferred to the WTCRF on ice on the day of collection. One sample was stored at -80°C prior to DNA extraction and storage in the Genetics Core Laboratory at the WTCRF. The other was centrifuged and the plasma stored at -80°C prior to transfer to a research laboratory in Glasgow for measurement of various inflammatory and rheological biomarkers.

Patients were followed up for recurrent stroke, myocardial infarction, death and disability by: asking patients to contact the ESS research team following any suspected stroke or myocardial infarction; posting questionnaires (about possible vascular events and disability) to participants at 6 months, 1 year and annually; asking GPs and hospital doctors of participants to inform the ESS research team of future vascular events or deaths (by way of letter to the GP and a sticker on the patients' hospital notes); obtaining any death certificates for patients in the cohort from the General Register Office for Scotland.

6.2.3 Sample Preparation

DNA was extracted using the Nucleon BACC3 kit from Tepnel. Extraction took place either immediately on receiving the samples, or up to three years later. The extracted DNA was initially stored in tubes. The tubes were spun for two weeks to re-suspend the DNA. Half of each DNA sample (500µl) was plated into deep 384-well plates (the remaining stock (500µl) was kept in the tubes, both stored at -80°C). The concentration of each sample in the 384-well stock plates was measured using PicoGreen® immediately after being plated. PicoGreen® contains a fluorescence stain for double-stranded DNA (dsDNA) that has a minimal fluorescent signal in solution, but a strong signal when bound to dsDNA. Based on these concentrations, DNA samples from the 384-well stock plates were normalised to 10ng/µl in 96-well plates, ready for genotyping (where concentrations were <10ng/µl the samples were transferred to the 96-well plates neat). On each 96-well normalised plate there were between one and four water controls.

6.2.4 Genotyping

The epsilon variant of APOE comprises two SNP mutations. The ε3ε3 wildtype genotype has a Cys amino acid at position 112 and an Arg at position 158. A change to Cys at position 158 represents the ε2 allele and a change to Arg at position 112 represents the ε4 allele. Genotyping was carried out on the 96-well normalised plates by the WTCRF Genetics Core Laboratory using Applied Biosystems TaqMan genotyping assays. The TaqMan assays for the two SNPs (rs7412 and rs429358) are c904973 and c3084793, respectively. TaqMan assays include PCR primers and allele specific probes. The PCR step amplifies the genome region of interest and

then allelic discrimination is achieved by selective annealing of fluorescent probes.

Genotype calling was performed by WTCRF Genetics Core laboratory staff using ABI 7900 and Applied Biosystems AutoCaller™ genotyping software. Genotyping was carried out blind to the identity and all clinical and phenotypic information of the subjects, and as the samples were plated in approximately the order they were taken, there should be no structure in the plating which could introduce bias.

6.2.5 Data Analysis Plan

I developed a plan for genotype-phenotype association analysis of the data as follows:

6.2.5.1 Genotype definition

I planned to combine the SNP data to produce epsilon (ε) genotypes (ε2ε2, ε2ε3, ε3ε3, ε3ε4, ε4ε4, ε2ε4). The primary comparison was to be between the following groups: E2 (ε2ε2, ε2ε3), E3 (ε3ε3) and E4 (ε3ε4, ε4ε4).

6.2.5.2 Phenotype definition

The primary comparison was to use a modified TOAST classification (appendix 10), comparing large artery ischaemic stroke (LAS, n=154) with small artery ischemic stroke (SAS, n=282). Secondary analyses were to include cardioembolic stroke (with LAS); include TIAs with a visible infarct

on brain imaging; and to use the OCSP classification, comparing TACI and PACI (total and partial anterior circulation infarcts) against LACI (lacunar infarcts) (see section 1.1.3). The 1st event (within the ESS) was to be counted for each patient included.

6.2.5.3 Covariates

The following covariates were to be considered:

Age at event, sex, hypertension, diabetes, smoking, excess alcohol intake, total cholesterol plasma concentration and internal carotid artery stenosis.

6.2.5.4 Analysis

APOE genotype and each covariate were to be tested separately for association with stroke subtype, using a t-test or Mann-Whitney U test (for continuous variables) or χ^2 test (for categorical variables).

The data were to be analysed using stepwise multiple logistic regression to obtain odds ratios for LAS vs SAS, comparing the genotypes in a stepwise manner – E2 to E3 to E4, using the following covariates in each model:

- Model 1. covariates: age & sex
- Model 2. covariates: age, sex & significant ($p<0.05$) covariates (excluding cholesterol & carotid stenosis)
- Model 3. covariates: age, sex, significant covariates & cholesterol
- Model 4. covariates: age, sex, significant covariates & carotid stenosis

Comparison of models 3 and 4 with model 2 would show whether any effects of APOE are mediated through effects on cholesterol concentration and/or carotid stenosis (a measure of atheroma).

6.3 Genotyping Problems

6.3.1 1st Round of Genotyping Results

This genotyping run included all 1858 samples that were collected at the recruitment stage, including from those patients that were later excluded from the study because they turned out not to have had a stroke or TIA. The additional 158 blank controls for each assay were all called as undetermined. Call rates were 1700/1858 (91%) and 1711/1858 (92%) for SNPS rs429358 and rs7412 respectively. Table 6.1 shows the numbers of patients with each call for both SNPs.

Table 6.1 Genotype frequencies for SNPs rs429358 and rs7412, with the corresponding epsilon genotypes.

rs429358 rs7412	TT	TC	CC	undetermined	subtotals
CC	1034 (ε3ε3)	279 (ε3ε4)	29 (ε4ε4)	100	1442
CT	206 (ε3ε2)	39 (ε2ε4)	-	11	256
TT	13 (ε2ε2)	-	-	-	13
undetermined	84	12	4	47	147
subtotals	1337	330	33	158	1858

Table 6.2 Observed and expected genotype frequencies, the heterozygote observed/expected ratio and Hardy-Weinberg equilibrium p-values for SNPs rs429358 and rs7412.

		CC	CT	TT	Heterozygote observed/expected	HWE p-value
rs429358	observed	33	330	1337		
	expected	23	350	1327	0.94	0.02
rs7412	observed	1442	256	13		
	expected	1441	259	12	0.99	0.66

Table 6.2 shows the observed and expected genotype frequencies for both SNPs. rs7412 did conform to Hardy-Weinberg equilibrium (HWE) expected proportions (p=0.66). rs429358 did not conform to HWE expected frequencies (p=0.02). This is a potential cause for concern, but as this is a selected sample of stroke patients, not a random population sample, I would not necessarily expect genotypes to be in HWE.

Figures 6.1 and 6.2 show the allelic discrimination plots for the two assays, produced by the Applied Biosystems AutoCaller™ genotyping software. Each axis of these graphs represents the reporter fluorescent signal intensity (Rn) for one of two probes (each relating to an allele). Therefore when the Rn for one probe is high and for the other is very low, this represents a homozygous genotype, and when both Rn are intermediate, this represents a heterozygous genotype. Samples from up to four 96-well plates are displayed on each graph, and so three clusters of individuals with the same genotypes are normally observed. The clusters observed in figures 6.1 and 6.2 are not distinct as one would generally hope for. The plots for both SNP

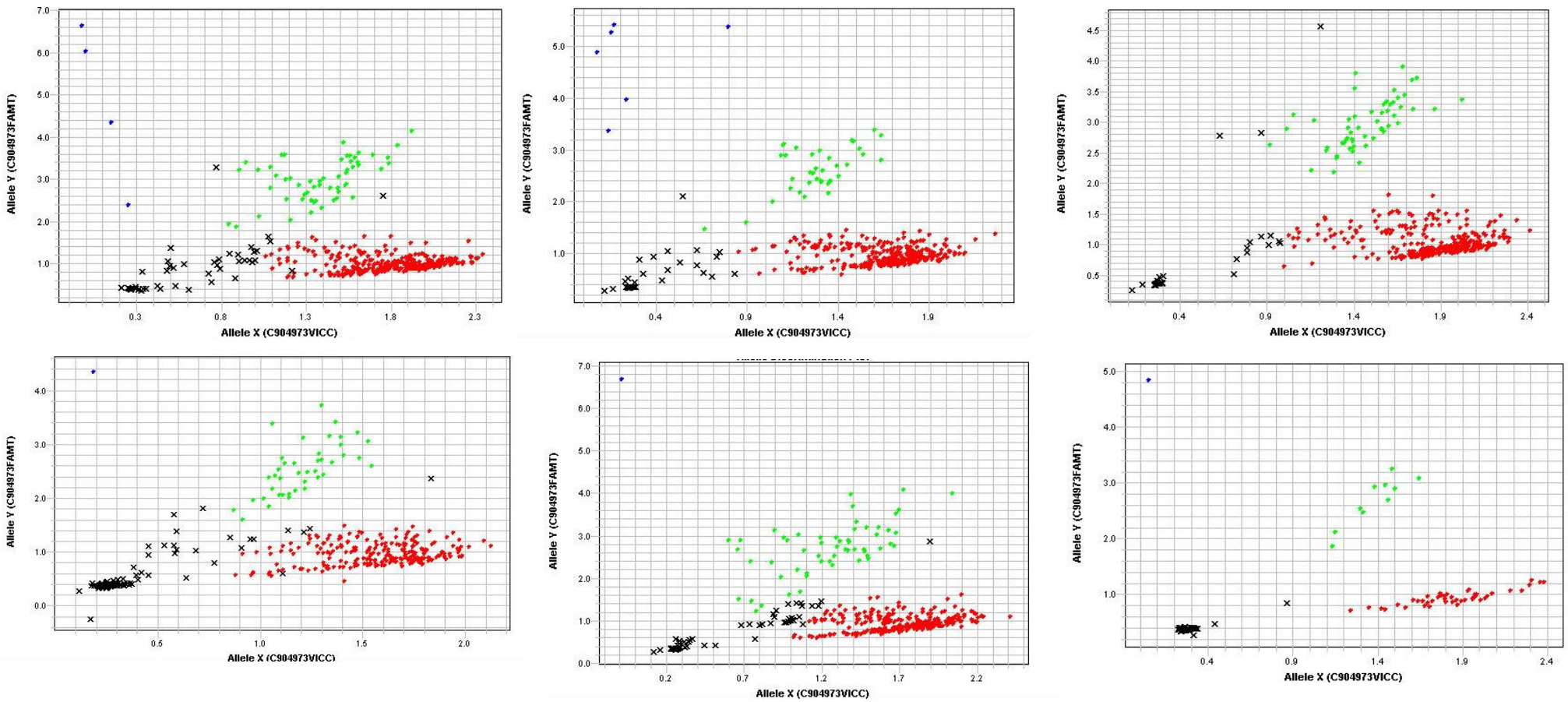


Figure 6.1 Allelic discrimination plots for assay c904973 (rs7412) from the 1st round of APOE genotyping in the Edinburgh Stroke Study.

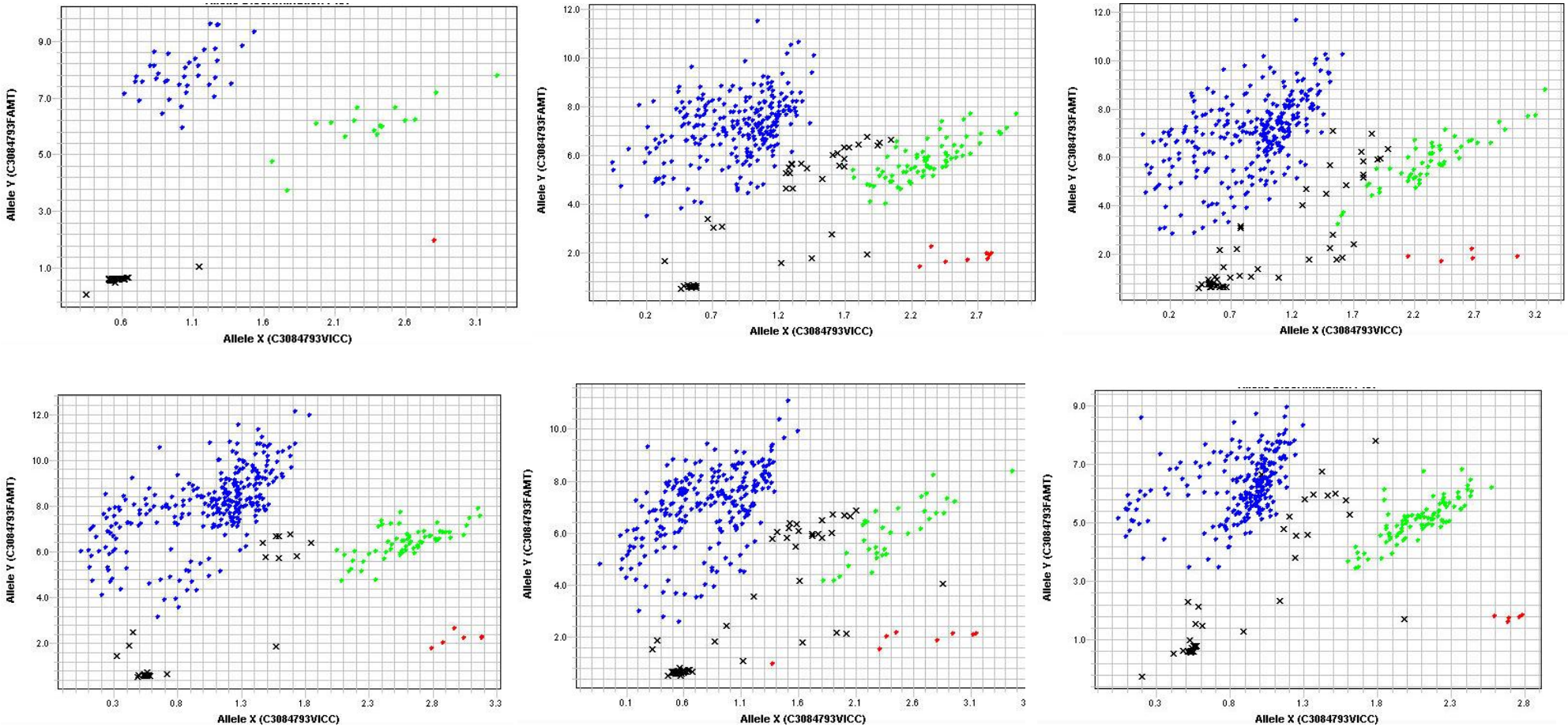


Figure 6.2 Allelic discrimination plots for assay c3084793 (rs429358) from the 1st round of APOE genotyping in the Edinburgh Stroke Study

assays show very large Rn ranges. For assay c904973 the FAM range is ~0.5 to ~7. For assay c3084793 the range is ~1 to ~12.

The clusters spread divergently and trail out from the origin radially, resulting in some overlapping of the clusters. The genotypes could not be called reliably using the automatic algorithm in the AutoCaller™ software and had to be called manually by the WTCRF technicians. In particular regions on the plots (where there was some overlap), the genotypes were called as 'undetermined'. This resulted in the systematic removal of particular samples (in this case, more heterozygotes than homozygotes are removed), which leads to bias. As the accuracy of genotyping calling using this method relies on the accuracy of the clustering, when this clustering is unsuccessful, all genotyping 'calls' should be disregarded.

6.3.2 Possible Reasons for Problems

The genotyping may have been unsuccessful due to a problem with the samples, or a problem with the assay.

6.3.2.1 Sample problems

- A problem with the samples can be detected by carrying out TaqMan genotyping on the ESS samples using an assay known to perform well in another sample collection. If the ESS samples genotype successfully with such an assay, then the problem is most likely to be with the APOE assays. If the ESS samples do not genotype successfully with

this 'good' assay, then the samples are probably the source of the problem.

- The large Rn ranges observed on the allelic discrimination plots could suggest that the DNA samples were not normalised correctly. This could be due to lab error at the normalisation stage or incorrect initial DNA concentration measurements, resulting in incorrect dilutions being applied. This could be checked by re-estimating the concentrations of the stock samples and comparing with the original concentration estimates, and by measuring the concentration of the diluted assay plates, to see if the samples are at 10ng/μl.

- Contamination or impurity of the samples may also be the source of the problem. This could be cross-contamination (although this is highly unlikely given the meticulous protocol used by the WTCRF technicians) or impurity caused by the presence of material other than DNA (including proteins, indicating problems at the DNA extraction phase). The samples can be checked for DNA purity using a nanodrop technique that measures the absorbance of the samples at particular wavelengths, which represent particular impurities.

6.3.2.2 Assay problems

- The other possible reason for poor clustering is that the assay does not work well. Using the same TaqMan assays on a separate sample collection could check this. If the genotyping is successful this

suggests that the assay is not the source of the problem. If the genotyping does not work well in this second sample, then this indicates that there is likely to be a problem with the assay.

- If there is a problem with the assay, there may be a mutation present within the primer or probe sites that prevent the assay from working as designed. This can be investigated by searching a SNP database, for potential problem SNPs. This problem could be present across all populations or be specific to particular population samples.

I carried out all of the investigations described above, to determine potential sources of the genotyping problem. The results of these investigations along with explanations and further investigations necessary are described.

6.3.3 Testing Samples With Good Assay

To test whether the ESS DNA samples were the source of the genotyping problem, TaqMan genotyping was carried out on four of the 96-well plates using assay c27915549. This assay has been previously used in the WTCRF laboratory for a separate study, and produced a clear cluster plot (see figure 6.3).

Figure 6.4 shows the cluster plot for genotyping of 380 ESS samples using assay c27915549. This plot does not have clear, distinct clusters (like figure 6.3) and many samples have been called as 'undetermined'. A large range of fluorescence intensity can be seen on the y-axis for the ESS samples (~2 to

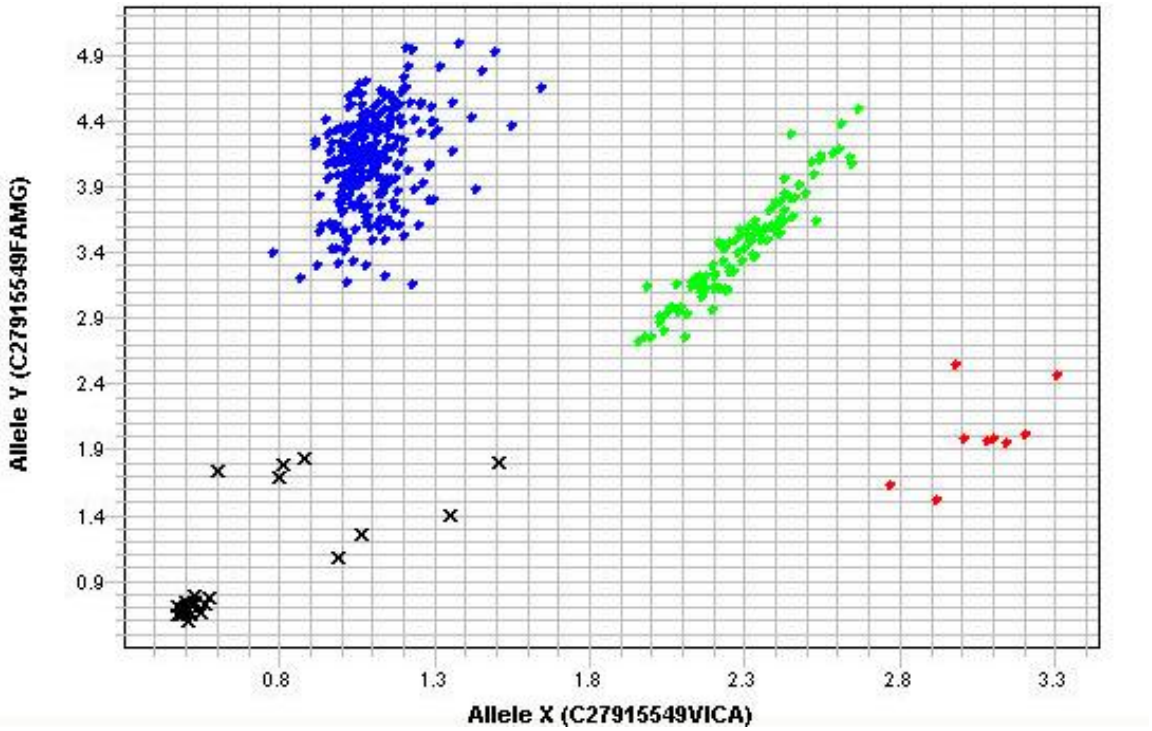


Figure 6.3 Previous genotyping of C27915549 in a different sample, indicating that it is a good assay

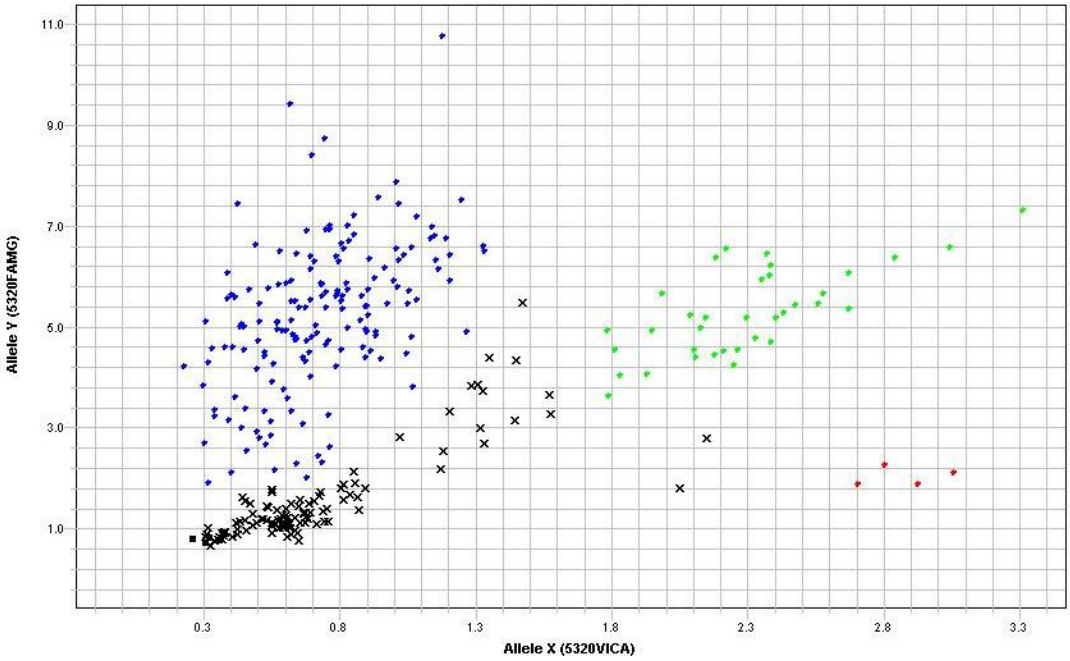


Figure 6.4 Genotyping of C27915549 on 380 Edinburgh Stroke Study samples

~11). The cluster plot for the previous genotyping using this assay has a much smaller fluorescent range on the y-axis (~3 to ~5). Figure 6.4 is similar to the cluster plots produced in our study using the APOE assays. This suggests that the genotyping problem encountered is probably due to poor quality samples, rather than poor assays. Poor samples could be due to incorrect normalisation, cross-contamination or problems during the extraction process leading to sample impurity (i.e. the samples might still contain protein, not just DNA).

6.3.4 Concentration Investigation

6.3.4.1 Testing if 96-well 'normalised' plates are at 10ng/µl

Two 96-well normalised plates were tested to check that the normalisation had been successful and that the samples were all at 10ng/µl using nanodrop (which uses a spectrophotometric method for quantifying DNA).

Figure 6.5 shows the distribution of concentrations from 190 of the samples (two 96-well plates). The estimates of the concentrations of the samples ranged from -143 to 1347ng/µl (the four minus values have been excluded from the histogram). The median was 40ng/µl and the inter quartile range (IQR) was 24 to 71ng/µl. It is clear that most samples' concentrations deviated markedly from the expected 10ng/µl. This suggests that the DNA samples were not successfully normalised. This could be due to the original concentration estimates on the 384-well stock plates being inaccurate (and therefore, inappropriate dilutions carried out) or another problem at the normalisation stage.

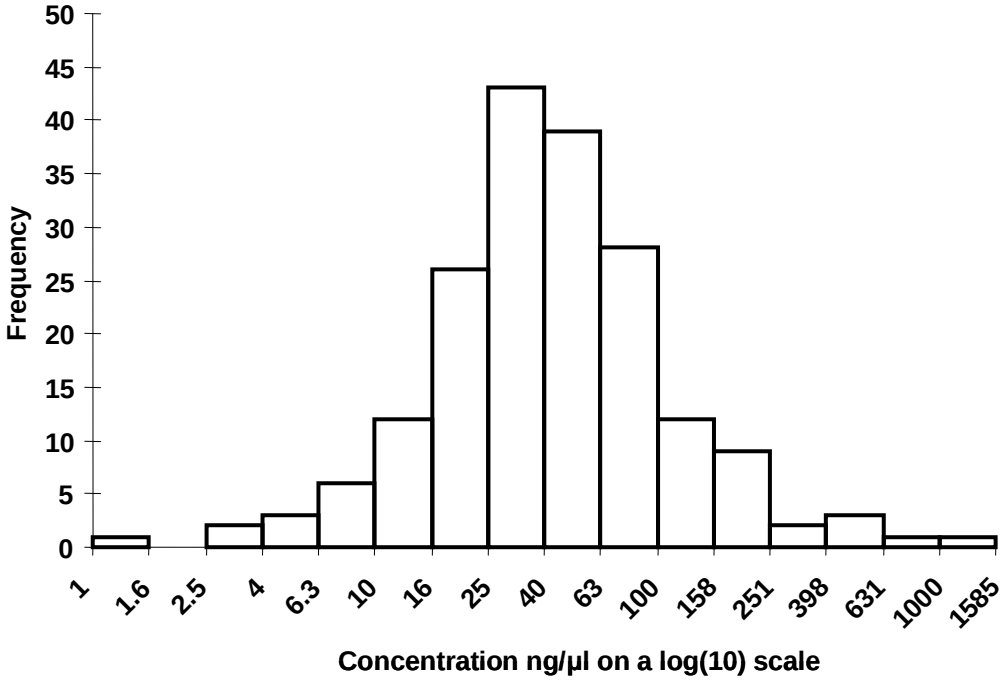


Figure 6.5 Nanodrop estimations of the concentrations of 190 of the Edinburgh Stroke Study 'normalised' samples. Samples should be normalised to 10ng/µl.

