META-ANALYSIS OF GENETIC ASSOCIATION STUDIES: OVERVIEW OF THE METHODOLOGICAL ISSUES AND PROPOSAL OF GUIDELINES

Thesis submitted for the degree of

Doctor of Philosophy

At the University of Leicester

by

Cosetta Minelli M.D. M.Sc

Department of Health Sciences

University of Leicester

September 2005

UMI Number: U208162

UMI U208162

META-ANALYSIS OF GENETIC ASSOCIATION STUDIES: OVERVIEW OF THE METHODOLOGICAL ISSUES AND PROPOSAL OF GUIDELINES

Cosetta Minelli M.D. M.Sc

Abstract

Genetic association studies have shown a disappointing failure to replicate and validate postulated associations, and evidence synthesis is advocated to overcome the problem of inadequate sample size and help address other methodological issues of primary studies. Despite the fast growth in the number of published meta-analyses of genetic association studies, their methodological aspects have received little attention, and the need for guidelines in this field is demonstrated by the poor quality of many published meta-analyses. This thesis focuses on four specific methodological aspects; 1) how to pool data across genotypes using assumptions about the underlying genetic model; 2) the use of subgroup analysis and the potential for borrowing information on secondary parameters across subgroups, even when the parameter of interest differs; 3) the evaluation of departures from Hardy-Weinberg equilibrium as a proxy for poor quality of primary studies, and how to handle studies with such departures; 4) the role of meta-analysis in Mendelian randomisation, where genetic data are used to derive an unconfounded estimate of the association between a risk factor and a disease. Methods currently used to deal with these issues are reviewed and discussed, and alternative approaches are proposed to overcome their limitations. These approaches are developed using Bayesian methods and are illustrated by re-analysing individual examples, or entire datasets, of published meta-analyses. Use of different methods in specific contexts is further investigated through simulation work. Finally, the findings presented in this thesis are combined with evidence available in the literature to provide a set of guidelines on how to deal with all methodological issues which require consideration in the meta-analysis of genetic association studies. These guidelines are developed at two levels of sophistication, one relatively simple although methodologically correct, the other more sophisticated but more efficient, in order to be of potential use to a wide range of investigators with varying statistical skills.

Acknowledgments

Contents

GLOSSARY

Allele = One of two or more states in which either copy of a gene can exist.

Assortative mating = A tendency for individuals with similar genotypes to mate.

Autosome = Any chromosome other than a sex chromosome. Humans have 22 pairs of autosomes.

Canalisation = The phenomenon for which a phenotype is kept within narrow boundaries even in the presence of disturbing environments or mutations.

Candidate gene = A gene, located in a chromosome region suspected of being involved in a disease, whose protein product suggests that it could be the disease gene in question.

Carrier = An individual who possesses one copy of a mutant allele that causes disease only when two copies are present (recessive mode of inheritance). Although carriers not affected by the disease, two carriers can produce a child who has the disease.

Chromosome = One of 46 double strands of DNA that form the human genome.

Co-dominant = A mode of inheritance in which two different alleles at the same locus are both expressed. Thus, each of the three genotypes has a distinct effect on the phenotype, as opposed to recessive and dominant. Often used to imply equal effect on the phenotype of the two alleles.

Complex disease = The term complex trait/disease refers to any phenotype that does not exhibit classic Mendelian inheritance attributable to a single gene; although they may exhibit familial tendencies (e.g., familial clustering, concordance among relatives). The contrast between Mendelian diseases and complex diseases involves more than just a clear or unclear mode of inheritance. Other hallmarks of complex diseases include known or suspected environmental risk factors; seasonal, birth order, and cohort effects; late or variable age of onset; and variable disease progression.

Co-segregation = A tendency for alleles at two linked loci to be transmitted from parent to offspring together.

DNA = Deoxyribonucleic acid, which makes up genes and chromosomes.

Dominant = A mode of inheritance in which individuals with a single copy of a disease susceptibility allele have the same risk of disease as those with two copies.

Gamete = A germ cell (sperm or ovum) that carries only one of each pair of the parent's genes. Also called reproductive cells or germ cells.

Gene = The fundamental unit of genetic information that is transcribed to form a single protein.

Gene-environment interaction = This term refers both to the modification of genetic risk factors by environmental risk and protective factors and to the role of specific genetic risk factors in determining individual differences in vulnerability to environmental risk factors. When gene-environment interaction is present, a specific environmental change influences the outcome in different ways depending on the genotype. Also called genotype-environment interaction.

Gene interaction = The collaboration of several different genes in the production of one phenotypic character (or related group of characters).

Genetic epidemiology = Genetic epidemiology is the epidemiological evaluation of the role of inherited causes of disease in families and in populations; it aims to detect inheritance pattern of a particular disease, localise the gene and find a marker associated with disease susceptibility. Gene-gene and gene-environment interactions are also studied in genetic epidemiology of a disease. In its broad context, genetic epidemiology includes family studies, molecular epidemiologic studies with genetic components, and more traditional cohort and case-control studies with family history components.

Genetic model = see *Mode of inheritance*

Genome = All the DNA contained in an organism or a cell.

Genotype = The (possibly unobserved) state of an individual's gene at one or more loci relating to the phenotype of interest.

Heterosis = see *Over-dominance* and *Under-dominance*.

Heterozygote = An individual who carries two different alleles for a particular gene.

Heterozygote advantage = Refers to a situation in a population where the heterozygous individuals have higher relative fitness than each of the homozygous individuals. A well established case of heterozygote advantage is that of the gene involved in sickle cell anaemia.

Heterozygote disadvantage = This unusual selection process occurs when heterozygotes are less fit than either homozygote. This situation is likely to arise when two adjacent populations are isolated and become homozygous for different alleles, and then come into secondary contact at the borders of their ranges. Also called heterozygous inferiority.

Homozygote = An individual who carries two identical copies of an allele for a particular gene.

Human Genome Project = An international research project to map each human gene and to completely sequence human DNA
(http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml)

Inbreeding = When two parents have one or more ancestors in common, which results in inbred offspring. Also known as consanguinity.

Linkage = A tendency for genes that are located nearby on the same chromosome to cosegregate; also used to indicate a form of analysis aimed at detecting linkage, linkage analysis.

Linkage analysis = The process of determining the approximate chromosomal location of a gene by looking for evidence of cosegregation with other genes whose locations are already known [i.e., marker genes].

Linkage disequilibrium (LD) = A tendency for certain pairs of alleles at two linked loci to be associated with each other in the population more than would be expected by chance.

Locus = The location on a chromosome of a particular gene (pl. *loci*).

Mendelian randomisation = A natural randomisation process that occurs at conception to determine a person's genotype. It is possible to use 'Mendelian randomisation' to derive an estimate of the association that is free of the confounding and reverse causation typical of classical epidemiology. According to the second law of Mendel (random assignment of genes), the inheritance of one trait is independent of the inheritance of other traits. The distribution of genetic polymorphisms is largely unrelated to the confounders (socioeconomic or behavioural) that distort interpretations of observational epidemiological studies. The basis of Mendelian randomisation is best seen in parent–offspring designs that study the way phenotype and alleles co-segregate during transmission from parents to offspring. This study design is closely analogous to that of randomised clinical trials as by Mendelian principles there should be an equal probability of either allele being randomly transmitted to the offspring. Due to Mendelian randomisation, genetic association studies are less prone to confounding than conventional risk-factor epidemiology, although pleiotropy and linkage disequilibrium can still produce confounding. Mendelian randomisation concept can be used as a tool for epidemiological inference on environmental risk factors by examining the genetic counterpart of a suspected environmental exposure association free of confounding by conventional confounders.

Mode of inheritance = The manner in which a particular genetic trait or disorder is passed from one generation to the next. Autosomal dominant or recessive, X-linked dominant or recessive are examples. Also called genetic model.

Mutation = Any change in the sequence of a DNA molecule, which may result in disease if it occurs in the small fraction of the genome that contains coding sequences, or it may be totally harmless. Mutation forms the basis for the differences between alleles. Some mutations are beneficial and may be the substrate for positive natural selection. The term mutation is sometimes limited to those polymorphisms that have a serious effect on the function of the gene, which are usually rare and deleterious.

Negative assortative mating = Preferential mating between phenotypically different partners.

Outbreeding = The mating of genetically unrelated individuals.

Over-dominance = A relationship in which the phenotypic expression of the heterozygote is greater than that of either homozygote. This is the opposite of underdominance. The term is sometimes used to mean heterozygote advantage. Also called positive heterosis.

Phenotype = The observable trait or disease status that may be influenced by a genotype.

Pleiotropy = The phenomenon whereby a single mutation affects several apparently unrelated aspects of the phenotype.

Polymorphism = A tendency for gene to exist in more than one form or the specific alleles thereof. The term polymorphism is often distinguished from mutation by requiring it to be relatively common (e.g., greater than 1%).

Population stratification = An example of 'confounding by ethnicity' in which the co-existence of different disease rates and allele frequencies within population sub-sections lead to an association between the two at a whole population level. Differing allele frequencies in ethnically different strata in a single population may lead to a spurious association or 'mask' an association by artificially modifying allele frequencies in cases and controls when there is no real association. For this to happen, the subpopulations should differ not only in allele frequencies but also in baseline risk to the disease being studied.

Positive assortative mating = A situation in which like phenotypes mate more commonly than expected by chance.

Product rule = The probability of two independent events occurring simultaneously is the product of the individual probabilities. The rule stating that the probability of the occurrence of independent events is the product of their separate probabilities.

Random mating = Mating between individuals where the choice of partner is not influenced by the genotypes (with respect to specific genes under study). The mating of individuals in a population such that the union of individuals with the trait under study occurs according to the product rule of probability.

Recessive = A mode of inheritance in which individuals with only one copy of the susceptibility allele have the same risk as those with none.

Segregation = The phenomenon in which an individual inherits half of his or her genes from each parent; one of the two fundamental Mendelian principles.

Selection = The process in nature whereby one genotype leaves more offspring than another genotype because of superior life history attributes (fitness) such as survival or fecundity. Also called natural selection.

Somatic cell = Any cell in the body other than germ cells.

Under-dominance = A relationship in which the phenotypic expression of the heterozygote is lower than that of either homozygote. This is the opposite of overdominance. The term is sometimes used to mean heterozygous disadvantage or inferiority. Also called negative heterosis.

Variant = Because of the ambiguity in the definitions of mutation and polymorphism, any genetic change is sometimes called a sequence variation and such alleles are called variant. Also called genetic variant.

X-linked gene = Located on the X chromosome, with a distinctive pattern of inheritance. X-linked (or sex-linked) diseases are generally seen only in males.

1   INTRODUCTION

1.1   Aim of the project

The identification of gene-disease associations represents not only an important step in understanding the pathological basis of diseases in populations, but also one of the main thrusts towards "personalised" medicine. Future therapeutic choices are likely to be based on the subject's genotype (Drews and Ryser, 1997), which might indicate different risks for the disease, degrees of severity or response to specific treatments. Therefore the importance of evaluating gene-disease associations and their impact on future public health is not in doubt. However, genetic association studies tend to show important flaws in their conduct, analysis, or reporting (Bogardus, Concato and Feinstein, 1999; Attia, Thakkinstian and D'Este, 2003), and these deficiencies may contribute to the failure to replicate and validate postulated associations, which represents the most serious issue in this field (Ioannidis *et al.*, 2001). Evidence synthesis of genetic association studies not only may overcome the problem of inadequate sample size, but can also help identify and address some of the other issues of primary studies, for example by investigating possible causes of heterogeneity in study results. However, although meta-analysis of genetic association studies is an area of rapidly growing importance, its methodological aspects have, to date, been researched less intensively than one might have expected (Attia, Thakkinstian and D'Este, 2003).

The aim of this work is to contribute to the development of valid approaches to the meta-analysis of genetic association studies, and the project consists of three elements:

1.   Investigation and overview of the study features that are most problematic when synthesising the evidence from genetic association studies. A review of the approaches that have so far been proposed and a discussion of advantages and disadvantages of the different methods will accompany the presentation of each issue.

2.   Development of alternative techniques aimed at overcoming some of the limitations of currently used methods, either by extending existing methods used in other meta-analysis contexts or by developing new approaches. Practical examples

represented by published meta-analyses are used to illustrate how to implement the methods proposed, to compare them with currently used methods, and to show the potential impact of inappropriate analyses on the estimation of the genetic effect.

3.   Development of guidelines which can be used to advise researchers and research-funders considering a meta-analysis of genetic association studies, as to which elements require consideration prior to the collection, synthesis, analysis and interpretation of the evidence. In order to make the results of the project potentially useful to a wider range of people with varying statistical skills, methods for dealing with each issue are proposed at two different levels of sophistication, one relatively simple and straightforward although methodologically correct, the other more sophisticated but more efficient.

In the following paragraphs background information on genes and diseases (§ 1.2), genetic association studies (§ 1.3) and meta-analysis (§ 1.4) is provided. Justification for the choice of a Bayesian framework to implement the analyses presented is given in § 1.5, where advantages and disadvantages of the Bayesian approach compared to the frequentist approach are discussed. An account of the relevance of meta-analysis of genetic association studies is presented in § 1.6, followed in § 1.7 by a brief presentation of the methodological issues involved. Finally, § 1.8 provides an overview of the contents of the thesis.

1.2   Genes and disease

In all cells of our body the basic genetic information is encoded in genes, which are located on chromosomes comprised of DeoxyriboNucleic Acid, DNA (Strachan and Read, 1999). The major functional endpoints of genetic information stored in *DNA* are proteins, which in turn regulate diverse cellular functions. However, only a small fraction of DNA actually codes for proteins, while the majority of it is non-functional. The *gene* represents the fundamental unit of genetic information, which is transcribed to form a single protein. There are 23 different *chromosomes* in human genome, among which 22 (numbered 1 to 22) are autosomes, and 1 is the sex chromosome (named X and Y). All somatic cells contain two copies of the genome, that is 46 chromosomes including 22 pairs of autosomes and 1 pair of sex chromosomes, which is XX in

females and XY in males. Gametes, i.e. sperm and ova, contain only one copy of each gene, which will be transmitted to offspring to recreate the double genome.

In the human race, DNA shows incredibly little variation, less than 1%. Those few DNA locations, or *loci*, that vary from person to person are called polymorphic, and the different DNA sequences found at a polymorphic locus are called *alleles*. People with two copies of the same allele are said to be *homozygous*, while people with different alleles on the two chromosomes are said to be *heterozygous*. The term *polymorphism* indicates a genetic variant that is relatively common, say present in at least 1% of individuals, while the term *mutation* is usually restricted to indicate a rare variant with a serious deleterious effect on the function of the gene. Although most genes are polymorphic, the majority of polymorphisms are silent, i.e. completely harmless. Only genetic variants occurring in the small fraction of the genome that contains coding DNA sequences may result in an increased risk of disease. The maximum effect of genetic variants is seen in classical Mendelian disorders, where the presence of a genetic mutation can by itself lead to disease, and the parent-offspring transmission follows an inheritance pattern predicted by Mendel's laws. The *mode of inheritance*, also called *genetic model*, determines the way that a trait, such as the disease outcome, is expressed according to the individual genotype. A disease is inherited as *dominant* if one copy of the disease allele is sufficient to cause the disease (i.e. the disease is present in heterozygotes), or *recessive*, if two disease alleles are necessary to cause the disease (i.e. the disease is not present in heterozygotes). Sometimes the disorder can be inherited as *co-dominant*, meaning that the disease is present in heterozygotes but in a form which is less severe than in homozygotes. Apart from Mendelian disorders, genetic polymorphisms can affect the risk of much more common complex diseases, which are not normally considered "genetic" since the genetic factor is only one among the many factors affecting the overall risk of the disease. It is on gene-disease associations of this type that this thesis will focus.

1.3   Genetic association studies

The evaluation of the association between a gene and a disease can be carried out using two categories of study designs, family-based studies and population-based studies, and

are often used in different phases of the evaluation process (Keavney, 2000; Risch, 2000; Thomas, 2004).

Family-based studies are usually carried out in the initial phase represented by gene discovery, and classically consist of studying large families which display a clear segregation pattern for a disease. These families are examined at a number of polymorphic sites, spread over the whole genome at known locations, in order to identify polymorphisms which co-segregate with the disease, i.e. are transmitted through families in a way that parallels the transmission of the disease. Co-segregating polymorphisms, analysed using linkage analysis, provide evidence for the general location of the gene on a chromosome. Linkage analysis is based on the principle that while two genes on different chromosomes are always inherited independently from one another, for two genes on the same chromosome the probability that they are inherited together increases with the decrease of the physical distance between them. This allows indirect location of the disease genes by studying the co-segregation pattern of a number of markers, i.e. polymorphisms for which the locations are known. From the tendency of the genetic marker to co-segregate with the disease, which means that the marker is inherited together with the disease gene, one can infer that the disease gene must be close to that marker.

Population-based studies, also called "genetic association studies" or "molecular genetic studies", are usually performed in the second phase that evaluates the association between a candidate gene and a disease on a population scale, and their study designs are those of conventional epidemiology, most often case-control study. Family-based association studies also exist and are based on the use of offspring trios (family-based trio analysis), where most commonly each trio comprises an affected offspring together with both parents. However, in this thesis we will refer to genetic association studies as to classical population-based association studies.

The main reason for this differential use of types of study in the evaluation of gene-disease associations is that family-based studies based on linkage analysis perform well, and with high efficiency, only in those cases where the disease has a substantial genetic determinant, while they have little power to detect genes with more modest

effects, such as in complex diseases. In these situations linkage studies can only narrow the location of the disease susceptibility gene to a rather large region. On the other hand, genetic association studies are much more resource consuming, due to the higher requirements in terms of sample size, compared to linkage studies, so that their use is usually restricted to the evaluation of genes previously identified as candidates. In the study of complex diseases, genetic association studies based on candidate genes represent the final step to reveal the aetiological relevance of genetic factors to disease, and it is on such studies that this thesis will focus.

Although genetic association studies can be performed using either a cohort or case-control design, the latter is by far the most widely used. The outcome in these studies is represented by traits, which include disease status, such as myocardial infarction; physiological continuous variables, such as blood pressure measurements; and response to environmental stimuli, such as drug efficacy or side effects (Cardon and Bell, 2001). Ideally, the exposure in genetic association studies is represented by the presence in a subject's genotype of a particular polymorphism, where one of the different forms of the gene (alleles) causes a modification of the disease risk through functional impairment of the gene. Thus, genetic association studies require a "candidate gene", that is a gene suspected to be associated with the disease, based on previous knowledge of the functional relevance of the gene products for the disease risk (Day *et al.*, 2001). In the absence of any candidate gene for the disease, a different type of association study, named a "genome-wide" study, has been recently developed, and which aims to map the genes for human diseases by making use of large sets of polymorphic markers throughout the whole genome. Such studies, however, pose the problem of multiple testing due to the great number of markers examined and tests required, so that their role is likely to be relegated to that of hypothesis-forming studies for the investigation of candidate genes, whose findings need to be confirmed using classical genetic association studies (Day *et al.*, 2001).

1.4   Meta-analysis

The synthesis of the evidence available on a specific research topic may be carried out either qualitatively in what is called *systematic review*, or quantitatively through the use

of appropriate statistical methods, which are referred to as *meta-analysis* and provide an overall pooled estimate of the effect of interest. If correctly performed, systematic reviews and meta-analyses represent a powerful tool, which may sometimes provide conclusive evidence on the effect of interest from the data already available. Moreover, evidence synthesis may help avoid misinterpretation of the results of individual studies and provide explanations for the variability, or "heterogeneity", between study results. Studies on the same topic often differ in terms of design, sample size, and other aspects such as characteristics of the study population or outcomes considered, and this might explain why findings can appear to be contradictory. When carrying out a meta-analysis, investigation of heterogeneity is crucial in order to decide what method should be used for pooling study results and, more importantly, whether study results should be combined at all (Thompson, 1994; Higgins *et al.*, 2003). Statistical tests to evaluate the presence of heterogeneity are available, but it has been argued that estimating the magnitude of the observed heterogeneity is more informative than testing alone, and measures aimed at quantifying heterogeneity in meta-analysis have been proposed (Higgins and Thompson, 2002).

Meta-analytical methods can be broadly classified into two types, the fixed effect model and the random effects model, based on the way the variability between study results is treated (Sutton *et al.*, 2000). The *fixed effect* model assumes no heterogeneity, that is all studies estimate the same true underlying effect size with the estimates differing only because of random fluctuation, and the combined effect is calculated as a weighted average of all estimates. When heterogeneity between study results is present, the use of the fixed effect model leads to an inflation of type I error, i.e. the probability of a "false positive" result. In the *random effects* model, a random term for the effect sizes is included in the model to account for the extra variability represented by the heterogeneity. The studies are assumed to estimate different underlying effect sizes, which vary at random usually following a normal distribution. Random effects models can be expressed algebraically as:

$$y_i \sim N(\theta_i, s_i^2) \ , \quad \text{with} \quad \theta_i \sim N(\mu, \tau^2)$$

where y_i is an estimate of effect size, θ_i is the true effect size and s_i^2 is the variance of y_i for study i; μ is the pooled estimate of effect size, and τ^2 is the random effects variance, which represents the between-study variance. When τ^2 is 0, the random

effects model will reduce to the fixed effect model (Sutton *et al.*, 2000). Compared to fixed effect models, random effects models are more conservative and provide wider confidence intervals, and they give relatively more weight to smaller studies than under a fixed effect model. The choice between fixed and random effects models is not straightforward due to the difficulty in assessing the presence of heterogeneity. In fact, the statistical power of tests for heterogeneity is often very low due to the small number of studies included in the meta-analysis, and usually a cut-off significance level higher than the conventional nominal level is used to compensate for this ($p<0.10$ instead of $p<0.05$). Some authors suggest that random effects models might be used in any case, i.e. disregarding the results of the test for heterogeneity, arguing that when low heterogeneity is present the two models will give similar results (Thompson and Pocock, 1991; Sutton *et al.*, 2000). One exception to such strategy might be represented by situations where the number of studies included in the meta-analysis is very low. Hedges and Vevea (1998) suggest that in these situations the random effects model should be regarded as only approximate, due to the difficulty of estimating the between-study variance. As a rule of thumb, these authors suggest a minimum of 5 studies for a reliable use of random effects models. It has to be noted that, although random effects models can accommodate heterogeneity, it is important to investigate possible causes of variability in study results using subgroup analyses or meta-regression (§ 3.2.1). In many situations it may be sensible not to combine studies at all, if they show extreme heterogeneity.

Meta-analyses are mostly performed using the aggregated summary results of the studies, where the combined estimate of the effect is calculated by pooling the estimates from the different studies, usually extracted directly from journal articles. On the other hand, meta-analyses may be carried out on individual patient data, where the full study dataset is requested from the original researchers for each study (Stewart and Tierney, 2002). In this kind of meta-analysis the data coming from the individual patients included in all studies are re-analysed as if they were coming from a large single study, although data for each study are still analysed separately. Meta-analyses on individual patient data are considered of higher quality compared to meta-analyses on aggregated data due to the higher feasibility of checking of the quality of the data, standardising inclusion and exclusion criteria, ensuring the appropriateness of the

analyses, and undertaking subgroup analyses, particularly in specific and relatively rare subgroups. Another key aspect of the use of individual patient data is that it allows heterogeneity to be correctly explained when patient-level covariates are involved (Lambert *et al.*, 2002; see also § 3.7.1). However, meta-analyses on individual patient data are extremely time consuming, have high costs, and rely heavily on the willingness to co-operate of the different research groups who carried out the studies. For these reasons they represent only a minority of all published meta-analyses, and they will not be considered in this thesis.

1.5   Bayesian versus frequentist approach to meta-analysis

The use of Bayesian methods has recently increased in meta-analyses as in many other areas of health care research, mainly as a result of advances in computational methods for carrying out Bayesian analyses (Spiegelhalter, Abrams and Myles, 2004). The basic feature which distinguishes the Bayesian from the frequentist approach is that while the classical statistical analysis of a study is based purely on the data collected, the Bayesian analysis combines the evidence coming from the data (generating what is called the likelihood function) with evidence external to the study and already available about the effect of interest (globally referred to as prior distribution). The external evidence may come from the results of previous studies or purely from the beliefs of experts in the field on the expected effect of interest. The two sources of evidence are combined to provide the updated belief (posterior distribution) about the effect size; in other words the prior belief about the effect size is updated in the light of the evidence coming from the data collected (Figure 1.1).

Two other important features which are exclusive to the Bayesian approach are the possibility to make direct probability statements regarding the effect of interest, such as what is the probability that the odds ratio is greater than 1, and to obtain a predictive distribution of the parameter of interest based on the posterior density, which allows predictive statement about future observations (Spiegelhalter, Abrams and Myles, 2004).

Figure 1.1 – Representation of the Bayesian framework

$$p(\theta \mid Y) = \frac{p(Y \mid \theta) \cdot p(\theta)}{p(Y)}$$
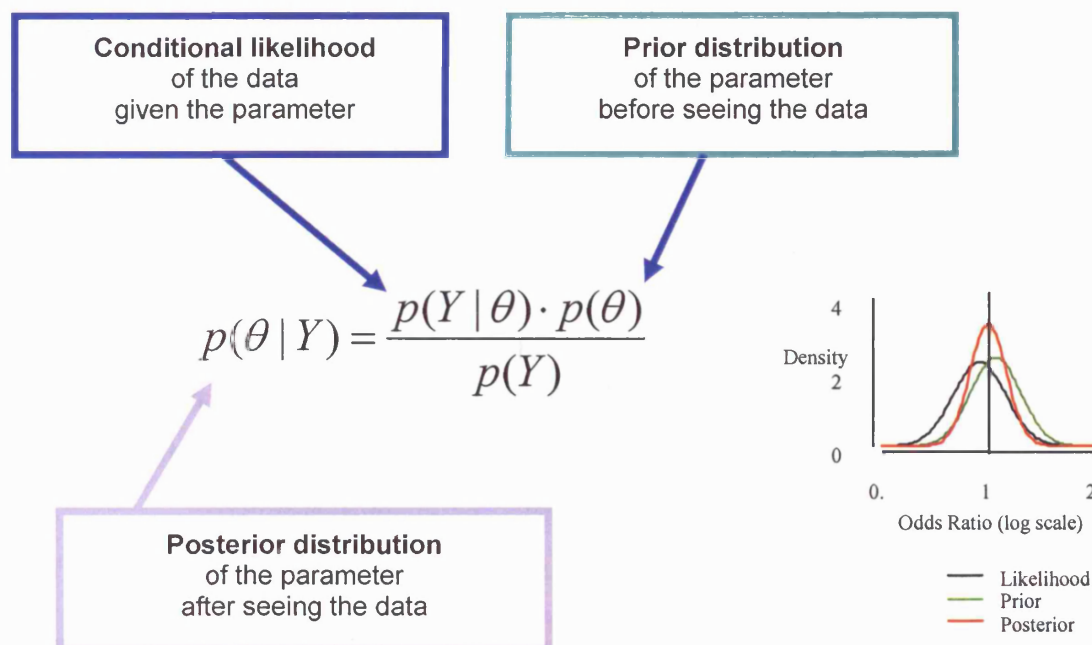
A Bayesian approach to meta-analysis is possible and sometimes desirable in alternative to the classical frequentist methods when dealing with relatively complex models. The random effects meta-analysis is an example of a two-level hierarchical model. By assuming exchangeability between studies, each can "borrow strength" from the others, causing the estimated effects for the individual studies to shrink towards the overall mean. More importantly, the precision of the estimate of that overall mean will allow for the variability between studies and the uncertainty in the between-study variance (Sutton and Abrams, 2001). Although both fixed and random effects models can be implemented in a Bayesian framework, usually random effects Bayesian models are performed. In comparison with conventional frequentist models, the confidence intervals, which in Bayesian terminology are called *credible intervals*, will be slightly wider in a meta-analysis context. The Bayesian approach to hierarchical modelling has been advocated for theoretical and practical reasons. Bayesian analysis allows the researchers to use external information in the form of prior opinion and enables them to produce subjective probability statements about the model parameters (Spiegelhalter, Abrams and Myles, 2004). Eliciting probability distributions from experts in the field under investigation is a delicate task, since people are prone to certain biases in the way they respond to situations involving uncertainty. A number of elicitation methods have

been proposed, which differ in the way of asking questions about uncertain quantities, and some perform better than others in their capability of adequately representing the opinion elicited (Garthwaite, Kadane and O'Hagan, 2005). Fitting Bayesian models by Markov chain Monte Carlo (MCMC) methods, in particular Gibbs sampling (Gelfand and Smith, 1990), is very flexible and makes it practical to handle relatively complex and computationally intensive analyses such as hierarchical models, while allowing for uncertainty in all parameters (Best *et al.*, 1996; Brooks, 1998). Although possible, inclusion of uncertainty for variance parameters in an equivalent likelihood analysis is not straightforward and this source of variability is sometimes ignored (Hardy and Thompson, 1996).

1.5.1   The choice of vague prior distributions

The main practical problem in undertaking a Bayesian meta-analysis is to specify appropriate prior distributions for the model parameters based on external information (Spiegelhalter, Abrams, and Myles, 2004). When expert opinion is available it can be difficult to use this to derive probability distributions, especially for parameters such as variances or correlations (Gokhale and Press, 1982; Garthwaite and Dickey, 1988). When no external evidence is considered worthwhile being taken into account, the prior belief can be represented by a non-informative (or vague) prior distribution, and the results of the Bayesian analysis will be close to those of the frequentist approach. In this case, however, we are left with the equally difficult problem of specifying non-informative prior distributions for all of the model parameters, which include the hyperparameters if the model is hierarchical (Sutton and Abrams, 2001). Although a number of such prior distributions have been proposed, and routinely used, strictly speaking "non-informative" prior distributions, i.e. prior distributions that formally represent ignorance, and thus do not favour any particular parameter values, do not exist (Berger, 1985; Irony and Singpurwalla, 1997). In fact, *any* prior distribution exerts some influence on the posterior distribution, the more so in the presence of sparse data. The real aim is to identify a prior distribution that has minimal effect on the final inference relative to the data (Bernardo and Smith, 1994). For this reason, the term non-informative prior distribution is better replaced by "vague" prior distribution (Kass and Wasserman, 1996; Lambert *et al.*, 2005), which indicates a density that is sufficiently diffuse and gives similar prior probability to a wide but plausible range of parameter

values. The problem of choosing vague prior distributions has been demonstrated to be particularly critical for hierarchical variance parameters, since prior distributions proposed as vague might in fact influence the analysis due to limited data (Spiegelhalter, 2001; Turner, Omar and Thompson, 2001; Gelman, 2004; Lambert *et al.*, 2005). Thus, sensitivity analyses that examine the robustness to the choice of prior distributions are an essential part of a Bayesian hierarchical analysis, including random-effect meta-analysis. Although this issue will be considered throughout all sections of this thesis, sensitivity analysis to the choice of prior distributions will be investigated and discussed in more detail in Chapter 2.

1.5.2 Bayesian implementation of meta-analysis methods

Although the meta-analytical methods presented in this thesis can be implemented using a frequentist approach, the Bayesian framework is an attractive alternative with both philosophical and practical advantages, and for these reasons has been chosen.

Bayesian analysis allows explicit inclusion of prior information on parameters representing the genetic effect, the genetic mode of inheritance and the presence/extent of Hardy-Weinberg disequilibrium, and because of the simplicity of the MCMC algorithm it is easy to implement. Indeed, one of the main driving forces behind the increasing use of Bayesian methods in medical research has been the combination of increased computing power together with the availability of free software, particularly WinBUGS (Spiegelhalter *et al.*, 2004), which is the most widely used statistical package for Bayesian analysis. This software is based on MCMC simulation methods for parameter estimation, in particular the Gibbs sampler, which samples only from the conditional densities in order to obtain the marginal posterior density of each unknown parameter and thus simplifies the process of sampling (Gilks, Richardson and Spiegelhalter, 1996).

Checking whether the Markov chain has converged to its equilibrium distribution is an important issue when using MCMC methods (Gilks, Richardson and Spiegelhalter, 1996; Brooks, 1998). Lack of convergence might be diagnosed in a number of ways. Monitoring of the sample values with visual inspection of the trace plot for a specific chain is often not sufficient, since the chain might be stuck in a particular area due to

the choice of initial values. For this reason, convergence is assessed via sensitivity analyses, by running multiple chains from a diverse set of initial values, and with different length of burn-in and length of sample. Specific tests have also been proposed, which include the Geweke, Heidelberger and Welch, and the Raftery and Lewis diagnostic tests (Cowles and Carlin, 1996) implemented in BOA (Smith, 2004).

All meta-analytical models presented in this thesis were carried out using WinBUGS 1.4.1 (Spiegelhalter *et al.*, 2004), and the WinBUGS codes for the models presented are provided in the Appendix.

1.6   Why is meta-analysis of genetic association studies important?

If in all fields of medical research the volume of scientific knowledge has massively increased with time, this is particularly true for genetic epidemiology. In the last few years, with the completion of the Human Genome Project, the identification of several millions of genetic polymorphisms has provided the opportunity for a huge number of evaluations of possible gene-disease associations (http://www.ncbi.nlm.nih.gov/projects/SNP/). This, coupled with an increased efficiency and cost reduction of genotyping, has resulted in a worldwide outpouring of genetic association studies to such an extent as to create a serious publishing problem for journals and scientists (The Lancet, 2003; Colhoun, McKeigue and Davey Smith, 2003). On a daily basis, one or more new gene discoveries are claimed to be associated with increased risk for some disease, promising to help the diagnosis, prevention or treatment of that condition based on the characterisation of a subject's genotype (Khoury, Little and Burke, 2004).

Unfortunately, studies of the effect of genetic polymorphisms on disease have often shown a disappointing lack of reproducibility (Gambaro, Anglani and D'Angelo, 2000; Cardon and Bell, 2001; Colhoun, McKeigue and Davey Smith, 2003). Two possible reasons for this inconsistency are multiple testing, which can lead to false positive results, and a lack of power, which can lead to false negative results. The problem of false positive results is due to the fact that a large number of different polymorphisms are likely to have been tested on the same biological samples and, by definition when

using a cut-off significance level of 0.05, 5% of them will result in statistically significant findings even in the absence of any genetic effect. On the other side, the issue of the lack of power is also particularly important in population-based genetic epidemiology, which often seeks to identify relatively small effects against a noisy background of biological and social complexity. Most genetic associations have odds ratios in the range of 1.1 to 1.4 (Ioannidis, 2003; Ioannidis, Trikalinos and Ntzani, 2004), and genetic association studies tend to be much smaller than would be ideal, and thus statistically underpowered to detect such relatively small effects (Clayton and McKeigue, 2001; Hirschhorn et al., 2002). Whilst the need for large-scale population-based association studies has recently been recognised, they are rarely available because they require a large number of individuals, are highly resource intensive, and can often be long term in nature (Cardon and Bell, 2001; Colhoun, McKeigue and Davey Smith, 2003). Thus, data from such studies will not be available in the near future, and in the meanwhile data from a relatively large number of small studies is likely to be the sole source of evidence. Under such circumstances, evidence synthesis has the potential to play an important role in advancing biomedical knowledge by increasing the statistical power (Lohmueller et al., 2003).

Heterogeneity in the size of the estimate of the genetic effect, as opposed to its significance, can be due to true variability between studies, that is, variability in excess of that which would be expected due to sampling error. An important role of any meta-analysis is the identification of possible sources of such inconsistency or heterogeneity, using subgroup analyses or the closely related technique of meta-regression (Sutton et al., 2000; see also § 3.2.1). These methods can help assess whether the heterogeneity might be due to systematic differences in study methodology or in the characteristics of the studies' populations.

The important role of evidence synthesis in the evaluation of gene-disease associations has been recognised, as showed by the incredible rate at which the number of published meta-analyses has increased in the last few years. While Attia and colleagues identified only 37 meta-analyses of genetic association studies published between 1991 and 2000 (Attia, Thakkinstian and D'Este, 2003), in March 2005 the HuGE website archive listed 243 meta-analyses published from 2000 onwards (§ 4.4.1). On the other hand, the

appropriate use of meta-analysis of genetic association studies has been researched less than might be anticipated, and the general methodological quality of published meta-analyses of genetic association studies is poor (Attia, Thakkinstian, and D'Este, 2003).

1.7   Meta-analysis of genetic association studies: methodological issues

Although, in general, meta-analysis is considered the strongest source of evidence and thus the ideal basis for medical decision-making, the validity of the results of any meta-analysis depends on the use of appropriate methodological approaches (§ 7.1). The meta-analysis of genetic association studies not only shares those methodological issues common to all applications of biomedical research (Sutton *et al.*, 2000), but also introduces problems that are specific to this field. Some of the general meta-analysis issues play a particularly important role in meta-analyses of genetic association studies. One example is publication bias, the seriousness of which is partly explained by the recent flood of genetic association studies submitted for publication and the consequent publishing problems for journals (The Lancet, 2003; Colhoun, McKeigue and Davey Smith, 2003). In this situation, it can be expected that only studies with "interesting" results, i.e. either statistically significant or with large effect sizes, might be published. However, in this thesis such general issues will only be discussed when insights on them are derived as a by-product from the application of methods proposed to deal with different problems (e.g. § 5.4.3 and 5.6.2), or when providing practical recommendations for the conduct of meta-analysis, as in Chapter 6. The decision to focus this thesis on the specific issues of the meta-analysis of genetic association studies has been taken in recognition of the fact that this is an area where less methodological work has been done, and for which guidelines could be potentially more useful (Attia, Thakkinstian, and D'Este, 2003). The absence of consensus and standardisation of the ways of conducting meta-analyses of genetic association studies are indeed reflected in the number of different approaches used to deal with the problems presented in the following chapters. These approaches are often based on very different assumptions, which only rarely are made explicit by the authors of the study, with a resulting lack of transparency for the reader who cannot critically evaluate them.

Perhaps the most fundamental difference in meta-analysis of genetic association studies is that there are always at least three possible genotype groups to compare, rather than the two treatment groups characteristic of most published meta-analyses in biomedical research, and that comparison must be sensitive to the mode of the gene's effect, or "genetic model" (Attia, Thakkinstian, and D'Este, 2003). The impact of different assumptions about the underlying genetic model on the estimation of the genetic effect and the development of an alternative approach are presented in Chapter 2.

An important issue in any meta-analysis is the identification of possible sources of inconsistency or heterogeneity. Although heterogeneity is often allowed for by using a random effects model, where systematic differences between studies are suspected subgroup analyses may be more appropriate (§ 3.2.1). In genetic association studies such subgroups allow for an interaction between the gene and a study characteristic, such as differences in ethnic origin of the study population. It has to be noted that when patient-level covariates are investigated, individual patient data might be required to correctly explain the observed heterogeneity (Lambert *et al.*, 2002; see also § 3.7.1). Although the issue of subgroup analyses is common in general meta-analysis, genetic association studies are peculiar in the sense that there are other parameters, in addition to the parameter of interest, which subgroups may or may not share. Thus, for instance, the same underlying genetic model could be assumed across subgroups without assuming any similarity in the magnitude of the genetic effect. This issue is dealt with in Chapter 3, where the impact of different assumptions on different parameters is shown, and an approach to the adjustment for study level interaction is proposed.

A specific problem of genetic association studies is the evaluation of Hardy-Weinberg Equilibrium (HWE) in the study population, particularly in the controls. In a large population in which there is a bi-allelic polymorphism, alleles G and g, if the frequency of allele G is p and the frequency of allele g is 1-p, then the frequencies of the three genotypes, GG, Gg, and gg will be p^2, $2p(1$-$p)$, and $(1$-$p)^2$, respectively, if there is random mating and no selection, that is if the population is in HWE. The methodological interest in HWE is in its use as proxy for possible problems associated with the quality of genetic association studies. These problems include; population

stratification, a bias caused by a mixture of different ethnic groups in the study population whenever the frequency of the polymorphism and the disease risk vary between ethnic groups and the study fails to match cases and controls for ethnicity; genotyping error, a mistake in the laboratory identification of a subject's genotype; selection of controls. The issues of how to evaluate departures from HWE and what might be the best strategy for dealing in practice with departures from HWE in studies included in the meta-analysis are presented in Chapter 4.

Finally, an interesting opportunity offered by genetic association studies is that of combining the information from studies evaluating the association between a genotype and a disease, with that from studies evaluating the association between the same genotype and an intermediate phenotype, thought to be on the causal pathway to the disease. This approach, known as "Mendelian randomisation", can be used to assess whether the association between the phenotype (risk factor) and the disease is truly causal, and to derive an estimate of the level of association (Davey Smith and Ebrahim, 2003). The term Mendelian randomisation derives from the fact that a natural randomisation process occurs at conception to determine a person's genotype, so that genotype can be used as an instrumental variable to derive an estimate of the phenotype-disease association which is free of the confounding and reverse causation typical of classical epidemiology. However, the uncertainty in the derived estimate of the phenotype-disease association can be very large, as it depends on uncertainty in both the estimates of the genotype-phenotype and genotype-disease association. Thus, an estimate of the phenotype-disease association which is sufficiently precise is only likely to be obtained through a meta-analysis of all available evidence. Despite the growing interest in Mendelian randomisation, very limited methodological work has been done to date in order to investigate methods aimed not only at providing precise estimates, but also at checking the basic assumptions on which Mendelian randomisation is based. These issues are addressed in Chapter 5, where a novel integrated approach to the meta-analysis of genetic association studies using Mendelian randomisation will be presented.

1.8   Overview of the thesis

Chapter 1 has provided background information on genetic association studies, meta-analysis, and the Bayesian approach, which will be adopted throughout this thesis. The four specific methodological issues of meta-analysis of genetic association studies briefly presented in this chapter are addressed in chapters 2 to 5; the choice of genetic model to pool data across genotypes; the use of subgroup analysis; the role of HWE; the use of Mendelian randomisation. The structure is similar for all of them. The problem is first described in detail; a review of the literature is then presented to illustrate what methods are available to deal with it, and how these methods have been used in published meta-analyses. This overview is accompanied by a discussion of the shortcomings of current methods, which leads to the proposal of novel approaches developed to address some of the limitations identified. The approaches proposed are illustrated by re-analysing either individual examples or entire datasets of published meta-analyses, and the results are compared with those obtained using current methods. Problematic issues encountered are further investigated through simulation work; in particular, simulations are used to evaluate the impact of different assumptions in the Bayesian implementation of the method in Chapter 2, and to compare different strategies for dealing with departures from HWE in Chapter 4. In Chapter 6 the findings presented in this thesis are combined with evidence available in the literature to provide a set of guidelines on how to deal with all issues which require consideration in the meta-analysis of genetic association studies. Finally, an overall discussion of the findings and the need for further work is presented in Chapter 7. An Appendix with the WinBUGS code for the models proposed and other material is also provided, together with a brief Glossary, which contains the definition of all the genetics terms used in the following sections, and an Addenda with papers published in relation with this PhD.

2   THE CHOICE OF GENETIC MODEL

2.1   Chapter overview

This chapter addresses the important issue of how results of genetic association studies can be pooled in a meta-analysis by including data on all genotype groups and taking into account the uncertainty in the underlying genetic model. A brief introduction to the problem is presented in § 2.2, while the different methods that have been used in the literature are reviewed in § 2.3. In § 2.4 an alternative "genetic model-free" approach is proposed with the aim of addressing some important limitations of existing methods, particularly the strong assumptions often made regarding the underlying genetic model. In § 2.5 the approach is applied to five examples of published meta-analyses and the results compared to the results of currently used methods. In § 2.6 the results of the genetic model-free approach based on the commonly used prospective likelihood are compared to those obtained using a retrospective likelihood; the latter reflects the sampling method of case-control studies, which represent by far the majority of genetic association studies. The prospective likelihood is preferred due to the simplicity of its implementation; however, although it has been shown to provide equivalent results to the retrospective likelihood within individual case-control studies, there is no evidence of equivalence in a meta-analysis context. The issue of the choice of vague prior distributions when implementing the genetic model-free approach within a Bayesian framework is illustrated in § 2.7, based on some of the examples; this issue is further investigated through simulation work in § 2.8. Finally, a discussion of advantages and disadvantages of the genetic model-free approach and its Bayesian implementation is presented in § 2.9.

2.2   Introduction

One of the most problematic issues of the meta-analysis of genetic association studies is the choice of the method for pooling study results across genotype groups. In meta-analysis of genetic association studies there are always at least three possible genotype groups to compare, rather than the two groups characteristic of most published meta-analyses in biomedical research, and the comparison must be sensitive to the mode of

inheritance, or genetic model (Attia, Thakkinstian and D'Este, 2003). In the simplest case of a polymorphism with two alleles (G and g), one of which is thought to be associated with a disease (G), association studies will collect information on the relative frequency of disease in subjects with each of the three genotypes (gg, Gg and GG). There are thus two odds ratios, or relative risks, to be estimated, that of GG and Gg, each compared with the wild genotype gg, and, more importantly, these two risks are correlated in a way that depends on the operating genetic model (e.g. recessive, co-dominant, dominant). In particular, the genetic model determines what is the risk of disease in the heterozygous group Gg (§ 1.2). Authors of published meta-analyses of genetic association studies have dealt with this issue in different ways, based on very different assumptions about the underlying genetic model which have not always been explicit, and in many cases unjustified.

2.3   Review of the literature

To date almost all meta-analyses of genetic association studies have reduced the three groups to two by; (a) assuming a recessive model to justify combining the gg and Gg genotypes and comparing gg+Gg with GG; (b) assuming a dominant model and comparing gg with Gg+GG, (c) assuming a per-allele effect that places Gg midway between gg and GG, also called the co-dominant model; (d) ignoring the heterozygotes and comparing gg with GG; (e) performing multiple pairwise comparisons. Although rare, heterosis has also been described (Williams, 1998; Juengel, 2000), where the risk of the Gg group can be higher (positive heterosis or "over-dominance") or lower (negative heterosis or "under-dominance") than either of the homozygous groups. When unsure which genetic model is operating, some investigators fit multiple models and/or perform multiple pairwise comparisons. However, adjustment for such multiple testing is seldom made, and the pairwise estimates of the odds ratio of GG vs. gg (subsequently referred to as OR_{GG}) and the odds ratio of Gg vs. gg (subsequently referred to as OR_{Gg}) are usually obtained by carrying out two separate meta-analyses, thus ignoring the within-study correlation between the two odds ratios induced by the common baseline group.

Among these meta-analysis approaches, the most efficient in terms of precision of the estimate of the genetic effect is that based on the assumption of a specific genetic model operating in all studies included in the meta-analysis (genetic model-based analysis). By collapsing all data from the three genotypes into two groups, this approach enables all available information to be used and has only one parameter to estimate for the genetic effect. Although its statistical efficiency explains why the genetic model-based analysis tends to be the most commonly used, unfortunately the choice of the specific genetic model is usually not supported by any prior knowledge, and assuming the wrong model can lead to biased estimates of the genetic effect. When unsure which genetic model might be operating, some investigators test multiple genetic models or apply both genetic model-based analyses and pairwise comparisons.

A review of 37 published meta-analyses of genetic association studies by Attia *et al.* (Attia, Thakkinstian and D'Este, 2003; see also § 6.2) showed that 24 of them (65%) used a genetic model-based analysis, although a biological justification for the choice of the genetic model was provided in only eight meta-analyses, while in 12 the uncertainty about the underlying genetic model was addressed by testing multiple modes of inheritance. Only 4 of the meta-analyses reviewed by Attia and colleagues avoided having to specify a genetic model by performing separate pairwise analyses, where the two odds ratios OR_{Gg} and OR_{GG} were estimated in two separate meta-analyses. This approach is less efficient than the genetic model-based analysis, since the correlation between the two odds ratios induced by the underlying genetic model is ignored. Overall, 16 meta-analyses performed multiple comparisons by assuming different genetic models or adopting different methods, but only two adjusted for multiple testing. Finally, 9 of the meta-analyses in Attia's review evaluated the gene-disease association solely by comparing the frequency of the allele of interest (G) between cases and controls. Apart from not being directly interpretable from a clinical point of view, given that individuals are characterised by genotypes rather than alleles, this per-allele approach has been criticised since it is based on two implicit assumptions, which are never evaluated before applying the method (Sasieni, 1997). The first assumption is that of a co-dominant genetic model, which means that, on a log scale, the risk of Gg is midway between the risks of gg and GG; the second assumption

is that of Hardy-Weinberg equilibrium, which defines the relation between the allele frequencies and the genotype frequencies (§ 4.2.1).

In this thesis a novel approach is presented, which does not assume prior knowledge on what is the underlying genetic model, but still analyses all genotypes simultaneously by assuming that there is an unknown genetic model which is common across studies.

2.4   Proposed genetic model-free approach

All of the methods of analysis in common use, with the exception of the pairwise comparisons, make the implicit assumptions that a particular genetic model applies in all studies, and, more importantly, that the model is known in advance; for instance, the gene might be assumed to be recessive in all populations. Here a genetic model-free approach to the meta-analysis of genetic association studies is proposed, which also assumes a common genetic model across studies but which does not specify the mode of inheritance in advance. The underlying genetic model is instead estimated from the data. Although no specific genetic model is assumed, the analyses are, of course, still based on an assumed statistical model. The model is based on a simple reparameterisation and uses the odds ratio between the homozygous genotypes (OR_{GG}) to capture the magnitude of the genetic effect, and λ, the ratio of $logOR_{Gg}$ and $logOR_{GG}$, to capture the genetic mode of inheritance. Lambda is assumed to be common across studies, but if this assumption is in doubt then pairwise comparisons obtained using bivariate random-effect meta-analysis methods, which take into account the correlation between OR_{GG} and OR_{Gg}, should be used (van Houwelingen, Arends and Stijnen, 2002; Nam, Mengersen and Garthwaite, 2003). A graphical way of investigating whether the assumption of a common λ is appropriate will be presented. Finally, allowing λ to take any value (unbounded analysis), is equivalent to allowing the possibility of heterosis, i.e. that the risk of the Gg group can be higher or lower than either of the homozygous groups. If this possibility can be excluded on biological grounds then it is better to constrain λ between 0 and 1 (bounded analysis); this restricts the mode of inheritance to the spectrum between dominant, through co-dominant, to recessive.

2.4.1 Model specification

Consider the meta-analysis of a bi-allelic polymorphism, in which G is the risk allele, and a dichotomous disease outcome ascertained for each genotype. Define two parameters; the odds ratio for the two homozygous genotypes, OR_{GG}, and λ, the ratio of $logOR_{Gg}$ and $logOR_{GG}$. In the bounded analysis, although λ can take any value between 0 and 1, values equal to 0, 0.5 and 1 correspond to the recessive, co-dominant and dominant genetic model, respectively. For the unbounded analysis, values of λ greater than 1 or smaller than 0 correspond to positive or negative heterosis.

The $logOR_{GG}$'s are modelled using a random effects meta-analysis model that allows for heterogeneity across studies (Sutton *et al.*, 2000), while λ is modelled as a fixed effect, that is, the genetic model is assumed to be the same in all studies. It is usually not possible to model both $logOR_{GG}$ and λ as random effects because, without extra information, it is very difficult to disentangle the heterogeneity of λ from that of $logOR_{GG}$.

Denoting by y_{0j} and y_{1j} the number of controls and cases, respectively, in the genotype group j, with $j=1,2,3$ (i.e. gg, Gg and GG), the prospective likelihood (L_P) for each study included in the meta-analysis will be:

$$L_P(\alpha, \delta \mid y) = \prod_{j=1}^{3} \prod_{d=0}^{1} \left\{ \frac{\alpha^d \exp(d\delta_j)}{\sum_{k=0}^{1} \alpha^k \exp(d\delta_j)} \right\}^{y_{dj}} \tag{1}$$

where the parameter α is the baseline odds of disease (no exposure), i.e. the odds of disease when $j=1$ (genotype gg), and δ is the log odds ratio of interest (δ_2 for $logOR_{Gg}$ and δ_3 for $logOR_{GG}$).

In the meta-analysis, the full likelihood is then obtained as the product of likelihoods (1) over the i studies, under the assumption of independence of the studies. The study-specific log odds ratios for GG vs. gg, δ_{3i}, are modelled as normally distributed random effects parameters, which vary about an overall mean, θ, with variance, τ^2:

$$\delta_{3i} \sim N(\theta, \tau^2) \qquad i=1, \dots, I$$

The study-specific log odds ratios for Gg vs. gg, δ_{2i}, are equal to the product of δ_{3i} and λ, i.e. $\lambda = \delta_{2i}/\delta_{3i}$, and the mode of inheritance, λ, is assumed constant across studies and thus modelled as a fixed effect parameter. However, if there are reasons to believe that λ differs across populations, the model could be generalised to include subgroups of studies within which λ is constant.

Prior distributions have to be specified for the unknown model parameters θ, τ and λ. A diffuse normal distribution is used for θ in all models, i.e. $\theta \sim Normal(0,10000)$. In the bounded analysis, where heterosis is excluded and λ is restricted to the range between 0 to 1, we need a vague prior distribution constrained to cover the range between 0 and 1, such as $\lambda \sim Beta(0.5,0.5)$. For the unbounded analysis, a normal prior distribution is used, $\lambda \sim Normal(0.5,10)$, where mean and variance are chosen to reflect a wide, but plausible, range of values of λ (i.e. λ between -6 and 7). On the other hand, the vague prior distribution for τ needs to be positive, such as $\tau \sim Uniform(0,2)$. Sensitivity analyses to the choice of prior distributions for λ and τ for the bounded analysis are presented in § 2.7.1.

Prior to model fitting, it may be useful to plot, for each study, the $logOR_{Gg}$ versus $logOR_{GG}$, as shown in Figure 2.1, and in which the slope of the association between $logOR_{Gg}$ and $logOR_{GG}$ represents λ. Such a plot may help check the consistency of λ across studies and identify outlying studies. Study-specific estimates of λ and 95% Credible Intervals (95% CrIs), as shown in Figure 2.2, help assess whether the variation in λ across studies might be explained by sampling error.

If the genetic model does not seem to be consistent across studies then it may be better to perform joint pairwise comparisons using a general bivariate meta-analysis model (Nam, Mengersen and Garthwaite, 2003) that does not assume that λ is common but still takes into account the correlation between the estimates of OR_{GG} and OR_{Gg}. The model is specified as follows:

$$r_{ij} \sim Bin(\pi_{ij}, n_{ij}) \qquad i=1, \dots, I \qquad j=1, 2, 3$$

where r_{ij} is the number of cases for genotype j in study i, which is binomially

distributed with n_{ij} being the total number of subjects for genotype j in study i.

The log odds of disease for each genotype is thus:

$$\log\left(\frac{\pi_{i1}}{1-\pi_{i1}}\right) = \mu_i$$

$$\log\left(\frac{\pi_{i2}}{1-\pi_{i2}}\right) = \mu_i + \delta_{i2}$$

$$\log\left(\frac{\pi_{i3}}{1-\pi_{i3}}\right) = \mu_i + \delta_{i3}$$

where: $\delta_{i2} \sim N(\theta_1,\tau_1^2)$ and $\delta_{i3} \sim N(\theta_2,\tau_2^2)$

The parameterisation of the model, i.e. the presence of μ_i in the three log odds,

accounts for the correlation between the two estimated log odds ratios. It will be

interesting for future work to extend the model to allow for correlated random effects,

that is correlated true log ORs rather than correlated estimates of the ORs, which might

be more appropriate. However, it is likely that there might be little information to

estimate such correlation, so that the impact of the choice of a prior distribution for the

correlation might heavily influence the results.

The same prior distributions used for the genetic model-free approach are adopted here;

$\theta_1 \sim Normal(0,10000)$, $\theta_2 \sim Normal(0,10000)$, $\tau_1 \sim Uniform(0,2)$

and $\tau_2 \sim Uniform(0,2)$

The number of simulations was varied and the traces were inspected for evidence of

non-convergence before deciding on a burn-in of 10,000 iterations, followed by chains

of length 50,000. Convergence was assessed via sensitivity analyses with respect to

initial values, length of burn-in and length of sample, using visual inspection of trace

plots (§ 1.5.2). Details of the WinBUGS code for the genetic model-free approach and

the bivariate meta-analysis model for the joint pairwise comparisons are given in

Appendix 1.

2.5   Illustration

2.5.1   Five examples

The implementation of the genetic model-free approach and its comparison to the conventional methods are illustrated using five published examples of the meta-analysis of genetic association studies. For each meta-analysis, the number of studies included, frequency of the risk allele, methods used by their authors and main reported results, are given in Table 2.1.

ACE gene and diabetic nephropathy

This meta-analysis was carried out to evaluate the controversial association of the I/D polymorphism of the *ACE* gene with diabetic microangiopathy, in particular nephropathy and retinopathy (Fujisawa *et al.*, 1998). Here we consider only the meta-analysis assessing the effect on nephropathy. In the original analysis a dominant model was assumed and 21 studies were pooled to give an odds ratio of 1.32 (95% CI: 1.15 to 1.51). The average allele frequency for the genetic variant was 0.46.

KIR6.2 gene and Type II diabetes

The K+ inwardly rectifier channel (KIR) is a protein that plays a major role in glucose-stimulated insulin secretion. Its encoding gene, *KIR6.2*, has been suggested as a candidate for inherited defects in Type II diabetes. This meta-analysis was carried out assuming dominant, recessive and co-dominant models with p-values corrected for multiple testing (Hani *et al.*, 1998). The result of the meta-analysis, based on four studies, was a significant association between *KIR6.2* and Type II diabetes. The average frequency for the risk allele was 0.34.

AGT gene and essential hypertension

The genetic variant Thr235 of the angiotensinogen (*AGT*) gene has been found to be associated with a predisposition to hypertension in some linkage and association studies. This meta-analysis of seven Japanese case-control studies reported an odds ratio for the Thr235 allele of 1.22 (95% CI: 1.05 to 1.42), with an average allele frequency of 0.75 (Kato *et al.*, 1999).

MTHFR gene and coronary heart disease

The 677C. T is a polymorphism of the *Methylene TetraHydroFolate Reductase* (*MTHFR*) gene involved in folate metabolism, which causes elevated homocysteine levels and has been associated with an increased risk of coronary heart disease. This meta-analysis of 49 studies reported an odds ratio of 1.21 (95% CI: 1.06 to 1.39) for the TT vs. CC comparison (Wald, Law and Morris, 2002), in close agreement with another meta-analysis published around the same time (Klerk *et al.*, 2002). The average frequency for the T allele was 0.32.

PON1 Q192R polymorphism and myocardial infarction

PON1 is one of the genes encoding for paraoxonase, a serum enzyme that has been implicated in the prevention of atherogenesis and coronary heart disease through its association with high-density-lipoprotein particles. This recent meta-analysis of 19 studies investigated the effect of the Q192R polymorphism in the *PON1* gene on the risk of myocardial infarction (Wheeler *et al.*, 2004). The reported per-allele relative risk was 1.08 (95% CI: 1.02 to 1.14), and the average allele frequency was 0.33.

TABLE 2.1 - Five published meta-analyses used for illustration, with methods and results reported in the original articles

Author, year	Association evaluated	Number of studies	Risk allele frequency	REPORTED ANALYSIS	
				Method	Results
Fujisawa, 1998	*ACE* gene and diabetic nephropathy	21	0.46	Assumed dominant genetic model	1.32 (1.15 to 1.51)
Hani, 1998	*KIR6.2* gene and Type II diabetes	4	0.34	Only *p* value, under dominant and recessive genetic models	Dominant: $p < 0.05$ Recessive: $p < 0.01$
Kato, 1999	*AGT* gene and essential hypertension	7	0.75	Per-allele analysis	1.22 (1.05 to 1.42)
Wald, 2002	*MTHFR* gene and coronary heart disease	49	0.32	Heterozygotes ignored, pairwise comparison for OR_{GG}	1.21 (1.06 to 1.39)
Wheeler, 2004	*PON1* Q192R polymorphism and myocardial infarction	19	0.33	Per-allele analysis	1.12 (1.15 to 1.51)

2.5.2 Results

Figure 2.1 shows, for each meta-analysis, a plot of $\log OR_{Gg}$ against $\log OR_{GG}$. All meta-analyses show variation in the genetic effect as represented by the two log odds ratios.

This might be explained by a number of factors, including sampling error, differences in the study methods and differences in the true genetic risk across study populations. In the absence of heterogeneity in the genetic model and sampling error, all studies would be expected to lie along a straight line with slope λ. The solid line in Figure 2.1 represents the slope, λ, estimated by the genetic model-free approach using a bounded analysis, while the three dotted lines corresponding to the dominant, co-dominant and recessive genetic models are plotted for comparison. The figure allows visual identification of any outliers or influential studies. Figure 2.2 plots the study-specific estimates of λ and their 95% CrIs, and is used to investigate whether any departures from linearity in Figure 2.1 are consistent with sampling error. Within individual studies, λ is often poorly estimated, but there is little indication in any of the meta-analyses that the genetic models are not common across studies.

Table 2.2 summarises the results for the different meta-analytical methods in common use, which consist of separate pairwise comparisons, where $logOR_{Gg}$ is pooled independently of $logOR_{GG}$, and methods based on assumed genetic models. The results are expressed in terms of median and 95% CrI. The result for the *ACE* example when assuming a dominant model (Table 2.2) differs from the published result, which also assumed a dominant model (Table 2.1), because the main result in the original paper was based on a fixed effect meta-analysis rather than the random effect meta-analysis used here, and the analysis was implemented using a classical approach. The use of a random effects model and the implementation within a Bayesian framework both contribute to explain why the 95% CrI of the estimate in Table 2.2 is sensibly wider than the corresponding 95% CI in Table 2.1 (see § 1.4 and 1.5). The choice of the genetic model in model-based methods can have a marked impact on the estimates of OR_{GG} and OR_{Gg}. For instance in the *KIR6.2* example the estimates of OR_{GG} vary between 1.41 (95% CI: 0.73 to 2.79) and 1.93 (95% CI: 0.94 to 3.96). Separate pairwise comparisons give an unbiased estimate of OR_{GG} of 2.21, but with an unnecessarily wide confidence interval (95% CI: 1.07 to 4.55) because they do not incorporate any of the information on OR_{Gg} when estimating OR_{GG}.