6.3.4.2 Re-measuring the 384-well stock plate concentrations

To determine if the original concentration estimates of the 384-well stock plates were accurate, these were re-estimated using PicoGreen®, after robotic remixing of the samples, to ensure they were in solution. Figure 6.6a shows the correlation between the 2nd measuring of the 384-well stock plate concentrations (A2) and the original 384-well stock plate concentration measurements (A1) for all samples. As can be seen from the figure there is very poor agreement between the two measurements. The range of concentration values in A2 was larger than the A1 estimates (max A1 = 320ng/µl, max A2 = 543ng/µl) but the mean was smaller (mean A1 = 79±50ng/µl, mean A2 = 69±74ng/µl). The Bland-Altman plot (figure 6.6b) shows the percentage differences between the A1 and A2 concentration

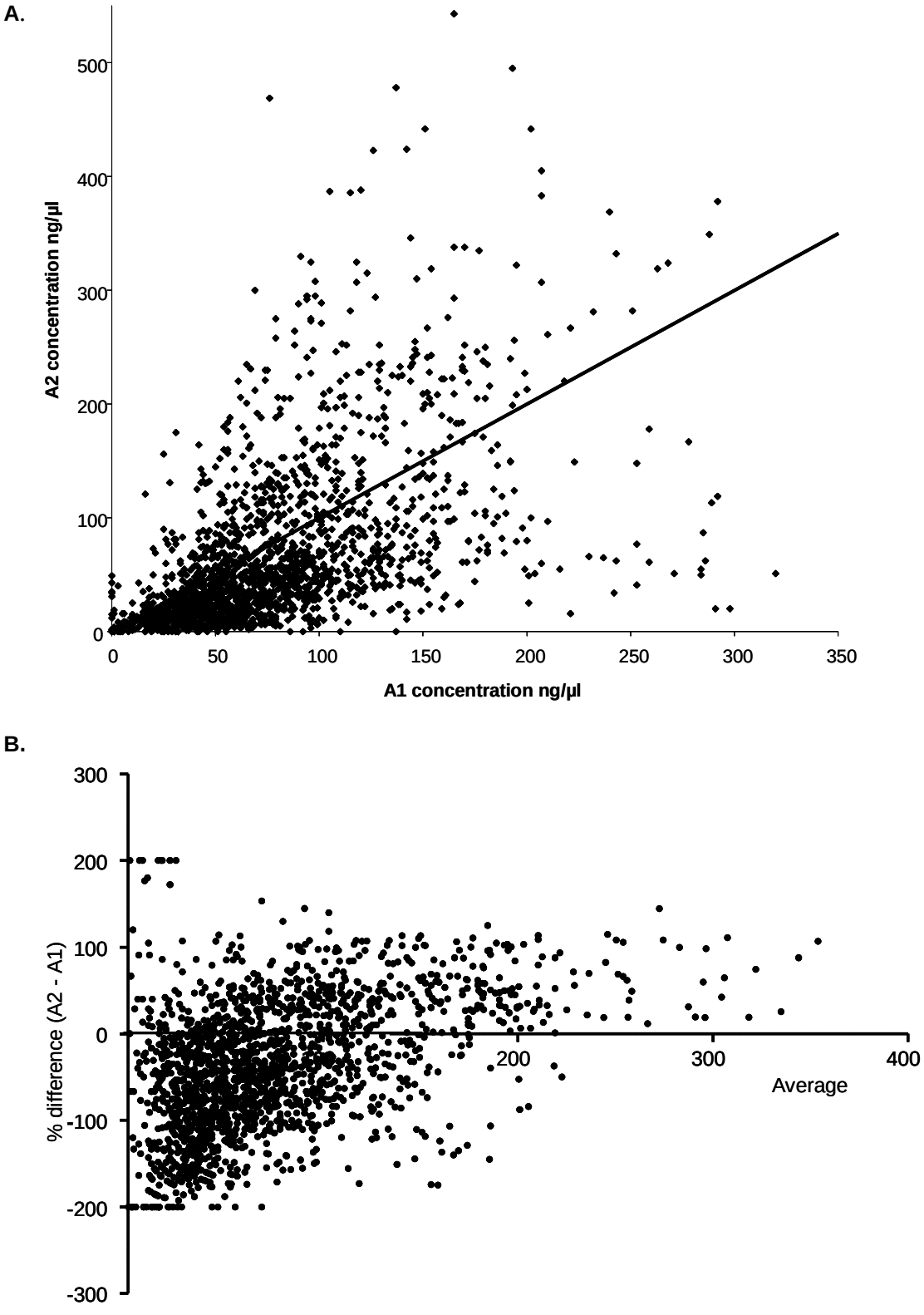


Figure 6.6 Comparison of A1 and A2 PicoGreen® concentration estimates. **A.** The two concentrations plotted against each other, the line represents A1=A2. **B.** A Bland-Altman plot of % difference against average estimate for each sample.

estimates ([(A2 – A1)/(A1 + A2)/2]*100) plotted against the average concentration estimate ((A1 + A2)/2) [Pollock *et al.*, 1992]. The 95% limit of agreement (LOA, +/-1.96 SD of the mean) is -184 to 109%. This implies that for most samples either one or both of the estimates are wrong. It is possible that the samples were not in solution for one or both estimates, and so when sampling the plates for PicoGreen® a 'glob' of DNA could have been sampled and the concentration estimated on that. The tube stock samples were originally spun on the wheel before plating out and the A1 concentration estimates being made, but they were possibly not spun for long enough. Before the A2 concentration estimates the samples were robotically mixed to get them back into solution. Assuming this was successful, the A2 concentration estimates may be more correct than the original A1 estimates. This could explain the contradicting estimates and the fact that the plates for genotyping do not appear to be at 10ng/μl (i.e. normalisation was carried out using incorrect concentration estimates).

6.3.4.3 Re-measuring the stock concentrations from tubes

Taking the A2 concentrations as correct 210, samples have a concentration of less than 10ng/μl. As the original tube stock samples may have not been in solution, these low concentrations for particular samples may be improved by going back to the tube stock. The remaining stock samples in tubes, were plated out into 384- well stock plates (I refer to these as 'B' plates), robotically mixed and the concentrations estimated using PicoGreen. 252 of the samples in the B 384-well stock plates have a concentration of less than 10ng/μl (but 130 of these had a concentration of more than 10ng/μl in A2). 88 samples

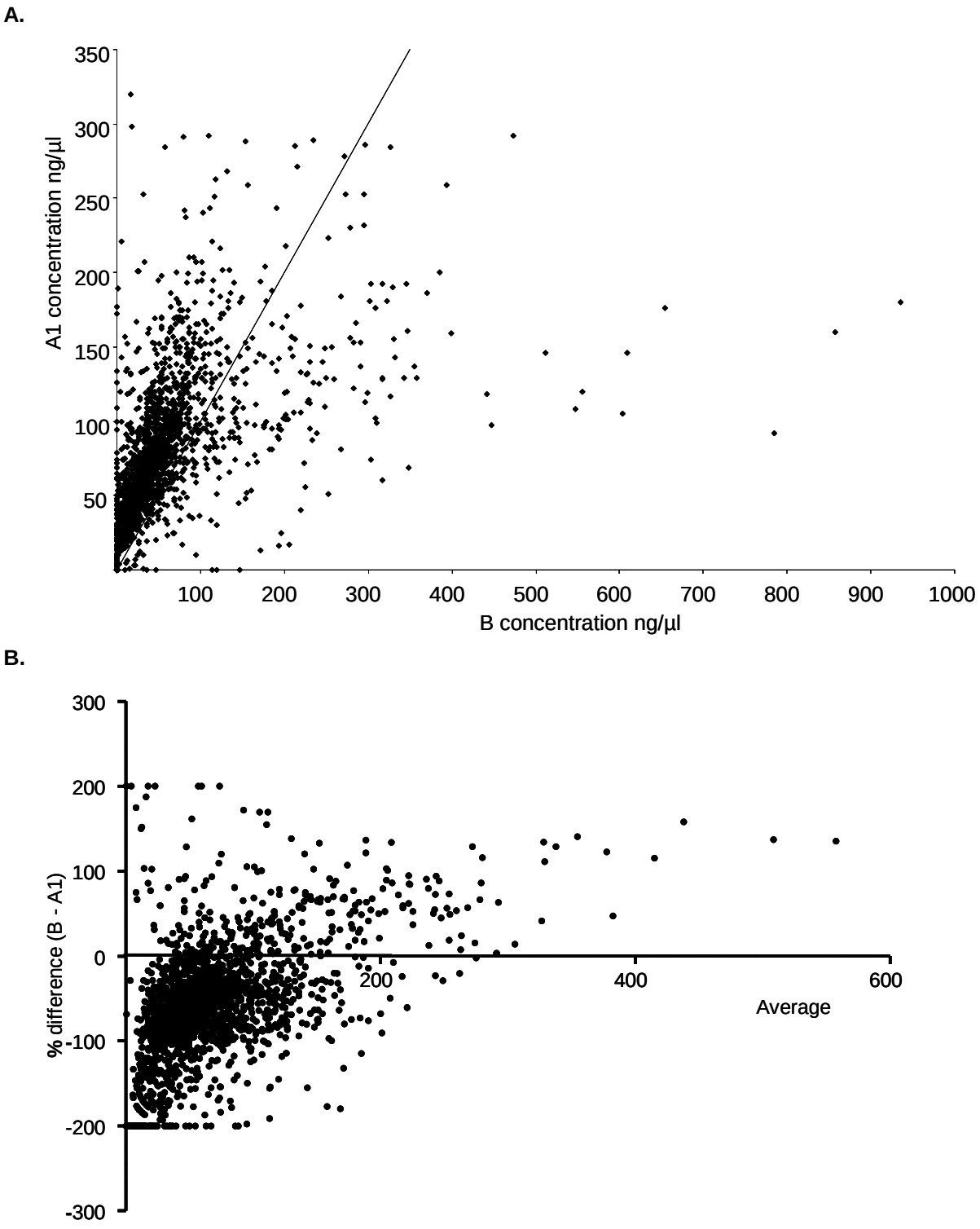


Figure 6.7 Comparison of A1 and B PicoGreen® concentration estimates. **A.** The two concentrations plotted against each other, the line represents A1=B. **B.** A Bland-Altman plot of % difference against average estimate for each sample.

which had a concentration of less than 10ng/μl in A2 had a concentration of at least 10ng/μl in the B 384-well stock plates.

The concentration estimates of the B plate samples had a range of values up to 935, with mean 60±77ng/μl. This is a lower mean but a higher maximum than the A1 and A2 estimates. Figure 6.7a shows the comparison between the original 384 –well stock plates concentration estimations (A1) and the second 384–well stock plates concentration estimations (B). There is quite substantial spread in the data indicating the agreement between the two measurements is not good. Figure 6.7b shows the Bland-Altman plot for this comparison. The limit of agreement is -192 to 80%. The agreement between A2 and B is also not good, LOA= -163 to 200% (figure 6.8).

These differences between the three concentration estimates can be explained by the tube samples not being in solution. Hence, when they were split into the two sub-samples (A and B), they were at different concentrations. The lab has taken subsequent steps, assuming that the robot mixing was successful in getting the samples into solution and so the A2 and B concentrations, although different, are assumed to be accurate.

6.3.5 2nd Round of Genotyping Results

The lab genotyped 1599 selected samples a second time (those with confirmed stroke in ESS database and with a deep well plate DNA concentration of >10ng/μl). Samples were selected from stock plates B in the first instance and this was supplemented with some samples from 96-well stock plates A (using the A2 concentrations), if the B concentration did not reach 10ng/μl, but the A2 concentration did.

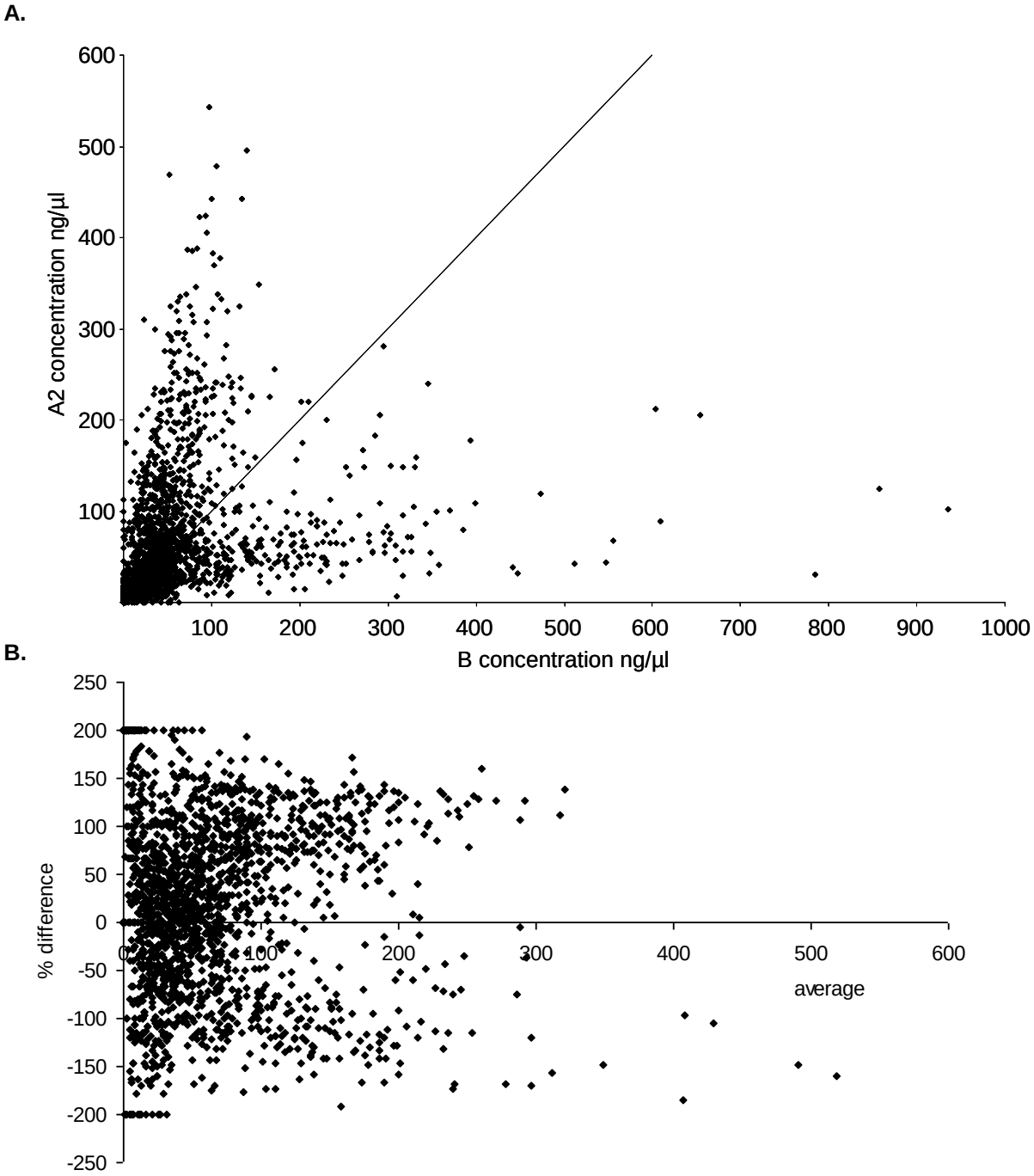


Figure 6.8 Comparison of A2 and B PicoGreen® concentration estimates. **A.** The two concentrations plotted against each other, the line represents A2=B. **B.** A Bland-Altman plot of % difference (A2-B) against average estimate for each sample.

Table 6.3 Genotype call frequencies for SNPs rs429358 and rs7412, with the corresponding epsilon genotypes, from the 2nd round of genotyping

rs429358 / rs7412	TT	TC	CC	Undetermined	Subtotals
CC	785 (ε3ε3)	253 (ε3ε4)	32 (ε4ε4)	57	1127
CT	163 (ε3ε2)	31 (ε2ε4)	-	3	197
TT	9 (ε2ε2)	-	-	1	10
undetermined	192	10	2	61	265
subtotals	1149	294	34	122	1599

Call rates were 1477/1599 (92%) and 1334/1599 (83%) for SNPS rs429358 and rs7412 respectively. Table 6.3 shows the number of individuals with each call for both SNPs. As in the 1st genotype round, rs7412 did conform to expected HWE proportions (p=0.67), but rs429358 did not (p=0.004).

A comparison of the results from the first and second genotyping attempts are shown in table 6.4. Of the samples genotyped in this second attempt,

Table 6.4 Comparison of the two genotyping rounds for both SNPs for n=1599 genotyped both times.

	rs429358	rs7412
Called same genotype in both	1304	1257
undetermined→genotyped	88	74
genotyped→ undetermined	93	238
Undetermined in both	29	27
Changed genotype	85	3
Change:	46 TT→CT 39 CT→TT	3 CT→CC

82% (1304/1599) were called the same in both attempts for rs419358 and 79% (1257/1599) were called the same for rs7412. 85/1599 (5%) were called as different genotypes in the two attempts for rs429358 (changing between TT and CT). 3/1599 (<1%) were called as different genotypes in the two attempts for rs7412 (changing from CT to CC).

Table 6.5 shows the distribution of genotypes that were called identically in both attempts. There were 1061/1599 (66%) samples that were called the same in both genotyping rounds, for both SNPs.

The allelic discrimination plots (see figures 6.9 and 6.10) show the same problems of large Rn ranges and overlapping clusters that were observed on the first genotyping attempt. This may mean that the second attempt at normalisation was no more successful than the first or that there is still some other underlying issue.

Table 6.5 Genotype call frequencies for SNPs rs429358 and rs7412 (with the corresponding epsilon genotypes) of the samples which were called identically for both SNPs, in both genotyping rounds.

rs429358 rs7412	TT	TC	CC	subtotals
CC	692 (ε3ε3)	171 (ε3ε4)	19 (ε4ε4)	882
CT	147 (ε3ε2)	23 (ε2ε4)	-	170
TT	9 (ε2ε2)	-	-	9
subtotals	848	194	19	1061

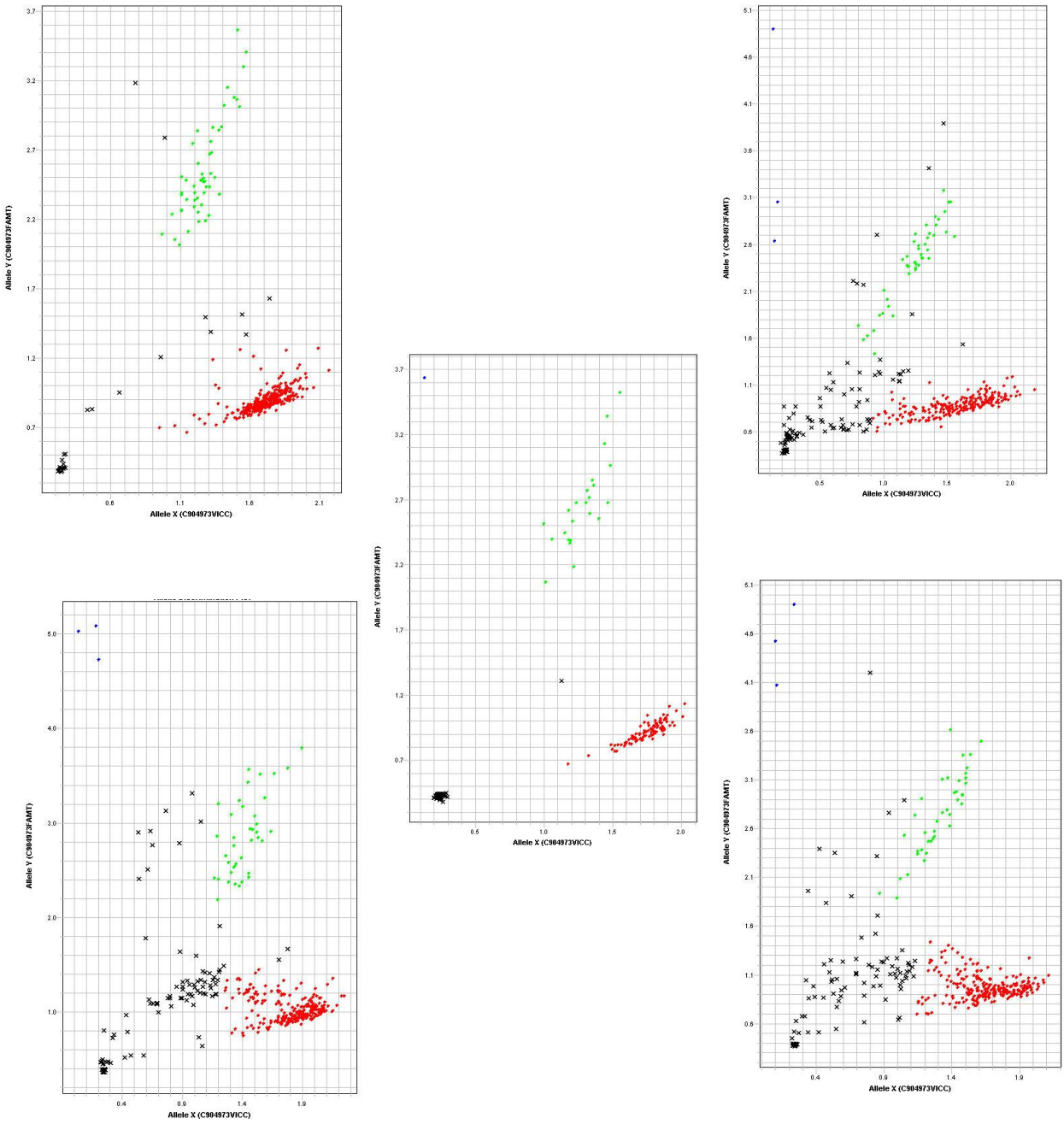


Figure 6.9 Allelic discrimination plots for assay c904973 (rs7412) from the 2nd round of APOE genotyping in the Edinburgh Stroke Study

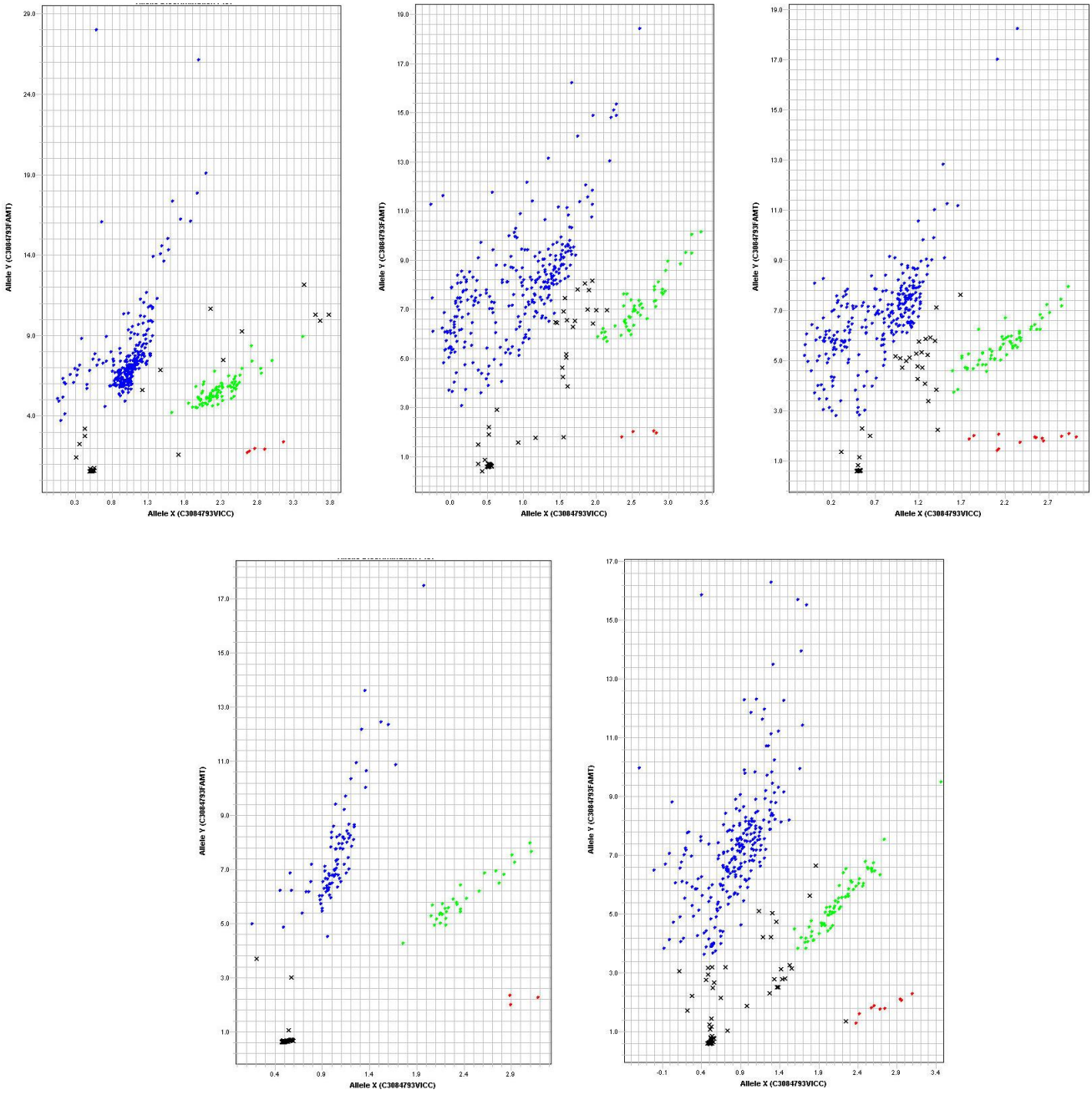


Figure 6.10 Allelic discrimination plots for assay c3084793 (rs429358) from the 2nd round of APOE genotyping in the Edinburgh Stroke Study.

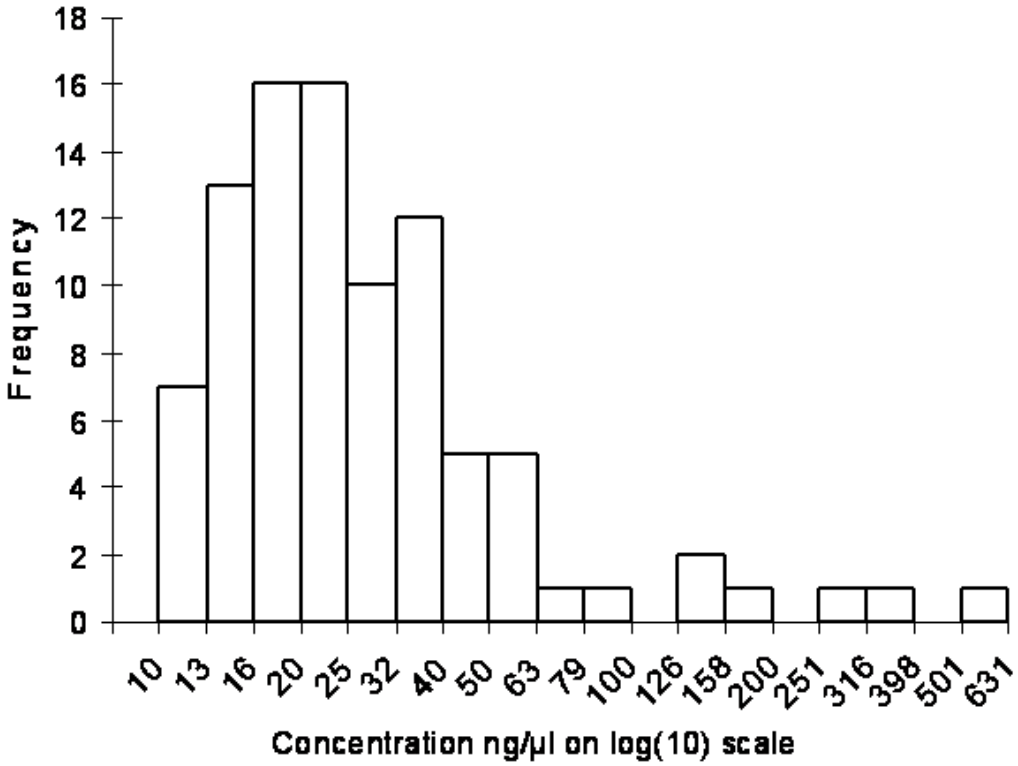


Figure 6.11 Nanodrop estimations of the concentrations of 92 of the Edinburgh Stroke Study 'normalised' samples. Samples should be normalised to 10ng/µl

6.3.5.1 Testing if genotyped plates are at 10ng/µl

Figure 6.11 shows the distribution of the nanodrop concentrations from 92 of the samples (one 96-well 'normalised' plate). All of the concentrations were >10ng/µl. The range was 11 to 504ng/µl and the median and IQR were 24ng/µl (16 to 35). These are better than seen in the first round, but are still far too variable, suggesting normalisation was unsuccessful again.