FIGURE 2.1 - Plot of the log OR$_{Gg}$ against the log OR$_{GG}$ for: a) *ACE* gene and diabetic nephropathy; b) *KIR6.2* gene and Type II diabetes; c) *AGT* gene and essential hypertension; d) *MTHFR* gene and coronary heart disease; e) *PON1* Q192R polymorphism and myocardial infarction. The solid line represents the slope λ estimated by the genetic model-free approach; the three dotted lines correspond to the dominant, co-dominant and recessive genetic models respectively
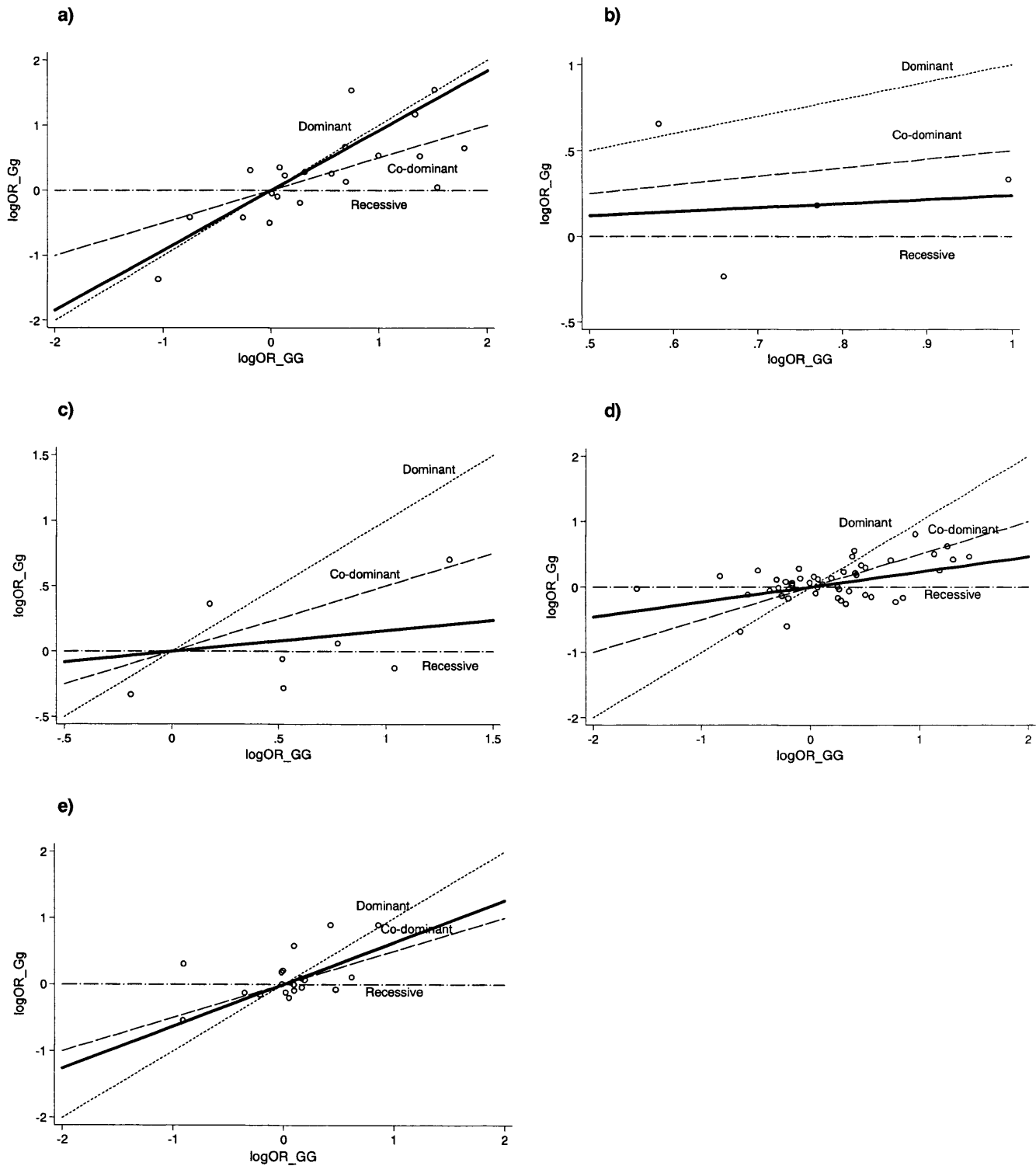
FIGURE 2.2 - Plot of the study-specific estimates of λ (with 95% CrI) for: a) *ACE* gene and diabetic nephropathy; b) *KIR6.2* gene and Type II diabetes; c) *AGT* gene and essential hypertension; d) *MTHFR* gene and coronary heart disease; e) *PON1* Q192R polymorphism and myocardial infarction. To better investigate the region in the middle, where the two lines correspond to the recessive and dominant models, the 95% CrIs have been truncated at ±5

a)

b)

c)

d)

e)
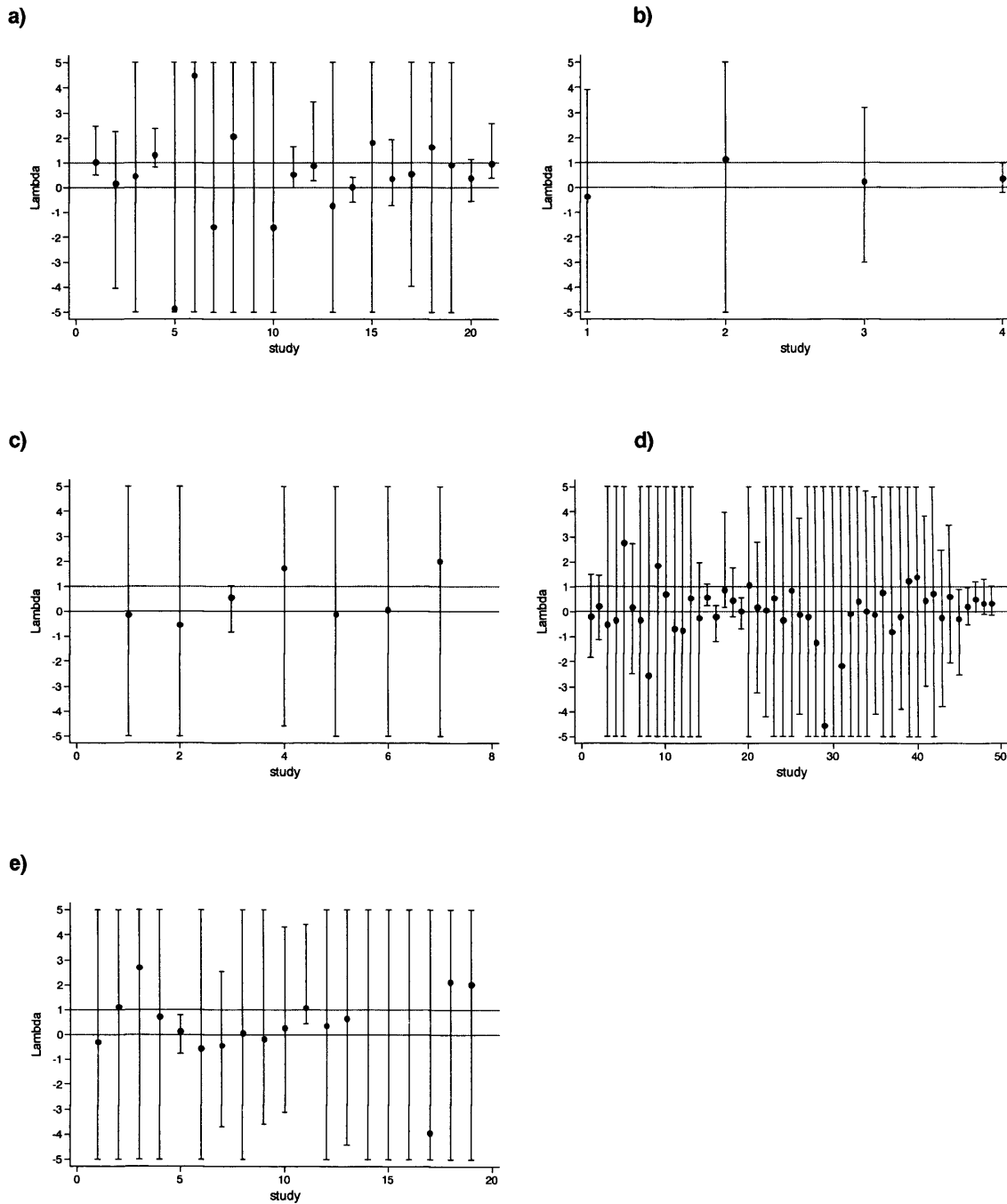
TABLE 2.2 - Results of the currently used meta-analytical methods for the five meta-analyses

Meta-analysis	Method	OR$_{Gg}$ (95% CrI)	OR$_{GG}$ (95% CrI)	Implicit λ
ACE gene and diabetic nephropathy	Separate pairwise comparisons	1.24 (0.92 to 1.69)	1.48 (1.08 to 2.11)	-
	Recessive model	-	1.21 (1.02 to 1.49)	0
	Co-dominant model	1.21 (1.04 to 1.43)	1.46 (1.08 to 2.04)	0.5
	Dominant model	-	1.31 (0.99 to 1.77)	1
KIR6.2 gene and Type II diabetes	Separate pairwise comparisons	1.24 (0.92 to 1.69)	2.21 (1.07 to 4.55)	-
	Recessive model	-	1.94 (0.97 to 3.87)	0
	Co-dominant model	1.39 (0.97 to 1.99)	1.93 (0.94 to 3.96)	0.5
	Dominant model	-	1.41 (0.73 to 2.79)	1
AGT gene and essential hypertension	Separate pairwise comparisons	1.14 (0.67 to 1.98)	1.68 (0.98 to 3.04)	-
	Recessive model	-	1.66 (1.06 to 2.78)	0
	Co-dominant model	1.49 (1.04 to 2.23)	2.21 (1.09 to 4.99)	0.5
	Dominant model	-	1.44 (0.88 to 2.41)	1
MTHFR gene and coronary heart disease	Separate pairwise comparisons	1.06 (0.99 to 1.13)	1.19 (1.03 to 1.38)	-
	Recessive model	-	1.16 (1.02 to 1.32)	0
	Co-dominant model	1.09 (1.02 to 1.16)	1.18 (1.04 to 1.34)	0.5
	Dominant model	-	1.08 (1.01 to 1.16)	1
PON1 Q192R polymorphism and myocardial infarction	Separate pairwise comparisons	1.10 (0.98 to 1.25)	1.13 (0.95 to 1.33)	-
	Recessive model	-	1.09 (0.94 to 1.26)	0
	Co-dominant model	1.08 (1.01 to 1.16)	1.17 (1.02 to 1.35)	0.5
	Dominant model	-	1.11 (1.00 to 1.24)	1

Table 2.3 presents the results of the genetic model-free approach with λ unbounded and bounded between 0 and 1, and the joint pairwise comparisons. The pooled estimates of λ obtained from the genetic model-free approach tend not to be very precise, but can usually rule out some of the commonly assumed genetic models. For example, the *KIR6.2* gene and the *ACE* gene examples rule out the dominant and recessive models respectively, whilst the *MTHFR* gene example suggests that λ is different from any of the values corresponding to the standard genetic models. In the example of the *ACE* gene the estimate of λ is very close to one, that is, close to dominant. Compared to an assumed dominant model the estimate of OR$_{GG}$ is very similar, but the confidence interval is wider reflecting our uncertainty surrounding whether the true underlying model is actually dominant. The bounded analysis, in which λ must lie between 0 and 1, did not materially alter the point estimates of any of the parameters in our examples

TABLE 2.3 - Results of the proposed genetic model-free approach and the joint pairwise comparisons, obtained using bivariate meta-analysis, for the five meta-analyses

Meta-analysis	Method		OR_{Gg} (95% CrI)	OR_{GG} (95% CrI)	λ (95% CrI)
ACE gene and diabetic nephropathy	Genetic model-free approach	Unbounded .	1.30 (0.99 to 1.74)	1.33 (0.99 to 1.84)	0.95 (0.64 to 1.39)
		Bounded .	1.31 (1.01 to 1.73)	1.35 (1.01 to 1.87)	0.92 (0.65 to 1.00)
	Joint pairwise comparisons		1.21 (0.94 to 1.57)	1.42 (1.10 to 1.91)	-
KIR6.2 gene and Type II diabetes	Genetic model-free approach	Unbounded .	1.19 (0.87 to 1.81)	1.97 (0.94 to 4.12)	0.27 (-0.25 to 0.85)
		Bounded .	1.16 (1.00 to 1.77)	2.01 (0.97 to 4.10)	0.24 (0.00 to 0.82)
	Joint pairwise comparisons		1.24 (0.61 to 2.63)	2.21 (1.08 to 4.46)	-
AGT gene and essential hypertension	Genetic model-free approach	Unbounded .	1.03 (0.69 to 1.71)	1.65 (1.03 to 3.51)	0.07 (-1.59 to 0.53)
		Bounded .	1.08 (1.00 to 1.74)	1.81 (1.05 to 3.66)	0.16 (0.00 to 0.54)
	Joint pairwise comparisons		1.10 (0.63 to 1.84)	1.79 (1.04 to 3.34)	-
MTHFR gene and coronary heart disease	Genetic model-free approach	Unbounded .	1.04 (1.00 to 1.11)	1.20 (1.04 to 1.38)	0.25 (0.03 to 0.47)
		Bounded .	1.04 (1.00 to 1.11)	1.19 (1.04 to 1.38)	0.23 (0.01 to 0.46)
	Joint pairwise comparisons		1.06 (0.99 to 1.12)	1.19 (1.04 to 1.36)	-
PON1 Q192R polymorphism and myocardial infarction	Genetic model-free approach	Unbounded .	1.08 (0.99 to 1.21)	1.13 (0.97 to 1.31)	0.65 (-1.06 to 2.55)
		Bounded .	1.08 (1.00 to 1.21)	1.15 (1.01 to 1.33)	0.63 (0.03 to 1.00)
	Joint pairwise comparisons		1.10 (0.99 to 1.23)	1.14 (0.98 to 1.33)	-

but affected their interval estimates. For instance, in the *AGT* example, where the fitted model is very close to recessive, the restriction on λ implies that OR_{Gg} cannot fall below 1.00 as this would either require a negative λ or a protective effect of the GG genotype; the bound rules out the former and the data contradict the latter.

2.6 Retrospective versus prospective likelihood

By far the majority of genetic association studies use a case-control design that requires a retrospective likelihood based on the probability of exposure given disease, since this reflects their method of sampling. Prentice and Pyke (1979) showed that a maximum likelihood analysis based on the corresponding prospective likelihood gives the same results as an analysis of the retrospective likelihood for a single study. The advantage of using the prospective likelihood is that the outcome variable, disease, is binary, whereas in the retrospective analysis the outcome, exposure, can have many levels. In the case of genetic association studies, the exposure, i.e. genotype, has three categories even in the simplest case of a bi-allelic polymorphism. Although the equivalence of the

two likelihoods for fixed effect meta-analyses follows from the analogy with a stratified case-control study, there is no reason to suppose that exactly equivalent results will be obtained with more complex hierarchical models. None-the-less where the heterogeneity is small or the data are not sparse we might expect the results to be similar. Equivalence within the Bayesian framework does not generally exist and has only been established for very particular choices of prior distributions (Seaman and Richardson, 2004). Although not exactly equivalent, the results of Prentice and Pyke would suggest that with vague prior distributions the retrospective and prospective Bayesian analysis should give similar answers. This issue is investigated in the following sections.

2.6.1 Model specification for retrospective likelihood

In the meta-analysis based on a retrospective likelihood, which mirrors the method of sampling in case-control studies, subjects are selected dependent on their disease status and then their exposure status is ascertained.

Denoting by y_{0j} and y_{1j} the number of controls and cases, respectively, in the genotype groups j, with j=1,2,3 (corresponding to gg, Gg and GG), the retrospective likelihood (L_R) for each study included in the meta-analysis is derived from a pair of multinomial distributions;

$$y_{0j} \sim Multinomial(n_0, p_{0j}) \qquad\qquad y_{1j} \sim Multinomial(n_1, p_{1j})$$

where n_0 and n_1 are the total number of controls and cases respectively,

$$p_{dj} = \frac{\beta_j \exp(d\delta_j)}{\sum_{k=1}^{3} \beta_k \exp(d\delta_k)} \qquad j=1, 2, 3$$

and d is an indicator of the disease status, taking the value of 0 for controls and 1 for cases. The probability that a control has exposure j is $\beta_j / \sum_{k=1}^{3}\beta_k$, with $\beta_1 = 1$. The log odds ratios of disease for the exposure groups Gg and GG compared to no exposure (gg) are represented by δ_2 and δ_3 respectively, while δ_1 is zero by definition. The likelihood for each study will thus take the form:

$$L_R(\beta, \delta \mid y) = \prod_{d=0}^{1} \prod_{j=1}^{3} \left\{ \frac{\beta_j \exp(d\delta_j)}{\sum_{k=1}^{3} \beta_k \exp(d\delta_k)} \right\}^{y_{dj}} \qquad (2)$$

In the meta-analysis the full likelihood is obtained as the product of the likelihoods (2) over the i studies, assuming that the studies are independent. As in the prospective meta-analysis (§ 2.4.1), the study-specific log odds ratios δ_{3i} are modelled as normally distributed random effects parameters, with an overall mean θ and between-study variance τ^2. The underlying study-specific log odds ratios δ_{2i} are again derived as the product of δ_{3i} and λ, and λ is modelled as bounded between 0 and 1 (bounded analysis). A diffuse prior normal distribution is used for θ, $\theta \sim Normal(0,10000)$, while $Uniform(0,2)$ and $Beta(0.5,0.5)$ are used for τ and λ, respectively.

Again, the number of simulations was varied and the traces were inspected for evidence of non-convergence before deciding on a burn-in of 5,000 iterations followed by chains of length 10,000. Convergence was assessed via sensitivity analyses with respect to initial values, length of burn-in and length of sample, and using visual inspection of trace plots (§ 1.5.2). Details of the WinBUGS code for fitting this model are given in Appendix 1.

2.6.2 Application and comparison

The similarity between models based on retrospective and prospective likelihood was investigated by applying them to the five meta-analyses described in § 2.5.

Table 2.4 shows the results for the retrospective and the prospective models, which are nearly identical for all meta-analyses and parameters of interest, both in terms of the point estimates (medians) and the width of the credible intervals. Different prior distributions for λ and τ gave similar results. The only difference in the two approaches was a tendency to a slower convergence for the prospective models, even after the use of hierarchical centring (see Appendix 1) in the specification of the prospective models (Gelfand, Sahu and Carlin, 1995). This is the reason why the results for the prospective models were based on longer burn-in (10,000 instead of 5,000) and chain length (50,000 instead of 10,000), compared to the retrospective models.

Table 2.4 – Comparison of the results of the genetic model-free approach applied to the five examples using a retrospective and a prospective likelihood

Meta-analysis	Likelihood	OR_{Gg} *(95% CrI)*	OR_{GG} *(95% CrI)*	λ *(95% CrI)*	τ *(95% CrI)*
ACE gene and diabetic nephropathy	Retrospective	1.31 (1.01 to 1.72)	1.36 (1.01 to 1.83)	0.92 (0.64 to 1.00)	0.54 (0.31 to 0.87)
	Prospective	1.31 (1.01 to 1.73)	1.35 (1.01 to 1.87)	0.92 (0.65 to 1.00)	0.54 (0.32 to 0.88)
KIR6.2 gene and Type II diabetes	Retrospective	1.16 (1.00 to 1.80)	2.03 (0.96 to 3.96)	0.24 (0.00 to 0.79)	0.29 (0.02 to 1.51)
	Prospective	1.16 (1.00 to 1.77)	2.01 (0.97 to 4.10)	0.24 (0.00 to 0.82)	0.31 (0.02 to 1.49)
AGT gene and essential hypertension	Retrospective	1.09 (1.00 to 1.71)	1.83 (1.06 to 3.60)	0.17 (0.00 to 0.54)	0.56 (0.24 to 1.37)
	Prospective	1.08 (1.00 to 1.74)	1.81 (1.05 to 3.66)	0.16 (0.00 to 0.54)	0.56 (0.24 to 1.43)
MTHFR gene and coronary heart disease	Retrospective	1.04 (1.00 to 1.11)	1.19 (1.04 to 1.38)	0.23 (0.01 to 0.46)	0.33 (0.19 to 0.50)
	Prospective	1.04 (1.00 to 1.11)	1.19 (1.04 to 1.38)	0.23 (0.01 to 0.46)	0.33 (0.19 to 0.50)
PON1 Q192R polymorphism and myocardial infarction	Retrospective	1.08 (1.00 to 1.21)	1.15 (1.02 to 1.34)	0.62 (0.03 to 1.00)	0.11 (0.01 to 0.33)
	Prospective	1.08 (1.00 to 1.21)	1.15 (1.01 to 1.33)	0.63 (0.03 to 1.00)	0.11 (0.01 to 0.33)

2.7 Sensitivity analyses to the choice of vague prior distributions

When adopting a Bayesian approach to a hierarchical model that allows heterogeneity in OR_{GG}, unless there is prior knowledge, we have the problem of specifying vague prior distributions for the between-study variance of $logOR_{GG}$ (τ) and for the parameter λ. The sensitivity of the analysis to such choice is investigated for three of the examples considered, in particular for the meta-analyses on *KIR6.2* gene and Type II diabetes (Hani), *AGT* gene and essential hypertension (Kato), and *PON1* Q192R polymorphism and myocardial infarction (Wheeler). The three examples differ in the number of studies included, frequency of the risk allele, estimate of λ, and estimate of the between-study variance. In the following models only analyses based on a bounded λ are considered.

2.7.1 Prior distributions for τ and λ

Three prior distributions were considered for the between-study standard deviation of $logOR_{GG}$, τ, and two for the parameter λ. Figure 2.3 shows the densities for the

different priors on both parameters, all presented on the standard deviation scale.

Prior distributions for τ

The value of the random effects standard deviation τ can be very influential in assessing the uncertainty about the parameter of interest. Because of the limited information in the data, the prior distribution for τ becomes particularly important, and yet there is no generally accepted reference prior for this parameter (Spiegelhalter, Abrams, and Myles, 2004; see also § 1.5.1). A number of suggestions have been made, among which the most popular are the following three reference prior distributions.

The first is a gamma distribution for the precision parameter (the inverse of the variance),

$$\frac{1}{\tau^2} \sim Gamma(0.001, 0.001)$$

This corresponds to an inverse-gamma distribution on the between-study variance, and has a "spike" of probability mass close to zero after which it rapidly decreases. Although this is perhaps the most commonly used vague prior distribution for the heterogeneity parameter, it has been recently criticised and prior distributions on the standard deviation parameter have been recommended, as they are more directly interpretable (Gelman, 2004; Spiegelhalter *et al.*, 2004).

The second prior for the standard deviation, τ, is a standardised half-normal distribution truncated at zero,

$$\tau \sim Half - Normal(0,1) \qquad \tau > 0$$

This prior distribution gives a low probability to values greater than 2.

Finally, the third prior distribution considered is a uniform distribution over the range 0 to 2, which excludes the possibility that the standard deviation can be over 2,

$$\tau \sim Uniform(0,2)$$

It is interesting to think of what these distributions mean in terms of average values of τ and what the different values of τ imply in terms of variability of the odds ratio. For example, a *Half-Normal(0,1)* for τ means a median τ of 0.76, which means that the

ratio of the 97.5% odds ratio to the 2.5% odds ratio is about 20, i.e. $\exp(3.92\tau)$, and this

roughly represents the range of odds ratios. In the same way, a *Uniform(0,2)* means a

median τ of 1 and a ratio of 50 (Spiegelhalter, Abrams and Myles, 2004).

Figure 2.3 – Density plots for the prior distributions considered for; a) τ; b) λ. In a) the densities for the three priors are all presented on the standard deviation scale; the gamma distribution has been re-scaled in order to show the shape of the distribution

a)
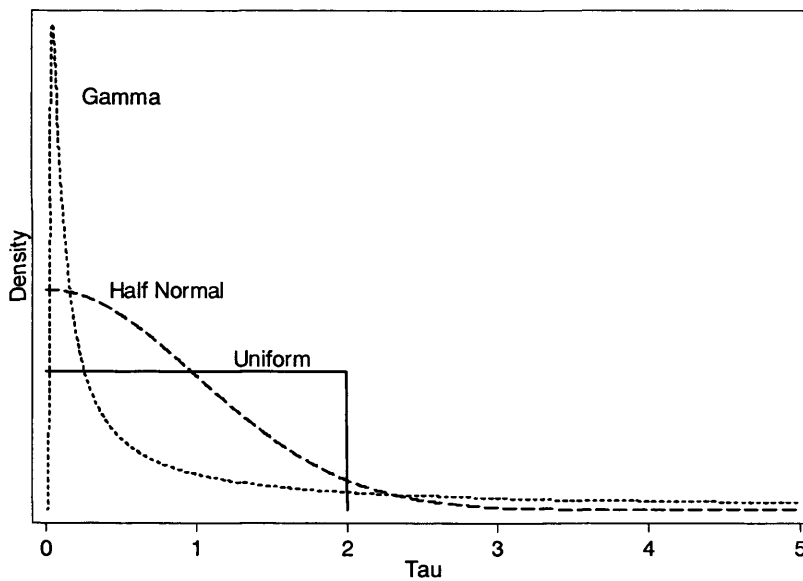
b)
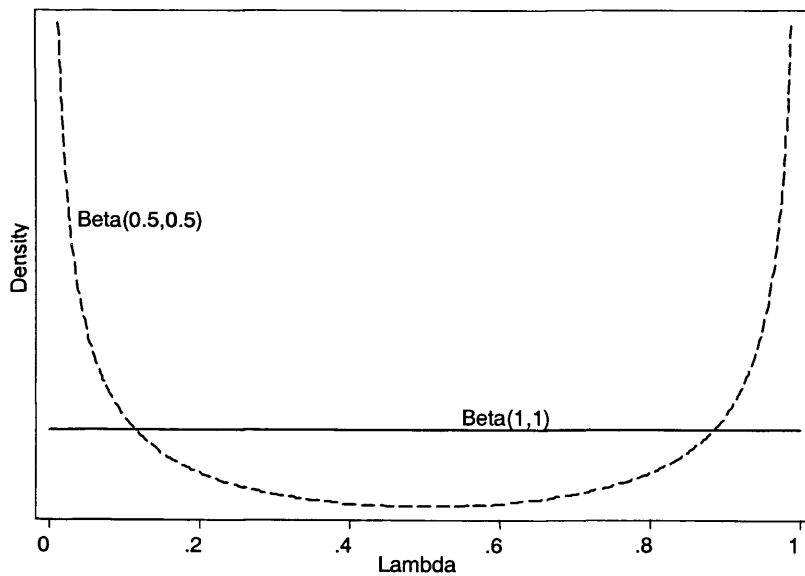
Prior distributions for λ

For the bounded λ, two beta prior distributions are considered over the range between 0 and 1, and which have been used for modelling vague prior beliefs about proportions (Spiegelhalter, Abrams and Myles, 2004).

The first prior distribution is a beta distribution with both parameters equal to one,

$\lambda \sim Beta(1,1)$

This distribution is uniform between 0 and 1. However, when parameters have values very close to the extremes 0 or 1 and the data are sparse, this prior will tend to pull the results towards the middle of the distribution. For instance, for a near recessive model when the true value of λ is very close to 0 the prior will tend to distort the results because it gives 90% prior probability to values greater than 0.1.

The second prior distribution is a beta distribution with both parameters equal to 0.5,

$\lambda \sim Beta(0.5, 0.5)$

which corresponds to a Jeffrey's prior for a binomial likelihood. This distribution gives greater prior probability to values of λ close to the extremes (Box and Tiao, 1973), i.e. to models that are close to recessive or dominant. However, if the model is actually close to co-dominant, i.e. $\lambda=0.5$, and the data are sparse, this distribution may tend to inflate the uncertainty surrounding λ.

2.7.2 Results of sensitivity analyses

The results for the 6 combinations of prior distributions for τ and λ for the meta-analyses of Hani, Kato and Wheeler are illustrated in Figure 2.4a, 2.4b, and 2.4c, respectively. Point estimates (medians) and 95% CrI of the four parameters of interest, OR_{Gg}, OR_{GG}, λ and τ, are plotted for each model.

The gamma distribution, with its spike close to zero, tends to produce lower estimates of τ with narrower credible intervals, which in turn tends to be reflected in the widths of the credible intervals for the odds ratios. This is particularly pronounced in Hani's meta-analysis, where the data are particularly sparse because there are only 4 studies. Here the estimate of τ is 53% and 51% lower with the gamma prior compared with the

Figure 2.4 – Plots of the results for OR_{GG}, OR_{Gg}, λ and τ, obtained by applying models with different prior distributions to the three meta-analyses; a) Kato (n=7); b) Hani (n=4); c) Wheeler (n=19)

a)
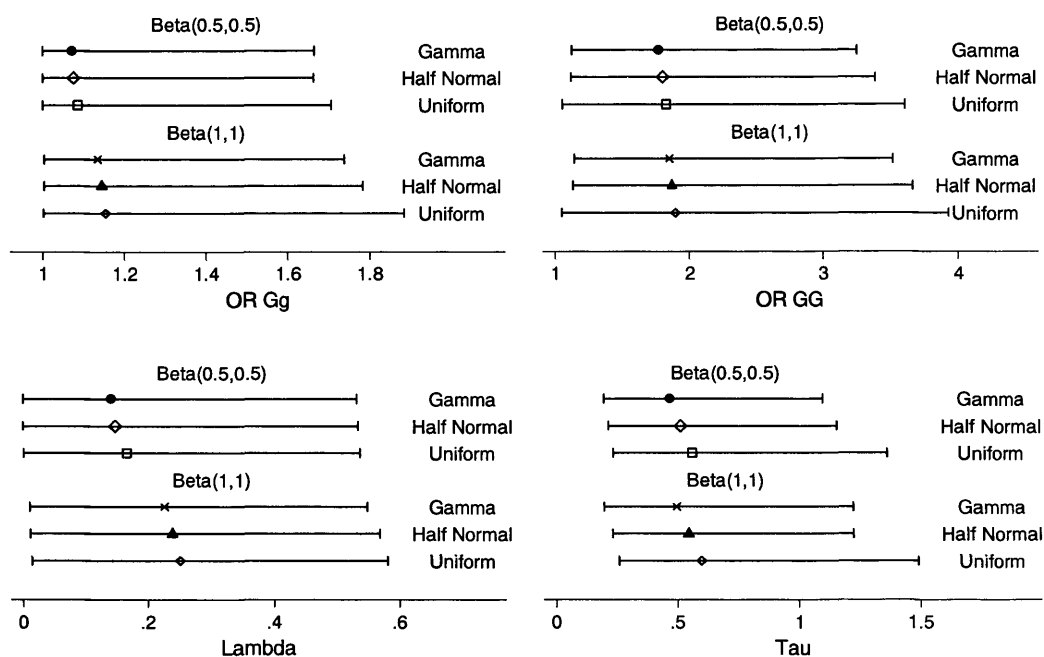
b)
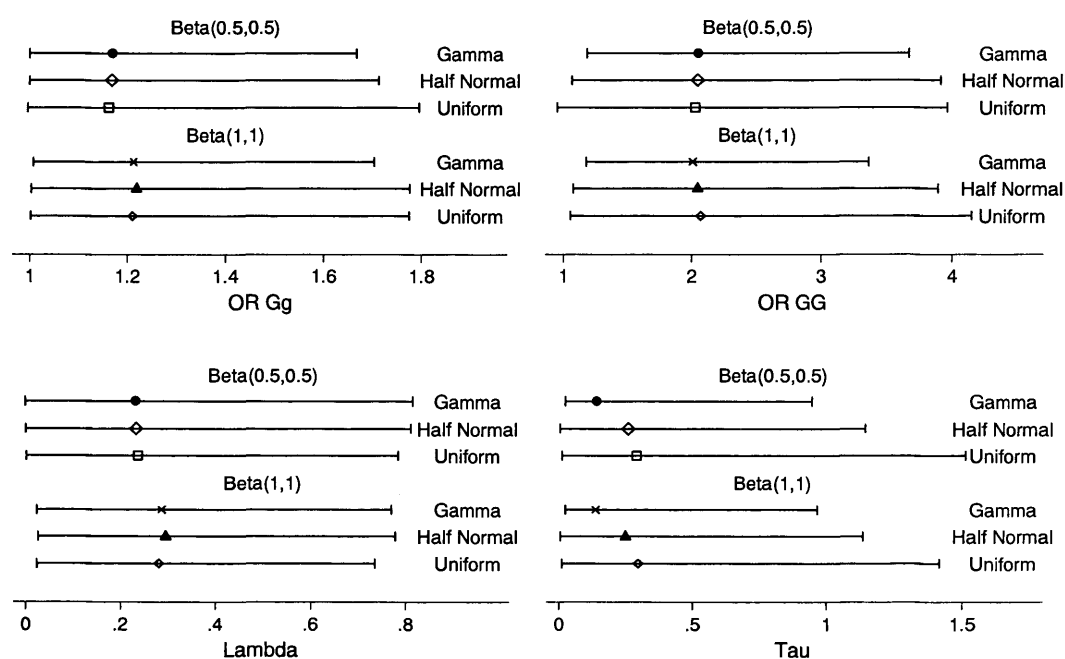
c)
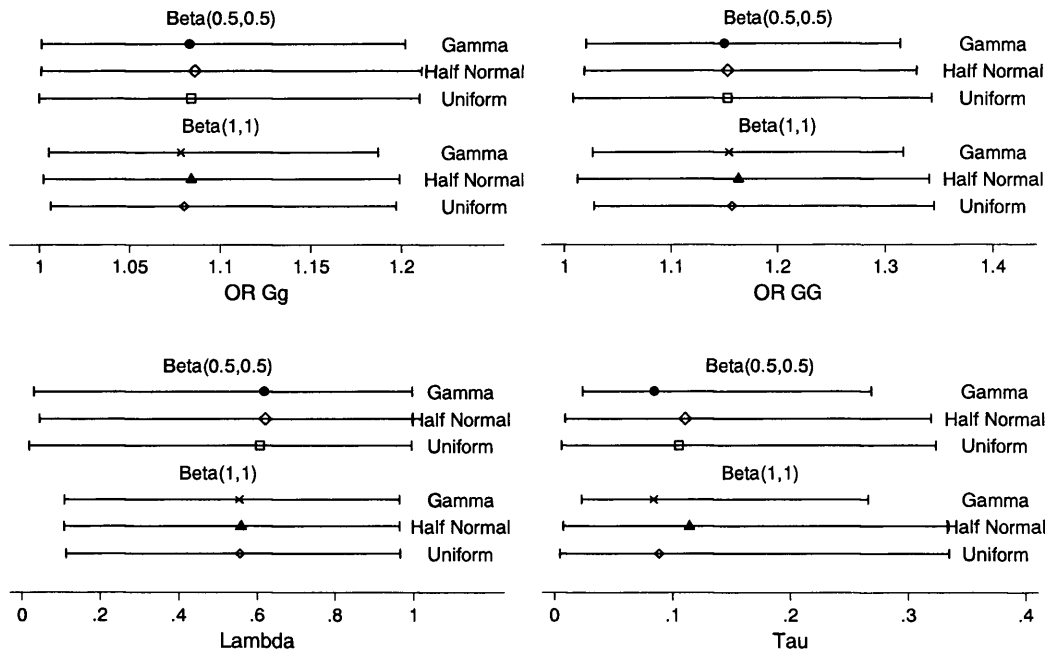
uniform, and the credible interval is 33% and 38% narrower, for *Beta*(1,1) and *Beta*(0.5,0.5) respectively.