6.3.5.2 Determining if samples are in solution

The concentrations of 135 samples from the second 384-deep well stock plates (B) were estimated using PicoGreen® on two consecutive days, to test whether the samples were in solution. Figure 6.12a shows the correlation between these two estimates. The two estimates do not appear to give similar results (and neither is similar to the original B concentration estimates - data not shown). Figure 6.12b shows the Bland-Altman plot for the comparison of the results from consecutive dates. The limit of agreement is -163 to 183%. This shows that when estimating the concentration of the same samples using the same method on two consecutive days, they appear to give very different results, suggesting that they are still not in solution. It may be that the robotic mixing method was not enough to get the samples into solution. If this is the case then none of the concentration estimates are reliable and attempting to get the samples into solution should be a matter of priority.

6.3.6 Impurity of Samples

DNA quantification may be difficult if there are protein (or other) impurities in the sample. Nanodrop UV light absorption data can be used to investigate the purity of DNA samples (Thermo Scientific, T009 Technical Bulletin). The peak of UV light absorption for DNA is at 260nm and the peak for proteins (and some other contaminants) is at 280nm. Therefore, the ratio of absorbance at 260nm and 280nm ($A_{260/280}$) can be used to assess the purity of DNA. A ratio of 1.8 is considered to indicate pure DNA, and any ratio between 1.7 and 1.9 is generally considered acceptable.

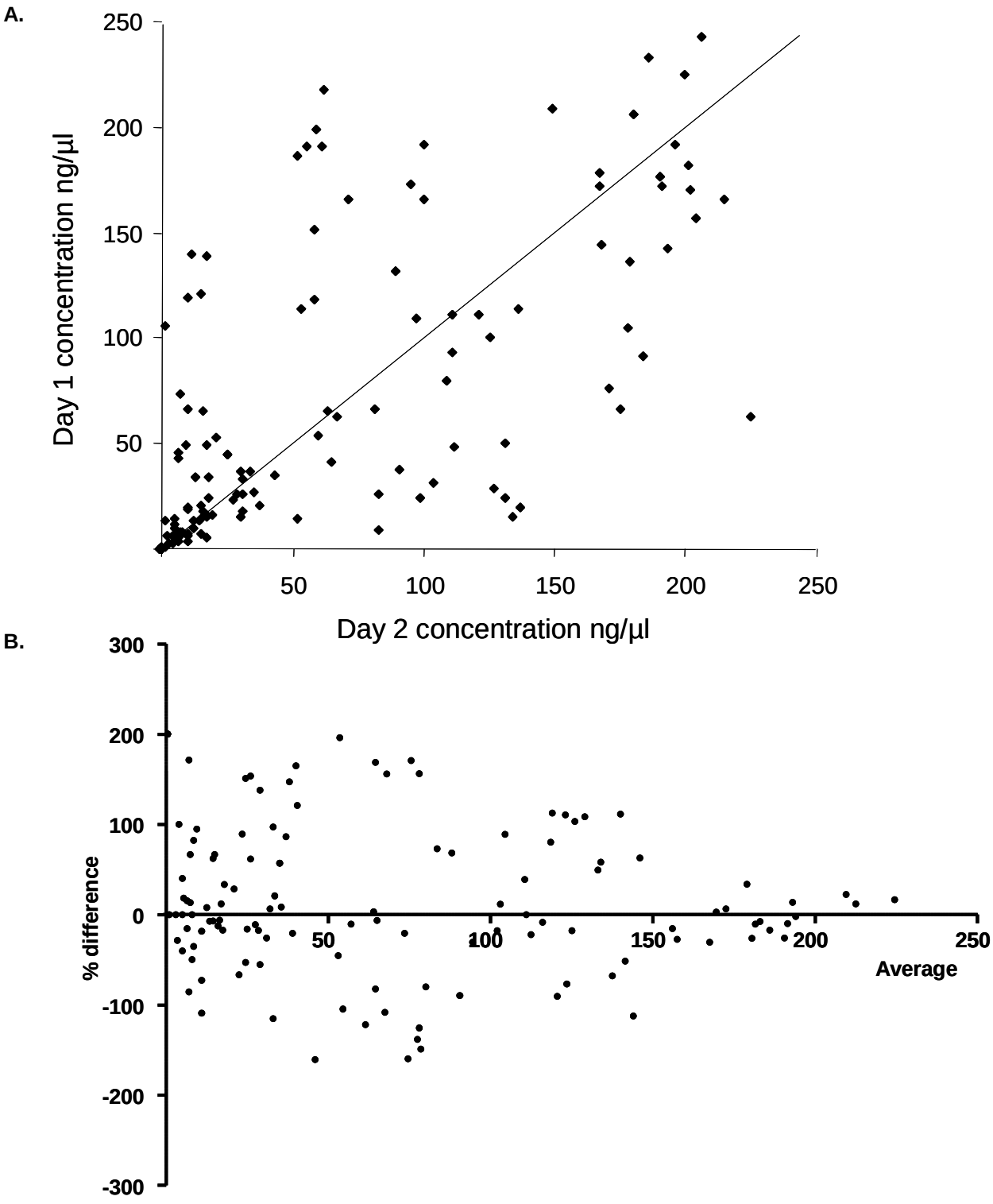


Figure 6.12 Comparison of PicoGreen® concentration estimates from the same samples from two consecutive days**. A**. The two concentrations plotted against each other, the line represents day1 = day 2. **B.** A Bland-Altman plot of % difference against average estimate for each sample.

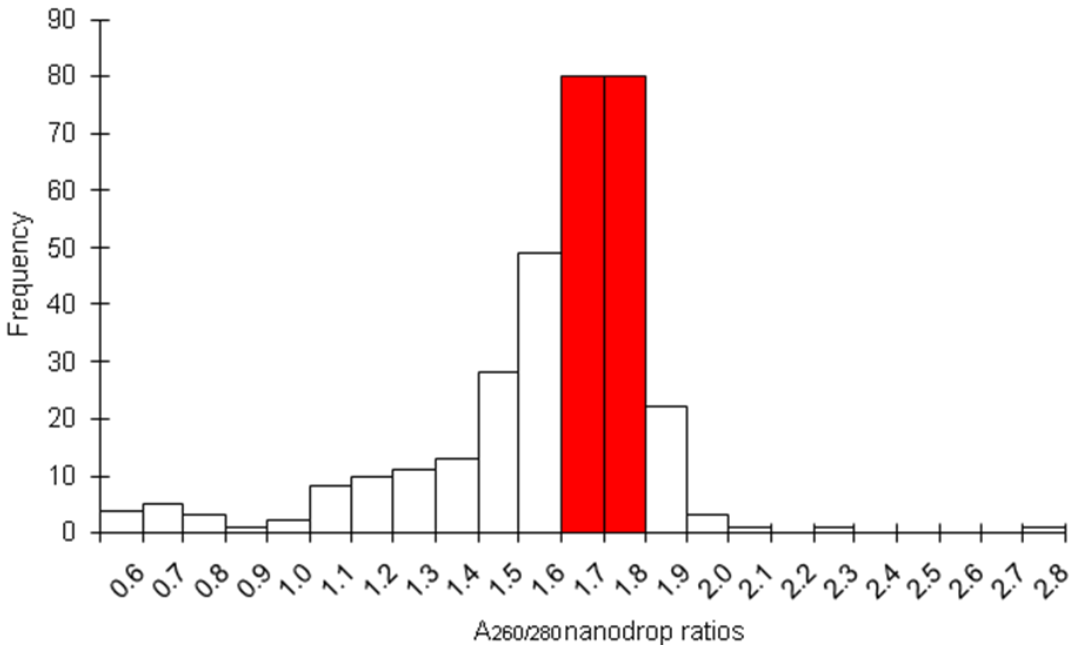


Figure 6.13 $A_{260/280}$ ratios of 322 samples from the Edinburgh Stroke Study. The acceptable range for DNA (1.7 to 1.9) is represented by the red bars.

Figure 6.13 shows the $A_{260/280}$ ratios from 322 of the ESS samples (four 96-well plates). Only 36% (116) of the samples were within the acceptable range. Most (200, 62%) were below 1.7, suggesting contamination with proteins. If DNA samples are impure, successful genotyping may be impossible. It may be worthwhile limiting the genotype calling to those samples with acceptable $A_{260/280}$ ratios. This would require access to the raw Rn data and attempting to re-cluster only the 'acceptable' samples. As I do not have access to the raw Rn data this is not something I could test and because so few samples are within the acceptable $A_{260/280}$ range, this would greatly reduce the usable dataset.

The lab staff have subsequently suggested that DNA extraction problems may be due to the small quantities of blood collected. Contamination with proteins or other impurities may be more likely if the quantity of blood collected is small. According to the Tepnel Nucleon® Genomic DNA Extraction Kit Manual, this technology has been designed for between 3 and 10ml. ESS blood samples sent for DNA extraction were 2ml or less.

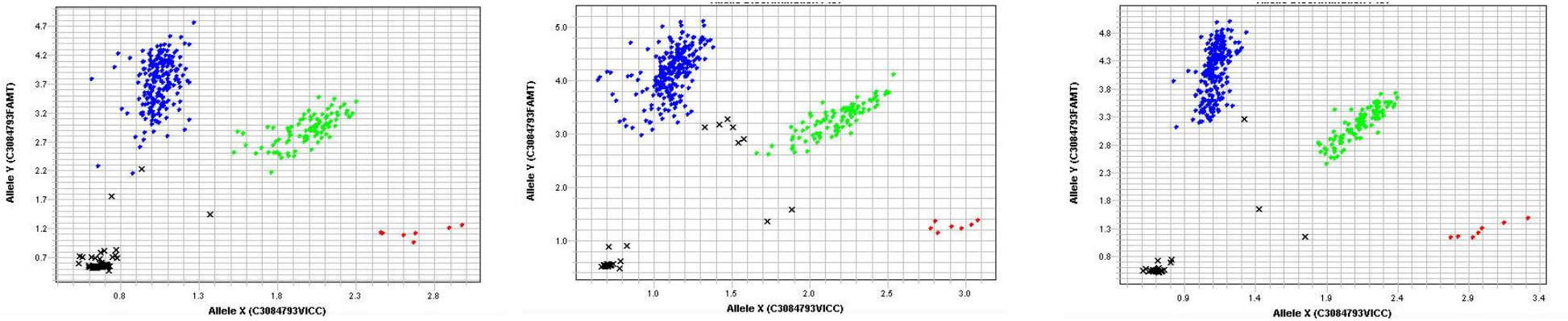
6.3.7 Validity of Assay

To check whether there was a specific problem with the ESS samples, or whether this assay did not work for other samples too, I wanted to check the assay use on a separate sample set. The WTCRF laboratory had previously genotyped the two APOE SNPs using the same assays in a Scottish population. The cluster plots produced from these samples are shown in figure 6.14. These plots show much tighter clusters and smaller Rn ranges compared to our data. However they still observed some spreading out of the clusters resulting in some samples being called as 'undetermined'.

6.3.8 Investigation of Mutation in the Primer or Probe Regions

As the primers and probes used in these assays are commercially owned I was unable to obtain the exact locations and sequences or the primers and probes. However, the probes will be within the ±25bp 'context sequences' that ABI provide. The ± 40bp region for both assays is shown in figure 6.15 along with all SNPs in this region. For the rs7412 SNP the closest SNPs were at +39 and -39 and so may not affect the c904973 assay probes. However, for the rs429358 SNP, two SNPs were very close to the SNP of interest and might be likely to affect the efficacy of the c3084793 assay probes: rs11542041 at

A. c3084793 – rs429358



B. c904973 – rs7412

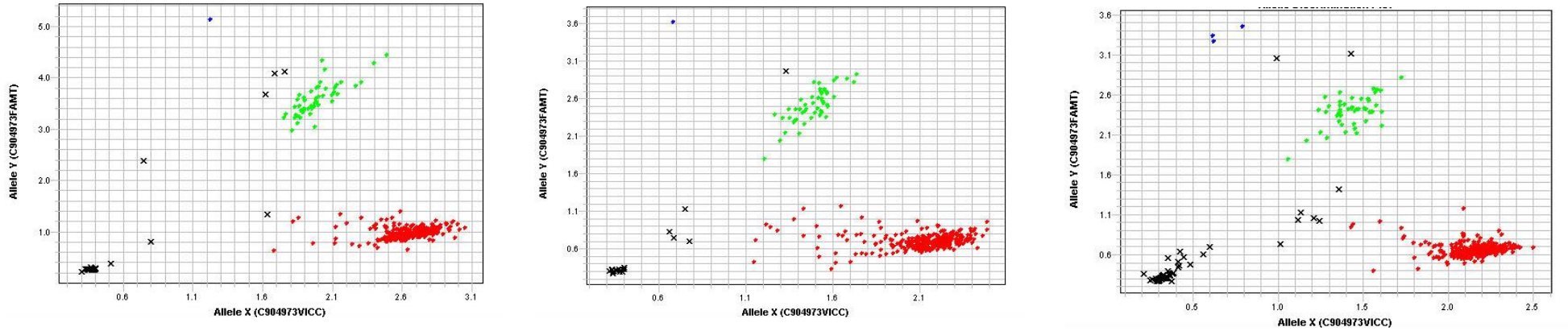


Figure 6.14 Allelic discrimination plots for assays c3084793 **(A)** and c904973 **(B)** in a different Scottish population.

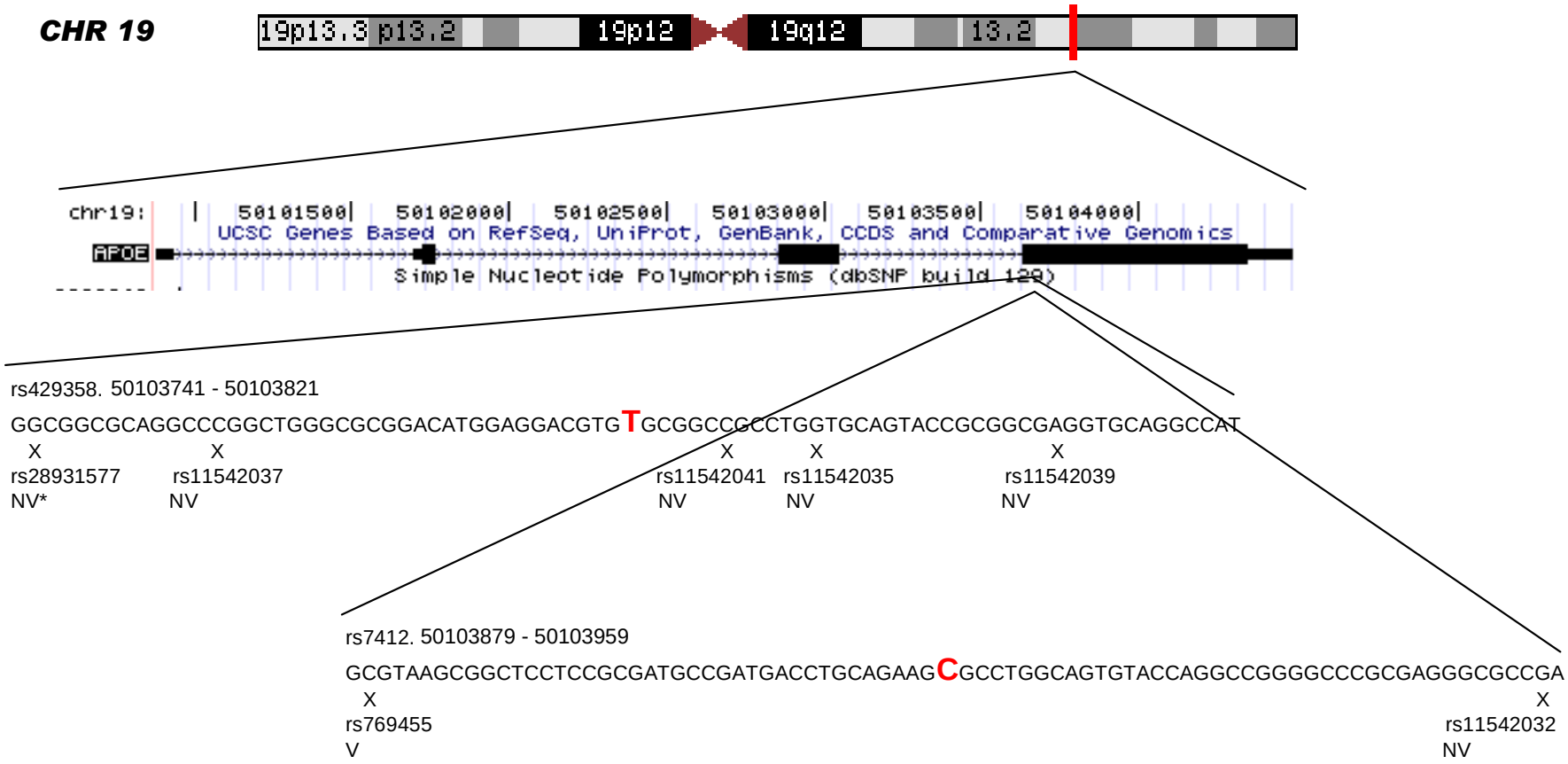


Figure 6.15 The ±40bp regions around the two SNPs typed in this study. The red letter denotes the SNP of interest, the X's denote other SNPs in the surrounding region. NV= non-validated SNP, V= validated SNP**.**

+6bp from the investigated SNP and rs11542035 at +12bp from the investigated SNP. Both of these are missense mutations but have not been extensively studied and no allele frequencies are reported in the Single Nucleotide Polymorphism database (dbSNP). There are three further SNPs within the surrounding 80bp region which may affect the efficacy of the primers used in the assay.

The previous example of assay use that the WTCRF provided me with for use as a comparison shows possible evidence of a SNP in the probe region (see cluster drifts in figure 6.14a). This drifting of clusters due to a SNP could explain some of the overlap of clusters seen for the ESS samples.

A similar data pattern has previously been observed when genotyping the Ile655Val variant of the ERBB2 gene [Benusiglio *et al.*, 2005]. The cluster plot produced from their genotyping is reproduced in figure 6.16. They observed a group of heterozygote samples that are shifted to the left (circled in the figure). They concluded that an Ile654Val variant close to the Ile655Val variant interferes with the correct binding of the probe, resulting in a cluster plot with more than 3 clusters. This kind of scenario leaves the investigator with a dilemma of what to do with these extra clusters. If they are classified as undetermined (as in the case of our data and the ERBB2 data), this could introduce bias, as a systematic group of samples are unclassified. The shift to the left in the heterozygotes results in unclassified samples, whilst a similar shift in the homozygotes may result in the homozygote genotype still being called. By systematically removing particular individuals (heterozygotes) from the analysis, bias is introduced.

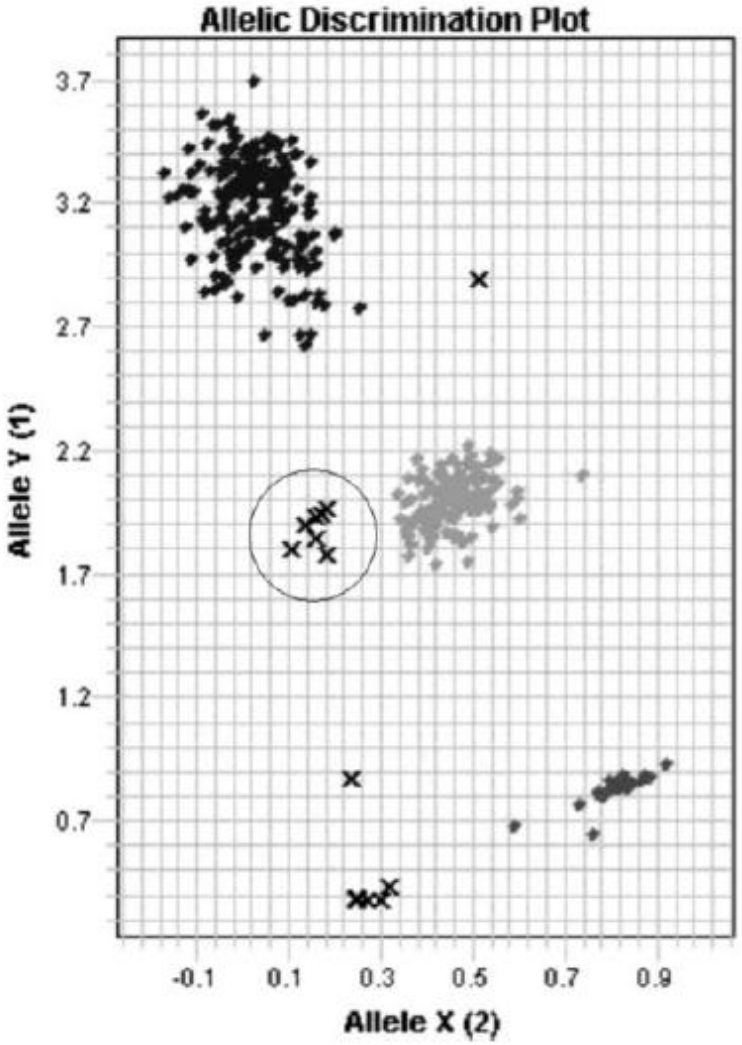


Figure 6.16 Allelic discrimination plot from a study [Benusiglio *et al.,* 2005] that reported an underlying SNP in the probe region as seen by the circled samples that have drifted to the left of the heterozygote cluster.

6.3.10 Future Directions

The lab have now combined the A and B 384-well stock plate samples and have robotically mixed these combined samples twice in an attempt to get the samples into solution. Preliminary indications from the lab are that they are getting repeatable PicoGreen® concentration estimates from the samples suggesting that they may finally be in solution.

It has been agreed between the lab staff and ESS investigators that although there may be sample problems, some samples will probably genotype ok. There are potentially some issues with the APOE assays, and so deciding which samples may genotype is difficult from the results of these assays, so other assays must be used to select the 'good quality' samples for future genotyping. A panel of 14 SNPs that have previously genotyped well (>93% call rate) in a Scottish population is going to be used to select the usable ESS samples.

6.4 Impact on Future Work and Other Studies

The ABI assays are designed and tested using a small specific population. It is quite likely that these may not work well in other populations. It may be sensible to design assays specifically for the study population, taking into account nearby SNPs. Although high through-put genotyping has its uses, this shows that it may not always be appropriate.

For future genotyping in the ESS, the samples need to be good quality (properly normalised and without contamination) and poor quality samples should be excluded. To increase the genotyping success rate in future studies it may be sensible to collect larger quantities of blood (between 3 and 10ml), to improve the DNA extraction step.

6.5 Conclusion

I have not carried out my planned analysis of the association between APOE genotype and stroke subtype because of the poor quality of the samples (and possibly also the problematic assays). The accuracy of the genotyping (even for those where calls were the same in both) is in question. This is reflected in the poor allelic discrimination plots and departure from Hardy-Weinberg equilibrium. These serve as an alert that there have been problems in the genotyping and further quality control checks need to be carried out on these samples before proceeding with data analysis.

References

Abbott RD, Donahue RP, MacMahon SW, Reed DM, Yano K (1987) Diabetes and the risk of stroke. The Honolulu Heart Program. *JAMA,* **257:** 949-952.

Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE, III (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke,* **24:** 35-41.

Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A (2000) ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. *Arterioscler.Thromb.Vasc.Biol.,* **20:** 484-492.

Allender S, Peto V, Scarborough P, Kaur A, Rayner M (2008) *Coronary Heart Disease Statistics*. BHF, London.

Altamura C, Squitti R, Pasqualetti P, Tibuzzi F, Silvestrini M, Ventriglia MC, Cassetta E, Rossini PM, Vernieri F, Vernieri F (2007) What is the relationship among atherosclerosis markers, apolipoprotein E polymorphism and dementia? *Eur.J.Neurol.,* **14:** 679-682.

Altman DG, Deeks JJ (2002) Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC.Med.Res.Methodol.,* **2:** 3.

Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. *Science,* **322:** 881-888.

Amar K, MacGowan S, Wilcock G, Lewis T, Scott M (1998) Are genetic factors important in the aetiology of leukoaraiosis? Results from a memory clinic population. *Int.J.Geriatr.Psychiatry,* **13:** 585-590.

Andrew PJ, Mayer B (1999) Enzymatic function of nitric oxide synthases. *Cardiovasc.Res.,* **43:** 521-531.

Arai K, Yamasaki Y, Kajimoto Y, Watada H, Umayahara Y, Kodama M, Sakamoto K, Hori M (1997) Association of methylenetetrahydrofolate reductase gene polymorphism with carotid arterial wall thickening and myocardial infarction risk in NIDDM. *Diabetes,* **46:** 2102-2104.

Ariyaratnam R, Casas JP, Whittaker J, Smeeth L, Hingorani AD, Sharma P (2007) Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. *PLoS.Med.,* **4:** e131.

Arnett DK, Borecki IB, Ludwig EH, Pankow JS, Myers R, Evans G, Folsom AR, Heiss G, Higgins M (1998) Angiotensinogen and angiotensin converting enzyme genotypes and carotid atherosclerosis: The atherosclerosis risk in communities and the NHLBI family heart studies. *Atherosclerosis,* **138:** 111-116.

Asakimori Y, Yorioka N, Tanaka J, Kohno N (2003) Effect of polymorphism of the endothelial nitric oxide synthase and apolipoprotein E genes on carotid atherosclerosis in hemodialysis patients. *Am.J.Kidney Dis.,* **41:** 822-832.

Attia J, Thakkinstian A, D'Este C (2003) Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J.Clin.Epidemiol.,* **56:** 297-303.

Atwood LD, Wolf PA, Heard-Costa NL, Massaro JM, Beiser A, D'Agostino RB, Decarli C (2004) Genetic variation in white matter hyperintensity volume in the Framingham Study. *Stroke,* **35:** 1609-1613.

Awad IA, Spetzler RF, Hodak JA, Awad CA, Carey R (1986) Incidental subcortical lesions identified on magnetic resonance imaging in the elderly. I. Correlation with age and cerebrovascular risk factors. *Stroke,* **17:** 1084-1089.

Bachmann G, Damian MS, Koch M, Schilling G, Fach B, Stoppler S (1996) The clinical and genetic correlates of MRI findings in myotonic dystrophy. *Neuroradiology,* **38:** 629-635.

Balkestein EJ, Ji GW, Struijker-Boudier HAJ, Barlassina C, Bianchi G, Birkenhager WH, Brand E, Den HE, Fagard R, Herrmann S-M, Van Bortel LM, Staessen JA (2002) Carotid and femoral intima-media thickness in relation to three candidate genes in a Caucasian population. *J.Hypertens.,* **20:** 1551-1561.

Bamford J, Sandercock P, Dennis M, Burn J, Warlow C (1991) Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet,* **337:** 1521-1526.

Barber R, Gholkar A, Scheltens P, Ballard C, McKeith IG, Morris CM, O'Brien JT (1999) Apolipoprotein E epsilon4 allele, temporal lobe atrophy, and white matter lesions in late-life dementias. *Arch.Neurol.,* **56:** 961-965.

Barley J, Markus H, Brown M, Carter N (1995) Lack of association between angiotensinogen polymorphism (M235T) and cerebrovascular disease and carotid atheroma. *J.Hum.Hypertens.,* **9:** 681-683.

Bartoli F, Angotti C, Fatini C, Conforti ML, Guiducci S, Blagojevic J, Melchiorre D, Fiori G, Generini S, Damjanov N, Rednic S, Pignone A, Castellani S, Abbate R, Matucci CM (2007) Angiotensin-converting enzyme I/D polymorphism and macrovascular disease in systemic sclerosis. *Rheumatology (Oxford).,* **46:** 772-775.

Bartres-Faz D, Junque C, Clemente IC, Serra-Grabulosa JM, Guardia J, Lopez-Alomar A, Sanchez-Aldeguer J, Mercader JM, Bargallo N, Olondo M, Moral P (2001) MRI and genetic correlates of cognitive function in elders with memory impairment. *Neurobiol.Aging,* **22:** 449-459.