The impact of the two different beta prior distributions for λ varies according to the meta-analysis. As expected the *Beta*(0.5,0.5) tends to pull the point estimates for λ towards the extremes, i.e. 0 and 1, and the *Beta*(1,1) tends to provide more precise estimates of λ, when λ is near 0.5, as in Wheeler.

2.8 Simulations to evaluate the effect of different vague prior distributions

2.8.1 Setting

Simulation work was undertaken in order to further investigate the sensitivity of the parameter estimates to the choice of vague prior distributions for τ and λ, in situations were the true values were known. Simulated datasets were created based on the three meta-analyses described in § 2.7, that is *KIR6.2* gene and Type II diabetes, *AGT* gene

and essential hypertension, and *PON1* Q192R polymorphism and myocardial infarction. The total number of studies and study sizes were kept the same as in the original meta-analyses, while values for the model parameters were taken from the profile maximum likelihood approach previously used to analyse these data (Minelli, Thompson *et al.*, 2005b – included in Addenda). For each of the three meta-analyses, 1,000 new datasets were randomly generated and each was analysed in WinBUGS with different prior distributions as described in § 2.7.1. It was not possible to check the convergence of all 18,000 analyses, so datasets that gave a large discrepancy in results when analysed with different prior distributions were selected, and for these convergence was checked by running longer chains (burn-in 50,000, chain length 100,000) with different starting values. In all cases the results confirmed the original analyses.

The median of the corresponding MCMC simulations was taken as the point estimate for each of the four parameters OR_{Gg}, OR_{GG}, λ and τ. The medians from the analyses with the different priors were compared in terms of their mean, their Root Mean Square Error (RMSE), and the coverage of the 95% CrIs, that is the percentage of intervals that included the true value. These three measures describe the average properties of the estimators across the 1,000 datasets.

2.8.2 Results

For the four parameters OR_{Gg}, OR_{GG}, λ and τ, the mean, RMSE, and coverage of the 95% CrIs are summarised in Table 2.5. For the scenarios based on Wheeler's and Kato's meta-analyses, the number of datasets effectively analysed was in fact 995 and 998 respectively, since a few simulated meta-analyses contained studies with 0 cells for *both* cases and controls in a genotype group, and the MCMC algorithm did not converge.

In all cases the half-normal and uniform prior distributions caused the heterogeneity, τ, to be over-estimated on average, although only in the case of the Hani-based simulations was the RMSE also appreciably larger. The beta prior distributions for λ caused the average estimate of λ to move towards 0.5, the more so in the presence of

Table 2.5 – Results of the sensitivity analyses to different prior distributions for λ and τ, for the simulated meta-analyses based on each of the 3 scenarios; a) Kato; b) Hani; c) Wheeler. RMSE = Root Mean Square Error

a)

Parameter		OR_{Gg}			OR_{GG}			Lambda			Tau		
Assumed value		1.077			1.770			0.130			0.460		
Statistics		Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Prior for λ	*Prior for τ*												
Beta(0.5,0.5)	Gamma	1.134	0.092	99.50	1.914	0.348	95.39	0.210	0.119	99.20	0.433	0.199	94.59
	Half-normal	1.144	0.098	99.30	1.931	0.357	97.39	0.222	0.126	99.00	0.511	0.174	96.59
	Uniform	1.152	0.104	99.60	1.945	0.366	98.10	0.234	0.134	98.50	0.564	0.210	94.89
Beta(1,1)	Gamma	1.186	0.118	98.09	1.979	0.386	93.88	0.273	0.148	97.39	0.454	0.209	94.48
	Half-normal	1.195	0.126	97.89	1.995	0.397	96.58	0.283	0.157	96.48	0.536	0.186	95.48
	Uniform	1.203	0.133	97.69	2.007	0.405	97.39	0.292	0.166	95.68	0.593	0.227	93.78

b)

Parameter		OR$_{Gg}$			OR$_{GG}$			Lambda			Tau		
Assumed value		1.177			2.030			0.230			0.134		
Statistics		Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Prior for λ	*Prior for* τ												
Beta(0.5,0.5)	Gamma	1.170	0.111	99.60	2.057	0.373	99.30	0.248	0.135	98.09	0.206	0.076	99.60
	Half-normal	1.164	0.110	99.90	2.058	0.378	99.80	0.245	0.133	98.30	0.361	0.227	97.60
	Uniform	1.161	0.110	99.90	2.062	0.382	99.90	0.244	0.133	98.20	0.436	0.302	96.19
Beta(1,1)	Gamma	1.208	0.097	99.60	2.064	0.373	99.20	0.293	0.117	98.90	0.206	0.076	99.50
	Half-normal	1.203	0.096	99.90	2.067	0.380	99.80	0.291	0.116	98.70	0.363	0.229	97.10
	Uniform	1.200	0.095	99.90	2.072	0.381	99.90	0.289	0.115	98.60	0.435	0.301	95.60

c)

Parameter		OR$_{Gg}$			OR$_{GG}$			Lambda			Tau		
Assumed value		1.091			1.150			0.620			0.088		
Statistics		Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Prior for λ	*Prior for* τ												
Beta(0.5,0.5)	Gamma	1.078	0.040	94.55	1.154	0.052	97.58	0.553	0.192	97.27	0.095	0.025	99.19
	Half-normal	1.076	0.041	94.74	1.155	0.053	97.88	0.54	0.195	96.26	0.113	0.040	98.69
	Uniform	1.076	0.041	94.43	1.154	0.053	97.87	0.539	0.196	96.46	0.114	0.040	98.79
Beta(1,1)	Gamma	1.078	0.037	94.75	1.16	0.054	97.17	0.532	0.153	97.98	0.097	0.026	99.19
	Half-normal	1.077	0.037	95.86	1.16	0.054	97.98	0.521	0.158	97.47	0.116	0.041	98.59
	Uniform	1.077	0.037	95.35	1.16	0.054	98.08	0.521	0.159	97.58	0.117	0.042	98.59

sparse data. This behaviour is caused by the constraint that λ must lie between 0 and 1, and the choice of symmetrical prior distributions such as the *Beta*(0.5,0.5) or the *Beta*(1,1). Such a situation is illustrated by the Kato-based simulations where the true value of λ is 0.13, so that under-estimates had to lie between 0 and 0.13 while over-estimates could lie between 0.13 and 1, and the average consequently tends to be too high. Under these circumstances the mean, or corresponding bias, is not an appropriate indicator of the quality of the estimator. A better indicator is the RMSE which favours the *Beta*(0.5,0.5) prior distribution when λ is close to 0 or 1, and the *Beta*(1,1) when λ is close to 0.5. On average, the odds ratios are relatively insensitive to the choice of prior distributions.

Good average performance is reassuring but may not be a reliable guide to the sensitivity to the choice of prior distributions for any particular single dataset. For this reason Bland-Altman style plots, originally described as a way to assess agreement between two methods of clinical measurement (Bland and Altman, 1986), were used in order to graphically evaluate the difference in results when comparing different prior distributions in all 1,000 datasets (Figure 2.5). The difference in estimates based on any two prior distributions is plotted against the average of the two estimates. Plots for the Kato-based simulations are shown in Figure 2.5 and include a line drawn at the mean difference. Two dotted lines are drawn at the mean difference plus and minus 1.96 times the standard deviation of the difference, in order to both quantify the difference that can be observed when using different prior distributions on the same dataset, and detect patterns in the difference which are related to the size of the parameter estimate. Plots for the other two scenarios showed similar results, as shown in Appendix 1, figures A and B.

Figure 2.5b shows the effects of the different prior distributions on the estimate of OR_{GG} in individual datasets generated under the conditions of the Kato meta-analysis. The posterior estimates of OR_{GG} with different prior distributions for τ are usually very close, mostly within ± 0.1 for estimates that are rarely over 3, and the agreement tends to be better in datasets where the posterior estimates of the odds ratio is close to 1. However on rare occasions the difference can be as large as 0.3 when the average estimate is 3, a 10% difference. Unfortunately there seems to be no way of distinguishing in advance if the prior distributions will have a large impact on a particular simulated dataset. If a 10%

Figure 2.5 – Bland-Altman plots of the difference in the estimates for Kato meta-analysis obtained by models with different prior against their average value for; a) OR_{Gg}; b) OR_{GG}; c) λ; d) τ.

a)

b)

c)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

d)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)
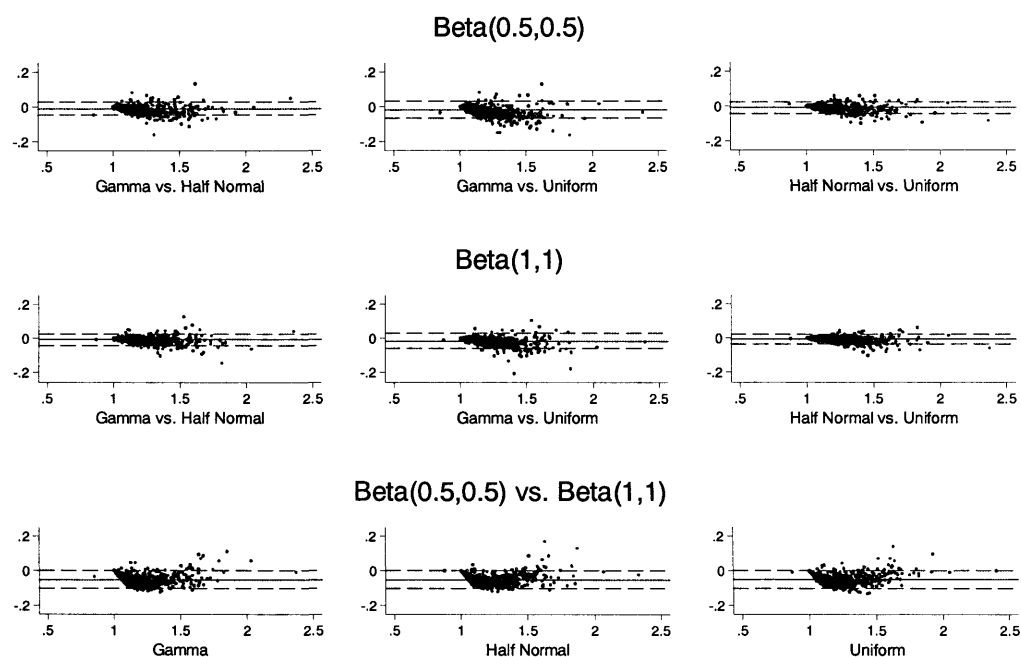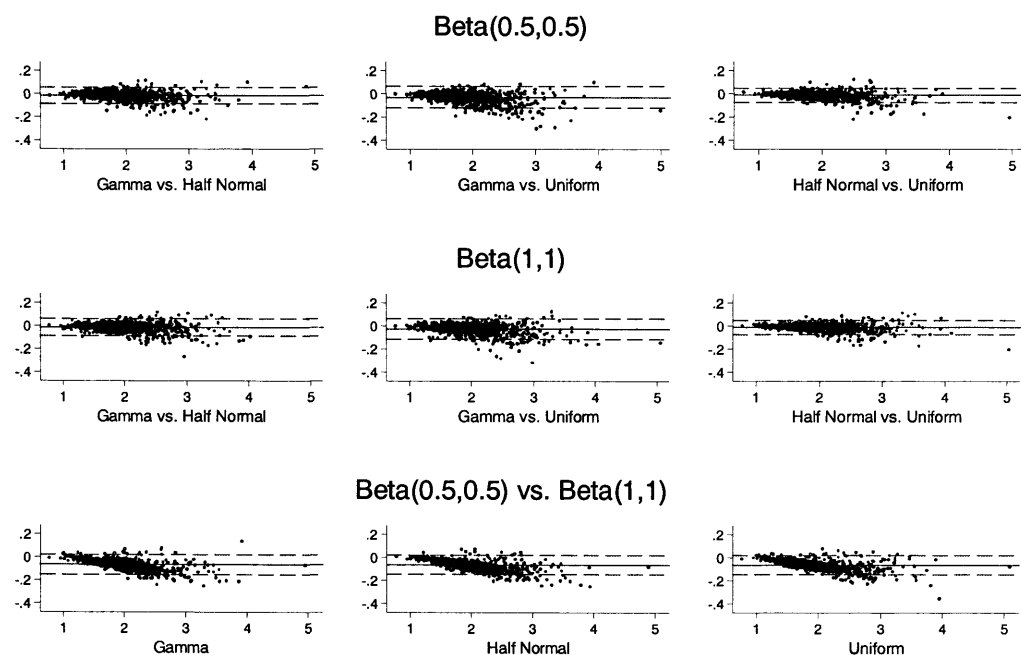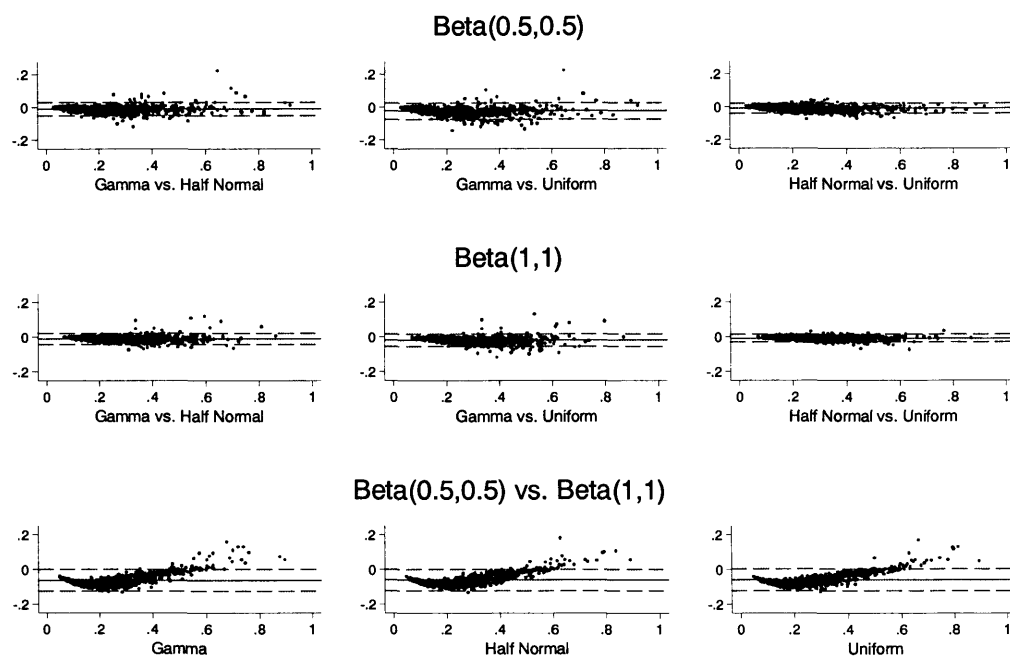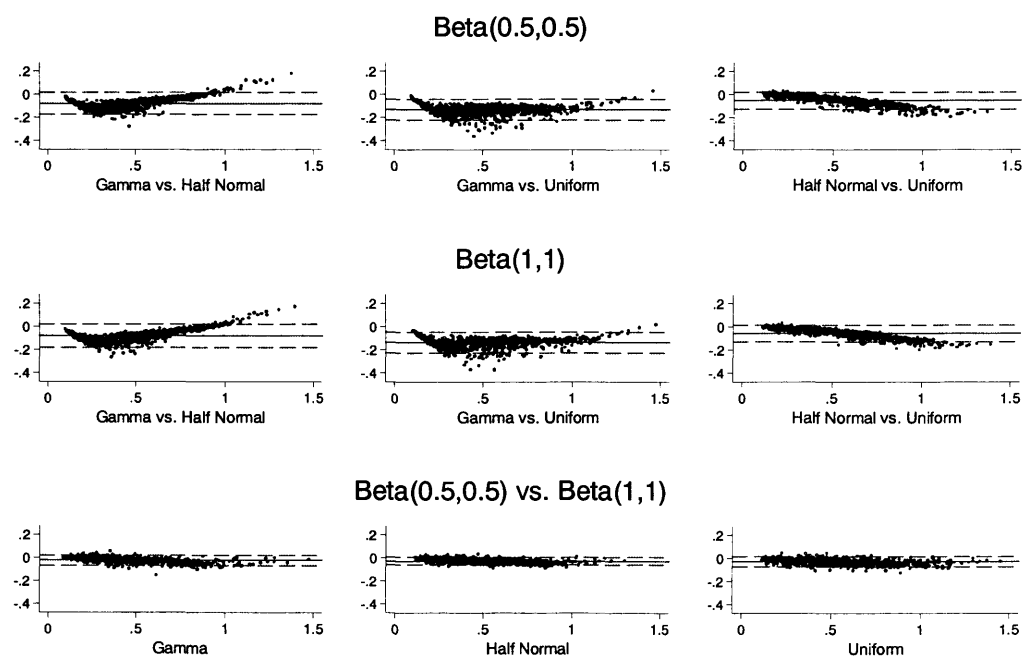
difference might be of importance, then sensitivity to the choice of prior distributions needs to be checked in any meta-analysis.

The impact of the choice of the prior on λ is shown in Figure 2.5c and once again the difference tends to be small but can, for particular datasets, be very large. For instance, datasets which produce estimates that average 0.6 can produce estimates that differ with prior by up to 0.2. Figure 2.5c also shows some of the systematic effects noted in the average results. As might be anticipated the impact of the choice of prior distribution is most marked in the estimate of τ, shown in Figure 2.5d. Not only are there strong systematic patterns, but the differences can also be large. For instance, when comparing a gamma prior distribution and a uniform prior distribution the estimates can vary by as much as 0.4 when the average estimate is 0.5, that is one estimate is 0.3 while the other is 0.7.

2.9 Discussion

2.9.1 A "genetic model-free" approach

When synthesising the evidence available on the association between a genetic polymorphism and a disease the main issue is the size of any association, but an important additional question is the mode of action of the gene. In practice, the estimate of the size of the association is influenced by what is assumed about the underlying genetic model. A review of the literature on meta-analysis of genetic association studies reveals how currently used approaches fail to address this issue (Attia, Thakkinstian and D'Este, 2003). Investigators often base their meta-analyses on the assumption of a specific genetic model and ignore their uncertainty about the mode of inheritance. Moreover, since it may be that no *a priori* biological evidence is available to justify the choice, different common genetic models are sometimes tested and the different results reported. Apart from the problem of multiple testing, this leaves the reader with a set of estimates and significance tests to interpret, all based on different assumptions. A number of investigators compare allele frequencies between cases and controls; however, this method yields a per-allele effect that is equivalent to assuming a co-dominant model and Hardy-Weinberg equilibrium (see Chapter 4). Additionally, the issue of whether the genetic model is common across populations does not seem to have been addressed. The results for the five meta-analyses examples show that adopting the wrong genetic model can lead to erroneous pooled estimates with deceptively high precision. The only meta-analytical approach currently in use which does not assume a common known

underlying genetic model is analysis by separate pairwise comparisons, i.e. independent meta-analyses comparing genotype groups two at a time. This method ignores the correlation between the two estimated odds ratios induced by the common baseline group and thus is inefficient, as the estimates cannot "borrow strength" from one another as they would do in a multivariate meta-analysis (van Houwelingen, Arends and Stijnen, 2002; Nam, Mengersen and Garthwaite, 2003). The genetic model-free approach is likely to be particularly beneficial compared to pairwise comparisons when either of the alleles is rare. Moreover, separate pairwise comparisons run into the problem of multiple testing, which becomes especially important when a polymorphism with more than two alleles is considered.

As Table 2.1 illustrates, published meta-analysis of genetic association studies have used a variety of methods for presenting their results. The genetic model-free approach offers a single unified method that could have been used in all of these examples giving a consistent presentation and avoiding the pitfall of overly strong assumptions about the genetic model or inefficient estimates.

The genetic model-free approach provides an integrated way of synthesising the evidence on genetic associations, which yields not only the magnitude of the genetic effect but also an indication of the operating genetic model based on the available data. The underlying genetic model is not constrained to correspond to one of the classical modes of inheritance (recessive, co-dominant, dominant), in recognition of the fact that the gene's mode of action in complex diseases might differ from that found in Mendelian traits, where the association between genotype and disease tend to be of a deterministic nature and hence the mode of inheritance is relatively clearly apparent. For example, a value of 0.23 for λ, as in the *MTHFR* meta-analysis, might be interpreted in two ways. The first explanation might be that the polymorphism is recessive in some studies and co-dominant in others, so that the average result is between the two. The second explanation could be that in complex diseases, the genotype is only one of many factors acting in a complex causal cascade leading to the disease. Although, at the molecular level, the polymorphism of interest might act in a clearly Mendelian manner on an intermediate phenotype, that Mendelian "signal" may be "diluted" or "distorted" when measured at the level of the final step in the cascade. Hence λ may be a more flexible and appropriate way to discuss

genetic models in complex disease. The genetic model-free approach has been used in practice by Thakkinstian and colleagues (2005 – included in Addenda) in a published meta-analysis of the association between β_2-Adrenoceptor polymorphism and asthma, where it provided insights on the underlying genetic model.

In the meta-analysis of genetic association studies there are two important types of heterogeneity that need to be addressed: heterogeneity in the genetic effect and heterogeneity in the genetic model. There are a number of reasons why we might see heterogeneity in the genetic effect, including differences in study methods and differences in the underlying genetic risk associated with gene-gene or gene-environment interactions. Heterogeneity of the genetic effect might also arise if the polymorphism under study does not act directly on the disease risk, i.e. it is not a "functional" or "causal" polymorphism but is simply a marker, which tends to be inherited together with the causal polymorphism (linkage disequilibrium). Populations may have different patterns of linkage disequilibrium, which lead to differences in the marker association with disease. It is important to note that causes of heterogeneity in the genetic effect will not necessarily cause heterogeneity in the genetic model. In fact, in order to act on the genetic model, interactions need to influence the disease risk in heterozygotes to a different extent to the risk in homozygotes.

The absence of heterogeneity in the genetic model is an important assumption of the genetic model-free analysis and, although this assumption is likely to hold in most cases, it still needs to be assessed. For example, the effect of genotype on allergy to pollens appears to follow different modes of inheritance for different ethnic groups and different forms of allergy (Sasazuki, 1983; Marsh and Huang, 1991). Although these are studies based on segregation analyses, which are relatively weak in the study of complex diseases (Thomas, 2004), they do raise the possibility that the mode of action may vary from study to study, perhaps due to complex gene-environment interactions which have different impact on the disease risk in heterozygotes compared to homozygotes for the polymorphism. Thus, the assumption of a common genetic model should be checked before applying the genetic model-free approach, for instance by using the graphs presented in figures 2.1 and 2.2. Should this assumption be in doubt, then the best approach would be to carry out joint pairwise comparisons using a multivariate meta-

analysis, where the correlations between the estimated odds ratios for the different genotype groups are taken into account. In general the random effects genetic model-free approach is easier to fit than the corresponding pairwise bivariate model because it contains two fewer parameters. Only in very large meta-analyses will it be possible to estimate the correlation in the heterogeneities required for the pairwise model. So, even when the assumptions of the genetic model-free analysis are not met exactly, such analysis may still be the best way of summarising the data and obtaining confidence intervals that are not falsely optimistic.

2.9.2 Bayesian implementation of the genetic model-free approach

The genetic model-free approach can be implemented using a maximum likelihood analysis, and indeed this route has been followed and did provide results very similar to those presented in § 2.5.2 (Minelli and Thompson, et al., 2005b – included in Addenda). However, the Bayesian framework might represent a desirable alternative for both philosophical and practical advantages. From a philosophical point of view, Bayesian analysis allows explicit inclusion of prior information on the genetic effect and on the genetic mode of inheritance. Although this possibility has not been explored in this thesis, the models presented could incorporate prior knowledge when it is available. Such knowledge might be based on evidence from studies not included in the meta-analysis or on expert opinion. While in the first instance the inclusion of prior information would often be straightforward, in the second case it can be difficult to use expert opinion to derive appropriate probability distributions (Gokhale and Press, 1982; Garthwaite and Dickey, 1988). From a practical point of view, the flexibility offered by Bayesian models estimated by MCMC algorithms makes it relatively straightforward to implement complex hierarchical models. The combination of increased computing power together with the availability of free software, particularly WinBUGS, to implement MCMC models has greatly contributed to the spread in the use of Bayesian methods (§ 1.5.2). However, the increased flexibility leads to a greater requirement to consider the issue of model choice (Gilks, Richardson and Spiegelhalter, 1996; Spiegelhalter, Abrams and Myles, 2004).

In situations where external information is not available, prior distributions still have to be specified for all parameters. Although such prior distributions may be intended to be

non-informative, this is in fact an impossible aim as Figure 2.3 illustrates. Rather, we must hope that the prior distributions will not be influential, in the sense that the use of alternative vague prior distributions will not change the conclusions. This may be an impossible aim if the data are sparse, especially when specifying prior distributions for scale parameters, such as the between-study heterogeneity in a random effects meta-analysis. If the results are sensitive to the choice of supposedly vague prior distributions, then we have no option but to consider that any prior distribution is informative and so must be chosen with care and subject to a sensitivity analysis (Spiegelhalter, Abrams and Myles, 2004). The problem of statistical inference in the presence of sparse data is not limited to the Bayesian approach, and an analogous non-Bayesian analysis would find a very flat likelihood and would produce wide confidence intervals. In such circumstances, the ideal solution might be to incorporate subjective prior information or other external evidence in the Bayesian analysis (Berger, 1985).

The hierarchical model for the meta-analysis of genetic association studies involving the ratio, λ, requires many large studies if the choice of prior distribution for λ is not to have an undue influence on the final estimates. In much the same way the prior distribution for the between-study variation, τ, can also influence the results. Recent research has suggested that the gamma prior distribution in not a good choice for the between-study precision in a hierarchical model and that the half-normal or uniform may be better (Gelman, 2004). The findings presented in this chapter do not support such a conclusion in the situation considered, where in general the gamma distribution performs well. Of course, part of the problem is to do with the scaling of the prior distributions. Had we taken a half-normal with a smaller standard deviation, or a uniform distribution over a shorter range, then the corresponding estimates would have been more similar to those obtained using a gamma prior distribution. This very fact emphasises the impossibility of defining a general vague prior distribution when data are sparse, and the importance of careful specification even when using vague prior distributions. An important issue when specifying a vague prior distribution is the need to take into account what values might be considered plausible in a particular situation (Spiegelhalter, Abrams and Myles, 2004).

The analysis presented illustrates the importance of an investigation of sensitivity to the choice of prior distributions in any Bayesian analysis in which the prior distributions are

not based on external knowledge. The sensitivity analysis will depend on the range of vague prior distributions that are considered reasonable in any given situation and the size of the change in the final estimate or its credible interval that is of practical importance. Thus, it will not be possible to find a single vague prior distribution that is always the least informative, so in complex models the desire to use vague prior distributions does not free the researchers from the need to tailor their prior distributions to their particular problem.

Although not investigated in this thesis, informative prior distributions, based on expert opinion or external evidence, could be used for the different model parameters. In a more general meta-analysis context, empirical data-based prior distributions have been advocated for the heterogeneity term, τ, and might be an attractive option, especially when the number of studies included in the meta-analysis is small (Higgins and Whitehead, 1996). For the parameter λ, there might well be data from studies evaluating the effect of the same polymorphism on similar disease pathways. The increase in the precision of the estimated λ due to the use of an informative prior distribution would in turn increase the precision in the estimates of the odds ratios of interest, OR_{GG} and OR_{Gg}, and so might be very beneficial.

In the analyses presented in this chapter, the prospective and retrospective likelihoods gave very similar results for all examples considered. It may well be that in practice the prospective likelihood could be used when synthesising evidence from case-control-studies. However, the approximate equivalence of the two likelihoods for a particular combination of dataset and model can only be established by using both, which rather removes the benefits of the simpler, but theoretically inappropriate, prospective model. The retrospective likelihood has the further advantage that it can easily incorporate the assumption of Hardy-Weinberg equilibrium in the controls (Thompson *et al.*, 2004; Cheng and Chen, 2005). Given these considerations it will often be more appropriate to use the retrospective likelihood unless there is considerable evidence of approximate equivalence from similar analyses.

2.9.3 Conclusions

In conclusion, the meta-analytical approach proposed is based on a re-parameterisation of the classical representation of genetic association studies, where the new parameters are biologically meaningful and informative. The approach makes maximum use of the information available by quantifying the magnitude of the genetic effect and estimating the genetic mode of action at the same time. The genetic model is estimated on the basis of the data rather than assumed, and this is important in all cases where no *a priori* knowledge about the underlying genetic model is available.

3   SUBGROUP ANALYSIS

3.1 Chapter overview

This chapter addresses the issue of how to handle subgroup analysis in the meta-analysis of genetic association studies. In § 3.2 an introduction to the general problem of subgroup analysis is presented, followed by an overview of those issues specific to meta-analysis of genetic association studies. Section 3.3 reviews and discusses the different approaches adopted for dealing with subgroups in 37 published meta-analyses of genetic association studies. In § 3.4 an alternative approach is proposed, based on the concept that a meta-analysis of genetic association studies might benefit from borrowing information on secondary parameters across subgroups, for example by assuming that such parameters are common across subgroups even if the size of the genetic effect differs. In particular, the impact of assuming a common genetic model and common heterogeneity are considered, and sensitivity analyses performed to investigate whether the benefit remains when the secondary parameters are assumed to be similar, but not identical. The method of analysis and method of sensitivity analysis are demonstrated in § 3.5 by re-analysing one of the meta-analyses reviewed, where the subgroups were defined by ethnicity. Three further meta-analyses, in which the subgroups were either defined by related disease outcomes or by gender, are discussed in detail in § 3.6, in order to illustrate the assumptions that might be made about shared parameters. Finally, the advantages and limitations of the approach proposed, compared to the commonly used methods for dealing with related subgroups, are discussed in § 3.7.

3.2 Introduction

3.2.1 Subgroup analysis in general meta-analysis literature

In meta-analysis, two types of subgroup analysis are encountered, which differ in their aims (Oxman and Guyatt, 1992; Sutton *et al.*, 2000). In the first type, subgroups are usually specified *a priori* with the purpose of estimating the magnitude of the effect of interest (e.g. treatment effect) in different types of patients; this reflects the belief that,

for instance, the underlying treatment effect may depend on certain baseline characteristics of the patients. In the second type, the subgroups are not interesting *per se*, but they are defined after the data have been combined, in order to investigate whether an observed variability between study results might be explained by specific characteristics of the studies. Although when performing a meta-analysis some variation, or heterogeneity, in study results might be expected by chance, often the variation observed is in excess of that, which might be explained by a number of reasons. Such heterogeneity may reflect not only differences in the characteristics of the patients included in each study, but also differences in the degree of the exposure (e.g. dose of a drug), length of follow-up, study design and quality, and even publication bias, whenever these factors are associated with the outcome of interest (Thompson, 1994). Although, to some extent, this variability can be accounted for by using a random effects meta-analysis model (§ 1.4), possible sources for the observed heterogeneity have to be investigated. Subgroup analysis is the approach most commonly used for this purpose (Assmann *et al.*, 2000), but other methods are available. While subgroup analysis consists of performing separate analyses of the effect of interest within each subgroup and then comparing estimates and *p*-values across subgroups, an alternative approach is to use a statistical test of interaction between effect of interest and subgroup (Oxman and Guyatt, 1992). Another technique closely related to subgroup analysis is meta-regression, where a specific characteristic of the study population or study methods that might explain heterogeneity in study results is modelled as a covariate in a regression-type analysis (Sutton *et al.*, 2000).

From a frequentist perspective, subgroup analysis poses the problem of multiple testing, which may result in false positive findings in one or more subgroups, and although adjustments such as Bonferroni correction have been proposed, their use is debatable. In fact, statistical adjustment modifies the *p*-value to account for the multiple tests performed, and thus control the probability of type I error, but, by doing so, it increases the probability of failing to detect a true effect, i.e. type II error (Perneger, 1998). On the other hand, the Bayesian approach to subgroup analysis is not concerned with the issue of multiplicity, since its philosophy repudiates the very idea of hypothesis testing and thus type I error has no relevance (Spiegelhalter, Abrams, and Myles, 2004). Moreover, the Bayesian approach allows to incorporate *a priori* beliefs

regarding how different the treatment effect in any specific subgroup can be from the overall treatment effect, whenever such prior knowledge is available (Donner, 1982; Pocock and Hughes, 1990; Dixon and Simon, 1991; Dixon and Simon, 1992; Spiegelhalter, Abrams, and Myles, 2004). The inclusion of such *a priori* structure causes shrinkage of both the point estimate of treatment effect and its credible interval. This may be very useful in all cases where the paucity of data within subgroups causes serious problems of estimation, which can lead to over interpretation of extreme results.

3.2.2 Subgroup analysis in meta-analysis of genetic association studies

Studies of the effect of genetic polymorphisms on disease have often showed a disappointing lack of reproducibility (Gambaro, Anglani and D'Angelo, 2000; Cardon and Bell, 2001; Colhoun, McKeigue and Davey Smith, 2003). In 2001, Ioannidis and colleagues reviewed 26 published meta-analyses of gene-disease associations, with the primary aim of evaluating how the degree of heterogeneity in study results may help explaining the problem of lack of reproducibility of genetic association studies (Ioannidis *et al.*, 2001). Using a cut-off significance level of 0.1, they found statistically significant heterogeneity in 39% of the comparisons, a figure that increased to 47% when they updated their dataset in 2003 to include a total of 55 meta-analysis papers (Ioannidis *et al.*, 2003). It is thus not surprising that meta-analyses of genetic association studies are often presented as organised into subgroups. In fact, although sometimes the interest of the investigators lie in specific subgroups of patients defined *a priori*, often subgroup analysis is performed in an attempt to explain the observed heterogeneity. Subgroups might be defined by studies on populations differing for ethnicity or origin (e.g. European or Asian), studies on men or women, or studies of the same gene on different disease outcomes. In such a subgroup analysis it is natural to ask whether it is possible to use information from one subgroup to improve the estimation in another. Subgroups are usually introduced because there appears to be systematic variation in the main outcome measure, i.e. the estimate of the genetic effect, but meta-analytic models contain other parameters and it may be that the precision of the estimate of the main outcome can be improved by learning about the secondary parameters using information from the other subgroups.

When modelling genetic associations there are several parameters that might or might

not vary across subgroups of studies. These include the magnitude of the genetic effect, the mode of inheritance (genetic model), the amount of heterogeneity in the size of the genetic effect and the allele frequency. Between-subgroup differences in the size of the genetic effect do not necessarily imply corresponding differences in the other parameters. Differing effect sizes in the subgroups may be the result of subject-level gene-gene or gene-environment interactions when the interacting factor has a different prevalence in each of the subgroups. However, for such interactions to be able to influence the underlying mode of inheritance, one has to imagine that the interaction influences the effect in the heterozygotes in a different way to the mutant homozygotes and this is much less likely. Consequently, one might want to assume that the subgroups share the same genetic model even though the sizes of the effects differ. Adding this assumption is potentially beneficial especially for a subgroup that contains few studies.

Another possible assumption is that the heterogeneity or variance of the effect size is the same in each subgroup. This could happen when a major gene-environment interaction modifies the genetic effect of a polymorphism, so necessitating a subgroup analysis, but when other less dramatic sources of variation are similarly distributed across subgroups. Conversely, allele frequency is a secondary parameter that is very likely to vary between studies, but even here, allele frequency might be modelled by a random effect with mean and/or variance that is common across subgroups.

3.3   Review of the literature

In order to evaluate what are the approaches commonly used for subgroup analysis in the evidence synthesis of genetic association studies, 37 published meta-analyses reported in a review by Attia and colleagues were re-evaluated (Attia, Thakkinstian and D'Este, 2003; see also § 6.2). For each meta-analysis information was extracted on; the gene-disease association under study, the number of studies, the presence of studies on ethnic subgroups and how they were dealt with, the presence of studies on different disease outcomes and how they were dealt with, and finally whether any other subgroups were considered. The information on how the investigators handled subgroups in each of the 37 meta-analyses is summarised in Table 3.1, and full references are reported in Appendix 2. The table shows how subgroup analysis tends to

be common in meta-analyses of genetic association studies and how the approach to handling subgroups is very diverse. In most instances, subgroups seemed to be defined after the data were combined with the aim of explaining heterogeneity of study results, rather than being of interest in themselves and thus specified *a priori*.

Frequently in the meta-analyses reviewed the definition of subgroups was based on the ethnicity and/or origin of the study population. In thirteen studies the meta-analysis was focused on a specific population, sometimes explicitly excluding studies on other ethnic groups on the grounds of maintaining genetic homogeneity (n=2), while one study performed separate meta-analyses on different ethnic subgroups (Furlong, 1999 b). Somewhat surprisingly, two of these meta-analyses gave the reason for excluding a study (Furlong, 1998), or for performing separate meta-analyses (Furlong, 1999 b), as the difference in allele frequency across ethnic groups. In 8 meta-analyses all ethnic groups were combined, and in three of them the decision was based on the absence of heterogeneity. In another 8 studies a combined meta-analysis across ethnic groups was reported together with separate meta-analyses for each ethnic group (n=6) or with a test for interaction between gene effect and ethnicity (n=2). In the remaining 7 meta-analyses no information on the presence of different ethnicities was reported.

Thirty one percent of the meta-analyses (n=11) considered more than one disease outcome, with the number of different outcomes ranging from 2 to 15. Apart from one meta-analysis in which the outcomes were lung cancer and Parkinson's disease, in all other cases the disease outcomes were more or less related to each other. The subgroups were represented by different forms of the same disease defined by severity and/or histological features (e.g. different types of lung cancer), or by the effect of the same disease on different organs (e.g. diabetic nephropathy and retinopathy). Again, the subgroups were analysed in different ways; separately (n=4), together (n=1), or both together and separately (n=5).

Table 3.1 also shows how other potential sources of inconsistency of study results were dealt with in the meta-analyses; these include male/female differences, possible gene-environment interactions (e.g., smoking), and proxy measures of study quality (e.g., small versus large studies). In most cases the different subgroups of studies were

Table 3.1 - Different approaches to handle subgroups or groups of studies in the 37 meta-analyses reviewed

Author, year	Association evaluated	Different ethnic groups		Different related outcomes		Other subgroups	
		Ethnicities	Method	Outcomes	Method	Groups	Method
Carlquist, 1991	HLA class II polym. & cardiomyopathy	Whites	/	/	/	/	/
Jenkins, 1992	HLA II polym. & Hashimoto's thyroiditis (3 polym.)	Caucasians	/	/	/	/	/
Pato, 1993	DRD2 gene & alcoholism	Not reported	/	/	/	/	/
Ordovas, 1995	*APOE* gene & response to pravastatin	Not reported	/	/	/	Male/Female	Together (similar results for the two)
McWilliams, 1995	GSTM1 gene & lung cancer	3 ethnic groups	Together + separate analyses	Different forms of lung cancer	Together + separate analyses	Phenotyping/ genotyping	Together + separate analyses
Shaikh, 1996	DRD3 Ser-9-Gly polym. & schizophrenia	Caucasians and Asians	Together + ethnicity tested as interaction	/	/	Different centres	Together + centre tested as interaction
Prasher, 1997	APOE gene & Alzheimer's disease in Down syndrome	Not reported	/	/	/	/	/
Kunz, 1997	*AGT* gene & essential hypertension	Whites	/	/	/	Family history and Severity	Together + separate analyses
Weston, 1997	H-*ras*-1 gene & breast/lung cancers	3 ethnic groups	Together + separate analyses	Lung & breast cancers	Separate analyses	Smokers/non -smokers (lung cancer)	Together + separate analyses
Staessen, 1997	*ACE* gene & cardio- vascular/renal disease	3 ethnic groups	Together + sensitivity analyses	15 outcomes considered	Separate analyses	Many subgroups	Together + separate analyses
Christensen, 1997	CYP2D6 gene & lung cancer	Mixed ethnic groups	Together	/	/	Small/large sample size	Together + only large studies
Hokanson, 1997	LPL gene & coronary disease (5 polym.)	Not reported	/	MI/ coronary sclerosis/ carotid sclerosis	Together	/	/
Iacoviello, 1998	PAI-1 promoter gene & myocardial infarction	Not reported	/	/	/	3 subgroups (by design & baseline risk)	Together + separate analyses
Sharma, 1998	*ACE* gene & ischaemic stroke	Whites	1 Japanese study excluded (to maintain homogeneity)	/	/	/	/
Arranz, 1998	5-HT$_A$ receptor gene & clozapine response (2 polym.)	Not reported	/	/	/	/	/
Furlong, 1998	Serotonin transporter gene & affective disorders	Whites	Japanese study excluded (different allele frequencies)	Bipolar/unipolar affective dis.	Together + separate analyses	/	/
Allison, 1998	ADRB3 Trp64Arg polym. & body mass index	>3 ethnic groups	Together + ethnicity tested as interaction	/	/	Diabetic status	Together + tested as interaction
Fujisawa, 1998 (a)	ADRB3 Trp64Arg polym. & body mass index	Whites and Japanese	Together (no lack of homogeneity)	/	/	Many subgroups	Together (no lack of homogeneity)
Fujisawa, 1998 (b)	*ACE* I/D polym. & diabetic nephropathy/retinopathy	Whites and Japanese	Together (no lack of homogeneity)	Nephropathy/ retinopathy	Separate analyses	IDDM/NIDDM diabetes	Together (no lack of homogeneity)
Williams, 1998	DRD3 *Bal*I polym. & schizophrenia	3 ethnic groups	Together	/	/	/	/
Dubertret, 1998	DRD3 *Bal*I polym. & schizophrenia	>3 ethnic groups	Together + separate analyses				
Hani, 1998	*KIR6.2* gene and Type II diabetes	Caucasians	/	/	/	/	/
Rostami- Hodjegan, 1998	CYP2D6 gene & lung cancer/Parkinson's disease	Mixed	Together + separate analyses	lung cancer/ Parkinson's disease	Separate analyses	Many subgroups	Together + separate analyses
Furlong, 1999 (a)	Tyrosine hydroxylase gene & affective disorders (2 polym.)	Whites	1 Japanese excluded (unclear allele specific.)	Bipolar/unipolar affective dis.	Together + separate analyses	/	/
Houlston, 1999	GSTM1 gene & lung cancer	>3 ethnic groups	Together	/	/	Phenotyping/ genotyping	Together + separate for typing method
Breen, 1999	DRD2 -141C D/I polym. & schizophrenia	Caucasians	/	/	/	/	/
Joost, 1999	CYP2D6 gene & Parkinson's disease	Europeans/ Europ.-derived North Americ.	/	/	/	/	/
Staessen, 1999	*AGT* gene & cardio- vascular/renal disease	3 ethnic groups	Together + sensitivity analyses	10 outcomes considered	Separate analyses + some in groups	Many subgroups	Together + separate analyses
Furlong, 1999 (b)	MAOA gene (2 polym.) & bipolar affective disorder	Caucasians and Japanese	Separate analyses (different allele freq.)	/	/	Male/Female	Together + separate analyses
Juo, 1999	APOAI promoter gene & apolipoprotein A-I level	Whites	1 Japanese study excluded (to remove statist. heterogeneity)	/	/	Male/Female	Together + separate analyses
McCarron, 1999	*APOE* gene & ischemic cerebrovascular disease	Whites and Japanese	Together (no statist. heterogeneity)	/	/	Age and Male/Female	Matched controls for each study
Kato, 1999	*AGT* gene and essential hypertension	Japanese	/	/	/	/	/
Rantala, 2000	APOB gene (5 polym.) & serum lipids	Whites + 1 study mixed	Together	6 correlated outcomes	Separate analyses	/	/
Preisig, 2000	MAOA gene (2 polym.) & bipolar affective disorder	Whites	/	/	/	Male/Female	Together + only females
Wong, 2000	CYP2E1 gene & alcoholic liver disease	Whites	/	Alcoholic liver disease/cirrhosis	Together (similar results for the two)	/	/
Keavney, 2000	*ACE* I/D polymorphism & myocardial infarction	Not reported	/	/	/	Small/large sample size	Separate analyses
Houlston, 2000	CYP1A1 gene (2 polym.) & lung cancer	>3 ethnic groups	Together	/	/	/	/

analysed separately in sensitivity analyses, while the main analyses were carried out on all studies.

3.4   The proposed approach

The method proposed for combining information on the mode of inheritance across subgroups is a development of the genetic model-free approach presented in § 2.4. This approach, whose formulation allows the data to estimate the genetic model rather than requiring it to be assumed *a priori*, parameterises the meta-analysis in terms of two parameters; the OR_{GG}, which captures the magnitude of the genetic effect, and λ, the ratio of $logOR_{Gg}$ to $logOR_{GG}$. The genetic model-free approach is summarised below.

The two log odds ratios from each study are modelled in a bivariate meta-analysis in which $logOR_{Gg}$ is derived as the product of $logOR_{GG}$ and λ. As almost all genetic association studies have a case-control design, the meta-analytical model is based on a retrospective likelihood. The model is described in detail in § 2.6.1. The $logOR_{Gg}$ and $logOR_{GG}$ are denoted by δ_2 and δ_3, and, over the studies i, the study-specific log odds ratios, δ_{3i}, are modelled as random effects in order to allow for heterogeneity in the genetic effect, so each δ_{3i} will vary about a mean, θ, with variance, τ^2:

$$\delta_{3i} \sim N(\theta, \tau^2)$$

For each study, the study-specific log odds ratios δ_{2i} are equal to the product of δ_{3i} and λ, and λ is modelled as a fixed effect, that is, the mode of inheritance is assumed constant across studies. The distribution of δ_{2i} is derived as the product of δ_{3i} and λ so that:

$$\delta_{2i} \sim N(\lambda\theta, \lambda^2\tau^2)$$

Lambda is modelled as bounded between 0 and 1, and so the rare situation of an over-dominant genetic model, in which λ can be higher than 1 or lower than 0, is not considered in the current examples. In all models vague prior distributions are used for parameters θ, τ and λ;

$$\theta \sim Normal(0,10000), \qquad \tau \sim Uniform(0,2), \qquad \lambda \sim Beta(0.5,0.5).$$

In order to combine information across subgroups, we allow the genetic model, described by the parameter λ, and the heterogeneity for the $logOR_{GG}$, described by the parameter τ,

to be the same in each subgroup. This reflects the situation where the researcher thinks that the different subgroups might have different genetic effect sizes, but is willing to assume that the subgroups are similar enough to share the same genetic model and/or the same within subgroup heterogeneity.

To test the sensitivity to the assumptions of common λ and common τ, we allow a difference between the subgroup specific λs and the τs. This models the situation when the researcher is willing to assume that the different subgroups might have a similar but not necessarily identical genetic model and/or within subgroup heterogeneity.

Since in our example we assume that λ is bounded between 0 and 1, for the sensitivity analyses on λ we define ε as equal to the difference in the *logits* of the λs. That is

$$\varepsilon_j = \text{logit}(\lambda_j) - \text{logit}(\lambda_1) \qquad j=2,3,\ldots$$

where j defines the subgroups. For the sensitivity analyses on τ we define ϕ as relating to the difference in the *log* of the τs. So that

$$\phi_j = \log(\tau_j) - \log(\tau_1) \qquad j=2,3,\ldots$$

The ε_j and ϕ_j are treated as having a prior distribution centred on 0 and with the same, fixed *known* variance τ_ε^2 and τ_ϕ^2, respectively

$$\varepsilon_j \sim N(0, \tau_\varepsilon^{\,2}) \qquad \text{and} \qquad \phi_j \sim N(0, \tau_\phi^{\,2})$$

The sensitivity analysis investigates the effect of increasing the two variances to reflect the situation in which the investigator is increasingly less certain that the genetic model and/or the within subgroup heterogeneity are common.

It is interesting to note that the models proposed for the sensitivity analyses for λ and ε require the arbitrary choice of a baseline subgroup, and the results of such analyses have to be interpreted considering this. Although the choice of a baseline group could be avoided by simply allowing the logits of λs or the log τs to be a random effect across subgroups, this approach is philosophically different, since it implies that the differences across studies for the genetic model or for the heterogeneity of the genetic effect come from a higher level distribution. While this approach would be preferable if we thought of the ethnic groups considered in the meta-analysis as representative of a larger population of many possible ethnicities, in our example it might not reflect the authors' point of

view, which seems to consider the three ethnic groups as the ones interesting in their own right.

Results of all meta-analysis models were obtained in WinBUGS using a burn-in of 10,000 and a chain length of 50,000. Convergence was assessed through visual inspection of trace plots by performing sensitivity analyses with respect to initial values, length of burn-in and length of sample (§ 1.5.2).

3.5 Illustration

3.5.1 Subgroups defined by ethnicity

Sethi and colleagues (Sethi, Nordestgaard and Tybjaerg-Hansen, 2003) reported a meta-analysis of the association between the genetic variant Thr235 of the angiotensinogen (*AGT*) gene and essential hypertension. In all, 37 studies were considered and the analyses were performed separately on subgroups of studies in three different ethnic populations; Whites (n=22 studies), Asians (n=12) and Blacks (n=6). The group of Whites included studies on European (n=14), North Americans (n=4), Australians (n=2), South Americans (n=1), and one study considered both Europeans and North Americans. The group of Asians included studies on Japanese (n=8), Chinese (n=3) and Middle Easterns (n=1), while the group of Blacks included studies on Caribbeans (n=2), North Americans (n=2), Africans (n=1) and Europeans (n=1).

The results reported for the two odds ratios OR_{GG} and OR_{Gg} were; Whites, 1.29 (95% Confidence Interval: 1.10 to1.50) and 1.08 (1.01 to 1.15); Asians, 1.60 (1.19 to 2.15) and 1.29 (0.96 to 1.74); Blacks, 1.16 (0.58 to 2.35) and 0.94 (0.45 to 1.96). Average allele frequencies of Thr235 (G) were 0.43, 0.76, and 0.83 in the three ethnic groups. Figure 3.1 shows the forest plots and pooled estimates for OR_{GG} and OR_{Gg}, separately for each of the three ethnic groups. This meta-analysis was re-analysed to illustrate the advantages of sharing information across subgroups. The results of the method proposed for handling related subgroups as compared to the most commonly used approach are reported in § 3.5.2, while the results of the sensitivity analyses performed to investigate whether the benefit remains when the secondary parameters are assumed to be similar rather than identical are shown in § 3.5.3.

3.5.2 Main results

The first row of Table 3.2 shows the results of separate random effects meta-analyses on OR_{GG} and OR_{Gg} within each ethnic group. This approach, which is described here as the current standard, does not make any assumptions about the genetic model and does not seek to share any information across subgroups. By contrast, the second row of Table 3.2 shows the results when the model is parameterised as described in § 3.4, and the heterogeneity and genetic model are assumed common across subgroups. With the standard approach the studies on Whites and Asians provide evidence of a genetic effect, while the 6 studies on Blacks are inconclusive and have wide credible intervals for the estimates of both odds ratios. When common heterogeneity and genetic models are assumed, the point estimates for the two odds ratios of Blacks become almost identical to those in Whites even though no assumptions have been made about the effect sizes being similar, and the credible intervals are much narrower. Indeed, the odds ratios are sufficiently similar that we might now also want to assume that they share the same size of genetic effect and pool these two subgroups. By analysing Black and White subgroups together the estimates for OR_{GG} and OR_{Gg} would become 1.30 (95% CrI: 1.09 to 1.54) and 1.13 (95% CrI: 1.04 to 1.26), with a common λ of 0.49 (95% CrI: 0.29 to 0.72). In the Asian subgroup, the second row of Table 3.2 shows that the precision of the estimates has increased, but which remain different from those in Whites and Blacks, suggesting that the genetic effect size might differ in Asians. In Whites, the results of the two approaches are similar because the assumptions are less critical in such a large subgroup (n=22).

3.5.3 Results of sensitivity analyses

Table 3.3 shows the results of the sensitivity analyses performed to investigate the impact of uncertainty as to whether the mode of inheritance (λ) is the same in the different subgroups, assuming that we are sure that the heterogeneity is common. The variance of the distribution of the prior on the difference in the logits of λs (τ_ε) was allowed to assume four different values. If λ in Whites is about 0.50, then; τ_ε=0 is equivalent to certainty that the mode of inheritance is the same in all subgroups; τ_ε=0.43 means that we have 95% belief that λ in the other two subgroups is between 0.30 and 0.70; τ_ε=1.12 means that λ in the two subgroups is between 0.10 and 0.90; τ_ε=+∞ corresponds to no

Figure 3.1 – Forest plots and pooled estimates for OR_{GG} and OR_{Gg} for each of the three ethnic groups; a) Whites; b) Asians; c) Blacks

a)

b)

Study	OR$_{GG}$	Odds ratio (95% CI)
Cheung		3.12 (0.32,30.66)
Chiang		0.90 (0.04,22.65)
Hata		2.83 (0.46,17.57)
Morise		0.83 (0.23,3.01)
Nishiuma		2.99 (1.04,8.60)
Kishimoto		1.35 (0.47,3.90)
Thomas		1.39 (0.51,3.79)
Kario		4.14 (1.62,10.56)
Sato		1.43 (0.52,3.96)
Iso		0.84 (0.33,2.13)
Kato		1.37 (0.75,2.53)
Frossard		1.72 (0.76,3.90)
Overall (95% CI)		1.61 (1.19,2.16)

Study	OR$_{Gg}$	Odds ratio (95% CI)
Cheung		2.64 (0.26,27.26)
Chiang		0.23 (0.01,6.22)
Hata		0.88 (0.14,5.74)
Morise		0.72 (0.19,2.77)
Nishiuma		1.73 (0.60,5.02)
Kishimoto		1.47 (0.50,4.33)
Thomas		1.12 (0.39,3.23)
Kario		1.84 (0.74,4.60)
Sato		0.74 (0.26,2.16)
Iso		1.03 (0.39,2.69)
Kato		1.48 (0.79,2.78)
Frossard		1.45 (0.67,3.16)
Overall (95% CI)		1.29 (0.96,1.75)

Odds ratio

c)

knowledge about the similarity of λ across subgroups. Even when this variance is infinite we still assume a common genetic model *within* the subgroups, but not that it is the same in each subgroup. For Blacks, it can be seen that the greater the *a priori* certainty that the modes of inheritance are the same across subgroups, the greater the precision of the odds ratios, particularly OR_{Gg}. However, the improved precision is still evident when $\tau_\varepsilon = 0.43$ and we are far from certain about the similarity in the modes of inheritance. In Asians, introducing the assumption of a common λ does not lead to an increased precision of either odds ratio, but changes the point estimates for the odds ratios, moving both estimates upwards and away from those of the other two groups. This might be due to the fact that we are forcing the λ for Asians to be the same as in the other two groups while in truth it is different. When repeating the analysis by assuming a common λ only for Blacks and Whites and monitoring the difference with the λ in Asians, although a difference of 0 was within the 95% credible interval, there was an 87% posterior probability that λ for Asians was smaller than λ for Blacks and Whites.
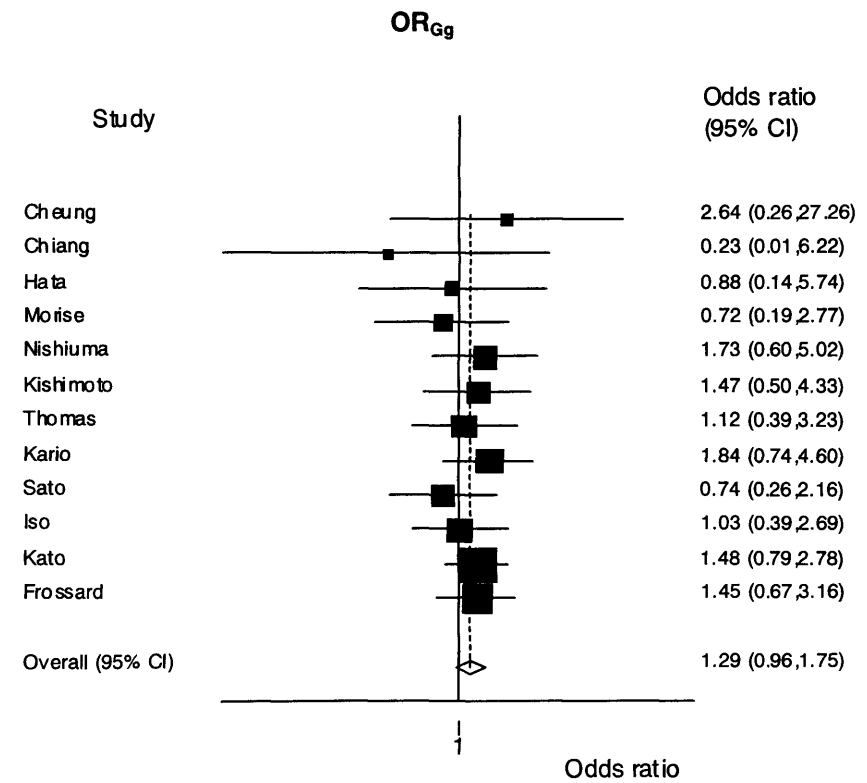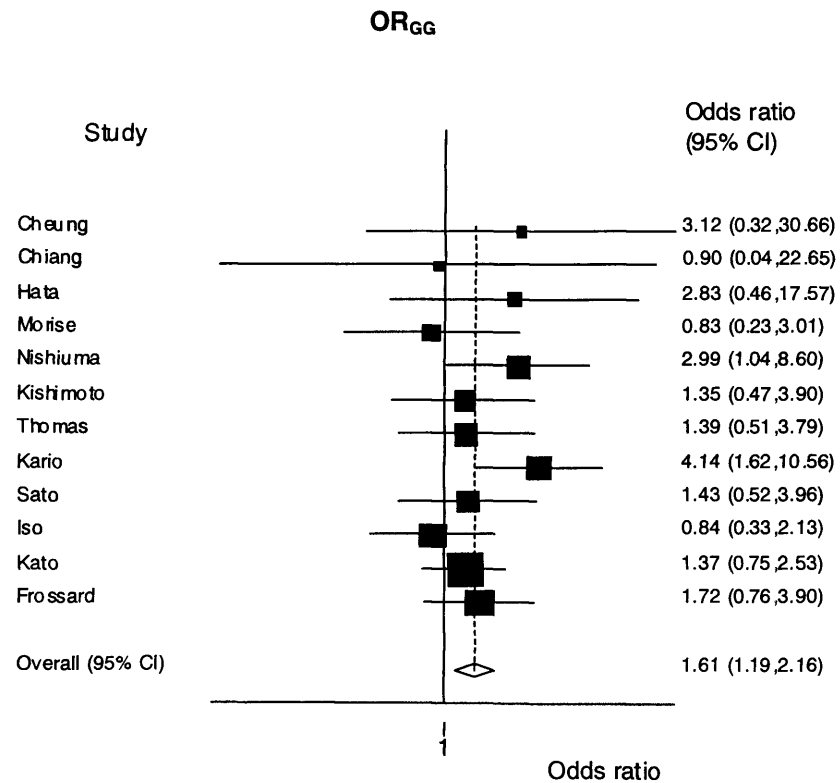
Table 3.4 shows the results of the sensitivity analyses performed to investigate the impact of uncertainty that the within subgroup heterogeneity (τ) is the same across subgroups when we are sure that the genetic model is common. The variance of the distribution of the prior on the difference in the log of the τs (τ_ϕ) was allowed to assume four different values. If τ in Whites is 0.30, then; $\tau_\phi = 0$ is equivalent to certainty that the within subgroup heterogeneity is the same in all subgroups; $\tau_\phi = 0.56$ means that we have 95% prior belief that τ in the other two subgroups vary between 0.22 and 0.38; $\tau_\phi = 1.01$, means that τ in the two subgroups can vary from 0.15 to 0.45; $\tau_\phi = +\infty$ corresponds to no assumptions on the similarity of τ across subgroups. It can be seen how introducing the assumption of the similarity of τ does not substantially affect the estimates for the two odds ratios in Blacks, while it does improve the precision of the estimates in Asians, with the improved precision still evident even when we are far from certain about the similarity in the within subgroup heterogeneity. In Asians, the assumption of common τ also changes the point estimates for the two odds ratios, bringing them towards those in Whites and Blacks. This is due to a change in the weights given to studies within each subgroup. In fact, among the 12 studies on Asians, there is one study (Chiang) with an OR_{GG} of 0.90 (95% CI: 0.04 to 22.65) and an OR_{Gg} of 0.23 (95% CI: 0.01 to 6.22), which, although inconclusive, would suggest a protective effect of the polymorphism on

the disease with an over-dominant mode of inheritance. When repeating the analysis with no assumptions on the similarity of τ (τ_ϕ=+∞) after excluding this outlier, the results for Asian studies are similar to those obtained from all studies with the assumption of common τ (τ_ϕ=0), with an OR_{GG} of 1.68 (95% CrI: 1.09 to 2.80) and an OR_{Gg} of 1.27 (95% CrI: 1.04 to 1.71). This shows how the assumption of common τ across subgroups makes the analysis more robust in the presence of outliers.