Bednarska MM, Rodo M, Markuszewski C, Rozenfeld A, Swiderska M, Habrat B, Wehr H (2005) Polymorphisms of apolipoprotein E and angiotensin-converting enzyme genes and carotid atherosclerosis in heavy drinkers. *Alcohol & Alcoholism,* **40:** 274-282.

Beilby JP, Hunt CCJ, Palmer LJ, Chapman CML, Burley JP, McQuillan BM, Thompson PL, Hung J (2003) Apolipoprotein E gene polymorphisms are associated with carotid plaque formation but not with intima-media wall thickening: Results from the Perth Carotid Ultrasound Disease Assessment Study (CUDAS). *Stroke,* **34:** 869-874.

Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, Keavney B, Collins R, Wiman B, de Faire U, Danesh J (2007) Association of Apolipoprotein E Genotypes With Lipid Levels and Coronary Risk. *JAMA: The Journal of the American Medical Association,* **298:** 1300-1311.

Benusiglio PR, Luccarini C, Ajai O, Ponder BA, Pharoah PD (2005) Reply: a bias in genotyping of the ERBB2 (HER2) Ile655Val variant. *Carcinogenesis,* **26:** 2213.

Bersano A, Ballabio E, Bresolin N, Candelise L (2008) Genetic polymorphisms for the study of multifactorial stroke. *Hum.Mutat.,* **29:** 776-795.

Bevan S, Dichgans M, Gschwendtner A, Kuhlenbaumer G, Ringelstein EB, Markus HS (2008) Variation in the PDE4D gene and ischemic stroke risk: a systematic review and meta-analysis on 5200 cases and 6600 controls. *Stroke,* **39:** 1966-1971.

Bhuiyan AR, Chen W, Srinivasan SR, Rice J, Mock N, Tang R, Bond MG, Boerwinkle E, Berenson GS (2007) Influence of nitric oxide synthase gene polymorphism (G894T) on carotid artery intima-media thickness in adults: the Bogalusa Heart Study. *J.Am.Soc.Hypertens.,* **1:** 362-368.

Bianchi G, Ferrari P, Staessen JA (2005) Adducin polymorphism: detection and impact on hypertension and related disorders. *Hypertension,* **45:** 331-340.

Bianchi G, Tripodi G, Casari G, Salardi S, Barber BR, Garcia R, Leoni P, Torielli L, Cusi D, Ferrandi M, . (1994) Two point mutations within the adducin genes are involved in blood pressure variation. *Proc.Natl.Acad.Sci.U.S.A,* **91:** 3999-4003.

Bigler ED, Lowry CM, Kerr B, Tate DF, Hessel CD, Earl HD, Miller MJ, Rice SA, Smith KH, Tschanz JT, Welsh-Bohmer K, Plassman B, Victoroff J (2003) Role of white matter lesions, cerebral atrophy, and APOE on cognition in older persons with and without dementia: the Cache County, Utah, study of memory and aging. *Neuropsychology,* **17:** 339-352.

Bilici A, Ulgen MS, Nazaroglu H, Ozturk O, Ekici F, Akgul C, Alan B (2006) The effect of ACE gene polymorphisms on doppler blood flow parameters of carotid and brachial arteries in patients with myocardial infarction. *Angiology,* **57:** 681-685.

Bleil ME, Ferrell RE, Sutton-Tyrrell K, Muldoon MF, Manuck SB (2006) Apolipoprotein E polymorphism and preclinical carotid artery disease in untreated hypertensive men. *Eur.J.Cardiovasc.Prev.Rehab.,* **13:** 98-100.

Bloem LJ, Foroud TM, Ambrosius WT, Hanna MP, Tewksbury DA, Pratt JH (1997) Association of the angiotensinogen gene to serum angiotensinogen in blacks and whites. *Hypertension,* **29:** 1078-1082.

Bornebroek M, Haan J, Backhovens H, Deutz P, Van Buchem MA, van den BM, Bakker E, Roos RAC, Van BC (1997a) Presenilin- 1 polymorphism and hereditary cerebral hemorrhage with amyloidosis, dutch type. *Ann.Neurol.,* **42:** 108-110.

Bornebroek M, Haan J, van Duinen SG, Maat-Schieman ML, Van Buchem MA, Bakker E, Van BC, Roos RA (1997b) Dutch hereditary cerebral amyloid angiopathy: structural lesions and apolipoprotein E genotype. *Ann.Neurol.,* **41:** 695-698.

Bozec E, Fassot C, Tropeano A-I, Boutouyrie P, Jeunemaitre X, Lacolley P, Dabire H, Laurent S (2003) Angiotensinogen gene M235T polymorphism and reduction in wall thickness in response to antihypertensive treatment. *Clin.Sci.,* **105:** 637-644.

Bracco L, Piccini C, Moretti M, Mascalchi M, Sforza A, Nacmias B, Cellini E, Bagnoli S, Sorbi S (2005) Alzheimer's disease: role of size and location of white matter changes in determining cognitive deficits. *Dement.Geriatr.Cogn.Disord.,* **20:** 358-366.

Brattstrom L, Wilcken DE, Ohrvik J, Brudin L (1998) Common methylenetetrahydrofolate reductase gene mutation leads to hyperhomocysteinemia but not to vascular disease: the result of a meta-analysis. *Circulation,* **98:** 2520-2526.

Brenner D, Labreuche J, Touboul P-J, Schmidt-Petersen K, Poirier O, Perret C, Schonfelder J, Combadiere C, Lathrop M, Cambien F, Brand-Herrmann S-M, Amarenco P (2006) Cytokine polymorphisms associated with carotid intima-media thickness in stroke patients. *Stroke,* **37:** 1691-1696.

Bronge L, Fernaeus SE, Blomberg M, Ingelson M, Lannfelt L, Isberg B, Wahlund LO (1999) White matter lesions in Alzheimer patients are influenced by apolipoprotein E genotype. *Dement.Geriatr.Cogn.Disord.,* **10:** 89-96.

Burdon KP, Langefeld CD, Beck SR, Wagenknecht LE, Carr JJ, Freedman BI, Herrington D, Bowden DW (2005) Association of genes of lipid metabolism with measures of subclinical cardiovascular disease in the Diabetes Heart Study. *J.Med.Genet.,* **42:** 720-724.

Burdon KP, Langefeld CD, Wagenknecht LE, Carr JJ, Freedman BI, Herrington D, Bowden DW (2006) Association analysis of genes in the renin-angiotensin system with subclinical cardiovascular disease in families with Type 2 diabetes mellitus: The Diabetes Heart Study. *Diabet.Med.,* **23:** 228-234.

Cambien F (2007) Genome-wide asociation studies: a new era in human genetics. *Current Cardiovascular Risk Reports,* **1:** 271-272.

Campo S, Sardo MA, Trimarchi G, Bonaiuto M, Castaldo M, Fontana L, Bonaiuto A, Bitto A, Saitta C, Saitta A (2004) The paraoxonase promoter polymorphism (-107)T>C is not associated with carotid intima-media thickness in Sicilian hypercholesterolemic patients. *Clin.Biochem.,* **37:** 388-394.

Cao H, Girard-Globa A, Serusclat A, Bernard S, Bondon P, Picard S, Berthezene F, Moulin P (1998) Lack of association between carotid intima-media thickness and paraoxonase gene polymorphism in non-insulin dependent diabetes mellitus. *Atherosclerosis,* **138:** 361-366.

Carluccio M, Soccio M, De CR (2001) Aspects of gene polymorphisms in cardiovascular disease: the renin-angiotensin system. *Eur.J.Clin.Invest,* **31:** 476-488.

Carmelli D, DeCarli C, Swan GE, Jack LM, Reed T, Wolf PA, Miller BL (1998) Evidence for genetic variance in white matter hyperintensity volume in normal elderly male twins. *Stroke,* **29:** 1177-1181.

Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD (2005) Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet,* **365:** 224-232.

Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, Hingorani AD (2006) Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *Am.J.Epidemiol.,* **164:** 921-935.

Casas JP, Hingorani AD, Bautista LE, Sharma P (2004) Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch.Neurol.,* **61:** 1652-1661.

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

Castellano M, Barlassina C, Muiesan ML, Beschi M, Cinelli A, Rossi F, Rizzoni D, Cusi D, gabiti-Rosei E (1997) Alpha-adducin gene polymorphism and cardiovascular phenotypes in a general population. *J.Hypertens.,* **15:** 1707-1710.

Castellano M, Muiesan ML, Rizzoni D, Beschi M, Pasini G, Cinelli A, Salvetti M, Porteri E, Bettoni G, Kreutz R, Lindpaintner K, Rosei EA (1995) Angiotensin-converting enzyme I/D polymorphism and arterial wall thickness in a general population: The Vobarno Study. *Circulation,* **91:** 2721-2724.

Cattin L, Fisicaro M, Tonizzo M, Valenti M, Danek GM, Fonda M, Da Col PG, Casagrande S, Pincetti E, Bovenzi M, Baralle F (1997) Polymorphism of the apolipoprotein E gene and early carotid atherosclerosis defined by ultrasonography in asymptomatic adults. *Arterioscler.Thromb.Vasc.Biol.,* **17:** 91-94.

Chapman CML, Beilby JP, Humphries SE, Palmer LJ, Thompson PL, Hung J (2003) Association of an allelic variant of interleukin-6 with subclinical carotid atherosclerosis in an Australian community population. *Eur.Heart J.,* **24:** 1494-1499.

Cheung EY, Bos MJ, Leebeek FW, Koudstaal PJ, Hofman A, de Maat MP, Breteler MM (2008) Variation in fibrinogen FGG and FGA genes and risk of stroke: the Rotterdam Study. *Thromb.Haemost.,* **100:** 308-313.

Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES (2007) Distinguishing protein-coding and noncoding genes in the human genome. *Proc.Natl.Acad.Sci.U.S.A,* **104:** 19428-19433.

Clarke MJ, Stewart LA (2001) Obtainign individual pateint ata from andomised controlled trials. In:*Systematic Reviews* (eds Egger M, Davey SG, Altman DG), pp. 109-121. BMJ Books, London.

Cordell HJ, Clayton DG (2005) Genetic association studies. *Lancet,* **366:** 1121-1131.

Cronin S, Furie KL, Kelly PJ (2005) Dose-related association of MTHFR 677T allele with risk of ischemic stroke: evidence from a cumulative meta-analysis. *Stroke,* **36:** 1581-1587.

Czarnecka D, Kawecka-Jaszcz K, Stolarz K, Olszanecka A, Kiec-Wilk B, mbinska-Kiec A, Narkiewicz K (2004) Genetic factors in hypertension. Angiotensin-converting enzyme polymorphism. *Kardiol.Pol.,* **61:** 1-10.

Czarnecka D, Kawecka-Jaszcz K, Stolarz K, Olszanecka A, Dembinska-Kiee A, Kiec-Wilk B (2005) Ambulatory blood pressure, left ventricular mass and vascular phenotypes in relation to the endothelial nitric oxide synthase gene Glu298Asp and intron 4 polymorphisms in a population-based family study. *J.Hum.Hypertens.,* **19:** 413-420.

Danesh J, Lewington S, Thompson SG, Lowe GD, Collins R, Kostis JB, Wilson AC, Folsom AR, Wu K, Benderly M, Goldbourt U, Willeit J, Kiechl S, Yarnell JW, Sweetnam PM, Elwood PC, Cushman M, Psaty BM, Tracy RP, Tybjaerg-Hansen A, Haverkate F, de Maat MP, Fowkes FG, Lee AJ, Smith FB, Salomaa V, Harald K, Rasi R, Vahtera E, Jousilahti P, Pekkanen J, D'Agostino R, Kannel WB, Wilson PW, Tofler G, rocha-Pinango CL, Rodriguez-Larralde A, Nagy E, Mijares M, Espinosa R, Rodriquez-Roa E, Ryder E, ez-Ewald MP, Campos G, Fernandez V, Torres E, Marchioli R, Valagussa F, Rosengren A, Wilhelmsen L, Lappas G, Eriksson H, Cremer P, Nagel D, Curb JD, Rodriguez B, Yano K, Salonen JT, Nyyssonen K, Tuomainen TP, Hedblad B, Lind P, Loewel H, Koenig W, Meade TW, Cooper JA, De SB, Knottenbelt C, Miller GJ, Cooper JA, Bauer KA, Rosenberg RD, Sato S, Kitamura A, Naito Y, Palosuo T, Ducimetiere P, Amouyel P, Arveiler D, Evans AE, Ferrieres J, Juhan-Vague I, Bingham A, Schulte H, Assmann G, Cantin B, Lamarche B, Despres JP, Dagenais GR, Tunstall-Pedoe H, Woodward M, Ben-Shlomo Y, Davey SG, Palmieri V, Yeh JL, Rudnicka A, Ridker P, Rodeghiero F, Tosetto A, Shepherd J, Ford I, Robertson M, Brunner E, Shipley M, Feskens EJ, Kromhout D, Dickinson A, Ireland B, Juzwishin K, Kaptoge S, Lewington S, Memon A, Sarwar N, Walker M, Wheeler J, White I, Wood A (2005) Plasma fibrinogen level and the risk of major cardiovascular diseases and

nonvascular mortality: an individual participant meta-analysis. *JAMA,* **294:** 1799-1809.

Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis,* **8:** 1-21.

Dawn TM, Barrett JH (2005) Genetic linkage studies. *Lancet,* **366:** 1036-1044.

de Leeuw FE, Richard F, de Groot JC, van Duijn CM, Hofman A, Van GJ, Breteler MM (2004) Interaction between hypertension, apoE, and cerebral white matter lesions. *Stroke,* **35:** 1057-1060.

de Maat MP, Bladbjerg EM, Drivsholm T, Borch-Johnsen K, Moller L, Jespersen J (2003) Inflammation, thrombosis and atherosclerosis: results of the Glostrup study. *J.Thromb.Haemost.,* **1:** 950-957.

Debette S, Lambert J-C, Gariepy J, Fievet N, Tzourio C, Dartigues J-F, Ritchie K, Dupuy A-M, Alperovitch A, Ducimetiere P, Amouyel P, Zureik M (2006) New insight into the association of apolipoprotein E genetic variants with carotid plaques and intima-media thickness. *Stroke,* **37:** 2917-2923.

Decarli C, Reed T, Miller BL, Wolf PA, Swan GE, Carmelli D (1999) Impact of apolipoprotein E epsilon4 and vascular disease on brain morphology in men from the NHLBI twin study. *Stroke,* **30:** 1548-1553.

Deeks JJ, Altman DG, Bradburn MJ (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In:*Statistical reviews in health car: meta-analysis in context* pp. 285-312. BMJ Publishing Group, London.

Demuth K, Moatti N, Hanon O, Benoit MO, Safar M, Girerd X (1998) Opposite effects of plasma homocysteine and the methylenetetrahydrofolate reductase C677T mutation on carotid artery geometry in asymptomatic adults. *Arterioscler.Thromb.Vasc.Biol.,* **18:** 1838-1843.

Department of Health (2005) *Reducing Brain Damage: Faster Access to Better Stroke Care*. The Stationery Office, London.

Dessi M, Gnasso A, Motti C, Pujia A, Irace C, Casciani S, Staffa F, Federici G, Cortese C (1999) Influence of the human paraoxonase polymorphism (PON1 192) on the carotid-wall thickening in a healthy population. *Coron.Artery Dis.,* **10:** 595-599.

Dessi-Fulgheri P, Catalini R, Sarzani R, Sturbini S, Siragusa N, Guazzarotti F, Offidani M, Tamburrini P, Zingaretti O, Rappelli A (1995) Angiotensin converting enzyme gene polymorphism and carotid atherosclerosis in a low-risk population. *J.Hypertens.,* **13:** 1593-1596.

DeStefano AL, Atwood LD, Massaro JM, Heard-Costa N, Beiser A, Au R, Wolf PA, Decarli C (2006) Genome-wide scan for white matter hyperintensity: the Framingham Heart Study. *Stroke,* **37:** 77-81.

Diamantopoulos EJ, Andreadis E, Kakou M, Vlachonikolis I, Vassilopoulos C, Giannakopoulos N, Tarassi K, Papasteriades C, Nicolaides A, Raptis S (2002) Atherosclerosis of carotid arteries and the ACE insertion/deletion polymorphism in subjects with diabetes mellitus type 2. *Int.Angiol.,* **21:** 63-69.

Dichgans M (2007) Genetics of ischaemic stroke. *Lancet Neurol.,* **6:** 149-161.

Dichgans M, Markus HS (2005) Genetic association studies in stroke: methodological issues and proposed standard criteria. *Stroke,* **36:** 2027-2031.

Dickersin K (2002) Systematic reviews in epidemiology: why are we so far behind? *Int.J.Epidemiol.,* **31:** 6-12.

Dijk JM, van der GY, Bots ML, Grobbee DE, Algra A (2006) Carotid intima-media thickness and the risk of new vascular events in patients with manifest atherosclerotic disease: the SMART study. *Eur.Heart J.,* **27:** 1971-1978.

Doody RS, Azher SN, Haykal HA, Dunn JK, Liao T, Schneider L (2000) Does APO epsilon4 correlate with MRI changes in Alzheimer's disease? *J.Neurol.Neurosurg.Psychiatry,* **69:** 668-671.

Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc.Natl.Acad.Sci.U.S.A,* **97:** 10483-10488.

Duggirala R, Gonzalez VC, O'Leary DH, Stern MP, Blangero J (1996) Genetic basis of variation in carotid artery wall thickness. *Stroke,* **27:** 833-837.

Durga J, Bots ML, Schouten EG, Kok FJ, Verhoef P (2005) Low concentrations of folate, not hyperhomocysteinemia, are associated with carotid intima-media thickness. *Atherosclerosis,* **179:** 285-292.

Eichner JE, Dunn ST, Perveen G, Thompson DM, Stewart KE, Stroehla BC (2002) Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am.J.Epidemiol.,* **155:** 487-495.

Elosua R, Ordovas JM, Cupples LA, Fox CS, Polak JF, Wolf PA, D'Agostino Sr RA, O'Donnell CJ (2004) Association of APOE genotype with carotid atherosclerosis in men and women: The Framingham Heart Study. *J.Lipid Res.,* **45:** 1868-1875.

Ezzat VA, Duncan ER, Wheatcroft SB, Kearney MT (2008) The role of IGF-I and its binding proteins in the development of type 2 diabetes and cardiovascular disease. *Diabetes Obes.Metab,* **10:** 198-211.

Fatah K, Hamsten A, Blomback B, Blomback M (1992) Fibrin gel network characteristics and coronary heart disease: relations to plasma fibrinogen concentration, acute phase protein, serum lipoproteins and coronary atherosclerosis. *Thromb.Haemost.,* **68:** 130-135.

Fazekas F, Barkhof F, Wahlund LO, Pantoni L, Erkinjuntti T, Scheltens P, Schmidt R (2002) CT and MRI rating of white matter lesions. *Cerebrovasc.Dis.,* **13 Suppl 2:** 31-36.

Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA (1987) MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *Am.J.Roentgenol.,* **149:** 351-356.

Fernandez MC, Aranda JL, Martin MA, Arenas J, Nunez V, Gomez De Lamara CA (2004) Apolipoprotein E polymorphism and carotid atherosclerosis in patients with coronary disease. *Int.J.Cardiol.,* **94:** 209-212.

Fernandez MC, Yebra M, Aranda JL, Gomez P, Martinez J, Nunez V, Gomez De La CA (2007) Effect of folic acid treatment on carotid intima-media thickness of patients with coronary disease. *Int.J.Cardiol.,* **118:** 345-349.

Ferrieres J, Elias A, Ruidavets J-B, Cantet C, Bongard V, Fauvel J, Boccalon H (1999) Carotid intima-media thickness and coronary heart disease risk factors in a low-risk population. *J.Hypertens.,* **17:** 743-748.

Flossmann E, Schulz UG, Rothwell PM (2004) Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke,* **35:** 212-227.

Fornage M, Mosley TH, Jack CR, de Andrade M, Kardia SL, Boerwinkle E, Turner ST (2007) Family-based association study of matrix metalloproteinase-3 and -9 haplotypes with susceptibility to ischemic white matter injury. *Hum.Genet.,* **120:** 671-680.

Fortunato G, Rubba P, Panico S, Trono D, Tinto N, Mazzaccara C, De MM, Iannuzzi A, Vitale DF, Salvatore F, Sacchetti L (2003) A paraoxonase gene polymorphism, PON 1 (55), as an independent risk factor for increased carotid intima-media thickness in middle-aged women. *Atherosclerosis,* **167:** 141-148.

Fox CS, Cupples LA, Chazaro I, Polak JF, Wolf PA, D'Agostino RB, Ordovas JM, O'Donnell CJ (2004a) Genomewide linkage analysis for internal carotid artery intimal medial thickness: evidence for linkage to chromosome 12. *Am.J.Hum.Genet.,* **74:** 253-261.

Fox CS, Larson MG, Corey D, Feng D, Lindpaintner K, Polak JF, Wolf PA, D'Agostino RB, Tofler GH, O'Donnell CJ (2004b) Absence of association between polymorphisms in the hemostatic factor pathway genes and carotid intimal medial thickness: the Framingham Heart Study. *Stroke,* **35:** e65-e67.

Fox CS, Polak JF, Chazaro I, Cupples A, Wolf PA, D'Agostino RA, O'Donnell CJ (2003) Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study. *Stroke,* **34:** 397-401.

Frost D, Pfohl M, Clemens P, Haring H-U, Beischer W (1998) Evaluation of the insertion/deletion ACE gene polymorphism as a risk factor for carotid artery intima-media thickening and hypertension in young type 1 diabetic patients. *Diabetes Care,* **21:** 836-840.

Garg UC, Arnett DK, Evans G, Eckfeldt JH (1998) No association between Factor V. Leiden mutation and coronary heart disease or carotid intima media thickness: The NHLBI family heart study. *Thromb.Res.,* **89:** 289-293.

Garin MC, James RW, Dussoix P, Blanche H, Passa P, Froguel P, Ruiz J (1997) Paraoxonase polymorphism Met-Leu54 is associated with modified serum concentrations of the enzyme. A possible link between the paraoxonase gene and increased risk of cardiovascular disease in diabetes. *J.Clin.Invest,* **99:** 62-66.

Gauderman WJ and Morrison JM (2006)
QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.

Georges JL, Loukaci V, Poirier O, Evans A, Luc G, Arveiler D, Ruidavets JB, Cambien F, Tiret L (2001) Interleukin-6 gene polymorphisms and susceptibility to myocardial infarction: the ECTIM study. Etude Cas-Temoin de l'Infarctus du Myocarde. *J.Mol.Med.,* **79:** 300-305.

Girerd X, Hanon O, Mourad JJ, Boutouyrie P, Laurent S, Jeunemaitre X (1998) Lack of association between renin-angiotensin system, gene polymorphisms, and wall thickness of the radial and carotid arteries. *Hypertension,* **32:** 579-583.

Glass GV (1976) Primary, secondary and meta-analysis of research. *Educational Research,* **5:** 3-8.

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science,* **298:** 2345-2349.

Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, Culebras A, Degraba TJ, Gorelick PB, Guyton JR, Hart RG, Howard G, Kelly-Hayes M, Nixon JV, Sacco RL (2006) Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group: the American Academy of Neurology affirms the value of this guideline. *Stroke,* **37:** 1583-1633.

Gormley K, Bevan S, Markus HS (2007) Polymorphisms in genes of the renin-angiotensin system and cerebral small vessel disease. *Cerebrovasc.Dis.,* **23:** 148-155.

Green SA, Turki J, Innis M, Liggett SB (1994) Amino-terminal polymorphisms of the human beta 2-adrenergic receptor impart distinct agonist-promoted regulatory properties. *Biochemistry (Mosc).,* **33:** 9414-9419.

Greenland P, Abrams J, Aurigemma GP, Bond MG, Clark LT, Criqui MH, Crouse JR, III, Friedman L, Fuster V, Herrington DM, Kuller LH, Ridker PM, Roberts WC, Stanford W, Stone N, Swan HJ, Taubert KA, Wexler L (2000) Prevention Conference V: Beyond secondary prevention: identifying the high-risk patient for primary prevention: noninvasive tests of atherosclerotic burden: Writing Group III. *Circulation,* **101:** E16-E22.

Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, Einarsson G, Gudjonsdottir HM, Valdimarsson EM, Einarsson OB, Thorgeirsson G, Hadzic R, Jonsdottir S, Reynisdottir ST, Bjarnadottir SM, Gudmundsdottir T, Gudlaugsdottir GJ, Gill R, Lindpaintner K, Sainz J, Hannesson HH, Sigurdsson GT, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. *Am.J.Hum.Genet.,* **70:** 593-603.

Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat.Genet.,* **35:** 131-138.

Guimaraes S, Moura D (2001) Vascular adrenoceptors: an update. *Pharmacol.Rev.,* **53:** 319-356.

Guo SW (2001) Does higher concordance in monozygotic twins than in dizygotic twins suggest a genetic component? *Hum.Hered.,* **51:** 121-132.

Gurol ME, Irizarry MC, Smith EE, Raju S, Diaz-Arrastia R, Bottiglieri T, Rosand J, Growdon JH, Greenberg SM (2006) Plasma beta-amyloid and white matter lesions in AD, MCI, and cerebral amyloid angiopathy. *Neurology,* **66:** 23-29.

Guz G, Ozdemir FN, Sezer S, Isiklar I, Arat Z, Turan M, Haberal M (2000) Effect of apolipoprotein E polymorphism serum lipid, lipoproteins, and atherosclerosis in hemodialysis patients. *Am.J.Kidney Dis.,* **36:** 826-836.

Hadjigeorgiou GM, Malizos K, Dardiotis E, Aggelakis K, Dardioti M, Zibis A, Dimitroulias A, Scarmeas N, Tsezou A, Karantanas A (2007) Paraoxonase 1 gene polymorphisms in patients with osteonecrosis of the femoral head with and without cerebral white matter lesions. *J.Orthop.Res.,* **25:** 1087-1093.

Hallman DM, Boerwinkle E, Saha N, Sandholzer C, Menzel HJ, Csazar A, Utermann G (1991) The apolipoprotein E polymorphism: a comparison of allele frequencies and effects in nine populations. *Am.J.Hum.Genet.,* **49:** 338-349.

Han J-H, Huang Y-Y, Luo Y-M, Ding D, Hong Z (2005) Association between intercellular adhesion molecule-1 K469E polymorphism and white matter lesions in asymptomatic elderly people. *Chin.J.Neurol.,* **. 38:** 247-250.

Hankey GJ (2006) Is plasma homocysteine a modifiable risk factor for stroke? *Nat.Clin.Pract.Neurol.,* **2:** 26-33.

Hanon O, Girerd X, Luong V, Jeunemaitre X, Laurent S, Safar ME (2000) Association between the apolipoprotein E polymorphism and arterial wall thickness in asymptomatic adults. *J.Hypertens.,* **18:** 431-436.

Haraki T, Takegoshi T, Kitoh C, Wakasugi T, Saga T, Hirai J-I, Aoyama T, Inazu A, Mabuchi H (2002) Carotid artery intima-media thickness and brachial artery flow-mediated vasodilation in asymptomatic Japanese male subjects amongst apolipoprotein E phenotypes. *J.Intern.Med.,* **252:** 114-120.

Hassan A, Gormley K, O'Sullivan M, Knight J, Sham P, Vallance P, Bamford J, Markus H (2004a) Endothelial nitric oxide gene haplotypes and risk of cerebral small-vessel disease. *Stroke,* **35:** 654-659.

Hassan A, Hunt BJ, O'Sullivan M, Bell R, D'Souza R, Jeffery S, Bamford JM, Markus HS (2004b) Homocysteine is a risk factor for cerebral small vessel disease, acting via endothelial dysfunction. *Brain,* **127:** 212-219.

Hassan A, Lansbury A, Catto AJ, Guthrie A, Spencer J, Craven C, Grant PJ, Bamford JM (2002) Angiotensin converting enzyme insertion/deletion genotype is associated with leukoaraiosis in lacunar syndromes. *J.Neurol.Neurosurg.Psychiatry,* **72:** 343-346.

Hassan A, Markus HS (2000) Genetics and ischaemic stroke. *Brain,* **123 (Pt 9):** 1784-1812.

Hatano S (1976) Experience from a multicentre stroke register: a preliminary report. *Bull.World Health Organ,* **54:** 541-553.

Heckbert SR, Hindorff LA, Edwards KL, Psaty BM, Lumley T, Siscovick DS, Tang Z, Durda JP, Kronmal RA, Tracy RP (2003) Beta2-adrenergic receptor polymorphisms and risk of incident cardiovascular events in the elderly. *Circulation,* **107:** 2021-2024.

Heinrich PC, Behrmann I, Haan S, Hermanns HM, Muller-Newen G, Schaper F (2003) Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem.J.,* **374:** 1-20.

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat.Genet.,* **36:** 233-239.

Henskens LH, Kroon AA, van Boxtel MP, Hofman PA, de Leeuw PW (2005) Associations of the angiotensin II type 1 receptor A1166C and the endothelial NO synthase G894T gene polymorphisms with silent subcortical white matter lesions in essential hypertension. *Stroke,* **36:** 1869-1873.

Herrick S, Blanc-Brude O, Gray A, Laurent G (1999) Fibrinogen. *Int.J.Biochem.Cell Biol.,* **31:** 741-746.

Higgins J (Nov, 2006).Personal Communication.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ,* **327:** 557-560.

Hindorff LA, Heckbert SR, Psaty BM, Lumley T, Siscovick DS, Herrington DM, Edwards KL, Tracy RP (2005) beta2-adrenergic receptor polymorphisms and determinants of cardiovascular risk: The Cardiovascular Health Study. *Am.J.Hypertens.,* **18:** 392-397.

Hirono N, Yasuda M, Tanimukai S, Kitagaki H, Mori E (2000) Effect of the apolipoprotein E epsilon4 allele on white matter hyperintensities in dementia. *Stroke,* **31:** 1263-1268.

Hogh P, Garde E, Mortensen EL, Jorgensen OS, Krabbe K, Waldemar G (2007) The apolipoprotein E epsilon4-allele and antihypertensive treatment are associated with increased risk of cerebral MRI white matter hyperintensities. *Acta Neurol.Scand.,* **115:** 248-253.

Horejsi B, Spacil J, Ceska R, Vrablik M, Haas T, Horinek A (2000) The independent correlation of the impact of lipoprotein(a) levels and apolipoprotein E polymorphism on carotid artery intima thickness. *Int.Angiol.,* **19:** 331-336.

Hu Y, Tian H, Liu R (2003) Gln-Arg192 polymorphism of paraoxonase 1 is associated with carotid intima-media thickness in patients of type 2 diabetes mellitus of Chinese. *Diabetes Research & Clinical Practice,* **61:** 21-27.

Huang X-H, Loimaala A, Nenonen A, Mercuri M, Vuori I, Pasanen M, Oja P, Bond G, Koivula T, Hiltunen TP, Nikkari T, Lehtimaki T (1999) Relationship of angiotensin-converting enzyme gene polymorphism to carotid wall thickness in middle-aged men. *J.Mol.Med.,* **77:** 853-858.

Humbert R, Adler DA, Disteche CM, Hassett C, Omiecinski CJ, Furlong CE (1993) The molecular basis of the human serum paraoxonase activity polymorphism. *Nat.Genet.,* **3:** 73-76.

Humphries SE, Luong LA, Ogg MS, Hawe E, Miller GJ (2001) The interleukin-6 -174 G/C promoter polymorphism is associated with risk of coronary heart disease and systolic blood pressure in healthy men. *Eur.Heart J.,* **22:** 2243-2252.

Humphries SE, Morgan L (2004) Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. *Lancet Neurol.,* **3:** 227-235.

Hung J, McQuillan BM, Nidorf M, Thompson PL, Beilby JP (1999) Angiotensin-converting enzyme gene polymorphism and carotid wall thickening in a community population. *Arterioscler.Thromb.Vasc.Biol.,* **19:** 1969-1974.

Ilveskoski E, Loimaala A, Mercuri MF, Lehtimaki T, Pasanen M, Nenonen A, Oja P, Bond MG, Koivula T, Karhunen PJ, Vuori I (2000) Apolipoprotein E polymorphism and carotid artery intima-media thickness in a random sample of middle-aged men. *Atherosclerosis,* **153:** 147-153.

Inamoto N, Katsuya T, Kokubo Y, Mannami T, Asai T, Baba S, Ogata J, Tomoike H, Ogihara T (2003) Association of methylenetetrahydrofolate reductase gene polymorphism with carotid atherosclerosis depending on smoking status in a Japanese general population. *Stroke,* **34:** 1628-1633.

Islam M, Lehtimaki T, Juonala M, Kahonen M, Hutri-Kahonen N, Kainulainen K, Miettinen H, Taittonen L, Kontula K, Viikari JSA, Raitakari OT (2006) Polymorphism of the angiotensin-converting enzyme (ACE) and angiotesinogen (AGT) genes and their associations with blood pressure and carotid artery intima media thickness among healthy Finnish young adults-the Cardiovascular Risk in Young Finns Study. *Atherosclerosis,* **188:** 316-322.

Jackson C, Hutchison A, Dennis MS, Wardlaw J, Lewis S, and Sudlow CLM (in press). Differences between ischaemic stroke subtypes in recurrent stroke and myocardial infarction support a distinct lacunar ischaemic stroke arteriopathy: a prospective, hospital-based study. *Lancet Neurol.*

Jackson C, Crossland L, Dennis M, Wardlaw J, Sudlow C (2008) Assessing the impact of the requirement for explicit consent in a hospital-based stroke study. *QJM,* **101:** 281-289.

Jackson C, Sudlow C (2005) Are lacunar strokes really different? A systematic review of differences in risk factor profiles between lacunar and nonlacunar infarcts. *Stroke,* **36:** 891-901.

Jackson CA, Hutchison A, Dennis MS, Wardlaw JM, Lindgren A, Norrving B, Anderson CS, Hankey GJ, Jamrozik K, Broadhurst RJ, Appelros P, and Sudlow CLM (in press). Risk factors for lacunar versus non-lacunar ischemic stroke: A collaborative individual patient data analysis. *Annals of Neurology.*

Jeng J-R (1999) Left ventricular mass, carotid wall thickness, and angiotensinogen gene polymorphism in patients with hypertension. *Am.J.Hypertens.,* **12:** 443-450.

Jeng J-R (2000) Carotid thickening, cardiac hypertrophy, and angiotensin converting enzyme gene polymorphism in patients with hypertension. *Am.J.Hypertens.,* **13:** 111-119.

Jerrard-Dunne P, Cloud G, Hassan A, Markus HS (2003a) Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke,* **34:** 1364-1369.

Jerrard-Dunne P, Sitzer M, Risley P, Steckel DA, Buehler A, Von KS, Markus HS (2003b) Interleukin-6 promoter polymorphism modulates the effects of heavy alcohol consumption on early carotid artery atherosclerosis: The carotid atherosclerosis progression study (CAPS). *Stroke,* **34:** 402-407.

Johnson JA, Terra SG (2002) Beta-adrenergic receptor polymorphisms: cardiovascular disease associations and pharmacogenetics. *Pharm.Res.,* **19:** 1779-1787.

Junyent M, Cofan M, Nunez I, Gilabert R, Zambon D, Ros E (2006) Influence of HDL cholesterol on preclinical carotid atherosclerosis in familial hypercholesterolemia. *Arterioscler.Thromb.Vasc.Biol.,* **26:** 1107-1113.

Juo SH, Lin HF, Rundek T, Sabala EA, Boden-Albala B, Park N, Lan MY, Sacco RL (2004) Genetic and environmental contributions to carotid intima-media thickness and obesity phenotypes in the Northern Manhattan Family Study. *Stroke,* **35:** 2243-2247.

Juo SH, Wyszynski DF, Beaty TH, Huang HY, Bailey-Wilson JE (1999) Mild association between the A/G polymorphism in the promoter of the apolipoprotein A-I gene and apolipoprotein A-I levels: a meta-analysis. *Am.J.Med.Genet.,* **82:** 235-241.

Kahraman S, Kiykim AA, Altun B, Genctoy G, Arici M, Gulsun M, Erdem Y, Yasavul U, Turgan C, Caglar S (2004) Apolipoprotein E gene polymorphism in renal transplant recipients: Effects on lipid metabolism, atherosclerosis and allograft function. *Clin.Transplant.*, **18:** 288-294.

Kardys I, Uitterlinden AG, Hofman A, Witteman JC, de Maat MP, Kardys I, Uitterlinden AG, Hofman A, Witteman JCM, de Maat MPM (2007) Fibrinogen gene haplotypes in relation to risk of coronary events and coronary and extracoronary atherosclerosis: the Rotterdam Study. *Thrombosis & Haemostasis*, **97:** 288-295.

Karvonen J, Kauma H, Kervinen K, Rantala M, Ikaheimo M, Paivansalo M, Savolainen MJ, Kesaniemi YA (2002a) Endothelial nitric oxide synthase gene Glu298Asp polymorphism and blood pressure, left ventricular mass and carotid artery atherosclerosis in a population-based cohort. *J.Intern.Med.*, **251:** 102-110.

Karvonen J, Kauma H, Kervinen K, Ukkola O, Rantala M, Paivansalo M, Savolainen MJ, Kesaniemi YA (2002b) Apolipoprotein E polymorphism affects carotid artery atherosclerosis in smoking hypertensive men. *J.Hypertens.*, **20:** 2371-2378.

Karvonen J, Kauma H, Paivansalo M, Kesaniemi YA (2004) Paraoxonase-1 gene Leu-Met55 and Gln-Arg192 polymorphisms are not associated with carotid artery atherosclerosis in a population-based cohort. *Eur.J.Cardiovasc.Prev.Rehab.*, **11:** 511-512.

Kauma H, Paivansalo M, Savolainen MJ, Rantala AO, Kiema T-R, Lilja M, Reunanen A, Antero KY (1996) Association between angiotensin converting enzyme gene polymorphism and carotid atherosclerosis. *J.Hypertens.*, **14:** 1183-1187.

Kawamoto R, Kohara K, Tabara Y, Miki T (2002) An interaction between systolic blood pressure and angiotensin-converting enzyme gene polymorphism on carotid atherosclerosis. *Hypertens.Res.*, **25:** 875-880.

Kawamoto R, Kohara K, Tabara Y, Miki T, Doi T, Tokunaga H, Konishi I (2001) An association of 5,10-methylenetetrahydrofolate reductase (MTHFR) gene polymorphism and common carotid atherosclerosis. *J.Hum.Genet.,* **46:** 506-510.

Keavney B, McKenzie C, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R (2000) Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. *Lancet,* **355:** 434-442.

Kelberman D, Hawe E, Luong LA, Mohamed-Ali V, Lundman P, Tornvall P, Aillaud MF, Juhan-Vague I, Yudkin JS, Margaglione M, di MG, Tremoli E, Humphries SE (2004) Effect of Interleukin-6 promoter polymorphisms in survivors of myocardial infarction and matched controls in the North and South of Europe. The HIFMECH Study. *Thromb.Haemost.,* **92:** 1122-1128.

Kelemen LE, Anand SS, Hegele RA, Stampfer MJ, Rosner B, Willett WC, Montague PA, Lonn E, Vuksan V, Teo KK, Devanesen S, Yusuf S (2004) Associations of plasma homocysteine and the methylenetetrahydrofolate reductase C677T polymorphism with carotid intima media thickness among South Asian, Chinese and European Canadians. *Atherosclerosis,* **176:** 361-370.

Khan U, Bevan S, Markus HS (2007) NADPH oxidase polymorphisms in cerebral small vessel disease. *Cerebrovasc.Dis.,* **24:** 135-138.

Kitsios G, Zintzaras E (2007) Genetic variation associated with ischemic heart failure: a HuGE review and meta-analysis. *Am.J.Epidemiol.,* **166:** 619-633.

Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, Schouten EG (2002) MTHFR 677C-->T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA,* **288:** 2023-2031.

Kogawa K, Nishizawa Y, Hosoi M, Kawagishi T, Maekawa K, Shoji T, Okuno Y, Morii H (1997) Effect of polymorphism of apolipoprotein E and angiotensin-converting enzyme genes on arterial wall thickness. *Diabetes,* **46:** 682-687.

Kohara K, Fujisawa M, Ando F, Tabara Y, Niino N, Miki T, Shimokata H, NILS-LSA Study (2003) 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,* **34:** 1130-1135.

Kuller LH, Shemanski L, Manolio T, Haan M, Fried L, Bryan N, Burke GL, Tracy R, Bhadelia R (1998) Relationship between ApoE, MRI findings, and cognitive function in the Cardiovascular Health Study. *Stroke,* **29:** 388-398.

Kuo HK, Yen CJ, Chang CH, Kuo CK, Chen JH, Sorond F (2005) Relation of C-reactive protein to stroke, cognitive disorders, and depression in the general population: systematic review and meta-analysis. *Lancet Neurol.,* **4:** 371-380.

Lange LA, Bowden DW, Langefeld CD, Wagenknecht LE, Carr JJ, Rich SS, Riley WA, Freedman BI (2002) Heritability of carotid artery intima-medial thickness in type 2 diabetes. *Stroke,* **33:** 1876-1881.

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 (2006) Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA,* **296:** 2703-2711.

Lekakis JP, Ikonomidis I, Tsibida M, Protogerou A, Papada A, Papapanagiotou A, Revela I, Papamichael CM, Kalofoutis AT, Kremastinos DT (2006) Genetic variations of the endothelial nitric oxide synthase gene are related to increased levels of C-reactive protein and macrophage-colony stimulating-factor in patients with coronary artery disease. *Thrombosis & Haemostasis,* **96:** 520-528.

Lembo G, De Luca N, Battagli C, Iovino G, Aretini A, Musicco M, Frati G, Pompeo F, Vecchione C, Trimarco B (2001) A common variant of endothelial nitric oxide synthase (Glu298Asp) is an independent risk factor for carotid atherosclerosis. *Stroke,* **32:** 735-740.

Lewis SJ, Ebrahim S, Davey SG (2005) 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,* **331:** 1053.

Leys D, Englund E, Del ST, Inzitari D, Fazekas F, Bornstein N, Erkinjuntti T, Bowler JV, Pantoni L, Parnetti L, De RJ, Ferro J, Bogousslavsky J (1999) White matter changes in stroke patients. Relationship with stroke subtype and outcome. *Eur.Neurol.,* **42:** 67-75.

Li M, Li C, Guan W (2008) Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet,* **16:** 635-643.

Li S-J, Sun N-L, Zhou S-M (2004) Carotid remodeling of hypertensive subjects and polymorphism of the angiotensin-converting enzyme gene. *Chin.Med.J.(Engl).,* **117:** 49-53.

Li X, Du Y, Du Y, Huang X (2003) Association of apolipoprotein E gene polymorphism with essential hypertension and its complications. *Clinical & Experimental Medicine,* **2:** 175-179.

Libby P, Ridker PM, Maseri A (2002) Inflammation and atherosclerosis. *Circulation,* **105:** 1135-1143.

Lim PS, Hung W-R, Wei Y-H (2001) Polymorphism in methylenetetrahydrofolate reductase gene: Its impact on plasma homocysteine levels and carotid atherosclerosis in esrd patients receiving hemodialysis. *Nephron,* **87:** 249-256.

Linnebank M, Moskau S, Farmand S, Fliessbach K, Kolsch H, Bos M, Grothe C, Becker D, Harbrecht U, Pohl C, Wullner U, Klockgether T (2006) Homocysteine and carotid intima-media thickness in a German population: Lack of clinical relevance. *Stroke,* **37:** 2840-2842.

Liu CS, Chen CH, Chiang HC, Kuo CL, Huang CS, Cheng WL, Wei YH, Chen HW (2007) B-group vitamins, MTHFR C677T polymorphism and carotid intima-media thickness in clinically healthy subjects. *Eur.J.Clin.Nutr.,* **61:** 996-1003.

Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M (2007) Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis. *Circulation,* **115:** 459-467.

Lunetta KL, Erlich PM, Cuenco KT, Cupples LA, Green RC, Farrer LA, Decarli C, for the MIRAGE Study Group. (2007) Heritability of magnetic resonance imaging (MRI) traits in Alzheimer disease cases and their siblings in the MIRAGE study. *Alz.Dis.Assoc.Dis,* **21:** 85-91.

Mackay J, Mensah G (2004) *Atlas of Heart Disease and Stroke*. World Health Organisation, Geneva.

Mackness B, Durrington PN, Mackness MI (1998) Human serum paraoxonase. *Gen.Pharmacol.,* **31:** 329-336.

Mackness M, Mackness B (2004) Paraoxonase 1 and atherosclerosis: is the gene or the protein more important? *Free Radic.Biol.Med.,* **37:** 1317-1323.

Mahley RW, Rall SC, Jr. (2000) Apolipoprotein E: far more than a lipid transport protein. *Annu.Rev.Genomics Hum.Genet.,* **1:** 507-537.

Maia LF, Vasconcelos C, Seixas S, Magalhaes R, Correia M (2006) Lobar brain hemorrhages and white matter changes: Clinical, radiological and laboratorial profiles. *Cerebrovasc.Dis.,* **22:** 155-161.

Majumder PP, Ghosh S (2005) Mapping quantitative trait loci in humans: achievements and limitations. *J.Clin.Invest,* **115:** 1419-1424.

Mannami T, Katsuya T, Baba S, Inamoto N, Ishikawa K, Higaki J, Ogihara T, Ogata J (2001) Low potentiality of angiotensin-converting enzyme gene insertion/deletion polymorphism as a useful predictive marker for carotid atherogenesis in a large general population of a Japanese City: The suita study. *Stroke,* **32:** 1250-1256.

Mannila MN, Eriksson P, Ericsson CG, Hamsten A, Silveira A (2006) Epistatic and pleiotropic effects of polymorphisms in the fibrinogen and coagulation factor XIII genes on plasma fibrinogen concentration, fibrin gel structure and risk of myocardial infarction. *Thromb.Haemost.,* **95:** 420-427.

Mannila MN, Eriksson P, Lundman P, Samnegard A, Boquist S, Ericsson CG, Tornvall P, Hamsten A, Silveira A (2005) Contribution of haplotypes across the fibrinogen gene cluster to variation in risk of myocardial infarction. *Thromb.Haemost.,* **93:** 570-577.

Manolio TA, Boerwinkle E, O'Donnell CJ, Wilson AF (2004) Genetics of ultrasonographic carotid atherosclerosis. *Arterioscler.Thromb.Vasc.Biol.,* **24:** 1567-1577.

Markus H, Kapozsta Z, Ditrich R, Wolfe C, Ali N, Powell J, Mendell M, Cullinane M (2001) Increased common carotid intima-media thickness in UK African Caribbeans and its relation to chronic inflammation and vascular candidate gene polymorphisms. *Stroke,* **32:** 2465-2471.

Markus HS (2008) Genes, endothelial function and cerebral small vessel disease in man. *Exp.Physiol,* **93:** 121-127.

Markus HS, Barley J, Lunt R, Bland JM, Jeffery S, Carter ND, Brown MM (1995) Angiotensin-converting enzyme gene deletion polymorphism: A new risk factor for lacunar stroke but not carotid atheroma. *Stroke,* **26:** 1329-1333.

Markus HS, Labrum R, Bevan S, Reindl M, Egger G, Wiedermann CJ, Xu Q, Kiechl S, Willeit J (2006) Genetic and acquired inflammatory conditions are synergistically associated with early carotid atherosclerosis. *Stroke,* **37:** 2253-2259.

Matarin M, Brown WM, Scholz S, Simon-Sanchez J, Fung HC, Hernandez D, Gibbs JR, De Vrieze FW, Crews C, Britton A, Langefeld CD, Brott TG, Brown RD, Jr., Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF (2007) A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol.,* **6:** 414-420.

Matsuoka Y, Li X, Bennett V (2000) Adducin: structure, function and regulation. *Cell Mol.Life Sci.,* **57:** 884-895.

Mayosi BM, Avery PJ, Baker M, Gaukrodger N, Imrie H, Green FR, Farrall M, Watkins H, Keavney B (2005) Genotype at the -174G/C polymorphism of the interleukin-6 gene is associated with common carotid artery intimal-medial thickness: Family study and meta-analysis. *Stroke,* **36:** 2215-2219.

Mazza A, Motti C, Nulli A, Pastore A, Andreotti F, Ammaturo V, Bianco P, Santoro E, Federici G, Cortese C (1999) Serum homocysteine, MTHFR gene polymorphism, and carotid intimal-medial thickness in NIDDM subjects. *J.Thromb.Thrombolysis,* **8:** 207-212.

McDonald S, Maguire G, Duarte N, Wang XL, Hoy W (2005) Homocysteine, renal disease and cardiovascular disease in a remote Australian Aboriginal community. *Intern.Med.J.,* **35:** 289-294.

McQuillan BM, Beilby JP, Nidorf M, Thompson PL, Hung J (1999) 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,* **99:** 2361-2363.

Mertens A, Holvoet P (2001) Oxidized LDL and HDL: antagonists in atherothrombosis. *FASEB J.,* **15:** 2073-2084.

Meyer JS, Kawamura J, Terayama Y (1992) White matter lesions in the elderly. *J.Neurol.Sci.,* **110:** 1-7.

Minelli C, Thompson JR, Abrams KR, Lambert PC (2005) Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Stat.Med.,* **24:** 3845-3861.

Moskau S, Golla A, Grothe C, Boes M, Pohl C, Klockgether T (2005) Heritability of carotid artery atherosclerotic lesions: an ultrasound study in 154 families. *Stroke,* **36:** 5-8.

Nakase T, Mizuno T, Harada S, Yamada K, Nishimura T, Ozasa K, Watanabe Y, Nagata K (2007) Angiotensinogen gene polymorphism as a risk factor for ischemic stroke. *Journal of Clinical Neuroscience,* **14:** 943-947.

Napoli C, Ignarro LJ (2007) Polymorphisms in endothelial nitric oxide synthase and carotid artery atherosclerosis. *J.Clin.Pathol.,* **60:** 341-344.

Nebes RD, Vora IJ, Meltzer CC, Fukui MB, Williams RL, Kamboh MI, Saxton J, Houck PR, DeKosky ST, Reynolds CF, III (2001) Relationship of deep white matter hyperintensities and apolipoprotein E genotype to depressive symptoms in older adults without clinical depression. *Am.J.Psychiatry,* **158:** 878-884.

Nergizoglu G, Keven K, Gurses MA, Aras O, Erturk S, Duman N, Ates K, Akar H, Akar N, Karatan O, Erbay B, Ertug AE (1999) Carotid intima-media thickness and ACE-gene polymorphism in hemodialysis patients. *J.Nephrol.,* **12:** 261-265.

Nieminen TK (2006) The effects of apoA-I/C-III/A-IV, apoE and apoB polymorphisms on carotid artery intima-media thickness. *Future Cardiology,* **2:** 179-186.

O'Leary DH, Polak JF, Wolfson SK, Jr., Bond MG, Bommer W, Sheth S, Psaty BM, Sharrett AR, Manolio TA (1991) Use of sonography to evaluate carotid atherosclerosis in the elderly. The Cardiovascular Health Study. CHS Collaborative Research Group. *Stroke,* **22:** 1155-1163.

Okura Y, Brink M, Zahid AA, Anwar A, Delafontaine P (2001) Decreased expression of insulin-like growth factor-1 and apoptosis of vascular smooth muscle cells in human atherosclerotic plaque. *J.Mol.Cell Cardiol.,* **33:** 1777-1789.

Olmer M, Renucci JE, Planells R, Bouchouareb D, Purgus R (1997) Preliminary evidence for a role of apolipoprotein E alleles in identifying haemodialysis patients at high vascular risk. *Nephrol.Dial.Transplant.,* **12:** 691-693.

Pall D, Settakis G, Katona E, Zatik J, Kollar J, Limburg M, Fulesdi B (2004) Angiotensin-Conversting Enzyme Gene Polymorphism, Carotid Intima-Media Thickness, and Left Ventricular Mass Index in Adolescent Hypertension. *J.Clin.Ultrasound,* **32:** 129-135.

Pallaud C, Sass C, Zannad F, Siest G, Visvikis S (2001) APOC3, CETP, fibrinogen, and MTHFR are genetic determinants of carotid intima-media thickness in healthy men (the Stanislas Cohort). *Clin.Genet.,* **59:** 316-324.

Pan WH, Lynn KS, Chen CH, Wu YL, Lin CY, Chang HY (2006) Using endophenotypes for pathway clusters to map complex disease genes. *Genet.Epidemiol.,* **30:** 143-154.

Paradossi U, Ciofini E, Clerico A, Botto N, Biagini A, Colombo MG (2004) Endothelial function and carotid intima-media thickness in young healthy subjects among endothelial nitric oxide synthase Glu298Asp and T-786C polymorphisms. *Stroke,* **35:** 1305-1309.

Passaro A, Vanini A, Calzoni F, Alberti L, Zamboni PF, Fellin R, Solini A (2001) Plasma homocysteine, methylenetetrahydrofolate reductase mutation and carotid damage in elderly healthy women. *Atherosclerosis,* **157:** 175-180.

Pearson K (1904) Report on certain enteric fever inoculation statistics. *BMJ,* **3:** 1243-1246.

Piao L, Tanaka Y, Nomiyama T, Nakajima K, Ogihara T, Miwa S, Urabe T, Watada H, Kawamori R (2002) Combined genotypes of ACE and NADPH oxidase p22phox associated with somatic mutation of mtDNA and carotid intima-media thickness in Japanese patients with type 2 diabetes mellitus. *Curr.Ther.Res.Clin.Exp.,* **63:** 842-852.

Pit'ha J, Krajickova D, Cifkova R, Hubacek J, Petrzilkova Z, Hejl Z, Stavek P, Skibova J, Poledne R (1999) Intima-media thickness of carotid arteries in borderline hypertensives. *J.Neuroimaging,* **9:** 19-22.

Pola R, Flex A, Gaetani E, Flore R, Serricchio M, Pola P (2003) Synergistic effect of -174 G/C polymorphism of the interleukin-6 gene promoter and 469 E/K polymorphism of the intercellular adhesion molecule-1 gene in Italian patients with history of ischemic stroke. *Stroke,* **34:** 881-885.

Pollex RL, Hegele R (2006) Genetic determinants of carotid ultrasound traits. *Curr.Atheroscler.Rep.,* **8:** 206-215.

Pollock MA, Jefferson SG, Kane JW, Lomax K, MacKinnon G, Winnard CB (1992) Method comparison--a different approach. *Ann.Clin.Biochem.,* **29 (Pt 5):** 556-560.

Pontremoli R, Ravera M, Viazzi F, Nicolella C, Berruti V, Leoncini G, Giacopelli F, Bezante GP, Sacchi G, Ravazzolo R, Deferrari G (2000) Genetic polymorphism of the renin-angiotensin system and organ damage in essential hypertension. *Kidney Int.,* **57:** 561-569.

Pruissen DM, Gerritsen SA, Prinsen TJ, Dijk JM, Kappelle LJ, Algra A (2007) Carotid intima-media thickness is different in large- and small-vessel ischemic stroke: the SMART study. *Stroke,* **38:** 1371-1373.

Pujia A, Motti C, Irace C, Cortese C, Biagiotti L, Mattioli PL, Federici G, Gnasso A (1996) Deletion polymorphism in angiotensin converting enzyme gene associated with carotid wall thickening in a healthy male population. *Coron.Artery Dis.,* **7:** 7-51.