These sensitivity analyses assume that our beliefs about the differences between Whites and Blacks are the same as, and independent of, our beliefs about the differences between Whites and Asians and Blacks and Asians. Given prior knowledge based on expert opinion, it would be simple to change these prior distributions and perhaps to introduce some correlation.

Additional sensitivity analyses investigating the assumptions of similar λ, when τ is not assumed common across subgroups and similarly τ, when λ is not assumed common (i.e. a two-way sensitivity analysis), show very similar results to those presented in Tables 3.3 and 3.4. Their results are reported in Appendix 2, Table A and B.

3.6   Other examples

3.6.1   Subgroups defined by related disease outcomes

As an example of related disease outcomes, the meta-analysis of the association between the I/D polymorphism of the *ACE* gene and two forms of diabetic microangiopathy, namely, diabetic nephropathy and diabetic retinopathy, will be considered (Fujisawa, Ikegami, and Kawaguchi, 1998; see also § 2.5.1). Fujisawa and colleagues performed separate fixed effect meta-analyses for the two outcomes basing both analyses on the assumption of a dominant genetic model. The results were an odds ratio of 1.32 (95% CI: 1.15 to 1.51) for nephropathy (n= 21 studies), and 0.91 (0.73 to 1.13) for retinopathy (n=9). The conclusion was that "the data suggest that the *ACE* I/D polymorphism affects the risk for diabetic nephropathy but not diabetic retinopathy". Rather questionably, the authors justified the choice of the dominant genetic model by the fact that it provided a higher odds ratio and marginally less statistical heterogeneity than a recessive model.

Table 3.2 - Results of our method based on the assumption of common λ and τ (heterogeneity for OR$_{GG}$) compared to the standard approach

META-ANALYTICAL APPROACH	White population (n=22)				Asian population (n=12)				Black population (n=6)			
	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)
Separate analyses for OR$_{GG}$ and OR$_{Gg}$	1.30 (1.08 to 1.56)	1.13 (1.03 to 1.26)	/	OR$_{GG}$: 0.29 (0.14 to 0.51) OR$_{Gg}$: 0.11 (0.03 to 0.23)	1.63 (1.12 to 2.36)	1.27 (0.90 to 1.78)	/	OR$_{GG}$: 0.24 (0.01 to 0.79) OR$_{Gg}$: 0.16 (0.01 to 0.60)	1.17 (0.49 to 2.84)	0.95 (0.40 to 2.27)	/	OR$_{GG}$: 0.35 (0.01 to 1.47) OR$_{Gg}$: 0.35 (0.02 to 1.46)
Assumption of common λ and τ	1.31 (1.07 to 1.59)	1.12 (1.03 to 1.25)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)	1.68 (1.25 to 2.35)	1.25 (1.08 to 1.54)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)	1.31 (0.78 to 2.23)	1.12 (0.90 to 1.48)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)

Table 3.3 - Sensitivity analyses for λ $[\varepsilon_j= \text{logit}(\lambda_j) - \text{logit}(\lambda_1), \text{ with } j=2,3; \ \varepsilon_j \sim N(0,\tau_\varepsilon^{2})]$

META-ANALYTICAL APPROACH	White population (n=22)				Asian population (n=12)				Black population (n=6)			
	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)
$\tau_\varepsilon=0$	1.31 (1.07 to 1.59)	1.12 (1.03 to 1.25)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)	1.68 (1.25 to 2.35)	1.25 (1.08 to 1.54)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)	1.31 (0.78 to 2.23)	1.12 (0.90 to 1.48)	0.43 (0.27 to 0.60)	OR$_{GG}$: 0.34 (0.20 to 0.53) OR$_{Gg}$: 0.15 (0.07 to 0.25)
$\tau_\varepsilon=0.43$ (λ: 0.3 to 0.7)	1.30 (1.07 to 1.58)	1.12 (1.03 to 1.25)	0.45 (0.27 to 0.63)	OR$_{GG}$: 0.34 (0.21 to 0.52) OR$_{Gg}$: 0.15 (0.07 to 0.26)	1.66 (1.24 to 2.32)	1.21 (1.06 to 1.54)	0.39 (0.20 to 0.61)	OR$_{GG}$: 0.34 (0.21 to 0.52) OR$_{Gg}$: 0.13 (0.06 to 0.25)	1.30 (0.75 to 2.34)	1.12 (0.86 to 1.62)	0.47 (0.22 to 0.72)	OR$_{GG}$: 0.34 (0.21 to 0.52) OR$_{Gg}$: 0.16 (0.06 to 0.30)
$\tau_\varepsilon=1.12$ (λ: 0.1 to 0.9)	1.30 (1.07 to 1.58)	1.13 (1.03 to 1.26)	0.46 (0.27 to 0.65)	OR$_{GG}$: 0.34 (0.21 to 0.51) OR$_{Gg}$: 0.15 (0.08 to 0.26)	1.60 (1.20 to 2.21)	1.14 (1.02 to 1.49)	0.29 (0.06 to 0.60)	OR$_{GG}$: 0.34 (0.21 to 0.51) OR$_{Gg}$: 0.10 (0.02 to 0.23)	1.29 (0.70 to 2.40)	1.12 (0.79 to 1.87)	0.56 (0.11 to 0.90)	OR$_{GG}$: 0.34 (0.21 to 0.51) OR$_{Gg}$: 0.18 (0.03 to 0.38)
$\tau_\varepsilon=+\infty$	1.30 (1.08 to 1.58)	1.13 (1.03 to 1.27)	0.48 (0.28 to 0.69)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.16 (0.08 to 0.26)	1.52 (1.17 to 2.14)	1.06 (1.00 to 1.42)	0.15 (0.00 to 0.55)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.05 (0.00 to 0.21)	1.26 (0.63 to 2.50)	1.11 (0.66 to 2.14)	0.72 (0.01 to 1.00)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.23 (0.00 to 0.45)

Table 3.4 – Sensitivity analyses for $\tau\,[\phi_j = \log(\tau_j) - \log(\tau_1),$ with $j=2,3;$ $\phi_j \sim N(0,\tau_\phi^{\ 2})]$

META-ANALYTICAL APPROACH	White population (n=22)				Asian population (n=12)				Black population (n=6)			
	OR_{GG} (95% CrI)	OR_{Gg} (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR_{GG} (95% CrI)	OR_{Gg} (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR_{GG} (95% CrI)	OR_{Gg} (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)
$\tau_\phi=0$	1.31 (1.07 to 1.59)	1.12 (1.03 to 1.25)	0.43 (0.27 to 0.60)	OR_{GG}: 0.34 (0.20 to 0.53) OR_{Gg}: 0.15 (0.07 to 0.25)	1.68 (1.25 to 2.35)	1.25 (1.08 to 1.54)	0.43 (0.27 to 0.60)	OR_{GG}: 0.34 (0.20 to 0.53) OR_{Gg}: 0.15 (0.07 to 0.25)	1.31 (0.78 to 2.23)	1.12 (0.90 to 1.48)	0.43 (0.27 to 0.60)	OR_{GG}: 0.34 (0.20 to 0.53) OR_{Gg}: 0.15 (0.07 to 0.25)
$\tau_\phi=0.56$ (τ: 0.20 to 0.35)	1.31 (1.08 to 1.57)	1.12 (1.03 to 1.24)	0.44 (0.27 to 0.60)	OR_{GG}: 0.31 (0.18 to 0.51) OR_{Gg}: 0.13 (0.06 to 0.24)	1.74 (1.22 to 2.66)	1.27 (1.08 to 1.64)	0.44 (0.27 to 0.60)	OR_{GG}: 0.46 (0.21 to 0.85) OR_{Gg}: 0.20 (0.08 to 0.41)	1.30 (0.79 to 2.18)	1.12 (0.90 to 1.47)	0.44 (0.27 to 0.60)	OR_{GG}: 0.24 (0.08 to 0.62) OR_{Gg}: 0.10 (0.03 to 0.29)
$\tau_\phi=1.01$ (τ: 0.20 to 0.55)	1.31 (1.09 to 1.57)	1.12 (1.03 to 1.24)	0.44 (0.27 to 0.60)	OR_{GG}: 0.29 (0.16 to 0.50) OR_{Gg}: 0.13 (0.06 to 0.23)	1.75 (1.20 to 2.75)	1.27 (1.07 to 1.66)	0.44 (0.27 to 0.60)	OR_{GG}: 0.51 (0.20 to 0.98) OR_{Gg}: 0.22 (0.08 to 0.48)	1.30 (0.79 to 2.13)	1.11 (0.90 to 1.44)	0.44 (0.27 to 0.60)	OR_{GG}: 0.16 (0.03 to 0.63) OR_{Gg}: 0.07 (0.01 to 0.29)
$\tau_\phi=+\infty$	1.31 (1.09 to 1.57)	1.12 (1.03 to 1.24)	0.45 (0.28 to 0.60)	OR_{GG}: 0.30 (0.15 to 0.51) OR_{Gg}: 0.13 (0.06 to 0.24)	1.81 (1.17 to 3.03)	1.30 (1.06 to 1.76)	0.45 (0.28 to 0.60)	OR_{GG}: 0.62 (0.29 to 1.23) OR_{Gg}: 0.27 (0.11 to 0.61)	1.31 (0.76 to 2.34)	1.12 (0.88 to 1.52)	0.45 (0.28 to 0.60)	OR_{GG}: 0.29 (0.01 to 1.11) OR_{Gg}: 0.10 (0.00 to 0.53)

Rather than assuming a dominant model without good prior reason, the genetic model could have been estimated using the genetic model-free approach, where the genetic model is represented by the parameter λ (see § 3.4). This parameter could be treated as common across the two subgroups and the results would automatically allow for uncertainty about the mode of inheritance. If appropriate, beliefs about the genetic model could be incorporated by specifying an informative prior distribution for λ.

It has to be noted that the choice of fixed effect models by the authors of the meta-analysis is debatable, and the routine use of random effects models might be a safer option (§ 1.4). In this example, heterogeneity was not statistically significant at the cut-off significance level of 0.1, but the power of the tests of heterogeneity are low and the odds ratios for nephropathy ranged widely; from 0.61 to 4.63. By analogy with the usual approach to analysis of variance (Snedecor and Cochran, 1989), a common variance for the random effect in the two disease subgroups could be assumed unless there was good reason not to. The frequency of the D allele in controls in the retinopathy studies varies between 24% and 62%. However, the low frequencies are all found in Asian populations, while the higher frequencies are found in ethnically European populations. Treating the allele frequencies as separate random effects within the two ethnic groups would cause shrinkage and could improve estimation in the smaller studies.

3.6.2 Subgroups defined by gender

In a primary study and a meta-analysis both Furlong *et al.* (Furlong, 1999 b) and Preisig *et al.* (Preisig, 2000) investigated the association between different polymorphisms of the *MAOA* gene and bipolar affective disorder. Because the *MAOA* gene is located on the X chromosome, a difference in the genetic effect might be expected between males and females, although the evidence on whether the association is only present in females or also in men is controversial. In both papers, meta-analyses were carried out on the overall sample and then separately on males and females, and the analyses were based on alleles (one per male subject and two per female) rather than on genotype. The studies showed a statistically significant association in females but not in males. The per-allele analysis implicitly assumes a co-dominant effect in females and relies on independence of alleles within women, which is equivalent to assuming Hardy-Weinberg equilibrium. The implicit genetic model was neither justified nor tested.

While an X-linked gene would undoubtedly lead us to check for differences between men and women, it does not follow automatically that a single copy of a disease allele will increase the risk of disease differently in men and women. In this example, we cannot learn about the genetic model in women from the male subgroup so there will be no corresponding gain in information across subgroups. However, the assumption of a common genetic model for women in all of the studies will, if appropriate, still improve precision to some extent. As in the previous example, the genetic model could be estimated using the genetic model-free approach (§ 3.4), though with an X-linked gene measured on males and females, where males can only be heterozygotes (Gg), rather than modelling the parameters $\log OR_{GG}$ and λ, it might be better to parameterise the model in terms of $\log OR_{Gg}$ and λ, so that $\log OR_{GG} = \log OR_{Gg}/\lambda$.

In the studies included in the two meta-analyses there is some indication of heterogeneity; again a model with common heterogeneity between studies would seem a reasonable choice, but this time the model would parameterise the heterogeneity on $\log OR_{Gg}$ and derive the distribution for $\log OR_{GG}$. The studies include Japanese and Caucasian populations so it may be difficult to learn about allele frequencies across studies but some increase in precision would result if the allele frequencies in controls could be assumed to be the same in men and women within the same study.

3.7   Discussion

3.7.1   Handling subgroups in meta-analysis of genetic association studies

An important stage in any meta-analysis is the identification of possible sources of inconsistency or heterogeneity. Heterogeneity is often allowed for by using a random effects model, but where systematic differences between studies are suspected, subgroup analyses may be more appropriate. Such subgroups allow for an interaction between the gene and some characteristic of the studies. It is important to note that heterogeneity arises because of gene-environment or gene-gene interaction at the individual level, but where the interacting factor is common to all subjects in a study the adjustment may be made at the study level. Thus, if a gene interacts with smoking and each study recruits a mixture of smokers and non-smokers, but does not distinguish them in their reports, then, unless individual patient data can be obtained, a meta-analysis will not be able to

investigate that interaction although it may find between study heterogeneity due to varying levels of smoking in the different populations under study. In fact, the attempt to explain heterogeneity due to individual level characteristics in meta-analysis of summary data using meta-regression techniques has not proved particularly useful, since failure to find any effect is not evidence of a lack of effect (Lambert *et al.*, 2002). However, if there are some studies that only recruit smokers and others that only recruit non-smokers, or if separate estimates of the genetic effect are reported for smokers and non-smokers within individual studies, then a meta-analysis with two subgroups will help us to evaluate the gene-smoking interaction. In general, the power to detect an interaction tends not only to be very low in primary studies (Brookes *et al.*, 2004), but it might be low even in large meta-analyses (Altman and Bland, 2003), so that the decision whether to use subgroups cannot be based on purely statistical grounds. This may be why apparently unnecessary subgroup analyses featured in some of the studies reviewed in Table 3.1.

The review of published meta-analyses presented in this chapter, which looked at different types of subgroups, showed considerable variation in the methods used. A particularly common criterion for defining subgroups is disease outcome; here investigators need to decide whether different diseases are "close enough" to be pooled. This has led some to combine all studies, others to carry out subgroup analyses, or to combine the two approaches in a form of sensitivity analysis. While complete pooling of all evidence might provide misleading results, subgroup analyses can be inefficient particularly when there are only a few studies in some of the subgroups. Reporting both overall and subgroup results, although seemingly more informative, does not help the reader when the results are materially different but within the limits of sampling error. In general, what happens is that investigators either assume that all parameters vary across subgroups and so carry out completely independent meta-analyses, or they make the assumption that every parameter is common and combine all studies into one large group, perhaps also assuming a particular genetic model. Often these assumptions are not explicitly discussed and so no attempt is made to justify them, and sensitivity analyses are frequently omitted.

When synthesising evidence on gene-disease associations, concern is frequently expressed about ethnic differences, but the logic behind this concern is not always clear. Following a review of published meta-analyses, Ioannidis, Trikalinos, and Ntzani (2004)

suggested that, much more often than not, the odds ratios were similar across racial groups and that reports of ethnic differences in odds ratios should be treated with caution. Population stratification is known to be a potential source of bias *within* individual studies, and is caused by a mixture of different ethnic groups in the study population whenever the frequency of the polymorphism and the disease risk vary between ethnic groups and the study fails to match cases and controls for ethnicity (Cardon and Palmer, 2003; Freedman, Reich, Penney, *et al.*, 2004; see also § 4.2.2). However, apart from being an indicator of quality for primary studies, this is not an issue when combining study results in a meta-analysis where the difference in ethnicity is *across* studies (or study groups), within which both cases and controls come from the same ethnic group. Thus, if population stratification is not adjusted for in the primary analysis, a meta-analysis can do little to overcome this. On the other hand, studies conducted in different parts of the world may show different results because of gene-environment interaction and this might justify the use of subgroup analysis. In some of the articles reviewed, the investigators excluded studies, or performed separate meta-analyses, on the grounds of differences in allele frequency in different ethnic groups. In itself, a difference in allele frequency should not be sufficient reason for a subgroup analysis, since it does not automatically imply a difference in genetic effect. Similar variation in the approach to the use of subgroups is present in relation to other factors, including gender, environmental and life-style factors such as smoking, and the study design, for example, typing method.

It is important to understand that in genetic association studies, although gene-environment interactions can have a strong influence, confounding should not be a problem. At the individual level, genotype is determined by a seemingly random process, which is analogous to a randomised trial, and precludes most forms of confounding within the study (§ 5.3). Confounding at the study level is theoretically possible, as factors that influence the risk of disease may vary between studies (Salanti, Sanderson, and Higgins, 2005), but this can be adjusted for by including study specific parameters if information on such factors is available, and is unlikely to be a problem.

Although there is little empirical evidence of the relative importance of the different potential sources of heterogeneity (Salanti, Sanderson, and Higgins, 2005), any observed heterogeneity needs to be investigated and appropriately dealt with. The approach to the adjustment for study level interaction proposed in this chapter is based on recognising

that there are many parameters in genetic association studies. Even when the magnitude of the genetic effect differs across subgroups, it may be that other parameters are common. The example of the effect of the *AGT* polymorphism on hypertension in different ethnic groups shows how the assumption of a common mode of inheritance and common within group heterogeneity of the genetic effect can greatly increase the precision of the estimates of the genetic effect in small subgroups. These assumptions also bring the size of the gene-disease association in Blacks into very close agreement with that found in Whites. Moreover, the sensitivity analyses show that the improvement in precision persists even when the secondary parameters are similar rather than identical. The re-analysis might lead us to question the conclusions of the authors of the original meta-analysis, who tentatively concluded that, unlike in Whites and Asians, "genotype did not predict hypertension in black subjects" (Sethi, Nordestgaard and Tybjaerg-Hansen, 2003).

The methods described in this paper could easily be extended to accommodate assumptions about other secondary parameters that might also improve the precision of the estimate for the gene-disease odds ratio. These include the assumption of Hardy-Weinberg equilibrium in control groups and the assumption of similarity of the allele frequencies in controls; these assumptions might be made within subgroups or across subgroups. None of the assumptions can be used indiscriminately, but all need to be justified by inspecting the data and calling on prior biological knowledge. However, it seems reasonable to suggest that parameters are assumed common, unless there is evidence to the contrary. Moreover, the models presented and used to assess the sensitivity of the results to different assumptions on the similarity of secondary parameters across subgroups could be adapted to incorporate prior beliefs. These prior beliefs would form the basis for informative prior distributions on the expected variation of the secondary parameters across subgroups.

One limitation of the approach presented in this chapter is that a convincing statistical measure to help judge the relative fit of each model could not be found. The attempt to use the Deviance Information Criterion (DIC) to compare models did not appear to be a successful approach. The DIC is a Bayesian analogue of Akaike's information criterion (AIC) and was proposed by Spiegelhalter et al. for comparing complex Bayesian models (Spiegelhalter *et al.*, 2002). For the *AGT* example, the maximum difference in DIC

between any two models was 1.9, while a minimum observed difference of 3 is advocated to be needed before a true difference is suggested (Spiegelhalter *et al.*, 2002). Although the DIC does not indicate a difference in fit between any of the models that we tried, we do not feel confident in stating that all models were equivalent since these findings might be due to a lack of sensitivity of the DIC. At present, there is still very limited information in the literature on the use of DIC and further experience is needed before its routine use can be advocated.

A natural extension to our models would be to allow for the correlation induced by those studies that provide data in more than one subgroup. This happened in the *AGT* re-analysis where three studies reported separately on both Blacks and Whites. Although the resulting correlation could easily be accommodated in our models, its impact needs further investigation. It might be anticipated that the precision of the estimate of such correlation would tend to be poor, so that the influence on the results of the choice of the prior distribution for the correlation parameter would not be negligible.

It would be possible, and quite in keeping with the usual practice in Bayesian analysis, to model the subgroups using a hierarchical model for the genetic effect size (Sutton *et al.*, 2000). A parameter that varies across subgroups can be thought of as a realisation from a higher-level distribution. Such a model would introduce shrinkage and enable estimates for one subgroup to benefit from the data in the others (Sutton and Abrams, 2001). However, this approach was not adopted since subgroups are often of interest in their own right and are not naturally viewed as being selected from some larger population. For instance, in the *AGT* example investigators might be specifically interested in the genetic effects in Whites, Blacks and Asians, and not think of these races as being selected as representative of some higher-level population of races. This is not to say that hierarchical structures do not have potential uses. One might, for instance, want to allow for differences between countries; for instance, Blacks in USA, Blacks in Nigeria, Blacks in South Africa etc. In these circumstances a hierarchical country effect would be very natural.

3.7.2 Conclusions

Investigators carrying out meta-analyses of genetic association studies should give more

thought to the assumptions that they are willing to make in the presence of subgroups. These assumptions should be made explicit so that the reader can critically evaluate them, and an effort should be made to test whether these assumptions hold, or, if this is not possible, sensitivity analyses should be performed. Apart from the parameters representing the genetic effect, there are other secondary parameters in the meta-analysis model for which information could be borrowed across subgroups, such as the genetic model and the between-study heterogeneity of the genetic effect. The work presented in this chapter suggests that assuming that such parameters are common across subgroups is a reasonable strategy, unless there is evidence of the contrary, and provides a potential gain in the precision of the estimate of the genetic effect. Whenever relevant expert opinion is available, stronger assumptions about the parameters of interest and/or secondary parameters should be incorporated through informative priors.

4 HARDY-WEINBERG EQUILIBRIUM

4.1 Chapter overview

This chapter addresses the issue of the assessment of Hardy-Weinberg Equilibrium (HWE) in a meta-analysis of genetic association studies, where departures from HWE are used as a proxy for problems associated with the quality of the studies included. In § 4.2 the general meaning and applications of HWE law are presented, the possible causes for departures from HWE discussed, and the implications for genetic association studies highlighted. A review of the methods used in the literature to assess departure from HWE, both in terms of testing the presence of departure and estimating its magnitude, is presented in § 4.3. In § 4.4 the practical implications of approaching the issue of HWE either from the perspective of hypothesis testing or from that of estimation are shown by re-analysing a dataset of 37 meta-analyses reported in 30 published papers, which included 488 primary studies evaluating 516 comparisons. Having investigated this, § 4.5 addresses the issue of what might be the best strategy in practice for dealing with departures from HWE in studies included in the meta-analysis. This is investigated through simulation work aimed at showing the impact, in terms of bias and precision of the estimate of the genetic effect, associated with different ways of identifying studies with departures from HWE and excluding them from the meta-analysis. Discussion of the results and conclusive remarks are presented in § 4.6.

4.2 Introduction

4.2.1 The Hardy-Weinberg equilibrium law

HWE law states that if the two alleles of a biallelic locus, G and g, with frequency p_G and p_g respectively and $p_G + p_g = 1$, are in equilibrium in a population, then the proportions of subjects with genotypes GG, Gg and gg follow the product rule, and thus will be p_G^2, $2p_G p_g$ and p_g^2 respectively (Sham, 2001). This law implies that in each subject alleles are inherited as statistically independent of one another if the study population is in HWE. HWE law is used in a variety of contexts. In evolutionary theory, departures from HWE, that is departures of observed from expected genotype

frequencies, are estimated to evaluate the impact of evolutionary forces in producing changes in allele frequencies of a population over time (Hartl and Clark, 1997). In forensic science, HWE law is used in the analysis of DNA evidence for human identification and paternity studies, where HWE is assumed in order to calculate expected genotype frequencies based on observed allele frequencies (Devlin, Risch, and Roeder, 1991). In the study of Mendelian recessive diseases such as sickle cell disease, hereditary haemochromatosis or congenital adrenal hyperplasia, where only subjects homozygous for the mutation are affected, HWE law is used to estimate the carrier rate in normal individuals in a population where all diseased individuals are observed. The number of carriers, i.e. subjects heterozygous for the mutation, might be important in order to implement and test screening methods for the disease (Thomas, 2004).

In the field of genetic association studies, HWE law is used for two different purposes. First, testing for departure from HWE in a gene bank of individuals affected by a disease has been proposed as a relatively efficient method for "rapid gene hunting", when searching for possible polymorphisms associated with the disease (Lee, 2003). In a population which is in HWE, and under the assumption that the operating genetic model is not co-dominant, the genotype proportions in affected individuals are expected to deviate from HWE as a result of the association itself, with the magnitude of the departure depending on the allele frequency and the underlying genetic model (Wittke-Thompson, Pluzhnikov and Cox, 2005). Second, HWE law has been used for study quality control; current practice views departures from HWE in controls as a proxy for poor quality and thus possible indication of threats to validity of study results. A number of problems in the design and conduct of genetic association studies can lead to departures from HWE, the most important being population stratification, genotyping error, and selection bias particularly in the choice of controls (Sham, 2001). These departures from HWE observed in genetic association studies have to be distinguished from genuine departures in the study population.

4.2.2 Causes of departures from HWE

Possible genuine causes of deviation from HWE in the study population of a genetic association study include non-random mating, migration and selection (Khoury, Beaty

and Cohen, 1993; Hartl and Clark, 1997; Sham, 2001). Non-random mating means that any two individuals in a population are not equally likely to mate, due to assortative mating, inbreeding, or outbreeding. *Assortative mating* indicates that individuals have a tendency to mate with others with similar phenotypes, being these either physiological variables (e.g. height and skin colour) or pathological conditions (e.g. deafness), and thus with similar genotypes. Such phenomenon, which leads to an increase in the frequency of homozygotes at the expense of heterozygotes over successive generations, occurs to some extent in nearly all communities. The opposite tendency is called "negative assortative" or "dissortative" mating, which leads to an increase in heterozygotes at the expense of homozygotes. *Inbreeding* means mating between relatives, that is mates related by ancestry; it is more likely to occur in small populations and leads to increased homozygosity. *Outbreeding* refers to mating between individuals from different populations that may have different genetic compositions, and leads to an increase in the frequency of heterozygotes. However, one generation of random mating in a population is sufficient to produce HWE. Another cause of departure from HWE directly related to outbreeding is *migration*, that is movement of individuals between populations with consequent migration of new alleles. Finally, genuine departure from HWE in a population may be caused by selection. *Selection* refers to a natural process by which genotypes that lead to either increased fertility (ability to procreate) or increased viability (ability to survive from fertilisation through to adulthood) are more likely to contribute to the next generation than genotypes that do not. An example of this is what is called "heterozygote advantage", where heterozygotes have, or have had at some point in the past, a reproductive advantage over wild homozygotes, as documented, for instance, for sickle cell anemia (Hartl and Clark, 1997).

To some extent, these genuine causes of departure from HWE do play a role in any real population, so that perfect HWE is unrealistic; however they usually result in minimal deviations. On the other hand, problems with the design and conduct of genetic association studies can lead to observed departures from HWE that do not reflect deviations in the study population, and the magnitude of which reflects the extent of the problem (Sham, 2001; Khoury, Little, and Burke, 2004). Such methodological problems, which include population stratification, genotyping error, and selection bias,

need to be investigated and addressed both in the context of primary studies and meta-analyses.

Population stratification is a well-known problem of genetic association studies caused by a mixture of different ethnic groups in the study population, when the frequency of the polymorphism and the disease risk vary between ethnic groups (Cardon and Palmer, 2003; Freedman, Reich, Penney, *et al.*, 2004). However, its actual confounding effect is a matter of intense debate. On one side, some authors suggest that sound epidemiological methods, such as matching cases and controls for ethnicity and proper selection of controls, address this problem, and therefore argue that well-designed and well-conducted genetic association studies are robust against bias from population stratification (Wacholder, Rothman, and Caporaso, 2002). On the opposite side, other authors view population stratification as a potentially important source of bias even in high-quality genetic association studies, due to the presence of residual "cryptic stratification". Cryptic stratification may be caused by a difficulty in defining suitable ethnicity categories or by the presence of mixed-ethnicity families. To address this problem they suggest either using genetic markers of ethnicity (genomic control methods) instead of the traditional self-reported ethnicity, or even replacing the use of unrelated controls with that of family-member controls in genetic association studies (Thomas and Witte, 2002; Freedman *et al.*, 2004). Genomic control methods are based on the principle that if disease frequency varies across ethnic groups, then *any* genetic variant associated with ethnicity will appear to be associated with the disease. By genotyping a number of unlinked markers, i.e. polymorphisms not associated with the disease, in addition to the polymorphism of interest, spurious associations due to population stratification could be estimated and adjusted for (Freedman *et al.*, 2004).

Genotyping error is another important cause of deviation from HWE (Gomes *et al.*, 1999; Hosking *et al.*, 2004). It is a mistake in the laboratory identification of a subject's genotype due to sample contamination, observer variability whenever visual inspection is required, or technical problems related to the specific technique (Xu *et al.*, 2002; Little *et al.*, 2002). The impact of genotyping error on the estimate of the genetic effect depends on whether the error is random, which leads to loss in precision of the estimate particularly in the presence of rare alleles (Kang, Gordon, and Finch, 2004), or

systematic, which causes bias. Systematic error may be due to identification failure rates differing between homozygous and heterozygous genotypes, preferential detection of one allele, or non-blinded reading of laboratory results. Reported figures of the frequency of genotyping errors vary from 1% up to 30% (Akey *et al.*, 2001). The difference in the error rate associated with different genotyping techniques is the main, but not the only, reason for such variability, which might also be explained by other differences occurring in the various steps of the genotyping process. In a study of the association between p53 polymorphism and risk of human papillomavirus-induced cervical cancer, the effect of inter-laboratory variation was investigated by dividing each specimen into aliquots to be analysed by three different laboratories. Although all laboratories adopted the same kind of technique, differences in the testing protocol and other conditions related to testing caused a disagreement in assigning genotype status which ranged from 22 to 29% (Makni *et al.*, 2000). This highlights the need for quality control protocols to be implemented in primary genetic association studies, where no family genotype data are available and thus no Mendelian inheritance checks can be performed on the results of genotyping. It has been suggested that investigators should routinely include blanks or duplicates on the plate for analysis and perform internal validation by checking the scoring of the alleles (Xu *et al.*, 2002; Little *et al.*, 2002). The availability of such information from a number of studies would greatly help clarify the extent of the problem and its impact on the estimates of the genetic effect. It has to be noted that the relative impact of even a low error rate on the estimate of the genetic effect might not be negligible in the study of gene-disease associations for complex diseases; here, gene effect sizes tend to be small with odds ratios in the range of 1.1 to 1.4 (§ 1.6), and a loss in precision due to genotyping error might decrease the power of the study to detect the association (Gordon *et al.*, 2002).

Selection bias can occur in genetic association studies based on a case-control study design, which is by far the most common, and can be due to inappropriate choice of either cases or controls (Edland, Slager and Farrer, 2004). The typical source of case selection bias is the use of prevalent cases ("prevalent case bias"), which, in chronic diseases, can lead to over-representation of genetic factors associated with survival with disease or under-representation of genetic prognostic factors. Control selection bias is more common and may be due to a number of problems, such as; controls chosen from

a geographically different population with a different mix of ethnic groups, which leads to population stratification; controls chosen from a different birth cohort ("cohort effect") in a population where migration patterns might have changed over time thus producing differences in gene frequencies due to demographic shifts.

4.2.3 Detecting departures from HWE in genetic association studies

The way HWE law has been used so far for quality control in genetic association studies consists of testing whether the genotype frequencies observed in the study population among disease-free subjects correspond to the proportions expected under HWE (i.e. p^2, $2pq$ and q^2). However, such an approach has been criticised for the reason that all tests for HWE have very low power to detect any departure from HWE (Salanti et al., 2005), and even when they can detect a statistically significant departure, they do not provide information on its magnitude. In general, there is a substantial literature that stresses the advantages of estimation over hypothesis testing in medical research, where clinically important effects may be statistically non-significant if the sample size is inadequate (Rothman, 1978; Gardner and Altman, 1986). Another important argument against hypothesis testing for departures from HWE is that real populations are never in perfect HWE anyway (§ 4.2.2). Moreover, if the observed departure from HWE in a study is caused by methodological problems of the study, then its magnitude is likely to depend on the extent of the problem. A number of measures of departure from HWE have been developed and could be used to estimate the magnitude of the departure, such as the inbreeding coefficient, the disequilibrium parameter, and the alpha parameter.