Purandare N, Oude Voshaar RC, Davidson Y, Gibbons L, Hardicre J, Byrne J, McCollum C, Jackson A, Burns A, Mann DM (2006) Deletion/insertion polymorphism of the angiotensin-converting enzyme gene and white matter hyperintensities in dementia: A pilot study. *J.Am.Geriatr.Soc.,* **54:** 1395-1400.

Ranade K, Kirchgessner TG, Iakoubova OA, Devlin JJ, DelMonte T, Vishnupad P, Hui L, Tsuchihashi Z, Sacks FM, Sabatine MS, Braunwald E, White TJ, Shaw PM, Dracopoli NC (2005) Evaluation of the paraoxonases as candidate genes for stroke: Gln192Arg polymorphism in the paraoxonase 1 gene is associated with increased risk of stroke. *Stroke,* **36:** 2346-2350.

Rantala M, Rantala TT, Savolainen MJ, Friedlander Y, Kesaniemi YA (2000) Apolipoprotein B gene polymorphisms and serum lipids: meta-analysis of the role of genetic variation in responsiveness to diet. *Am.J.Clin.Nutr.,* **71:** 713-724.

Rauramaa R, Vaisanen SB, Luong LA, Schmidt-Trucksass A, Penttila IM, Bouchard C, Toyry J, Humphries SE (2000) Stromelysin-1 and interleukin-6 gene promoter polymorphisms are determinants of asymptomatic carotid artery atherosclerosis. *Arterioscler.Thromb.Vasc.Biol.,* **20:** 2657-2662.

Ravera M, Viazzi F, Berruti V, Leoncini G, Zagami P, Bezante GP, Rosatto N, Ravazzolo R, Pontremoli R, Deferrari G (2001) 5,10-methylenetetrahydrofolate reductase polymorphism and early organ damage in primary hypertension. *Am.J.Hypertens.,* **14:** 371-376.

Reitz C, Berger K, de Maat MP, Stoll M, Friedrichs F, Kardys I, Witteman JC, Breteler MM (2007) CRP gene haplotypes, serum CRP, and cerebral small-vessel disease: the Rotterdam Scan Study and the MEMO Study. *Stroke,* **38:** 2356-2359.

Rigat B, Hubert C, henc-Gelas F, Cambien F, Corvol P, Soubrier F (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J.Clin.Invest,* **86:** 1343-1346.

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science,* **273:** 1516-1517.

Roest M, Rodenburg J, Wiegman A, Kastelein JJP, Voorbij HAM (2006) Paraoxonase genotype and carotid intima-media thickness in children with familial hypercholesterolemia. *Eur.J.Cardiovasc.Prev.Rehab.,* **13:** 464-466.

Rubattu S, Di AE, Stanzione R, Zanda B, Evangelista A, Pirisi A, De PP, Cota L, Brunetti E, Volpe M (2004) Gene polymorphisms of the renin-angiotensin-aldosterone system and the risk of ischemic stroke: a role of the A1166C/AT1 gene variant. *J.Hypertens.,* **22:** 2129-2134.

Rundek T, Elkind MS, Pittman J, Boden-Albala B, Martin S, Humphries SE, Juo S-HH, Sacco RL (2002) Carotid intima-media thickness is associated with allelic variants of stromelysin-1, interleukin-6, and hepatic lipase genes: The Northern Manhattan Prospective Cohort Study. *Stroke,* **33:** 1420-1423.

Sacco RL, Benjamin EJ, Broderick JP, Dyken M, Easton JD, Feinberg WM, Goldstein LB, Gorelick PB, Howard G, Kittner SJ, Manolio TA, Whisnant JP, Wolf PA (1997) American Heart Association Prevention Conference. IV. Prevention and Rehabilitation of Stroke. Risk factors. *Stroke,* **28:** 1507-1517.

Sakai T, Matsuura B, Onji M (1998) Serum paraoxonase activity and genotype distribution in Japanese patients with diabetes mellitus. *Intern.Med.,* **37:** 581-584.

Salanti G, Sanderson S, Higgins JP (2005) Obstacles and opportunities in meta-analysis of genetic association studies. *Genet.Med.,* **7:** 13-20.

Samdani AF, Dawson TM, Dawson VL (1997) Nitric oxide synthase in models of focal ischemia. *Stroke,* **28:** 1283-1288.

Sarzani R, Cusi D, Salvi F, Barlassina C, Macciardi F, Pietrucci F, Cola G, Catalini R, Dal FC, Dessi-Fulgheri P, Rappelli A (2006) The 460Trp allele of a-adducin increases carotid intima-media thickness in young adult males. *J.Hypertens.,* **24:** 697-703.

Sass C, Zannad F, Herbeth B, Salah D, Chapet O, Siest G, Visvikis S (1998) Apolipoprotein E4, lipoprotein lipase C447 and angiotensin-I converting enzyme deletion alleles were not associated with increased wall thickness of carotid and femoral arteries in healthy subjects from the Stanislas cohort. *Atherosclerosis,* **140:** 89-95.

Sawada H, Udaka F, Izumi Y, Nishinaka K, Kawakami H, Nakamura S, Kameyama M (2000) Cerebral white matter lesions are not associated with apoE genotype but with age and female sex in Alzheimer's disease. *J.Neurol.Neurosurg.Psychiatry,* **68:** 653-656.

Sayed-Tabatabaei FA, Houwing-Duistermaat JJ, van Duijn CM, Witteman JC (2003) Angiotensin-converting enzyme gene polymorphism and carotid artery wall thickness: a meta-analysis. *Stroke,* **34:** 1634-1639.

Sayed-Tabatabaei FA, Schut AF, Hofman A, Bertoli-Avella AM, Vergeer J, Witteman JC, van Duijn CM (2004) A study of gene--environment interaction on the gene for angiotensin converting enzyme: a combined functional and population based approach. *J.Med.Genet.,* **41:** 99-103.

Scaglione L, Gambino R, Rolfo E, Lillaz E, Gai M, Cassader M, Pagano G, Cavallo-Perin P (2002) Plasma homocysteine, methylenetetrahydrofolate reductase gene polymorphism and carotid intima-media thickness in Italian type 2 diabetic patients. *Eur.J.Clin.Invest.,* **32:** 24-28.

Scheltens P, Barkhof F, Leys D, Pruvo JP, Nauta JJ, Vermersch P, Steinling M, Valk J (1993) A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *J.Neurol.Sci.,* **114:** 7-12.

Scheltens P, Erkinjunti T, Leys D, Wahlund LO, Inzitari D, Del ST, Pasquier F, Barkhof F, Mantyla R, Bowler J, Wallin A, Ghika J, Fazekas F, Pantoni L (1998) White matter changes on CT and MRI: an overview of visual rating scales. European Task Force on Age-Related White Matter Changes. *Eur.Neurol.,* **39:** 80-89.

Schmidt H, Schmidt R, Fazekas F, Semmler J, Kapeller P, Reinhart B, Kostner GM (1996) Apolipoprotein E e4 allele in the normal elderly: neuropsychologic and brain MRI correlates. *Clin.Genet.,* **50:** 293-299.

Schmidt R, Schmidt H, Fazekas F, Kapeller P, Roob G, Lechner A, Kostner GM, Hartung HP (2000) MRI cerebral white matter lesions and paraoxonase PON1 polymorphisms : three-year follow-up of the austrian stroke prevention study. *Arterioscler.Thromb.Vasc.Biol.,* **20:** 1811-1816.

Schmidt R, Schmidt H, Fazekas F, Launer LJ, Niederkorn K, Kapeller P, Lechner A, Kostner GM (2001) Angiotensinogen polymorphism M235T, carotid atherosclerosis, and small-vessel disease-related cerebral abnormalities. *Hypertension,* **38:** 110-115.

Schmoelzer I, Renner W, Paulweber B, Malaimare L, Iglseder B, Schmid P, Schallmoser K, Wascher TC (2003) Lack of association of the Glu298Asp polymorphism of endothelial nitric oxide synthase with manifest coronary artery disease, carotid atherosclerosis and forearm vascular reactivity in two Austrian populations. *Eur.J.Clin.Invest.,* **33:** 191-198.

Schork NJ (1997) Genetics of complex disease: approaches, problems, and solutions. *Am.J.Respir.Crit Care Med.,* **156:** S103-S109.

Schulz UG, Flossmann E, Rothwell PM (2004) Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies. *Stroke,* **35:** 819-824.

Schut AFC, Janssen JAMJ, Deinum J, Vergeer JM, Hofman A, Lamberts SWJ, Oostra BA, Pols HAP, Witteman JCM, van Duijn CM (2003) Polymorphism in the promoter region of the insulin-like growth factor I gene is related to carotid intima-media thickness and aortic pulse wave velocity in subjects with hypertension. *Stroke,* **34:** 1623-1627.

Seifert T, Lechner A, Flooh E, Schmidt H, Schmidt R, Fazekas F (2006) Lack of association of lobar intracerebral hemorrhage with apolipoprotein E genotype in an unselected population. *Cerebrovasc.Dis.,* **21:** 266-270.

Seminara D, Khoury MJ, O'Brien TR, Manolio T, Gwinn ML, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam AP, Danesh J, Davey SG, Dolan S, Duncan R, Gruis NA, Hashibe M, Hunter D, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, Riboli E, Salanti G, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Ioannidis JP (2007) The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology,* **18:** 1-8.

Sethi AA, Nordestgaard BG, Tybjaerg-Hansen A (2003) Angiotensinogen gene polymorphism, plasma angiotensinogen, and risk of hypertension and ischemic heart disease: a meta-analysis. *Arterioscler.Thromb.Vasc.Biol.,* **23:** 1269-1275.

Shih DM, Gu L, Xia YR, Navab M, Li WF, Hama S, Castellani LW, Furlong CE, Costa LG, Fogelman AM, Lusis AJ (1998) Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis. *Nature,* **394:** 284-287.

Sierra C, Coca A, Gomez-Angelats E, Poch E, Sobrino J, de la SA (2002) Renin-angiotensin system genetic polymorphisms and cerebral white matter lesions in essential hypertension. *Hypertension,* **39:** 343-347.

Skoog I, Hesse C, Aevarsson O, Landahl S, Wahlstrom J, Fredman P, Blennow K (1998) A population study of apoE genotype at the age of 85: relation to dementia, cerebrovascular disease, and mortality. *J.Neurol.Neurosurg.Psychiatry,* **64:** 37-43.

Sleegers K, den HT, van Dijk EJ, Hofman A, Bertoli-Avella AM, Koudstaal PJ, Breteler MM, van Duijn CM (2005) ACE gene is associated with Alzheimer's disease and atrophy of hippocampus and amygdala. *Neurobiol.Aging,* **26:** 1153-1159.

Slooter AJC, Bots ML, Havekes LM, Iglesias del SA, Cruts M, Grobbee DE, Hofman A, Van BC, Witteman JCM, van Duijn CM (2001) Apolipoprotein E and carotid artery atherosclerosis: The Rotterdam study. *Stroke,* **32:** 1947-1952.

Spoto B, Benedetto FA, Testa A, Tripepi G, Mallamaci F, Maas R, Boeger RH, Zoccali C, Parlongo RM, Pisano A (2005) Atherosclerosis and the Glu298Asp polymorphism of the eNOS gene in white patients with end-stage renal disease. *Am.J.Hypertens.,* **18:** 1549-1555.

Srinivasan SR, Li S, Chen W, Tang R, Bond MG, Boerwinkle E, Berenson GS (2004) Q192R polymorphism of the paraoxanase 1 gene and its association with serum lipoprotein variables and carotid artery intima-media thickness in young adults from a biracial community: The Bogalusa Heart Study. *Atherosclerosis,* **177:** 167-174.

StataCorp. (2001) Statistical Software: Release 7.0. College Station, TX.

Steffens DC, Trost WT, Payne ME, Hybels CF, MacFall JR (2003) Apolipoprotein E genotype and subcortical vascular lesions in older depressed patients and control subjects. *Biol.Psychiatry,* **54:** 674-681.

Sudlow C, Martinez Gonzalez NA, Kim J, Clark C (2006) Does apolipoprotein E genotype influence the risk of ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage? Systematic review and meta-analyses of 31 studies among 5961 cases and 17,965 controls. *Stroke,* **37:** 364-370.

Sudlow CL, Warlow CP (1997) Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. International Stroke Incidence Collaboration. *Stroke,* **28:** 491-499.

Sutton AJ, Higgins JP (2008) Recent developments in meta-analysis. *Stat.Med.,* **27:** 625-650.

Svensson PJ, Dahlback B (1994) Resistance to activated protein C as a basis for venous thrombosis. *N.Engl.J.Med.,* **330:** 517-522.

Szolnoki Z, Melegh B (2006) Gene-gene and gene-environment interplay represent specific susceptibility for different types of ischaemic stroke and leukoaraiosis. *Curr.Med.Chem.,* **13:** 1627-1634.

Szolnoki Z, Somogyvari F, Kondacs A, Szabo M, Fodor L, Bene J, Melegh B (2004) Specific APO E genotypes in combination with the ACE D/D or MTHFR 677TT mutation yield an independent genetic risk of leukoaraiosis. *Acta Neurol.Scand.,* **109:** 222-227.

Szolnoki ZK (2007) Evaluation of the roles of the A185C and C406T kinesin light-chain 1 variants in the development of leukoaraiosis. *Neurosci.Lett.,* **429:** 18.

Tabara Y, Kohara K, Nakura J, Miki T (2001) Risk factor-gene interaction in carotid atherosclerosis: Effect of gene polymorphisms of renin-angiotensin system. *J.Hum.Genet.,* **46:** 278-284.

Takami S, Imai Y, Katsuya T, Ohkubo T, Tsuji I, Nagai K, Satoh H, Hisamichi S, Higaki J, Ogihara T (2000) Gene polymorphism of the renin-angiotensin system associates with risk for lacunar infarction. The Ohasama study. *Am.J.Hypertens.,* **13:** 121-127.

Tanriverdi H, Evrengul H, Mergen H, Acar C, Seleci D, Kuru O, Tanriverdi S, Kaftan A, Tanriverdi H, Evrengul H, Mergen H, Acar C, Seleci D, Kuru O, Tanriverdi S, Kaftan A (2007) Early sign of atherosclerosis in slow coronary flow and relationship with angiotensin-converting enzyme I/D polymorphism. *Heart & Vessels,* **22:** 1-8.

Taute B-M, Seifert H, Taute R, Glaser C, Podhaisky H (2000) Angiotensin-coverting enzyme gene insertion/deletion polymorphism and carotid artery wall thickness in patients with peripheral arterial occlusive disease. *Int.Angiol.,* **19:** 337-344.

Terry CF, Loukaci V, Green FR (2000) Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. *J.Biol.Chem.,* **275:** 18138-18144.

Terry JG, Howard G, Mercuri M, Bond MG, Crouse III, Jr. (1996) Apolipoprotein E polymorphism is associated with segment-specific extracranial carotid artery intima-media thickening. *Stroke,* **27:** 1755-1759.

Tesauro M, Thompson WC, Rogliani P, Qi L, Chaudhary PP, Moss J (2000) Intracellular processing of endothelial nitric oxide synthase isoforms associated with differences in severity of cardiopulmonary diseases: cleavage of proteins with aspartate vs. glutamate at position 298. *Proc.Natl.Acad.Sci.U.S.A,* **97:** 2832-2835.

Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J (2005) A method for meta-analysis of molecular association studies. *Stat.Med.,* **24:** 1291-1306.

The Cochrane Collaboration (2006) Review Manager (RevMan) Version 4.3. Copenhagen: The Nordic Cochrane Centre.

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* **447:** 661-678.

Tiret L, Bonnardeaux A, Poirier O, Ricard S, Marques-Vidal P, Evans A, Arveiler D, Luc G, Kee F, Ducimetiere P, . (1994) Synergistic effects of angiotensin-converting enzyme and angiotensin-II type 1 receptor gene polymorphisms on risk of myocardial infarction. *Lancet,* **344:** 910-913.

Torzewski M, Rist C, Mortensen RF, Zwaka TP, Bienek M, Waltenberger J, Koenig W, Schmitz G, Hombach V, Torzewski J (2000) C-reactive protein in the arterial intima: role of C-reactive protein receptor-dependent monocyte recruitment in atherogenesis. *Arterioscler.Thromb.Vasc.Biol.,* **20:** 2094-2099.

Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Desvarieux M, Ebrahim S, Fatar M, Hernandez HR, Kownator S, Prati P, Rundek T, Taylor A, Bornstein N, Csiba L, Vicaut E, Woo KS, Zannad F (2004) Mannheim intima-media thickness consensus. *Cerebrovasc.Dis.,* **18:** 346-349.

Turner ST, Fornage M, Jack CR, Jr., Mosley TH, Kardia SL, Boerwinkle E, de AM (2005) Genomic susceptibility loci for brain atrophy in hypertensive sibships from the GENOA study. *Hypertension,* **45:** 793-798.

Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de AM (2004) Heritability of leukoaraiosis in hypertensive sibships. *Hypertension,* **43:** 483-487.

Um JY, Jeong HJ, Park RK, Hong SH, Kim HM (2005) Aspects of gene polymorphisms in cerebral infarction: inflammatory cytokines. *Cell Mol.Life Sci.,* **62:** 824-833.

Vaessen N, Heutink P, Janssen JA, Witteman JC, Testers L, Hofman A, Lamberts SW, Oostra BA, Pols HA, van Duijn CM (2001) A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. *Diabetes,* **50:** 637-642.

Van Bortel LM (2005) What does intima-media thickness tell us? *J.Hypertens.,* **23:** 37-39.

Van Himbergen T, Roest M, De WF, De GJ, Voorbij H, Van TL, Stalenhoef A (2004) Paraoxonase genotype, LDL-oxidation and carotid atherosclerosis in male life-long smokers. *Free Radic.Res.,* **38:** 553-560.

van Rijn MJ, Bos MJ, Isaacs A, Yazdanpanah M, Arias-Vasquez A, Stricker BH, Klungel OH, Oostra BA, Koudstaal PJ, Witteman JC, Hofman A, Breteler MM, van Duijn CM (2007) Polymorphisms of the renin-angiotensin system are associated with blood pressure, atherosclerosis and cerebral white matter pathology. *J.Neurol.Neurosurg.Psychiatry,* **78:** 1083-1087.

van Rijn MJ, Bos MJ, Yazdanpanah M, Isaacs A, Arias-Vasquez A, Koudstaal PJ, Hofman A, Witteman JC, van Duijn CM, Breteler MM (2006) Alpha-adducin polymorphism, atherosclerosis, and cardiovascular and cerebrovascular risk. *Stroke,* **37:** 2930-2934.

Varda NM, Peterlin B, Bradac SU, Gregoric A (2005) Carotid artery intima-media thickness and angiotensin-converting enzyme gene polymorphism in the offspring of parents with premature stroke. *Acta Paediatr.,* **94:** 33-37.

Vauhkonen I, Niskanen L, Ryynanen M, Voutilainen R, Partanen J, Toyry J, Mercuri M, Rauramaa R, Uusitupa M (1997) Divergent association of apolipoprotein E polymorphism with vascular disease in patients with NIDDM and control subjects. *Diabet.Med.,* **14:** 748-756.

Verpillat P, Alperovitch A, Cambien F, Besancon V, Desal H, Tzourio C (2001) Aldosterone synthase (CYP11B2) gene polymorphism and cerebral white matter hyperintensities. *Neurology,* **56:** 673-675.

Vickers MA, Green FR, Terry C, Mayosi BM, Julier C, Lathrop M, Ratcliffe PJ, Watkins HC, Keavney B (2002) Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. *Cardiovasc.Res.,* **53:** 1029-1034.

Visvikis S, Sass C, Pallaud C, Grow MA, Zannad F, Siest G, Erlich HA, Cheng S (2000) Familial studies on the genetics of cardiovascular diseases: The stanislas cohort. *Clin.Chem.Lab.Med.,* **38:** 827-832.

Volcik KA, Barkley RA, Hutchinson RG, Mosley TH, Heiss G, Sharrett AR, Ballantyne CM, Boerwinkle E (2006) Apolipoprotein E polymorphisms predict low density lipoprotein cholesterol levels and carotid artery wall thickness but not incident coronary heart disease in 12,491 ARIC study participants. *Am.J.Epidemiol.,* **. 164**.

Wahlund LO, Barkhof F, Fazekas F, Bronge L, Augustin M, Sjogren M, Wallin A, Ader H, Leys D, Pantoni L, Pasquier F, Erkinjuntti T, Scheltens P (2001) A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke,* **32:** 1318-1322.

Walsh MF, Barazi M, Pete G, Muniyappa R, Dunbar JC, Sowers JR (1996) Insulin-like growth factor I diminishes in vivo and in vitro vascular contractility: role of vascular nitric oxide. *Endocrinology,* **137:** 1798-1803.

Wang D, Yang H, Quinones MJ, Bulnes-Enriquez I, Jimenez X, De La RR, Modilevsky T, Yu K, Li Y, Taylor KD, Hsueh WA, Hodis HN, Rotter JI (2005) A genome-wide scan for carotid artery intima-media thickness: the Mexican-American Coronary Artery Disease family study. *Stroke,* **36:** 540-545.

Watanabe Y, Ishigami T, Kawano Y, Umahara T, Nakamori A, Mizushima S, Hibi K, Kobayashi I, Tamura K, Ochiai H, Umemura S, Ishii M (1997) Angiotensin-converting enzyme gene I/D polymorphism and carotid plaques in Japanese. *Hypertension,* **30:** 569-573.

Welch GN, Loscalzo J (1998) Homocysteine and atherothrombosis. *N.Engl.J.Med.,* **338:** 1042-1050.

Wen HM, Baum L, Cheung WS, Mok V, Lam WW, Tomlinson B, Wong KS, Ng HK (2006) Apolipoprotein E epsilon4 allele is associated with the volume of white matter changes in patients with lacunar infarcts. *Eur.J.Neurol.,* **13:** 1216-1220.

Wendelhag I, Gustavsson T, Suurkula M, Berglund G, Wikstrand J (1991) Ultrasound measurement of wall thickness in the carotid artery: fundamental principles and description of a computerized analysing system. *Clin.Physiol,* **11:** 565-577.

Wheeler JG, Keavney BD, Watkins H, Collins R, Danesh J (2004) Four paraoxonase gene polymorphisms in 11212 cases of coronary heart disease and 12786 controls: meta-analysis of 43 studies. *Lancet,* **363:** 689-695.

Wikstrand J, Wiklund O (1992) Frontiers in cardiovascular science. Quantitative measurements of atherosclerotic manifestations in humans. *Arterioscler.Thromb.,* **12:** 114-119.

Williams SM, Addy JH, Phillips JA, III, Dai M, Kpodonu J, Afful J, Jackson H, Joseph K, Eason F, Murray MM, Epperson P, Aduonum A, Wong LJ, Jose PA, Felder RA (2000) Combinations of variations in multiple genes are associated with hypertension. *Hypertension,* **36:** 2-6.

Wiszniewska M, Devuyst G, Bogousslavsky J, Ghika J, van MG (2000) What is the significance of leukoaraiosis in patients with acute ischemic stroke? *Arch.Neurol.,* **57:** 967-973.

Wohlin M, Sundstrom J, Lannfelt L, Axelsson T, Syvanen AC, Andren B, Basu S, Lind L (2007) Apolipoprotein E epsilon4 genotype is independently associated with increased intima-media thickness in a recessive pattern. *Lipids,* **42:** 451-456.

Wolff B, Braun C, Schluter C, Grabe HJ, Popowski K, Volzke H, Ludemann J, John U, Cascorbi I (2005) Endothelial nitric oxide synthase Glu298Asp polymorphism, carotid atherosclerosis and intima-media thickness in a general population sample. *Clin.Sci.,* **109:** 475-481.

World Health Organisation (2004) *The Global Burden of Disease: 2004 Update*. WHO, Geneva.

Xiang G-D, Hu T-H, Wang Y-L (2003) Apolipoprotein E genotypes and carotid artery atherosclerosis in type 2 diabetes mellitus. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi,* **20:** 66.

Yamasaki Y, Katakami N, Sakamoto K, Kaneto H, Matsuhisa M, Sato H, Hori M, Haneda M, Kashiwagi A, Tanaka Y, Kawamori R, Kuno S-I (2006) Combination of multiple genetic risk factors is synergistically associated with carotid atherosclerosis in Japanese subjects with type 2 diabetes. *Diabetes Care,* **29:** 2445-2451.

Yazdanpanah M, Sayed-Tabatabaei FA, Hofman A, Aulchenko YS, Oostra BA, Stricker BHC, Pols HAP, Lamberts SWJ, Witteman JCM, Janssen JAMJ, van Duijn CM (2006) The a-adducin gene is associated with macrovascular complications and mortality in patients with type 2 diabetes. *Diabetes,* **55:** 2922-2927.

Ye Z, Liu EH, Higgins JP, Keavney BD, Lowe GD, Collins R, Danesh J (2006) Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls. *Lancet,* **367:** 651-658.

Zhang SH, Reddick RL, Piedrahita JA, Maeda N (1992) Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science,* **258:** 468-471.

Zhang Y, You K, Zhang L, Jiang Y (1998) Effect of apolipoprotein E polymorphism on serum lipids, coronary heart disease, and carotid artery atherosclerosis. *Chin.J.Cardiol.,* **26:** 443-447.

Appendix 1

Stata code:

Step 1 – Meta-ANOVA

Have data in long format:

study	genotype	n	mean	sd
1	0	x	x	x
1	1	x	x	x
1	2	x	x	x

Derive SE using:

```
gen se = sd / sqrt (n)
```

Meta-ANOVA:

```
xi: regress mean i.genotype i.study [aweight=1/se^2]
testparm_Igenotype*
```

Gives the p-value for genotype in the ANOVA model.

Step 2 – Determine genetic model

Have data in wide format:

study	n11	x11	s11	n12	x12	s12	n22	x22	s22
xxxx	x	x	x	x	x	x	x	x	x

Carry out MD1 and MD2 meta-analyses saving the _ES and _seES variables generated:

```
metan n12 x12 s12 n11 x11 s11, random nostandard
gen MD1 = _ES
gen se1 = _seES

metan n22 x22 s22 n11 x11 s11, random nostandard
gen MD2 = _ES
gen se2 = _seES
```

Calculate the average of the standard errors to get a study-wide estimate:

```
gen se = (se1 + se2) / 2
```

Finally carry out the weighted linear regression:

```
regress MD1 MD2 [aweight=1/se^2], noconstant
```

Step 3 – Calculating the mean difference using the appropriate genetic model

Based on the regression coefficient estimated from step 2, the appropriate genetic model was chosen. 0= recessive, 0.5=co-dominant, 1=dominant.

The data should be in wide format as for step 2.

The following `gametan`* commands are used to estimate the appropriate mean difference:

```
gametan n22 x22 s22 n12 x12 s12 n11 x11 s11, codominant
```
Or
```
gametan n22 x22 s22 n12 x12 s12 n11 x11 s11, recessive
```
Or
```
gametan n22 x22 s22 n12 x12 s12 n11 x11 s11, dominant
```

*gametan is a STATA program written by Julian Higgins to carry out genetic association meta-analyses. The code was distributed to me via personal communication and has not yet been released.

Appendix 2. Terms used in CIMT gene-specific searches

Gene	Medline terms	Embase terms
APOE	apolipoproteins/ or apolipoproteins e/ ((apolipoprotein$ adj e) or (apoprotein$ adj e) or apo-e or apo e or apoe).tw.	apolipoprotein/ or apolipoprotein e/ or apolipoprotein e2/ or apolipoprotein e3/ or apolipoprotein e4/ or apolipoprotein e5/ or apolipoprotein e7/ ((apolipoprotein$ adj e) or (apoprotein$ adj e) or apo-e or apoe or apo e).tw.
ACE	exp Peptidyl-Dipeptidase A/ge [Genetics] (angiotensin converting enzyme or ace or peptidyl-dipeptidase).tw.	exp Dipeptidyl Carboxypeptidase/ (angiotensin converting enzyme or ace or peptidyl-dipeptidase).tw.
MTHFR	exp "Methylenetetrahydrofolate Reductase (NADPH2)"/ge [Genetics] (MTHFR or c677t or nadph2 or methylene tetrahydrofolate or methylenetetrahydrofolate).tw.	exp "5,10 methylenetetrahydrofolate reductase (fadh2)"/ (MTHFR or c677t or nadph2 or meythlene tetrahydrofolate or methylenetetrahydrofolate).tw.
NOS3	exp Nitric Oxide/ or exp Nitric Oxide Synthase/ (eNOS or ecNOS or NOS or NOS3 or nitric oxide synthase or T-786C or T786C or Glu298Asp or NO synthase or G894T).tw.	exp Endothelial Nitric Oxide Synthase/ or exp Nitric Oxide Synthase/ (eNOS or ecNOS or NOS or NOS3 or nitric oxide synthase or T-786C or T786C or Glu298Asp or NO synthase or G894T).tw.
ADD1	 (adducin$ or add1 or add-1 or addA or add-A or alpha-adducin).tw.	exp ADDUCIN/ (adducin$ or add1 or add-1 or addA or add-A or alpha-adducin).tw.
PON1	(paraoxonase or paraoxon esterase or PON1).tw.	(paraoxonase or paraoxon esterase or PON1).tw.
IL6	exp Interleukin-6/ge [Genetics] exp Interferon-beta/ge [Genetics]	exp Interleukin 6/ exp Beta Interferon/

	(interleukin 6 or interferon beta 2 or IL6 or IL-6 or IL 6 or interleukin-6 or interferon beta-2 or BSF-2 or B-cell simulatory factor 2 or CDF or hybridoma growth factor or CTL differentiation factor).tw.	(interleukin 6 or interferon beta 2 or IL6 or IL-6 or IL 6 or interleukin-6 or interferon beta-2 or BSF-2 or B-cell simulatory factor 2 or CDF or hybridoma growth factor or CTL differentiation factor).tw.
IGF1	exp Insulin-Like Growth Factor I/ge [Genetics] (insulin-like growth factor I or IGF-I or IGF-1).tw.	 (insulin-like growth factor I or IGF-I or IGF-1).tw.
ADRB2	exp Receptors, Adrenergic, beta-2/ge [Genetics] (ADRB2 or beta 2 adrenergic receptor or ADRB2R or ADRBR or B2AR or BAR or BETA2AR or beta 2 adrenoceptor or catecholamine receptor).tw.	exp Beta 2 Adrenergic Receptor/ (ADRB2 or beta 2 adrenergic receptor or ADRB2R or ADRBR or B2AR or BAR or BETA2AR or beta 2 adrenoceptor or catecholamine receptor).tw.
CRP	exp C-Reactive Protein/ge [Genetics] (CRP or reactive protein).tw.	exp C Reactive Protein/ (crp or reactive protein).tw.
FGG/FGA	(fibrinogen or FGG or FGA).tw.	(fibrinogen or FGG or FGA).tw.
AGT	exp Angiotensinogen/ge [Genetics] exp Receptors, Angiotensin/ge [Genetics] (agt$ or angiotensin$).tw.	exp angiotensin derivative/ or angiotensinogen/ exp Angiotensin 1 Receptor/ or exp Angiotensin Receptor/ (agt$ or angiotensin$).tw.
FV	exp Factor V/ge [Genetics] (facV or factor V or FVL or Leiden or factor 5).tw.	exp Blood Clotting Factor 5/ (facV or factor V or FVL or Leiden or factor 5).tw.

Appendix 3. Data transformations of CIMT papers.

Often it was necessary to transform the data presented in a paper in order to get the data in the required format for analysis (i.e.. number of subjects, mean and standard deviation of CIMT).

3.1 Example of combining groups

(e.g. men and women, genotype groups, age across genotypes)

$$\text{mean}_{\text{total}} = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}$$

$$\text{variance}_{\text{total}} = \frac{n_1(\sigma_1^2 + \mu_1^2) + n_2(\sigma_2^2 + \mu_2^2)}{n_1 + n_2} - \mu_{\text{total}}^2$$

Elosua 2004:

	E2			E3			E4		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Men	169	0.77	0.17	874	0.78	0.20	272	0.77	0.20
Women	204	0.70	0.14	908	0.71	0.16	296	0.72	0.19
All*	373	0.73	0.16	1782	0.74	0.18	568	0.74	0.20

* means and SDs to be used in the analysis, calculated using the above formulas.

3.2 Combining data from left and right CCA

Some papers had reported the mean CIMT per genotype separately for left and right CCAs. The ideal measurement was the mean of the right and left, so this was calculated.

Note, this sort of combination of data (where the resulting sample size is the same as in the individual groups), is different to that described above (where there is a combination of data across individuals).

Example, Bilici 2006:

	DD			ID			II		
	n	$\bar{x}$	SD	n	$\bar{x}$	SD	n	$\bar{x}$	SD
Left	28	1.27	0.32	28	1.32	0.35	8	1.30	0.22
Right	28	1.31	0.27	28	1.25	0.30	8	1.33	0.36
Mean	28	1.29	0.30	28	1.29	0.33	8	1.32	0.29

3.3 Example of estimating numbers of subjects

Linnebank 2006:

In this paper the sample sizes per genotype were not reported. Instead the genotype proportions were reported. As the overall sample size is known it is possible to calculate the sample size per genotype. However, the total must equal the overall sample size and there

are several genotype sample sizes possible. Here I report the possible values and show which were selected:

	Genotype proportions	Total	Mean	Possible sample sizes						
TT	0.48	714	343	340	341	342	343	344	345	**346**
CT	0.44	714	314	311	312	313	314	315	**316**	317
CC	0.07	714	50	47	48	49	50	51	**52**	53
Total	0.99		707						714	

3.4 Bleil 2006

There was no SD data per genotype group in this paper. However, the overall SD was reported to be 0.16mm, so the SD of the three genotypes were assumed to be equal and 0.16 was used as the SD for each genotype.

3.5 Cattin 1997

This paper reports CIMT as the sum of the right and left CCAs. As other papers report either only one CCA, or the mean of both I transformed the data to be more similar to these. I divided the mean for each genotype by 2, to represent the average CCA CIMT. To estimate the standard deviation of this measurement I used the following formula:

$$\mathrm{Var}(X) = \mathrm{Var}(2X) / 4$$

The presented and transformed data are shown below:

	n	Mean (sum R&L)	SD	Mean (avg R&L)	Var (2X)/4	SD (avg R&L)
E2	32	3.6	0.2	1.80	0.01	0.10
E3	177	3.7	0.3	1.84	0.0225	0.15
E4	45	3.9	0.9	1.95	0.2025	0.45

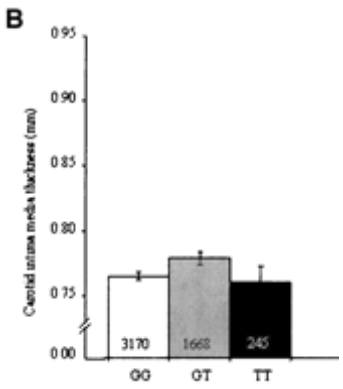
3.6 Slooter 2001

The CIMT (per genotype) data were reported as the mean difference from the reference group (ε3ε3), with 95%CIs, as shown in the table below. In the text, the median of ε3ε3 was reported to be 0.77mm (10th centile: 0.63; 90th centile: 1.00 mm). I therefore took the mean for ε3ε3 to be 0.77mm and calculated the standard deviation as if the data were normally distributed (i.e. (1.00 - 0.63)/ 2.564 = 0.14). From the estimated ε3ε3 mean, I could calculate the mean of all the other genotypes. Using the formula escribed above I converted each of the CIs to SDs. The before and after data are shown in the table below:

	ε2ε2	ε2ε3	ε3ε3	ε3ε4	ε4ε4
mean difference	-0.04	-0.02	0	0.00	0.01
CI	-0.08 to 0.00	-0.03 to -0.01	-	-0.01 to 0.01	-0.01 to 0.04
n	46	704	3122	1258	134
mean	0.73	0.75	0.77	0.77	0.78
SD	0.14	0.14	0.14	0.18	0.15

3.7 Yazdanpanah 2006

There were no data in the text or the tables relating to CIMT per genotype. However the mean and standard error data were presented in a graph. I therefore, estimated the means and standard deviations from the graph:



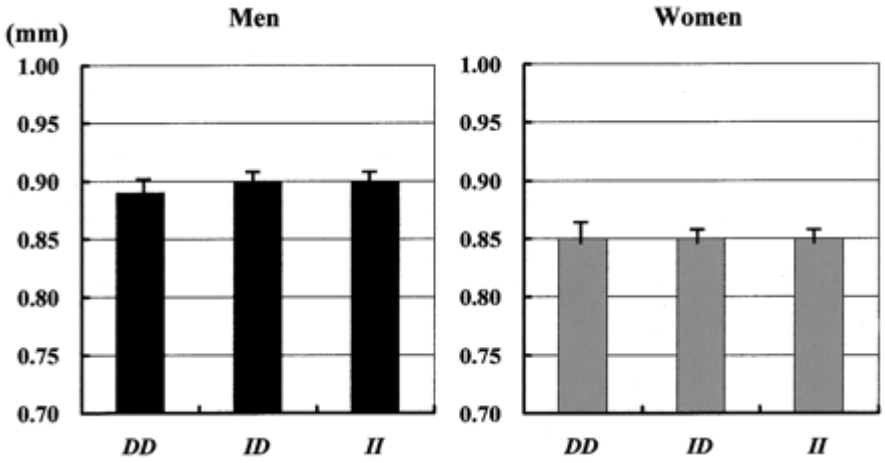
Using the scale on the graph: 1mm (on graph) = 0.0043

Genotype	n	Mean*	se*	sd†
GG	3170	0.77	0.00645	0.36
GT	1668	0.78	0.01075	0.44
TT	245	0.76	0.02365	0.37

*data estimated from graph, † standard deviation calculated from standard error and n.

3.8 Mannami 2001

There were no data in the text or the tables relating to CIMT per genotype. However the mean and standard error data was presented in a graph. I therefore, estimated the means and standard deviations from the graph:

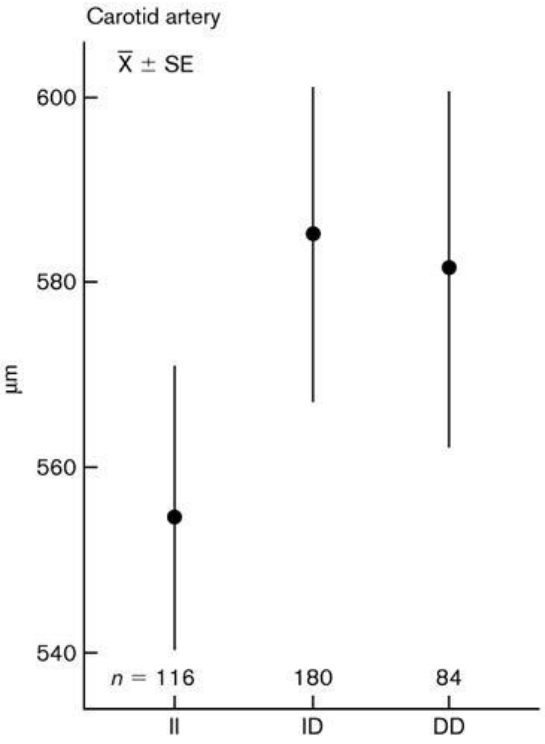


Gender	Genotype	n	Mean*	SE*	SD†
Men	DD	215	0.89	0.01	0.158
	DI	791	0.90	0.01	0.281
	II	694	0.90	0.01	0.263
Women	DD	262	0.85	0.02	0.324
	DI	849	0.85	0.01	0.291
	II	846	0.85	0.01	0.291

*data estimated from graph, † standard deviation calculated from standard error and n.

3.9 Balkestein 2002

There were no data in the text or the tables relating to CIMT per genotype. However the mean and standard error data was presented in a graph. I therefore, estimated the means and standard deviations from the graph:



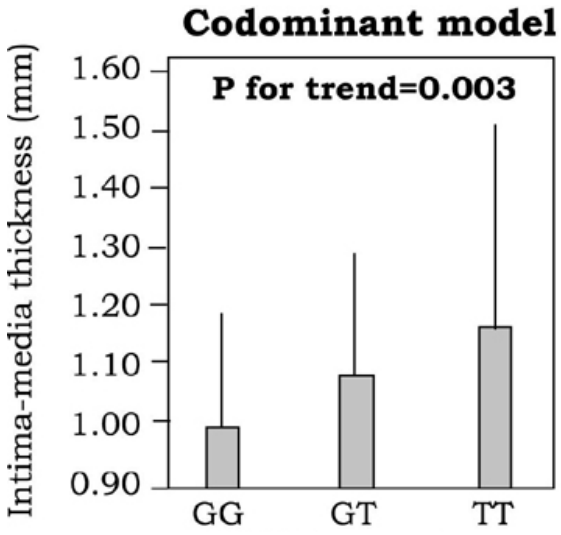
1mm = 0.95

Genotype	n	Mean*	se*	sd†
II	116	555	15	0.16
ID	180	585	17	0.23
DD	84	582	19	0.17

*data estimated from graph, † standard deviation calculated from standard error and n.

3.10 Spoto 2005

There were no data in the text or the tables relating to CIMT per genotype. However the mean and standard deviation data was presented in a graph.



Genotype	n	Mean*	SD*
GG	59	0.98	0.10
GT	56	1.07	0.23
TT	16	1.16	0.36

*data estimated from graph

Appendix 4. Example of data collection letter – including letter and forms.

Division of Clinical Neuroscience
Western General Hospital
Edinburgh
EH4 2XU

Tel. +44 131 537 2546
Fax. +44 131 332 5150

Dr Jan Staessen
Study Coordinating Centre,
Laboratory of Hypertension,
University of Leuven,
Campus Gasthuisberg,
Herestraat 49,
B3000 Leuven,
Belgium

07 October 2009

Dear Dr Staessen,

Re: Systematic review and meta-analyses of the association of commonly studied genes with carotid intima-media thickness.

We are carrying out a systematic review and series of meta-analyses of the association between genotype and carotid intima-media thickness, focusing on those genes studied in large numbers of subjects.

We have identified you as a principal investigator for the following study:

Balkestein EJ, Wang JG, Struijker-Boudier HAJ, Barlassina C, Bianchi G, Birkenhäger WH, Brand E, Den Hond E, Fagard R, Herrmann S-M, Van Bortel LM, Staessen JA. 2002. Carotid and femoral intima-media thickness in relation to three candidate genes in a Caucasian population. J. Hypertension 20:1551-61.

This study seems very relevant to our review, and so we would be most grateful if you could help us by providing some basic information about it. Attached is a short data collection form, on which we have noted as much information as possible. It would be very helpful if you could check the information in the boxes (making any necessary changes) and complete any boxes that remain blank The easiest and quickest way to do this is probably to complete the form, save it to your PC, and then email it back to us as an attachment. If you prefer, however, you could print it out, complete it and then fax or post it to us. We intend to carry

out the analyses in June and would therefore be most grateful if you could return the form to us before the end of May.

We very much appreciate your help with this. The review will be much more reliable if we are able to include data from all relevant studies identified. We will of course send you a copy of the results once we have completed the analyses and will acknowledge your contribution in publications arising from this work.

If you have any questions or comments, please do not hesitate to contact us by email, phone, post or fax.

With many thanks,
Yours sincerely,

Lavinia Paternoster
PhD Student

Dr Cathie Sudlow
Clinical Senior Lecturer, Wellcome Trust Clinician Scientist and Honorary Consultant Neurologist

Please return this form to Lavinia Paternoster preferably by email, L.Paternoster@sms.ed.ac.uk, or by post or fax.

CIMT measurement method

Please tell us how the CIMT values in the analysis were obtained. If you have analysis data relating to more than one type of CIMT measurement, we would prefer to have data relating to only the **mean** CIMT from the both the **right and left**, **far** walls of the **common** carotid artery (or close to this ideal).

Segment measured		Side measured		Wall measured	
(cross all that apply)					
Common carotid	☒	Right	☒	Near	☐
Bifurcation	☐	Left	☐	Far	☒
Internal carotid	☐	Both	☐	Both	☐
External carotid	☐				

Value used in analysis

Mean ☒ Maximum ☐

Briefly describe how many measurements were taken and how they were combined to create the final value used in the analysis (eg. the mean of the maximum from the right and the maximum from the left artery, 3 measurements taken from each side)

CIMT mean and standard deviation results per ADD1 genotype

Please complete the boxes.

Genotype	Number of subjects with ACE and CIMT data (N)	Mean CIMT (mm)	Standard deviation of CIMT
GG			
GT			
TT			

Appendix 5. Full table of studies identified in CIMT genetic search.

Gene	Total Subjects	Number of publications	Largest Study
APOE	37493	47	12491
ACE	23935	51	5321
MTHFR	14205	33	3247
NOS3	9434	19	2448
PON1	8921	27	1786
ADD1	8535	5	6471
AGT	7515	19	737
IL6	7190	10	2421
CRP	6603	3	4641
CD14	5943	7	1110
FAC V	5828	5	3750
TLR4	5638	6	2955
APOA1/C3	5363	8	2265
HFE	5288	4	2932
ADRB2	5249	1	5249
AGTR1	5117	14	737
CETP	4387	7	2632
FGG / FGA	4274	1	4274
IGF1	4239	2	3769
LPL	4178	10	2445
ADIPOQ	4035	4	1745
APOB	3386	7	326
LIPC	3181	4	2268
TLR2	3000	2	2955
PPARA	2991	2	2301
PPARG	2963	2	1379
TNFR1	2737	1	2737
MMP3	2531	5	1111
MCP1	2490	7	610
APOA5	2430	2	157
HGF	2412	1	2412
GJA4	2290	2	1440
APOA4	2276	2	2057
PTPN22	2268	1	2268
GSTM1	2228	4	1394
FABP2	2224	3	1621
FAC VII	2178	4	810
IL1	2142	3	1000
MTTP	2138	1	2138
GSTT1	2039	3	1394
CX3CR1	2038	3	1256
CYBA	2038	3	1361
MMP9	2005	4	1000
CYP7B1	1980	1	1980
APOA1	1950	3	1856
ABCA1	1817	4	969

Gene	Total Subjects	Number of publications	Largest Study
FGB	1804	1	1804
LTA	1778	2	1088
GP2B	1693	2	1292
CCR5	1691	3	380
PDE4D	1670	1	1670
APOA	1634	5	826
ADRB3	1488	2	731
LEP	1428	1	1428
PAI1	1378	3	218
UCP2	1334	1	1334
EDN1	1320	3	690
TGFB1	1312	4	80
FAC II	1307	4	407
CYB11B2	1270	3	420
FUT3	1238	1	1238
MMP1	1224	2	1000
ADH3	1181	1	1181
ITGB2	1160	1	1160
MPO	1160	1	1160
NPY	1152	2	966
MT-ND2	1148	1	1148
IL3	1109	1	1109
ECE1	1100	3	630
TNF	1036	4	332
ICAM1	1022	2	332
IRS1	1018	1	1018
OAZ1	1001	1	1001
IL5	1000	1	1000
MMP2	1000	1	1000
TIMP2	1000	1	1000
TIMP3	1000	1	1000
SOD2	989	1	989
ARG1	963	1	963
MT-TL1	935	2	673
GNB3	932	1	932
IL10	883	3	121
CCR2	850	2	531
THBD	803	2	333
ITGB3	792	3	161
SELE	788	4	332
PTGDS	782	1	782
ESR1	778	2	88
HTR2A	757	2	690
PON2	734	3	310
LTC4S	732	1	732
NRG1	706	1	706
BDKRB1	690	1	690
GDLM	690	1	690
GLUT1	690	1	690
IL18	690	1	690

Gene	Total Subjects	Number of publications	Largest Study
MGP	690	1	690
MARS	690	1	690
THPO	690	1	690
VEGF	690	1	690
VWF	690	1	690
GSTP1	645	2	605
CD40	620	1	620
TP53	605	1	605
PTPN1	590	1	590
ITGA2	537	1	537
APOC3	530	2	369
CLU	525	1	525
MT2A	506	1	506
SREBF2	497	1	497
VDR	471	1	471
ALOX5	470	1	470
CAV3	470	1	470
CD31	470	1	470
CD36	470	1	470
CTSG	470	1	470
FAC III	470	1	470
FAC VIII	470	1	470
NPPA	470	1	470
SPP1	470	1	470
P2RY1	470	1	470
PTAFR	470	1	470
PROC	470	1	470
SELL	470	1	470
SELPLG	470	1	470
TLR9	384	1	384
INS	331	1	331
PAFAH	330	2	190
CBS	322	2	161
ALDH2	304	1	304
OPG	234	2	175
COX2	220	1	220
SREBP1A	204	1	204
GPX1	184	1	184
SDF1	183	1	183
GP1B	158	1	158
GP1A	157	1	157
HPA1	156	1	156
PCK2	150	1	150
ADRA2B	148	1	148
LDLR	113	2	82
LCAT	105	1	105
HLA	86	1	86
GR	46	1	46
MMP4	42	1	42

Appendix 6. Terms used in the WMH gene-specific searches.

Gene	Medline terms	Embase terms
APOE	apolipoproteins/ or apolipoproteins e/ ((apolipoprotein$ adj e) or (apoprotein$ adj e) or apo-e or apo e or apoe).tw.	apolipoprotein/ or apolipoprotein e/ or apolipoprotein e2/ or apolipoprotein e3/ or apolipoprotein e4/ or apolipoprotein e5/ or apolipoprotein e7/ ((apolipoprotein$ adj e) or (apoprotein$ adj e) or apo-e or apoe or apo e).tw.
ACE	exp Peptidyl-Dipeptidase A/ge [Genetics] (angiotensin converting enzyme or ace or peptidyl-dipeptidase).tw.	exp Dipeptidyl Carboxypeptidase/ (angiotensin converting enzyme or ace or peptidyl-dipeptidase).tw.
MTHFR	exp "Methylenetetrahydrofolate Reductase (NADPH2)"/ge [Genetics] (MTHFR or c677t or nadph2 or methylene tetrahydrofolate or methylenetetrahydrofolate).tw.	exp "5,10 methylenetetrahydrofolate reductase (fadh2)"/ (MTHFR or c677t or nadph2 or meythlene tetrahydrofolate or methylenetetrahydrofolate).tw.
AGT	exp Angiotensinogen/ge [Genetics] exp Receptors, Angiotensin/ge [Genetics] (agt$ or angiotensin$).tw.	exp angiotensin derivative/ or angiotensinogen/ exp Angiotensin 1 Receptor/ or exp Angiotensin Receptor/ (agt$ or angiotensin$).tw.

Appendix 7. Data collection form for WMH systematic review

Reviewer initials

LEUKOARAIOSIS DATA EXTRACTION FORM

First author and year

Study or group name

Inclusion/Exclusion

Gene	APOE	MTHFR	ACE
Polymorphism	epsilon other:	677 other:	I/D other:
Leukoaraiosis	leukoaraiosis WM lesions other:	White matter hyperintensities (WMH) WM changes	grade: volume other:
Study Design	cohort other:	case/control	family-based

Included | Excluded | Variables measured, association not tested | Include pending correspondence

type? mean volume | dichotomized grading | mean grading

Reasons

Data Extraction

cases

Number of cases: (total number genotyped) Sex: (total number genotyped) Age: mean+ SD for each gender if given) Ethnicity: (total number genotyped)	case population: (what degree of WMH) (clinical features) stroke other: excl:

controls

Number of controls: (total number genotyped) Sex: (total number genotyped) Age: mean+ SD for each gender if given) Ethnicity: (total number genotyped)	ca-co matching: (details)	control population: (what degree of WMH) (clinical features) population patients other: excl:

Cohort

Size of cohort: total genotyped: (total number genotyped) Sex: (total number genotyped) Age: mean+ SD for each gender if given) Ethnicity: (total number genotyped)	cohort population: population stroke other: excl:

Geographic location:

Genotyping Technique	RFLP:	secondary:		HWE departure from?	detected / not detected / not stated
	other:				
blinding	scanning staff / genotyping staff	perhaps / not stated		LD	yes / no / not stated / which?
leukoariosis scanning	MRI / CT	periventricular / subcortical	not stated / other:	other info?	
volume or grade?	**volume** YES NO analysis measurement?	normalized YES	NO	**grade** which?	YES NO
				Mean?	dichotomized:
covariates	tested?				
	associated with genotype?				
	associated with phenotype?				
genetic model	dominant	recessive	codominant	not stated	

Results (N above and below cutoff or N, mean + sd for continuous data) A=wildtype	Population	Genotype aa	Genotype Aa	Genotype AA	key	Allele a	Allele A	adjusted for?
additional results odds ratios after adjustment								
reasons to contact authors								
any other comments? relevant papers in reference list								

Appendix 8. WMH data transformations.

Transformations similar to those carried out for CIMT were performed, such as combining data from two genotype groups, calculating sample size from proportions and calculating standard deviations from standard errors or confidence intervals. For examples of these types of transformations see appendix 5 and section 3.2.5 of the main text.

Some other specific transformations were required and are detailed below:

8.1 Bornebroek 1997

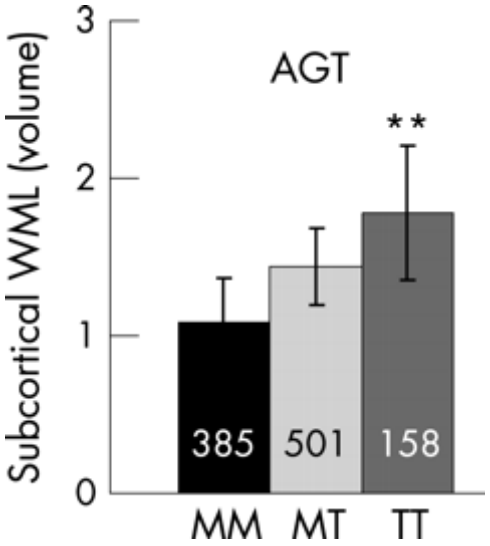
This paper reported the mean and range of WMH scores per genotype. As sample sizes were quite small it was possible for some genotypes to determine the actual individual values by using the range to provide two values and then determining which combination of values gave the reported means, so that standard deviations could be estimated:

genotype	n	mean	range	Individual values	SD
ε2ε2	1	22.0		22.0	-
ε3ε3	12	19.7	14 to 24	14.0; ? ; ? ; ? ; ? ; ? ; ? ; ? ; ? ; ? ; ? ; 24.0	?
ε2ε4	2	20	19 to 21	19.0; 21.0	1.4
ε3ε4	7	20	15 to 24	15.0; ? ; ? ; ? ; ? ; ?; 24.0	?
ε4ε4	3	20.7	14 to 24	14.0; 24.0; 24.0	5.8

For ε3ε3 and ε3ε4, this was not possible, so these SDs were estimated by treating the range as a 99 percentile as using the following formula to calculate the SD:
SD = (upper - lower) / 2 x 2.576
The SDs for ε3ε3 and ε3ε4 were estimated to be 1.94 and 1.75, respectively.

8.2 Van Rijn 2007

This paper reports the sample sizes and the mean WMH lesion volume, per genotype in the text. However, does not report the SDs in the text. A graph is displayed showing the CIs, from which I estimated the SDs:



1mm = 0.09

Genotype	n	Mean	CI ± *	sd†
MM	385	1.09	0.27	2.7
MT	501	1.45	0.23	2.7
TT	158	1.78	0.41	2.6

*data estimated from graph,
† standard deviation calculated from CI and n.