What happens in practice in genetic association studies, however, is that often no information at all is reported regarding HWE. The evidence on reporting rates ranges from 20% to 69% in primary genetic association studies published in different non-genetics journals over the last few years (Xu et al., 2002; Nemeth et al., 2004; Kocsis et al., 2004a; Kocsis et al., 2004b; Gyorffy, Kocsis, and Vasarhelyi, 2004a; Gyorffy, Kocsis, and Vasarhelyi, 2004b; Bardoczy et al., 2004). A reporting rate of 29% for any information on HWE has recently been shown even in studies published in three high-profile genetics journals in 2002 (Salanti et al., 2005). The same problem is reflected in meta-analyses of genetic association studies, as suggested by a review of 37 meta-

analyses by Attia and colleagues, where information on HWE in the studies included was reported only in 24% of the meta-analyses (Attia, Thakkinstian, and D'Este, 2003). Moreover, in the primary studies and meta-analyses reviewed by these authors where information on HWE was reported, such information was always limited to hypothesis testing. It is clear that there is a contradiction between the perception of the importance of departures from HWE as a proxy for quality problems in genetic association studies, and the fact that, in practice, there is little attempt to assess HWE in published primary studies and meta-analyses.

4.3   Review of the literature

4.3.1   Testing for HWE

A number of alternatives for testing for HWE have been proposed in the literature, the majority based on a frequentist approach (Emigh, 1980; Hernandez and Weir, 1989), and others more recent based on sophisticated Bayesian methods which allow appropriate modelling of the data (Shoemaker, Painter, and Weir, 1998; Montoya-Delgado et al., 2001; Rogatko, Slifker, and Babb, 2002).

Within the framework of classical hypothesis testing, most of the methods proposed are based on either Pearson's chi-square or Fisher's exact test. The chi-square test is problematic in the presence of sparse data, which can be due to rare alleles with frequency of one or more genotypes close to zero, small sample size, multi-allelic polymorphisms, or a combination of these. In these situations the chi-square asymptotic distribution becomes inadequate and adjustments of the test, such as the use of Yate's correction, have to be adopted to improve the approximation of the sampling distribution to chi-square. However, most authors agree that in such situations an exact test should be used instead (Emigh, 1980; Hernandez and Weir, 1989), although computation is tedious. Among the different forms of exact test that have been proposed, the most well known is Haldane's exact test, or conditional exact test, which is analogous to Fisher's exact test for contingency tables. A third method proposed for testing for HWE is the likelihood ratio test (Elston and Forthofer, 1977), but this suffers of the same problem of the chi-square test since it approximates discrete data by a continuous distribution (chi-square distribution) and thus is based on asymptotic results

which require the sample size to be large. All frequentist methods presented suffer, to different extent, from the additional problem of the dependence of the power of the test, and thus the statistical inference, on the nuisance parameter p indicating the frequency of the allele of interest.

It has to be noted that, although much research has been carried out to develop appropriate methods for testing HWE, in practice the only method used remains the chi-square test. Exact tests are not used either in primary genetic association studies (Salanti et al., 2005) or in meta-analyses of genetic association studies (Attia, Thakkinstian, and D'Este, 2003), despite the fact that studies investigating polymorphisms with rare alleles are relatively common. Among the 239 associations recently evaluated by Salanti and colleagues, 74 (31%) comprised data with low allele frequencies, but in none of these the investigators discussed the limitations of chi-square asymptotic inference (Salanti et al., 2005).

The main problem of testing for HWE is that the power of any test, which mainly depends on the sample size of the study and the frequency of the allele of interest, tends to be very low in most genetic association studies. Salanti and colleagues (2005) showed how, for example, only about 6% of the tested associations had a power greater than 80% to detect a departure from HWE corresponding to an inbreeding coefficient of ±0.10 (§ 4.3.2). Therefore, small studies will rarely show statistically significant departures even in the presence of considerable deviation from HWE. On the other hand, studies with large datasets might show statistically significant departures even when the magnitude of the deviation is limited, which is of no interest in any of the applications of HWE law, be it in the field of genetic association studies, forensic science or any other. For this reason the need for methods aimed at estimating the magnitude of deviations from HWE has long been recognised and a number of methods have been proposed (§ 4.3.2), but, again, estimates of departures from HWE are usually not reported in published primary studies nor meta-analyses of genetic association studies (Attia, Thakkinstian, and D'Este, 2003; Salanti et al., 2005).

4.3.2 Assessing the magnitude of departures from HWE

Assessing the magnitude of the departure from HWE observed in a genetic association study is very important in order to evaluate the presence and extent of methodological problems which might cause it. In practice, though, there seems to be a fundamental contradiction in the way investigators of genetic association studies handle this issue. On one side, the concept that departures from HWE, independently from their causes, can bias the estimate of the genetic effect is widely accepted. However, this has only been demonstrated for the per-allele analysis (Sasieni, 1997; Schaid and Jacobsen, 1999), where HWE is indeed an assumption of the analysis (§ 2.3), while for genotype-based analyses the possibility of bias introduced by departures from HWE is mostly based on theoretical considerations. On the other side, there is clearly no attempt to estimate the magnitude of departures from HWE in published primary studies and meta-analyses (Attia, Thakkinstian, and D'Este, 2003; Salanti *et al.*, 2005). If departures from HWE can indeed bias the study results, then surely such an influence depends on the extent, and possibly direction, of the departure, and methods aimed at correcting for departures should be considered in the analysis (§ 4.6.3).

Different measures of departures from HWE have been proposed, which again can be estimated using a frequentist or a Bayesian approach. Among these, the main are:

1) Inbreeding coefficient (f)

Despite its name, this parameter is used to indicate departure from HWE due to any cause, not only inbreeding (Weir, 1996). Denoting the genotype frequencies for the two homozygous groups as P_{GG} and P_{gg}, and the allele frequencies p_G and p_g, then the inbreeding coefficient, f, is:

$$f = P_{gg}/p_g + P_{GG}/p_G - 1$$

The lower bound of f depends on the allele frequencies:

$$\max[-p_g/p_G, -p_G/p_g] \le f \le 1$$

2) Disequilibrium parameter (D)

The disequilibrium parameter, described by Hernandez and Weir (1989), is defined as:

$$D = P_{gg} - p_g^2$$

The bounds on D are functions of the allele frequencies:

$$\max[-p_g^2, -p_G^2] \leq D \leq p_g p_G$$

3) Alpha

This parameter, though less well known than the previous two, has been proposed by Lindley in the context of a Bayesian approach (Lindley, 1988):

$$\text{alpha} = \frac{1}{2}\log (4P_{gg}P_{GG}/P_{Gg}^2)$$

The bounds on alpha are independent from the allele frequencies:

$$-\infty < \text{alpha} < \infty$$

The lack of lower and upper bounding and the independence from the allele frequencies make this parameter particularly suitable to statistical modelling.

For all the three parameters, zero corresponds to perfect HWE, positive values to excess of homozygotes and negative values to deficiency of homozygotes relative to HWE. Although real populations are never in exact HWE, the extent of genuine deviations from HWE is usually minimal. In a report for use in the evaluation of forensic DNA evidence in 1996, the United States National Research Council suggested an inbreeding coefficient of 0.03 as the upper limit of departure from HWE in human populations, from which the disequilibrium parameter can be calculated as $D=0.03p_g p_G$ (National Research Council, 1996).

4.4   Evidence of departure from HWE in meta-analysis

4.4.1   A dataset of 37 published meta-analyses

A dataset of 37 meta-analyses published from 2000 and including 488 genetic association studies was obtained searching the HuGE Reviews archive (http://www.cdc.gov/genomics/hugenet/reviews_arch.htm). The Human Genome Epidemiology Network (HuGE Net) is an international collaboration of individuals and organisations from different background, founded in 1998 with the aim of developing and disseminating population-based human genome epidemiological information (Centre for Disease Control and Prevention, 1999). From 2000 the collaboration has maintained a list of meta-analyses published under its auspices, preferentially in the *American Journal of Epidemiology* among other journals (HuGE reviews), as well as any other published genetic meta-analyses. Among those listed on the HuGE Net

Website on 15 March 2005, papers were selected if they; 1) were written in English; 2) considered a binary disease outcome; 3) reported the raw data on genotype frequencies for cases and controls; 4) included more than 5 studies or comparisons, when studies reported separate figures for different subgroups. Comparisons with zero cells for *both* cases and controls in a specific genotype group were not considered, since no meta-analytical methods are available to deal with such situations when using a relative measure of the effect size such as the odds ratio (Sweeting, Sutton, and Lambert, 2004). Comparisons with zeros for both cases and controls were observed when the frequency of one allele was extremely low (≤ 0.05) or when a relatively low allele frequency was coupled with very small sample size. If more meta-analyses were available assessing the same association, the one either more recent or which included the greater number of studies was chosen. Inclusion criteria were evaluated based either on the abstract or, more often, on the full text of the paper when the abstract did not contain sufficient information to establish eligibility. Among all 243 papers listed, only 30 fulfilled the inclusion criteria, and the reasons for exclusions are reported in Figure 4.1. It is interesting to note how 71% of the papers (n=172) were excluded because they did not report raw data on the genotype frequencies for cases and controls in the studies included. The 30 papers considered reported on 37 meta-analyses, which included 488 primary studies. After excluding comparisons with zero cells for both cases and controls in one or more genotype groups, the number of comparisons evaluated was 516. The characteristics of the 37 meta-analyses in terms of number of studies included and comparisons considered, average allele frequency for the allele of interest and average number of cases and controls are reported in Table 4.1 (complete references for the studies in Table 4.1 are reported in Appendix 3).

All meta-analyses considered focused on bi-allelic polymorphisms; in the following sections the two alleles will be referred to as G and g, where G indicates the allele thought to be associated with the disease. The three genotypes GG, Gg and gg provide two odds ratios of interest, the odds ratio of GG versus gg (referred to as OR_{GG}) and the odds ratio of Gg versus gg (referred to as OR_{Gg}).

4.4.2 Statistical methods

Hypothesis testing for HWE in the 516 gene-disease associations was performed using Fisher exact test, implemented by the `genhwcci` command in Stata 8 (Cui, 2000). Statistical significance was considered at the nominal level of 0.05. The parameter alpha was used for estimating the magnitude of departures from HWE, due to the fact that this parameter is not bounded, that is it has not got bounds dependent on allele frequency (§ 4.3.2). However, the inbreeding coefficient and the disequilibrium parameter were also calculated in order to assess the degree of agreement between the three measures of departure. The parameter alpha was estimated in WinBUGS, due to the simplicity of obtaining its standard error using MCMC methods, using vague prior distributions for all parameters. In particular, a diffuse normal prior distribution, $Normal(0,10000)$, was used for alpha, and the parameter estimates were obtained using a burn-in of 10,000 iterations followed by chains of length 50,000; convergence was assessed using visual inspection of trace plots and running multiple chains with different initial values, length of burn-in and length of sample (§ 1.5.2). The point estimates for the inbreeding coefficient and the disequilibrium parameter used for comparison with the point estimates of alpha were calculated using Stata.

4.4.3 Results

The distribution of p-values for the 516 associations, with values expressed as logarithms to base 10, is shown in Figure 4.2. One study with $\log_{10}$ (p-value) of –23.3, corresponding to a p-value of 5.0^{-24}, was removed in order to better show the distribution of the other 515. This was a very large study with more than 2,000 patients, which explains the extremely low p-value, and had an alpha value of 0.67. The percentage of genetic association studies showing statistically significant deviation from HWE ($p \leq 0.05$) was only 10.5% (54/516). These are the studies that researchers might want to exclude, either in their main analysis or in sensitivity analyses, on the assumption that they are studies of poor quality due to genotyping error, population stratification or problems with selection of cases or controls. However, because the nominal level chosen for statistical significance is 5%, only 5.5% out of the observed 10.5% of the studies will be truly statistically significant, with a false discovery rate of 47.6%. In other words, among the 54 studies with statistical significant deviation from
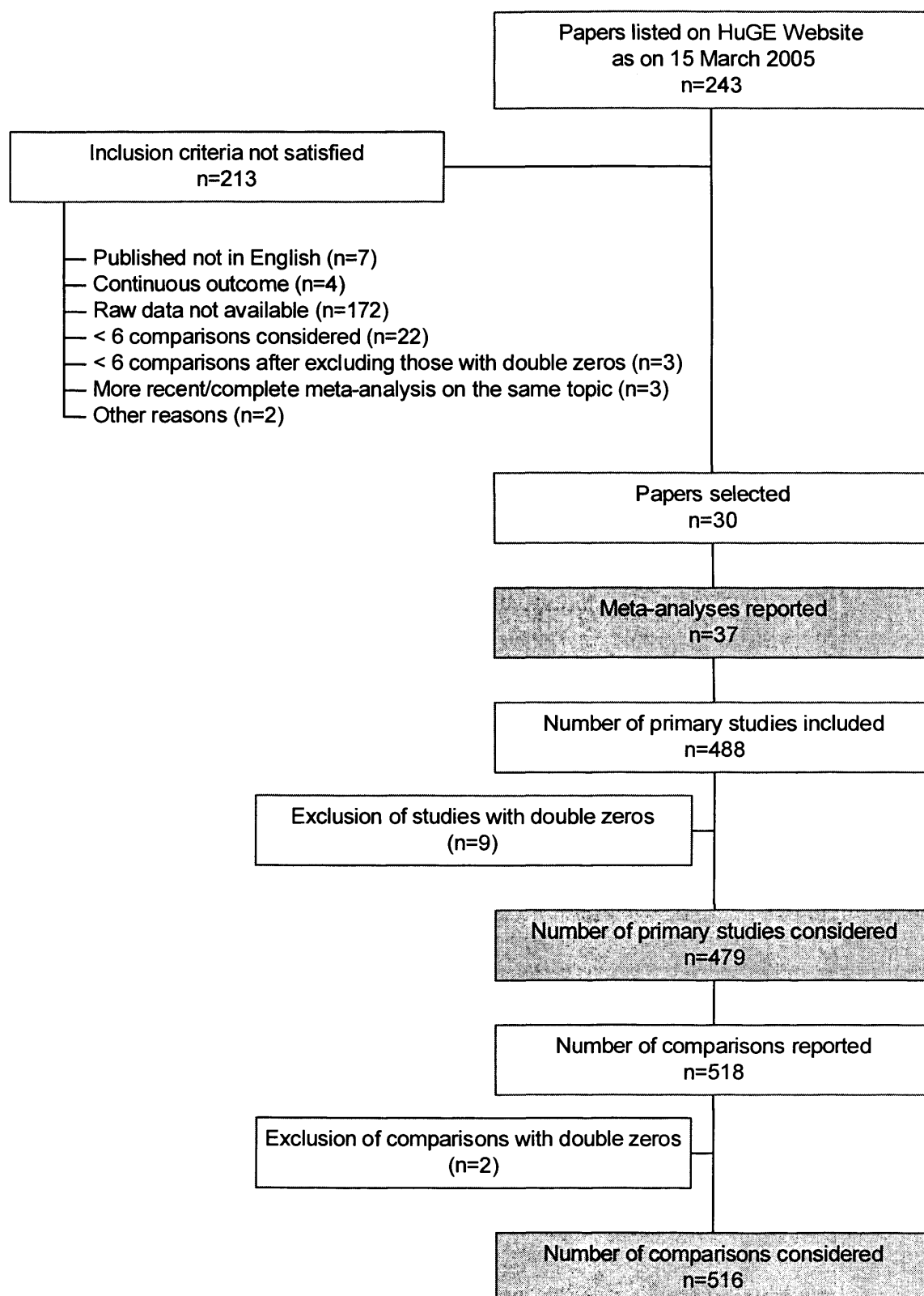
Figure 4.1 - Flow chart of the inclusion and exclusion of papers from HuGE Reviews archive

TABLE 4.1 - Characteristics of the 37 meta-analyses in the dataset, which included 479 studies and 516 comparisons. For complete references see Appendix 3

Author, year	Association evaluated	N. of studies	N. of comparisons	Average allele frequency	Average N. cases/controls
Anguelova, 2003	5-HT2A 102 T/C & suicidal behaviour	9	9	0.56	66/111
	5-HTT & suicidal behaviour	12	12	0.45	86/153
Benhamou, 2005	ERCC2*/XPD* Asp312Asn & lung cancer	6	6	0.31	481/514
	ERCC2*/XPD* Lys751Gln & Lung cancer	7	7	0.33	482/554
Botto, 2000	MTHFR C677T & spina bifida in people affected	12	12	0.31	104/219
	MTHFR C677T & mothers of patients with spina bifida	7	7	0.31	86/193
Brennan, 2004	ADH1C & Head and neck cancer	7	7	0.60	189/251
Camp, 2002	Ser217Leu & prostate cancer	6	6	0.29	371/325
Di Castelnuovo, 2001	GPIIIa PlA & coronary risk	26	29	0.14	258/261
Efstathiadou, 2001	COLIA1 Sp1 & fractures	13	13	0.16	73/208
Ertekin-Taner, 2005	PLAU_1 & Alzheimer's disease	6	6	0.26	138/146
Feyler, 2002	MPO & lung cancer	7	10	0.23	269/333
Ioannidis, 2002	PvuII & fractures	7	8	0.45	82/197
Jee, 2000	MTHFR C677T(V) & CAD + myocardial infarction	16	18	0.35	230/305
Kaklamani, 2003	TGFBR1 & cancer	7	7	0.08	305/230
Kavvoura, 2005	NEUROD1 Ala45Thr & Type I diabetes mellitus	6	6	0.23	152/178
	NEUROD1 Ala45Thr & Type II diabetes mellitus	6	6	0.19	231/215
Kehoe, 2003	ACE Alu indel & Alzheimer's disease	18	21	0.45	155/168
Lawlor, 2004	PON1 Q192R & coronary heart disease	38	39	0.35	281/438
Lee, 2002	mEH exon 3 & lung cancer	9	11	0.39	189/280
	mEH exon 4 & lung cancer	9	11	0.20	189/279
Lin, 2004	5-HTTLPR & suicide	17	17	0.53	89/143
Matakidou, 2003	TP53 & lung cancer	13	17	0.36	205/229
Noble, 2003	DRD2 TaqI A & alcoholism	20	20	0.15	87/71
Ntais, 2003 (a)	CYP17 & prostate cancer	10	12	0.38	200/230
Ntais, 2003 (b)	SRD5A2 V89L & prostate cancer	9	12	0.35	213/279
Ntais, 2003 (c)	VDR TaqI & prostate cancer	14	17	0.33	110/167
	VDR Poly(A) & prostate cancer	5	7	0.33	56/99
Ntais, 2004	CTSD & Alzheimer's disease	12	13	0.08	198/206
Ray, 2002	C677T MTHFR & venous thromboembolism	31	31	0.33	159/254
Rujescu, 2003	A218C TPH & suicidal behaviour	12	12	0.47	172/99
Schena, 2001	ACE I/D & IgA nephropathy	7	7	0.48	100/100
Sethi, 2003	AGT M235T & hypertension	38	40	0.58	316/310
	AGT M235T & ischemic heart disease	21	21	0.54	342/630
Ye, 2002	CYP17 MspA1 & breast cancer	14	15	0.40	280/312
Zhu, 2000	PlA1/A2 & myocardial infarction	16	18	0.14	228/272
Zintzaras, 2005	GLUT1 & diabetic nephropathy	6	6	0.34	118/125

HWE, we would expect to see 26 which are false positives. As for the estimate of the magnitude of departures from HWE, the degree of agreement between the three measures (inbreeding coefficient, disequilibrium parameter and alpha), which is represented graphically in Figure 4.3, appeared to be good for all combinations, with the exception of a few outliers.

Figure 4.2 – Distribution of the p-values (expressed as logarithms to base 10) in 515 associations. The vertical line corresponds to a p-value of 0.05
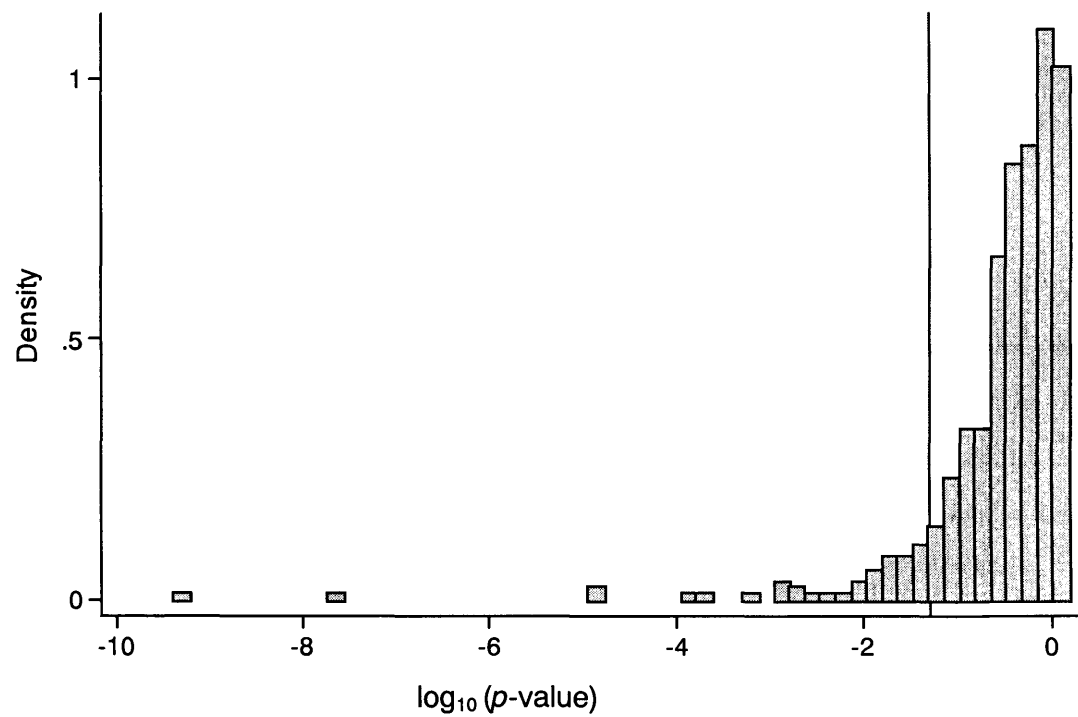
Figure 4.3 – Agreement between the three measures of departure from HWE
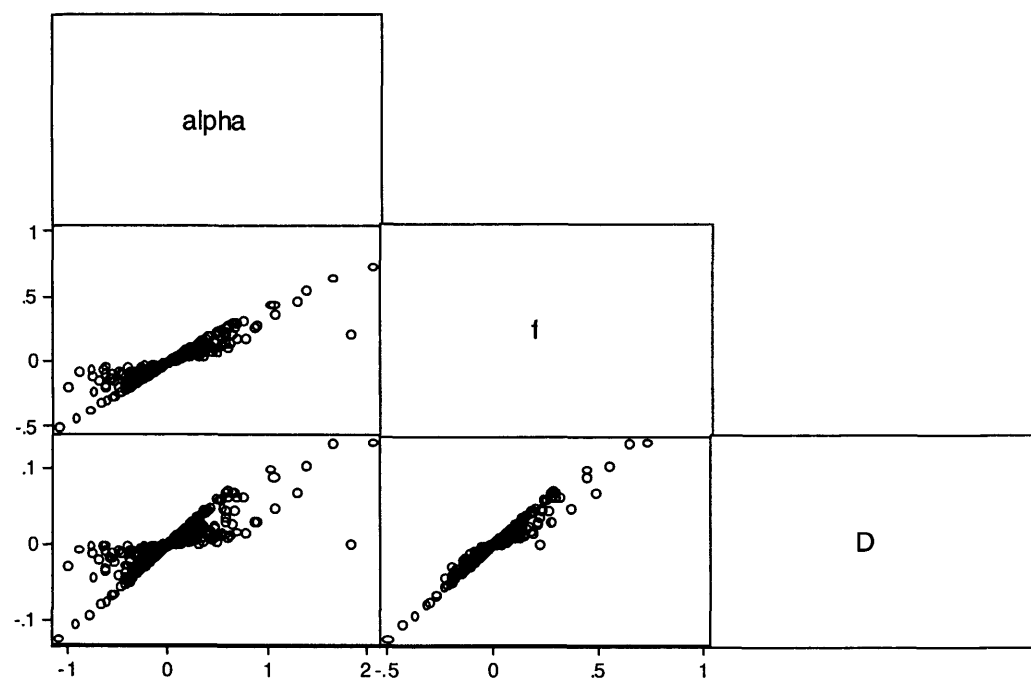
The values of alpha in the 516 comparisons showed a distribution very close to normal and centred on zero (mean 0.01; standard deviation: 0.34), as illustrated in Figure 4.4. The vast majority of studies appeared to be either in HWE or very close to it, with only four studies (0.8%) showing a value lower than −1 and five studies (1.0%) higher than 1. However, these estimates of alpha were very imprecise in almost all studies. The average standard error for alpha was 0.24, ranging from 0.02 to 0.98.

Given that the magnitude of departures from HWE reflects the extent of methodological problems of the study, it might be expected that large studies, which are usually associated with higher quality, might tend to show lower values of alpha. In order to investigate this aspect, the values of alpha observed in the 516 associations were plotted against the sample size of the studies, as shown in Figure 4.5. The scatter plot indicates a clear association; very large studies show values of alpha very close to 0, while smaller studies show either positive or negative values of alpha, the absolute magnitude of which increases with the decrease in sample size.

Figure 4.4 – Distribution of departures from HWE (alpha) in the 516 associations
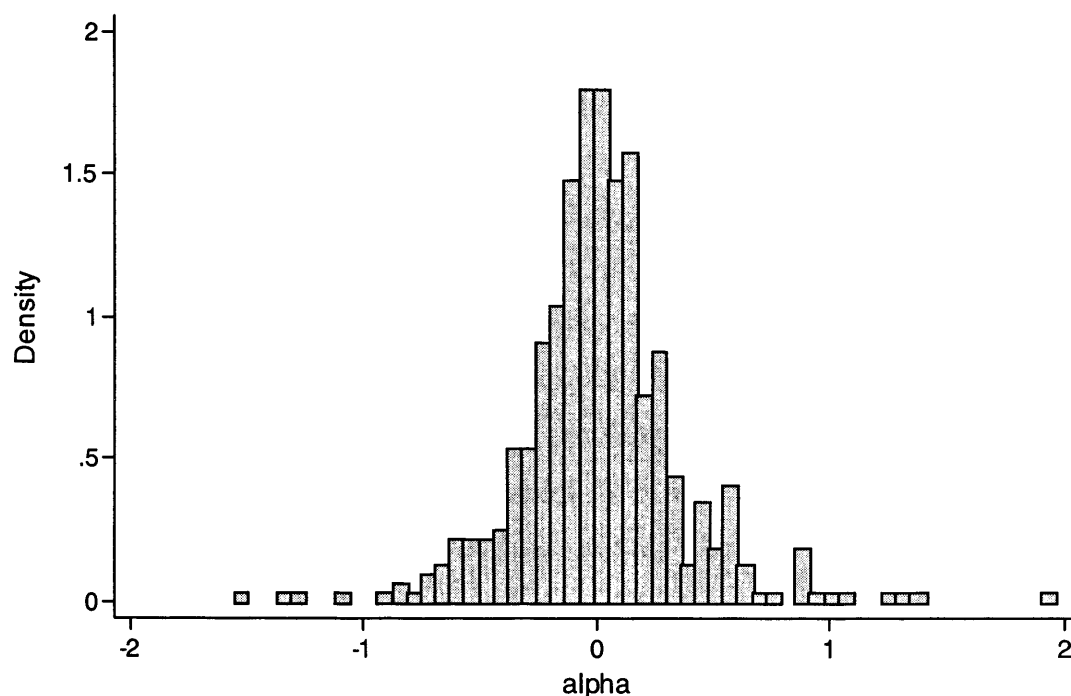
Figure 4.5 – Dependence of alpha on the total sample size in the 516 associations
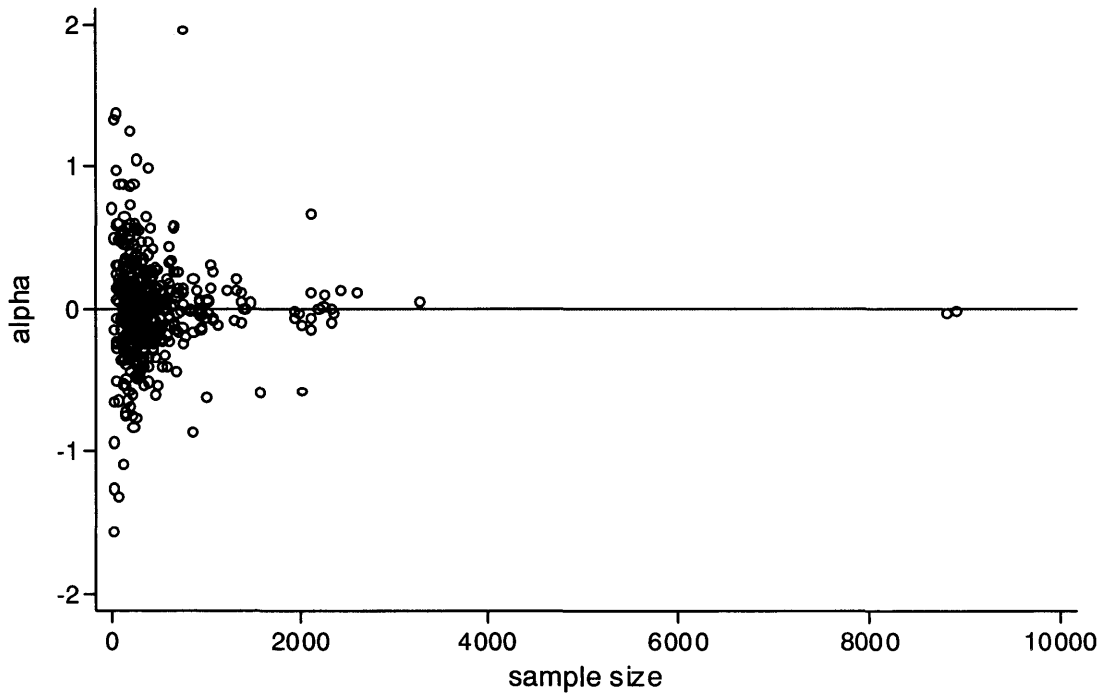
Figure 4.6 shows the plot of alpha against p-value for the 516 comparisons. The vertical line corresponds to a p-value of 0.05, the horizontal line to perfect HWE (alpha=0). The same outlier mentioned in Figure 4.2, with $\log_{10}$ (p-value) of –23.3, was removed from Figure 4.6 to better show the distribution of all the other studies. The plot indicates a horse-shoe shape for the relationship between alpha and p-value, where larger values of alpha, either positive or negative, are more likely to be associated with statistical significant tests. This corresponds to what might be expected. However, among the 516 studies there are a few cases where the decision to exclude a study based on the p-value might be completely inappropriate. The arrows in Figure 4.6 show two examples of this. Arrow a indicates a study where the test provides a statistically significant result but where there is hardly any deviation from HWE, as suggested by a value of alpha very close to 0. On the contrary, arrow b shows a study where the result of the test is far from being significant although the value of alpha indicates a large departure from HWE.