8.3 Hassan 2004

This paper did not report the genotype frequencies for WMH patients. It only reported two Odds Ratios, neither were the Odds Ratio of interest:

OR_1= (WMH & lacunar infarct patients) vs. controls	2.79 (1.36 to 5.7)
OR_2= (lacunar infarct patients) vs. controls	1.79 (0.72 to 4.5)

Control genotype frequencies were reported as follows: TT= 16; TC&CC= 154.
The WMH & lacunar infarct group contained 90 patients. Using trial-and-error, I apportioned these 90 patients to genotype groups to get the OR as close to 2.79 as possible:

	TC & CC	TT	Total
WMH & lacunar infarcts	70	20	90
Controls	154	16	170

OR= 2.75 (1.34 to 5.62)

The lacunar infarct only group contained 52 patients. Using trial-and-error, I apportioned these 52 patients to genotype groups to get the OR as close to 1.79 as possible.

	TC & CC	TT	Total
lacunar infarcts	44	8	52
Controls	154	16	170

OR= 1.75 (0.70 to 4.36)

Therefore the numbers for the WMH+ vs WMH- for lacunar patients is as follows:

	TC & CC	TT	Total
WMH+	70	20	90
WMH-	44	8	52

Appendix 9. ESS data collection forms

NIH Stroke Scale *(Please circle the most appropriate response for each section. See supplementary notes attached. If untestable please state reason. Add the scores for each item to get the total, and do not count untestable items)*

1a Level of Consciousness (LOC)	0 1 2 3	Alert – *keenly responsive* Drowsy – *arousable by minor stimulation to obey, answer, or respond* Stuporous – *requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped)* Comatose – *responds only with reflex motor or autonomic effects or totally unresponsive*
1b LOC Questions	0 1 2	Answers both correctly Answers one correctly Incorrect Patient is asked to state the month & his/her age. No credit for partly correct answers.
1c LOC Commands	0 1 2	Obeys both correctly Obeys one correctly Incorrect Patient is asked to close & open eyes, grip & release normal hand
2. Best Gaze	0 1 2	Normal Partial gaze palsy – *gaze is abnormal in one or both eyes, no forced deviation/total gaze paresis* Forced deviation – *or total gaze paresis not overcome by oculocephalic maneouvre*
3. Visual Fields	0 1 2 3	No visual loss Partial hemianopia or visual inattention Complete hemianopia Bilateral hemianopia – *including cortical blindness*
4. Facial Palsy	0 1 2 3	Normal Minor - *flattened nasolabial fold, asymmetry on smiling* Partial – *total or near total paralysis of lower face* Complete - *absent facial movement in upper and lower face on one or both sides*
5. Best Motor RIGHT ARM	0 1 2 3 4 x	No drift – *holds limb at 90 degrees for full 10 seconds* Drift - *drifts down but does not hit bed* Some effort against gravity No effort against gravity No movement *Untestable (only for amputation or shoulder joint fusion – please state which)*
6. Best Motor LEFT ARM	0 1 2 3 4 x	No drift – *holds limb at 90 degrees for full 10 seconds* Drift - *drifts down but does not hit bed* Some effort against gravity No effort against gravity No movement *Untestable (only for amputation or shoulder joint fusion – please state which)*
7. Best Motor RIGHT LEG	0 1 2 3 4 x	No drift – *holds limb at 45 degrees for full 5 seconds* Drift - *drifts down but does not hit bed* Some effort against gravity No effort against gravity No movement *Untestable (only for amputation or hip joint fusion – please state which)*
8. Best Motor LEFT LEG	0 1 2 3 4 x	No drift – *holds limb at 45 degrees for full 5 seconds* Drift - *drifts down but does not hit bed* Some effort against gravity No effort against gravity No movement *Untestable (only for amputation or hip joint fusion – please state which)*
9. Limb Ataxia	0 1 2 x	Absent Present in 1 limb Present in 2 or more limbs ***Untestable*** *(only for amputation or joint fusion – please state which)*
10. Sensory	0 1 2	Normal Partial loss - *patient feels pinprick is less sharp or is dull on affected side* Dense loss - *patient is unaware of being touched on face, arm, leg*
11. Best Language	0 1 2 3	No dysphasia Mild to moderate dysphasia - *obvious loss of fluency or comprehension, without significant limitation in ideas expressed or form of expression. Conversation about provided material difficult or impossible but examiner can identify items from patient's response.* Severe dysphasia - *all communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener who carries burden of communication. Examiner cannot identify items provided from patient response.* Mute - *no usable speech or auditory comprehension.*
12. Dysarthria	0 1 2 x	Normal articulation Mild to moderate dysarthria - *patient slurs some words, can be understood with some difficulty.* Unintelligible or worse - *speech is so slurred as to be unintelligible (absence of or out of proportion to dysphasia) or is mute/anarthric* *Untestable (intubation or other physical barrier to producing speech – please state)*
13. Neglect	0 1 2	No neglect Partial neglect - *Visual, tactile, auditory, spatial, or personal inattention or extinction to bilateral simultaneous stimulation in one of the sensory modalities* Complete neglect - *Profound hemi-inattention (e.g. does not recognise own hand or orients to only one side of space) or hemi-inattention to more than one sensory modality (e.g. visual + tactile).*

Codes for boxes [__]:	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Lothian Stroke Care Audit / ESS Form – Clinical Inpatient Information

Address label
Chi No. Sex
Unit No. WG
Name Title
Address
Dob
Postcode
Telephone

Date of admission : ____/____/____
Date of assessment : ____/____/____
Time of assessment: ____:____
Consultant in charge : _____

Final diagnosis and status *(Please tick all that apply)*

Cerebral[1]	Stroke (not SAH)	☐	Transient ischaemic attack	☐	Subarachnoid haemorrhage ☐
Eye[1]	Retinal artery occlusion	☐	Transient monocular blindness	☐	
Other	Possibly cerebrovascular[2]	☐	Details: _____		
	Definitely non-cerebrovascular	☐	Details: _____		

Casemix assessment

Was the patient independent in ADL[3] before event ? [__] Are they oriented in time, place and person ? [__]
Was the patient living alone at the time of event ? [__] Can the patient lift both arms off the bed ? [__]
Can the patient talk[4] ? [__] Able to walk without help from another person ? [__]
NIH stroke scale score *(0-42; please complete scoring sheet on back page and see supplementary notes attached)* [_____]

Clinical assessment –presenting event(s) presenting event(s), past history & related signs

Date of onset of symptoms *(or best estimate)* ____/____/____
Time of onset of symptoms *(enter ? if unknown)* ____:____
On aspirin at onset ? [__]
On other antiplatelet drug at onset ? [__]
On warfarin at onset? [__]
Side of brain/eye lesion *(please circle)*
Right / Left / Cerebellar or brainstem / Bilateral / Uncertain
Blood pressure at time of assessment [_____/_____]
Height (cm) [_____] *or* half - armspan[5] (cm) [_____]
Weight (to nearest kg) [_____]

Prior stroke (before presenting event(s)) ?[__]
Prior TIA (before presenting event(s)) ? [__]
History of ischaemic heart disease[6] ? [__]
History of treated hypertension ? [__]
History of diabetes mellitus ? [__]
Peripheral arterial disease[7] ? [__]
Cardiac failure[8] ? [__]
Clear history of atrial fibrillation ? [__]
(paroxysmal or persistent)

[1] Use these categories for **definite** or **probable** (>50% certain) cerebrovascular diagnoses
[2] Use if presentation could have cerebrovascular cause but < 50% certain and give details (e.g. lone vertigo)
[3] Independent in **walking, dressing, washing, feeding,** and **toileting**, not necessarily bathing, shopping or climbing stairs
[4] Able to utter understandable words even if quiet or slurred
[5] Mid-sternal notch to tip of middle finger with (non-paretic) arm outstretched at right angles to body and palm facing forward
[6] MI (including ECG evidence of silent MI) / angina /CABG/coronary angioplasty or stent etc.
[7] History of claudication / rest pain / peripheral arterial intervention, or definite signs (absent foot pulses / femoral arterial bruit)
[8] Definite clinical signs of heart failure or taking at least two drugs for its treatment (eg. ACE-inhibitor and loop diuretic)

Codes for boxes [__]:	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Clinician please complete as much of shaded area as possible at time of first assessment

Clinical assessment - social and family history

Cigarette smoker ? *(please circle)* : Never / Ex>12 m / Current or ex<12 m / Unknown

If current or ex <12 m cigarette smoker, cigarettes / day ? [_____]

Alcohol intake (units/week) [_____]

1st degree relative with stroke / TIA ? [__] Mother / Father / Sibling(s) / Child(ren)
(please circle all that apply) :

If yes, how many 1st degree relatives in total? *(please circle)* : 1 / 2 / >2

1st degree relative with IHD / PAD[9] ? [__] Mother / Father / Sibling(s) / Child(ren)
(please circle all that apply):

If yes, how many 1st degree relatives in total? *(please circle)* : 1 / 2 / >2

Edinburgh Stroke Study information / consent

*(if full consent already available following a previous event, simply write **"CONSENT OBTAINED ALREADY"** across this box)*

Patient / relatives given info pack with consent form? [__]

Consent for ESS: Patient / Relative / Witnessed / Waiver / Refused / None yet
(please circle- see patient information sheet and consent forms for details)

Consent date: ___/___/___

Use of data for research purposes [__] Blood for research [__]

Contact GP / examine medical records [__] Follow-up [__]

Clinical classification of stroke / TIA syndrome[10] *(please circle)*: LACS / PACS / POCS / TACS / Eye[11] / Uncertain

Other risk factors or unusual cause of stroke or TIA ? [__] *(circle any options that apply and give details)*

coronary catheterisation / carotid endarterectomy / cardiac valve disease / haematological illness / hereditary e.g. CADASIL / arterial dissection / coagulopathy / other

Details : _____

Clinical prediction of dependency at six months *(clinician's 'gut feeling' : 0-6 on Oxford Handicap scale[12])* [_____]

Assessing clinician *(please circle)*

MSD / RIL / PAGS / CPW / CLMS / BW / VC / Other *(please initial)* : _____

[9] ischaemic heart disease / peripheral arterial disease
[10] Based on clinical assessment *before* results of imaging or other investigations
[11] Use for transient monocular blindness / retinal artery occlusion
[12] Oxford Handicap Scale:
0 = no symptoms;
1 = minor symptoms which do not interfere with lifestyle;
2 = some restriction to lifestyle but look after themselves;
3 = significant restriction to lifestyle, preventing total independence;
4 = severe handicap preventing independent existence but not requiring constant attention;
5 = severe handicap, totally dependent, requiring attention night and day
6 = dead

Codes for boxes [__]:	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

As much of unshaded area as possible to be completed at (or before) weekly stroke register meeting

Admission blood tests	Taken ? *(please tick)*	Date taken:	Result *(leave blank if not yet known)*	
FBC	☐	___/___/___	Haemoglobin (g/l)	[_____]
			Haematocrit (ratio)	[0●____]
			White cell count ($x10^9$/l)	[____●__]
			Platelets ($x10^9$/l)	[_____]
ESR	☐	___/___/___	(mm/hour)	[_____]
U&E	☐	___/___/___	Urea (mmol/l)	[____●__]
			Creatinine (µmol/l)	[_____]
Glucose	☐	___/___/___	(mmol/l)	[____●__]
Lipids	☐	___/___/___	Total cholesterol (mmol/l)	[____●__]
			HDL cholesterol (mmol/l)	[____●__]
Research	☐	___/___/___		

Brain imaging and final classification

CT done ? [__] Date : ___/___/___ Evidence of new haemorrhage on CT/MRI[13] ? [__]

MRI done ? [__] Date : ___/___/___ Visible relevant infarct on CT/MRI ? [__]

Final syndrome classification : *(using all clinical and imaging information)* LACS / PACS / POCS / TACS / Uncertain / Eye[14]

Cardiac investigations

ECG since event available ? [__] AF on ECG ? [__]

LVH on ECG[15] ? [__]

Echocardiogram done ?*(please circle)*: None / TTE no contrast / TTE+contrast / TOE no contrast / TOE+contrast

Date of first echocardiogram : ___/___/___ LVH on echo ? [__]

Carotid imaging

Carotid Duplex examination performed ? [__] Date of 1st Duplex ___/___/___

2nd Carotid Duplex performed ? [__] Date of 2nd Duplex ___/___/___

MR Angiography performed ? [__] Date of MRA ___/___/___

CT Angiography performed ? [__] Date of CTA ___/___/___

Conventional Angiography performed ? [__] Date of angiography ___/___/___

Carotid imaging results	*Right*	*Left*
ICA % stenosis on Duplex[16] ?	[_____]	[_____]
Post-stenotic collapse (equivalent on Duplex) ?	[__]	[__]
Plaque instability / irregularity (on Duplex or MRA) ?	[__]	[__]

[13] Include harmorrhagic transformation of infarct but **NOT** petechial haemorrhage / microbleeds
[14] Use for transient monocular blindness / retinal artery occlusion
[15] **Don't rely on automatic report.** Use voltage criteria – sum of S wave in V_1 or V_2 + R wave in V_5 or $V_6 \geq 3.5$ mV (35 mm)
[16] Record discrete figure or range. If >1 result, record most severe. If result 'normal' record 0%; if 'minor atheroma' record 30%.

Codes for boxes [__]:	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Unshaded area to be completed by clinician in charge of patient when results become available

Brain imaging and final classification

CT done ? [__] Date : ____/____/____ Evidence of new haemorrhage on CT/MRI[14] ?[__]

MRI done ? [__] Date : ____/____/____ Visible relevant infarct on CT/MRI ? [__]

Final syndrome classification : LACS / PACS / POCS / TACS / Uncertain / Eye[15]
(using all clinical and imaging information)

ICD 10 final diagnosis :
(for definite or probable cerebrovascular diagnoses only) _____

Cardiac investigations

ECG since event available ? [__] AF on ECG ? [__]

LVH on ECG[16] ? [__]

Echocardiogram done ?*(please circle)*: None / TTE no contrast / TTE+contrast / TOE no contrast / TOE+contrast

Date of first echocardiogram : ____/____/____ Patent foramen ovale on echo? [__]

LVH on echo ? [__]

Data to audit carotid intervention service

Carotid Duplex examination performed ? [__] Date of 1st Duplex ____/____/____

2nd Carotid Duplex performed ? [__] Date of 2nd Duplex ____/____/____

MR Angiography performed ? [__] Date of MRA ____/____/____

CT Angiography performed ? [__] Date of CTA ____/____/____

Conventional Angiography performed ? [__] Date of angiography ____/____/____

Referred to vascular surgeons/interventional radiologist ? [__] Date referred ____/____/____

If not referred, why ? (please circle reason): patient choice / clinically not worthwhile (doctors decision)
mutual agreement / not appropriate (no severe stenosis)

If referred – intervention considered *(please circle)*: surgery / angioplasty ± stent

Seen by surgeon / radiologist ? [__] Date seen ____/____/____

Intervention performed ? [__]

If yes Side *(please circle)* Right / Left / Both Date of (first) procedure ____/____/____

Stroke within 30 days of intervention ? [__]

Other complication(s) of intervention ? [__] *(please specify)* _____

Reviewed in NV clinic after intervention ? [__] Date reviewed ____/____/____

Carotid imaging results	*Right*	*Left*
ICA % stenosis on Duplex[17] ?	[_____]	[_____]
Post-stenotic collapse (equivalent on Duplex) ?	[__]	[__]
Plaque instability / irregularity (on Duplex or MRA) ?	[__]	[__]

[14] Include harmorrhagic transformation of infarct but **NOT** petechial haemorrhage / microbleeds
[15] Use for transient monocular blindness / retinal artery occlusion
[16] **Don't rely on automatic report.** Use voltage criteria – sum of S wave in V_1 or V_2 + R wave in V_5 or $V_6 \geq 3.5$ mV (35 mm)
[17] Record discrete figure or range. If >1 result, record most severe. If result 'normal' record 0%; if 'minor atheroma' record 30%.

Codes for boxes [__]:			
	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

CS 06/10/09 *Clinician please complete shaded area in clinic*

LOTHIAN STROKE CARE AUDIT / REGISTRATION FORM – Outpatients

Address label
Chi No. Sex
Unit No. WG
Name Title
Address Dob
Postcode
Telephone

GP Initials_____ GP Surname_____

GP Postcode _____ GP Phone _____

Date of assessment ____/____/____

Responsible consultant _____ Unit : WGH

Date of referral ____/____/____ From GP ? [__]

Date referral received ____/____/____

Date of first appointment offered ____/____/____

Final diagnosis (of presenting event(s)) *(please tick all that apply)*

Cerebral[1] Stroke (not SAH) ☐ Transient ischaemic attack ☐ Subarachnoid haemorrhage ☐

Eye[1] Retinal artery occlusion ☐ Transient monocular blindness ☐

Other Possibly cerebrovascular[2] ☐ Details: _____

Definitely non-cerebrovascular ☐ Details: _____

Complete remainder of form only if definite / probable cerebrovascular diagnosis within last 6 months

Casemix assessment (complete for STROKE PATIENTS ONLY - refers to most recent event)

Was the patient independent in ADL[3] before event ? [__] Are they oriented in time, place and person ? [__]

Was the patient living alone at the time of event ? [__] Can the patient lift both arms off the bed ? [__]

Can the patient talk[4] ? [__] Able to walk without help from another person ? [__]

NIH stroke scale score *(0-42; please complete attached scoring sheet)* [_____]

Clinical assessment – presenting event(s), past history & related signs

Date of most recent stroke / TIA / eye attack ____/____/____ *(or best estimate)*

Number of TIAs (not strokes) in the last 3 months [_____]

Any stroke symptoms lasting > 7 days[5] ? [__]

Side of brain/eye lesion *(please circle)*
Right / Left / Cerebellar or brainstem / Bilateral / Uncertain

Have there been carotid *and* vertebral events ? [__]

Residual neurological signs from presenting event(s) ? [__]

Any symptomatic neck bruit ? [__]

Blood pressure [_____/_____]

Height (cm) [_____]

Weight (to nearest kg) [_____]

Prior stroke - before presenting event(s) ? [__]

Prior TIA - before presenting event(s) ? [__]

History of ischaemic heart disease[6] ? [__]

History of treated hypertension ? [__]

History of diabetes mellitus ? [__]

Peripheral arterial disease[7] ? [__]

Cardiac failure[8] ? [__]

Clear history of atrial fibrillation ? [__]
(paroxysmal or persistent)

[1] Use these categories for **definite** or **probable** (>50% probability) cerebrovascular diagnoses within last 6 months
[2] Use and give details: if < 50% probability cerebrovascular cause (e.g. lone vertigo); or if presenting event(s) not within last 6/12
[3] Independent in **walking, dressing, washing, feeding,** and **toileting**, not necessarily bathing, shopping or climbing stairs
[4] Able to utter understandable words even if quiet or slurred
[5] Only count focal neurological symptoms. If too soon to be sure, please code as unassessable
[6] MI (including ECG evidence of silent MI) / angina /CABG/coronary angioplasty or stent etc.
[7] History of claudication / rest pain / peripheral arterial intervention, or definite signs (absent foot pulses / femoral arterial bruit)
[8] Definite clinical signs of heart failure or taking at least two drugs for its treatment (eg. ACE-inhibitor and loop diuretic)

Codes for boxes [__]:			
	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Clinical assessment - social and family history

Cigarette smoker ? *(please circle)* : Never / Ex>12 m / Current or ex<12 m / Unknown

If current or ex <12 m cigarette smoker, cigarettes / day ? [_____]

Alcohol intake (units/week) [_____]

1st degree relative with stroke / TIA ? [__] Mother / Father / Sibling(s) / Child(ren)
(please circle all that apply) :

If yes, how many 1st degree relatives in total? *(please circle)* : 1 / 2 / >2

1st degree relative with IHD / PAD[9] ? [__] Mother / Father / Sibling(s) / Child(ren)
(please circle all that apply) :

If yes, how many 1st degree relatives in total? *(please circle)* : 1 / 2 / >2

Data to audit use of 2[ary] preventative drugs
(for each column, please tick all that apply or confirm NONE at foot)

Use of following drugs :	At time of event for which referred	At time of first assessment	Recommended following NV assessment	**But** record if patient known not to tolerate
Aspirin	☐	☐	☐	☐
Dipyridamole (Persantin/Asasantin)	☐	☐	☐	☐
Clopidogrel (Plavix)	☐	☐	☐	☐
Warfarin	☐	☐	☐	☐
ACE inhibitor	☐	☐	☐	☐
Diuretic	☐	☐	☐	☐
Other antihypertensive	☐	☐	☐	☐
Statin / lipid lowering agent	☐	☐	☐	☐
NONE	☐	☐	☐	n/a

Edinburgh Stroke Study consent
*(if full consent already available following a previous event, simply write **"CONSENT OBTAINED ALREADY"** across this box)*

Consent obtained for ESS ? *(please circle - see patient information sheet and consent forms for details)* Patient / Relative / Witnessed / None

Consent date : ____/____/____

Use of data for research purposes [__] Blood for research [__]

Contact GP / examine medical records [__] Follow-up [__]

[9] ischaemic heart disease / peripheral arterial disease

Codes for boxes [__]:			
	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Clinician please complete shaded area in clinic

Blood tests	Done prior to clinic *(please 3)*	Taken in clinic *(please 3)*	Date blood taken:	Result *(leave blank if not yet known)*	
FBC	☐	☐	____/____/____	Haemoglobin (g/l)	[_____]
				Haematocrit (ratio)	[0●____]
				White cell count (x10[9]/l)	[____●__]
				Platelets (x10[9]/l)	[_____]
ESR	☐	☐	____/____/____	(mm/hour)	[_____]
U&E	☐	☐	____/____/____	Urea (mmol/l)	[____●__]
				Creatinine (µmol/l)	[_____]
Glucose	☐	☐	____/____/____	(mmol/l)	[____●__]
Lipids	☐	☐	____/____/____	Total cholesterol (mmol/l)	[____●__]
				HDL cholesterol (mmol/l)	[____●__]
Research[10]		☐	____/____/____		
NONE[11]	☐				

Clinical classification
of presenting stroke / TIA syndrome[12] *(please circle)* LACS / PACS / POCS / TACS / Eye[13] / Uncertain

Other risk factors or unusual cause of stroke or TIA ? [__] *(circle any options that apply and give details)*

coronary catheterisation / carotid endarterectomy / cardiac valve disease / haematological illness / hereditary e.g. CADASIL / arterial dissection / coagulopathy / other

Details : _____

Clinical prediction of outcome *(clinician's 'gut feeling')*

Probability of stroke (%) at one year [_____] at five years [_____]

Probability of vascular event (%) *(stroke, MI or vascular death)* at one year [_____] at five years [_____]

Dependency *(for **STROKE PATIENTS ONLY**)* *(0-6 on Oxford Handicap scale[14])* at six months [_____]

Assessing clinician *(please circle)* MSD / RIL / PAGS / CPW / CLMS / BW / Other *(please initial)* : _____

[10] For research bloods, please fill completely: two normal 2.7 ml EDTA tubes (red top). Label with hospital patient stickies. Research samples should be kept in the ice box provided and will be sent to the Wellcome Trust Clinical Research Facility at the end of the clinic.

[11] Tick this box if no blood tests done since referral event(s)

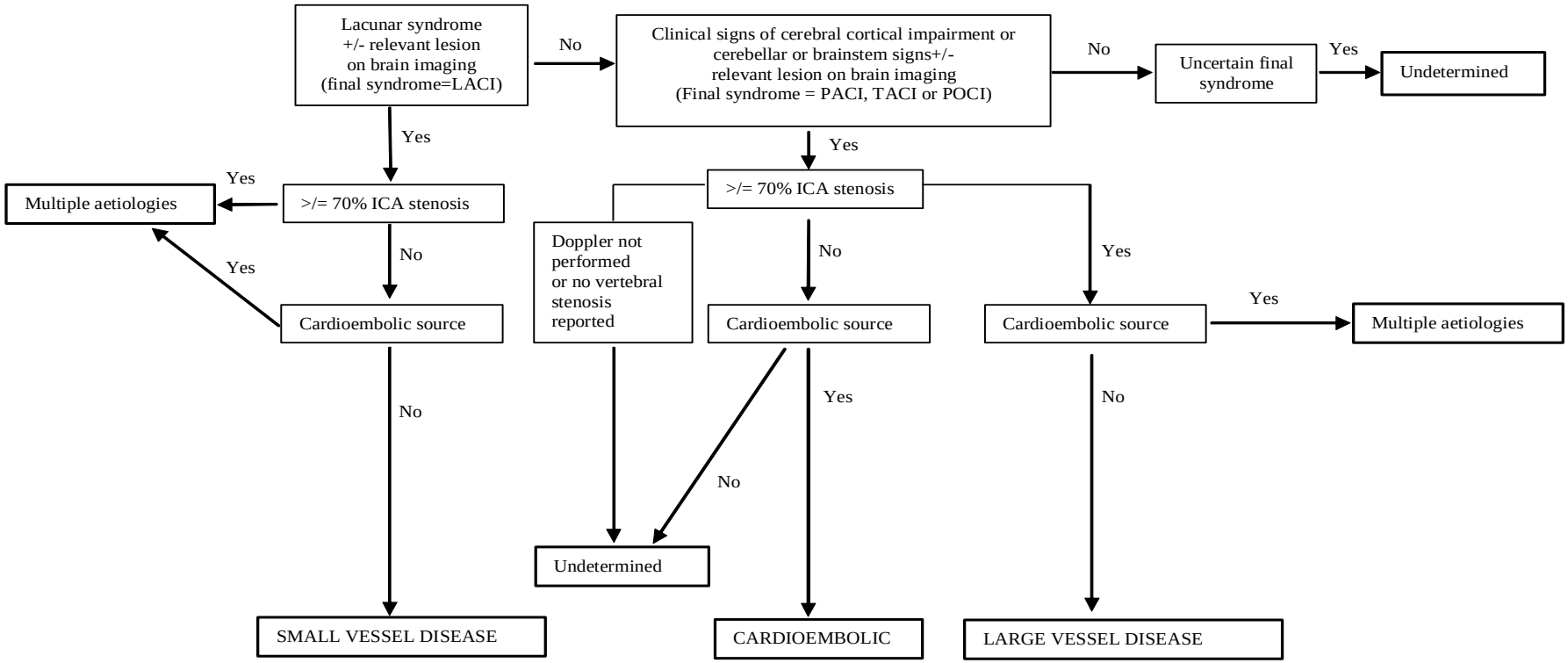
[12] Based on clinical assessment *before* results of imaging / other tests. If patient presents with TIA *and* stroke, classify the stroke.

[13] Use for transient monocular blindness / retinal artery occlusion

[14] Oxford Handicap Scale:
0 = no symptoms;
1 = minor symptoms which do not interfere with lifestyle;
2 = some restriction to lifestyle but look after themselves;
3 = significant restriction to lifestyle, preventing total independence;
4 = severe handicap preventing independent existence but not requiring constant attention;
5 = severe handicap, totally dependent, requiring attention night and day
6 = dead

Codes for boxes [__]:			
	Yes	Y	Wider boxes are for **numbers**
	No	N	Please use ? for unknown
	Unknown	?	**Dates**: please use ?? for unknown, and
	Unassessable	=	complete what you can (e.g. ??/12/1980)

Appendix 10. Modified TOAST algorithm used to assign aetiological ischemic stroke subtype classifications.



LACI = lacunar infarction; PACI = partial anterior circulation infarction; TACI = total anterior circulation infarction; POCI = posterior circulation infarction; ICA = internal carotid artery stenosis. Cardioembolic source defined as history of atrial fibrillation, atrial fibrillation on electrocardiogram or cardiac valve disease.