Figure 4.6 – Plots of the values of alpha against p-values for 515 comparisons. The vertical line corresponds to a p-value of 0.05, the horizontal to perfect HWE
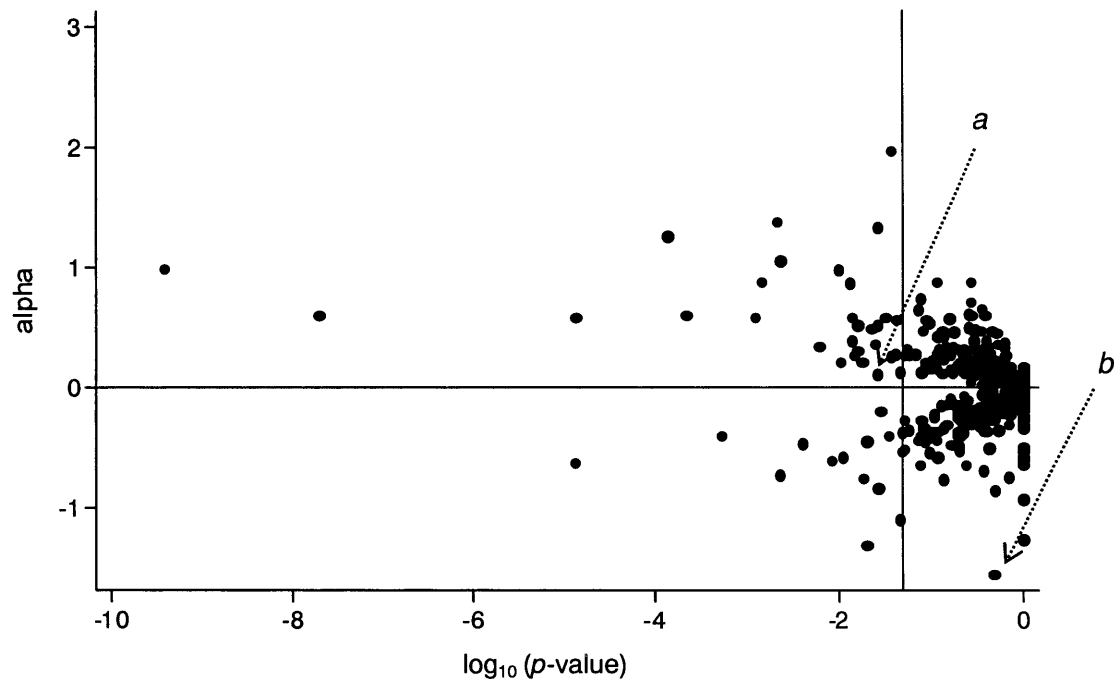
4.5   Comparing strategies to deal with HWE in meta-analysis

There is consensus on the fact that departures from HWE should be detected, quantified, and reported, both in primary studies and in meta-analysis, since they can signal the presence of specific problems, such as genotyping errors and population stratification, and thus represent a proxy for poor methodology. A number of authors have tried to assess the impact that genotyping errors or population stratification might have on the estimate of the genetic effect in terms of loss of power to detect the association and possible bias. This has been based on simulation work and assumptions on what might be, for instance, the error structure for genotyping errors (Gordon *et al.*, 2002; Kang, Gordon, and Finch, 2004), or the characteristics of the populations to be mixed and the sampling framework for population stratification (Wacholder, Rothman and Caporaso, 2002; Heiman *et al.*, 2004; Gorroochurn *et al.*, 2004).

Based on these theoretical considerations, a number of authors have suggested that meta-analyses of genetic association studies should assess HWE in the studies included using Fisher's exact test. Sensitivity analyses should then be carried out to assess

whether studies found as being not in HWE (p-value < 0.05) provide a different

estimate of the genetic effect, which would indicate bias and thus suggest the exclusion

of such studies from the main analysis (Attia, Thakkinstian, and D'Este, 2003; Salanti,

Sanderson and Higgins, 2005). Other authors suggest that studies not in HWE should

be directly excluded from the meta-analysis (Munafo and Flint, 2004). What is not

considered in this strategy is that there might be a trade-off between the possibility of

bias in the pooled estimate of the genetic effect introduced by studies not in HWE, and

the loss in precision associated with the exclusion of these studies from the meta-

analysis. Estimating such a trade-off seems particularly important, given that the

studies which are likely to be excluded based on a statistically significant result of the

test, are often large studies with sometimes modest departures from HWE, as suggested

by figures 4.5 and 4.6. Therefore, it might be preferable to approach the problem from a

more pragmatic perspective, while waiting until strong empirical evidence becomes

available to elucidate the role of genotyping error and population stratification in

current studies, their impact on the estimate of the genetic effect, and the validity of

HWE departures used as a proxy. Once investigators have addressed the issue of HWE

in meta-analyses of genetic association studies by performing an appropriate test for

HWE in each study, estimating the magnitude of departures from HWE, and reporting

these results, what should they do next? In fact, there are two different but related

questions which need to be answered;

1) Should studies showing departures from HWE be excluded?

2) How can investigators identify studies with departures from HWE; based on

statistical test, magnitude of the departure, or both? And if both, what thresholds should

they use?

In order to address both questions, simulation work was undertaken to evaluate

advantages and disadvantages of different possible strategies of identifying studies with

departures from HWE and dealing with them. In particular, four possibilities were

considered for the identification of such studies. The first is the method suggested by

most authors, which is based on the result of an exact test with a p-value cut-off of

0.05. The second is a strategy, still based on the result of an exact test, but using a cut-

off of 0.10; this higher cut-off significance level might be considered to compensate for

the lack of power of statistical tests for HWE, in analogy with the p-value cut-off

routinely used for the test of heterogeneity in meta-analysis (§ 1.4). The other two strategies for identifying studies with departures from HWE are based on a combination of the result of the test for HWE, at the cut-off level of 0.05 or 0.10, and a value of alpha, used to measure the magnitude of the departure, higher than 0.5 or lower than -0.5. The choice of this threshold for alpha was based on the following considerations: perfect HWE in human populations does not exist, and values of alpha between -0.1 and 0.1 might be considered the limits of true departure from HWE for allele frequency of 0.32, roughly corresponding to an inbreeding coefficient of 0.03 (§ 4.3.2). However, because of sampling error, values outside this range may not necessarily indicate departure from HWE caused by methodological problems of the study. Indeed, in the 516 published associations previously described the average standard error for alpha was 0.24, with 98% of the values of alpha being between -1.0 and +1.0 and 87% between -0.5 and 0.5 (§ 4.4.3). Finally, two identification strategies based purely on the magnitude of alpha were considered, one with the threshold values for alpha described above, ±0.5, and the other with threshold values of ±1.

In terms of how to deal with studies showing departure from HWE, the possibilities of either doing nothing, i.e. including all studies in the meta-analysis in any case, or excluding studies identified with any of the strategies described above, were considered. Thus, a total of seven different strategies for dealing with departures from HWE in meta-analysis were compared in terms of bias and precision of the estimates of the genetic effect, in order to identify the strategy with the best trade-off between validity and accuracy.

To strengthen the potential value of this approach in providing recommendations on how to deal with HWE, the assumptions on a number of parameters needed for the simulation work were chosen to reflect the values of those parameters observed in the 516 gene-disease associations (§ 4.4). However, no empirical evidence is available on one of the most crucial parameter in these simulations, the magnitude of the possible impact on the genetic effect of departures from HWE. Therefore, sensitivity analyses were performed to consider different possibilities, from no impact at all to a doubling of the estimate of the log odds ratio for the genetic effect per unit change in alpha. Moreover, if there is an impact at all, which is what most authors believe, then a

specific form for the relationship between the magnitude of the departure from HWE and the estimate of the genetic effect has to be assumed. Again, since there is no evidence on what form such relationship might take, sensitivity analyses were performed by considering two possible scenarios; the log odds ratio for the genetic effect could increase proportionally with the increase in the value of alpha (positive or negative), or the odds ratio could increase proportionally with the increase in the absolute value of alpha, which means that positive and negative values of the same magnitude would impact the log odds ratio in the same way.

4.5.1 Methods

In order to identify the best strategy to handle departures from HWE in meta-analysis, 10,000 datasets, each containing 10 genetic association studies, were randomly simulated, and on each dataset seven meta-analyses were performed that differed for the studies included, reflecting the different strategies of identifying and dealing with studies showing departure from HWE. The pooled genetic effect was estimated using a random effects meta-analyses model, implemented in Stata using the metan command, but sensitivity analyses were also performed using a fixed effect meta-analysis model, implemented with the program. For simplicity, the analysis was performed considering only the odds ratio between the two homozygous groups, GG vs. gg (OR_{GG}) to represent the effect of the polymorphism on the disease. The size of 10 for the simulated datasets was chosen based on the median number of studies included in the 37 published meta-analyses (Table 4.1). All studies were assumed to have an equal total number of cases and controls. This number, which was fixed for each of the 10 studies and varied from study to study, was chosen as to reflect the observed distribution of the sample sizes in the 516 published studies. In particular, the sample size of each study corresponded to a percentile of such distribution: 80, 150, 190, 230, 270, 330, 400, 500, 690, and 1390. The number of cases and controls in each of the three genotype groups for a study was simulated by assuming:

a) A true OR_{GG} of 1.5, which is a value towards the upper end of the range of odds ratios observed in the 516 published associations, where the mean odds ratio was 1.2. A higher odds ratio has been chosen to increase the possibility of

detecting a difference in the estimates of the odds ratio when applying the different strategies.

b) A co-dominant genetic model, which corresponds to a true OR_{Gg} of 1.22.

c) An allele frequency of the allele of interest, G, of 0.32, chosen to reflect the mean allele frequency observed in the 516 published associations.

d) A value of alpha randomly generated, at each simulation, by sampling from a normal distribution with mean 0 and standard deviation depending on the size of the study. In fact, as shown in § 4.4.3, the values of alpha in published studies are distributed symmetrically around 0 (Figure 4.4), with the absolute size of alpha depending on the size of the study, where larger studies show smaller values of alpha (Figure 4.5). Based on the relationship between alpha and sample size observed in the 516 studies, the true value of alpha for each of the 10 simulated studies (α_i, with $i=1,...,10$) was sampled from:

$$\alpha_i \sim Normal(0,\sigma_{\alpha i}^{2}) \qquad \text{with} \qquad \sigma_{\alpha i} = 40/(ncase_i + 40)$$

where $ncase_i$ represents the total number of cases for study i, which corresponds to half the sample size. The function for σ_i was chosen by identifying in the scatter plot of alpha against sample size for the 516 associations (Figure 4.5) the two lines that delimited a region containing about 95% of the points, as shown in Figure 4.7. These lines were drawn by considering that in the distribution of the observed values of alpha, centred on zero, there are two components of the variability which have to be added; the variability explained by the relationship with sample size, and the random variation due to sampling error. While the random variation was estimated with WinBUGS (§ 4.4.2), the form of the dependence of alpha on the sample size was estimated by trial and error, by evaluating different functions of alpha on sample size until the two lines in the scatter plot, one line for positive and one for negative values of alpha, included 95% of the 516 associations. This function was the one used to define σ_i, i.e. the variability in the distribution of alpha from which alpha values for the 10 studies were sampled at each simulation.

Since the underlying value of alpha simulated for each study was expected to play a major role in determining which strategy might perform better, a sensitivity analysis to assess the assumption of dependence of alpha on sample size was performed. In this

analysis, a mixture model was used, where the possibility that the value of alpha depended on sample size as described above was given a 90% probability, while a 10%

Figure 4.7 – Derivation of the function used for the simulations, based on the relationship between alpha and sample size observed in the 516 published associations. The two lines, which correspond to ±1.96s, where $s = \sqrt{\sigma_\alpha^2 + SE(alpha)^2}$, include approximately 95% of the points
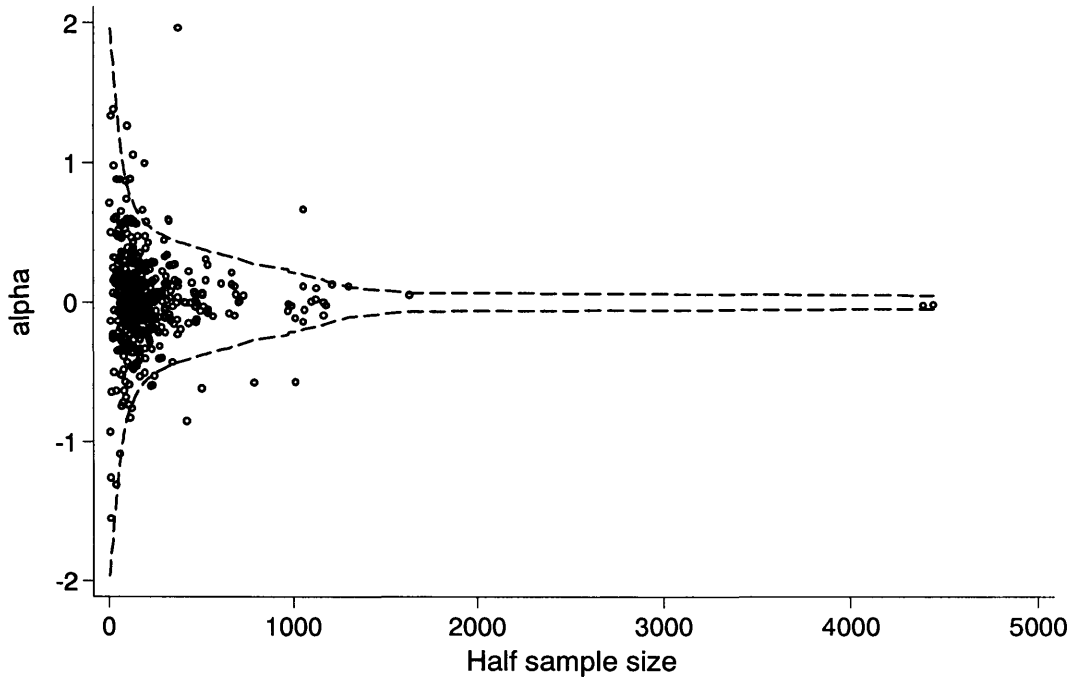
probability was given to the possibility that alpha might not depend on the sample size at all. Values for alpha were thus randomly sampled from a distribution represented by a mixture of two normal distributions, contributing to 90% and 10% in the definition of the shape of the overall distribution. The details of the Stata code for this model are reported in Appendix 3. The normal distribution representing the possibility of alpha being independent from sample size was defined as having a mean of 0 and a fixed standard deviation of 0.5, which is equivalent to assuming that the true values of alpha lie between –1.00 and +1.00. Thus:

$$Normal\_1(0,\sigma_i^2) \qquad and \qquad \sigma_i = 40/(ncase_i + 40)$$

$$Normal\_2(0,0.5^2)$$

e) Three different values were assumed for θ, the parameter representing the slope of the association between alpha and the log OR_{GG}, i.e. the impact of departures from HWE on the estimate of the genetic effect. In fact, as mentioned in the previous section, this is another crucial parameter in the comparisons of the different strategies to identify and deal with studies that show departures from HWE. Since no empirical evidence is available on this parameter, sensitivity analyses were performed assuming:

$\theta = 0$, corresponding to no impact at all (OR_{GG} = true OR_{GG} =1.50); $\theta = 0.2$, corresponding to a 50% increase of the estimate of log OR_{GG} per unit change in alpha ($OR_{GG} = 1.83$ when alpha=1); and $\theta = 0.4$, corresponding to doubling the estimate of log OR_{GG} per unit change in alpha ($OR_{GG} = 2.24$ when alpha=1). As regards the form of the relationship between log OR_{GG} and θ, two possible scenarios were considered, which are graphically represented in Figure 4.8:

- *Scenario 1*: A linear trend where log OR_{GG} increases with the increase in alpha, from negative to positive values of alpha.

Figure 4.8 – Illustration of two possible forms of the relationship between departure from HWE and magnitude of the estimated genetic effect; a) Scenario 1; b) Scenario 2. The two lines represent different values of θ; dashed line: θ=0.2; solid line: θ=0.4. The values of the OR are plotted on the log scale
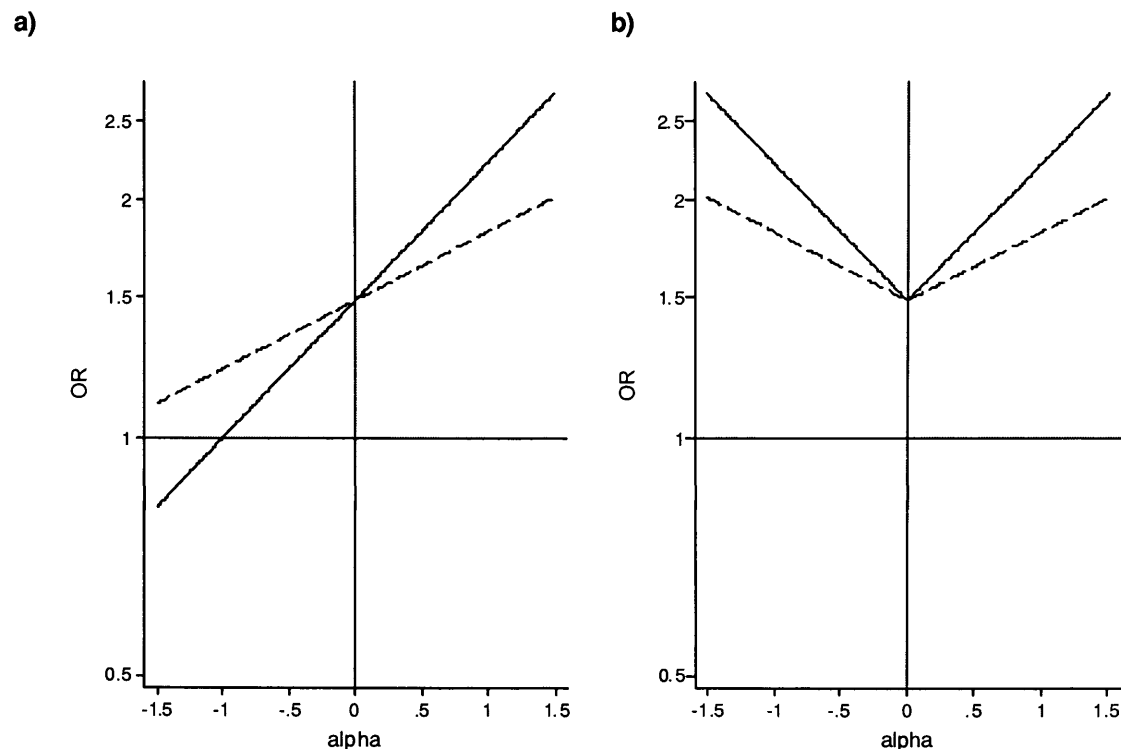
- *Scenario 2*: A linear trend where log OR_{GG} increases with the increase in the absolute value of alpha, which means that the trend for negative values of alpha mirrors the trend for positive values of alpha.

The following 7 strategies for identifying and handling studies with departures from HWE were evaluated:

1) Do nothing. All studies are included in the meta-analysis independently from any evidence of departure from HWE.

2) Exclude if $p<=0.05$. Studies are excluded if the p-value of the exact test for HWE is lower than, or equal to, 0.05.

3) Exclude if $p<=0.1$. As in 2) but with a cut-off significance level of 0.1.

4) Exclude if absolute value of alpha>1. Studies are excluded if the observed absolute value of alpha is higher than 1 (alpha lower than -1 or higher than $+1$).

5) Exclude if absolute value of alpha>0.5. As in 4) but with a threshold value of 0.5.

6) Exclude if $p<=0.1$ and absolute value of alpha>0.5. Studies are excluded if two conditions are met: the p-value of the exact test for HWE is lower than, or equal to, 0.1 and the observed absolute value of alpha is higher than 0.5.

7) Exclude if $p<=0.05$ and absolute value of alpha>0.5. As in 6) but with a cut-off significance level of 0.05.

The estimates of the pooled $logOR_{GG}$ for the simulated datasets analysed under the different scenarios were compared in terms of its mean, its Root Mean Square Error (RMSE), that is the square root of the mean squared errors, and the coverage of the 95% CIs, that is the percentage of intervals that included the true value ($OR_{GG} =1.5$; $logOR_{GG} =0.405$). These three measures describe the average properties of the estimator across the 10,000 datasets.

4.5.2 Results

Given the importance of the simulated magnitude of alpha in determining what might be the best strategy for handling HWE, the variation of alpha with sample size in the

simulations based on both the main and the mixture model was compared with that observed in the 516 published associations. Such evaluation consisted of simulating 516 studies using the two models, summarising the values of alpha with different scatter plots for each sample size, and comparing the two graphs with that of the 516 published associations, where scatter plots were drawn for each percentile of sample size, to which the 10 different sample sizes of the simulated studies correspond. The three graphs, shown in Figure 4.9, indicate a satisfying degree of similarity of both simulated datasets with the observed 516 published associations. The simulated studies based on the mixture model show more random variation of alpha compared to the main model for all sample sizes, and in particular for the largest sample sizes. This behaviour is what would be expected given the fact that in the mixture model the alpha values of 10% of the studies were sampled independently from the sample size of the study.

Tables 4.2 and 4.3 show the results of the simulations for the comparison of the different strategies for each of the two models respectively: the *main model*, with alpha sampled from the distribution with variance dependent on the sample size of the study, and the *mixture model*, with alpha sampled from the mixture distribution, where 10% probability is assigned to alpha being independent from sample size. For each model, the results of the simulations for the pooled OR_{GG} are reported for each of the three values of θ, representing different possible impacts of HWE departures on the estimates of the odds ratio, and under each of the two scenarios, representing different shapes of the possible association between HWE departures and estimates of the odds ratio. In tables 4.2 and 4.3, in addition to the mean, RMSE and coverage of the 95%CIs for the pooled log OR_{GG}, also reported are the percentage of bias, that is the percentage difference between the mean estimate of log OR_{GG} and its true value, the average standard error (SE) of log OR_{GG} estimates, calculated as the root of the average variance, and the percentage of studies excluded from the meta-analysis by each strategy, over the 10,000 simulations.

When using the main model and assuming scenario 1, i.e. a linear trend for the impact of alpha on the log OR_{GG} estimate with the extent of the impact represented by the three values of θ, all strategies seem to perform well in terms of bias even when the

Figure 4.9 – Box plots of the observed values of alpha against sample size for: a) the 516 studies in the database; b) a sample of 516 simulated studies under the main model; c) a sample of 516 simulated studies under the mixture model
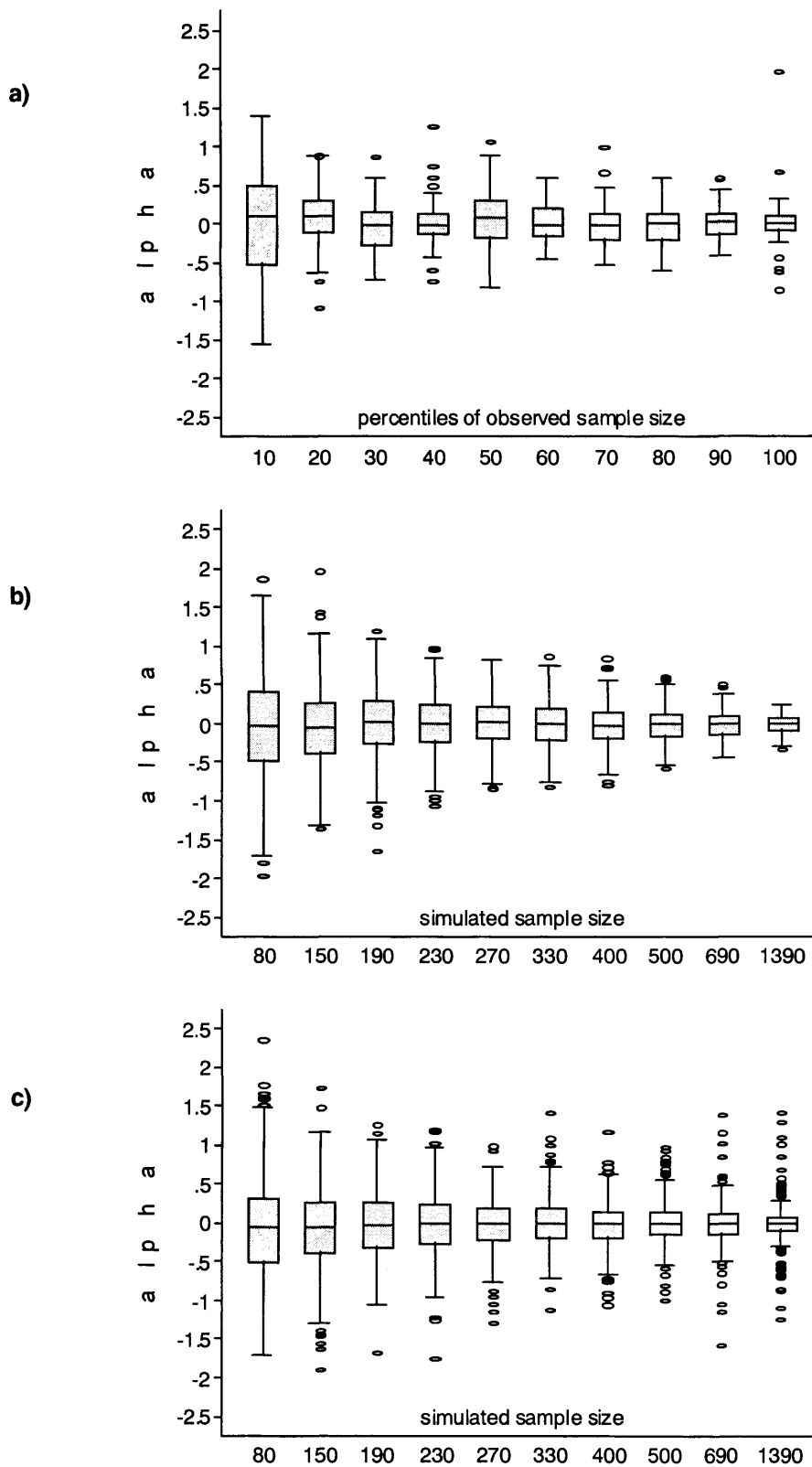
largest value of θ is assumed. In terms of precision, all strategies perform similarly, with the exception of the second and the third strategy, for which exclusion of studies is based only on p-value ($p<0.05$ and $p<0.1$, respectively), with the third strategy causing the highest loss in precision. This is explained by the fact that those studies showing statistically significant deviation from HWE are the largest studies, so that their exclusion from the meta-analysis does have an impact on the precision of the pooled estimate. A cut-off significance level of 0.1 implies that more studies can reach statistical significance for departure from HWE, as shown in Table 4.2 where the average percentage of studies excluded is 24% compared to 16% for a cut-off of 0.05. In comparison, the two strategies based on a combination of p-value and magnitude of alpha show an average percentage of studies excluded of 11% and 10% when the threshold for the p-value is 0.1 (sixth strategy) and 0.05 (seventh strategy), respectively. Moreover, the studies excluded will not necessarily be the largest ones, since large studies, although more likely to overcome the problem of lack of power of HWE tests, will also tend to have smaller values of alpha. The behaviour described also apply to scenario 2, where the estimates of log OR_{GG} are assumed to increase linearly with the increase in the absolute value of alpha, with the extent of the association dictated by the assumed value of θ.

The situation is not very different when using a mixture model for both scenarios, where again the performance of all strategies is similar in terms of bias, with only a slightly higher degree of bias for the first strategy, i.e. do nothing, for both scenarios when the highest impact of departure from HWE on the estimate of log OR_{GG} is assumed ($\theta=0.4$). In terms of precision, again the second and third strategies are those performing worse, with an average percentage of studies excluded of 19% and 27% respectively. These higher percentages compared to the results for the main model, and the consequent higher loss in precision, are due to the fact that in 10% of the simulated studies the magnitude of alpha does not depend on the size of the study, so there are more large studies with alpha value large enough to be statistically significant.

The measures of average performance commonly used to present the results of simulations, as those shown in tables 4.2 and 4.3, may not be a reliable guide to the

Table 4.2 – Results for the $logOR_{GG}$ under the different strategies obtained using the *main model* in the two scenarios; a) scenario 1; b) scenario 2.

a)

Strategy	Theta = 0						Theta =0.2						Theta = 0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.403	-0.7	0.115	0.083	96.4	0.0	0.406	0.1	0.116	0.083	96.5	0.0	0.409	0.9	0.117	0.084	96.3
Exclude if: p<=0.05	16.5	0.405	-0.1	0.127	0.089	96.7	16.3	0.407	0.5	0.127	0.089	96.4	16.3	0.409	0.9	0.129	0.091	96.2
Exclude if: p<=0.1	23.8	0.405	0.0	0.135	0.094	96.8	23.7	0.407	0.4	0.135	0.094	96.6	23.7	0.409	0.8	0.137	0.096	96.4
Exclude if: |α|>1	1.9	0.402	-0.8	0.115	0.083	96.4	1.9	0.405	-0.1	0.116	0.083	96.5	1.9	0.408	0.6	0.117	0.084	96.3
Exclude if: |α|>0.5	12.6	0.400	-1.2	0.118	0.085	96.6	12.5	0.402	-0.7	0.119	0.085	96.5	12.5	0.405	-0.2	0.120	0.086	96.2
Exclude if: p<=0.1 & α>0.5	10.9	0.402	-1.0	0.118	0.085	96.6	10.8	0.404	-0.4	0.119	0.085	96.5	10.8	0.406	0.1	0.120	0.086	96.3
Exclude if: p<=0.05 & α>0.5	9.6	0.402	-0.8	0.118	0.084	96.6	9.5	0.405	-0.2	0.119	0.084	96.5	9.5	0.407	0.3	0.120	0.086	96.2

b)

Strategy	Theta = 0						Theta =0.2						Theta = 0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.403	-0.6	0.115	0.082	96.5	0.0	0.428	5.6	0.115	0.084	96.1	0.0	0.455	12.2	0.116	0.092	94.6
Exclude if: p<=0.05	16.3	0.405	-0.1	0.127	0.089	96.6	16.3	0.427	5.2	0.127	0.090	96.4	16.4	0.450	11.1	0.127	0.096	95.4
Exclude if: p<=0.1	23.6	0.406	0.0	0.135	0.093	96.7	23.6	0.426	5.1	0.135	0.095	96.4	23.9	0.449	10.7	0.135	0.100	95.8
Exclude if: |α|>1	1.9	0.403	-0.7	0.116	0.082	96.5	1.9	0.427	5.4	0.115	0.084	96.1	2.0	0.453	11.7	0.116	0.092	94.9
Exclude if: |α|>0.5	12.5	0.400	-1.3	0.118	0.084	96.5	12.5	0.422	4.1	0.118	0.085	96.4	12.5	0.446	9.9	0.119	0.091	95.5
Exclude if: p<=0.1 & α>0.5	10.8	0.402	-1.0	0.118	0.084	96.6	10.8	0.423	4.4	0.118	0.085	96.3	10.9	0.447	10.3	0.119	0.091	95.4
Exclude if: p<=0.05 & α>0.5	9.4	0.402	-0.8	0.118	0.084	96.6	9.4	0.425	4.7	0.118	0.085	96.3	9.5	0.449	10.6	0.119	0.091	95.4

Table 4.3 – Results for the logOR_{GG} under the different strategies obtained using the *mixture model* in the two scenarios; a) scenario 1; b) scenario 2.

a)

Strategy	Theta = 0						Theta = 0.2						Theta = 0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.403	-0.6	0.116	0.083	96.7	0.0	0.407	0.3	0.116	0.083	96.5	0.0	0.411	1.4	0.119	0.086	95.8
Exclude if: p<=0.05	19.4	0.406	0.2	0.132	0.091	96.8	19.2	0.406	0.1	0.133	0.092	96.6	19.2	0.408	0.6	0.135	0.095	96.4
Exclude if: p<=0.1	26.7	0.406	0.1	0.141	0.096	96.9	26.6	0.406	0.1	0.142	0.097	96.6	26.6	0.408	0.5	0.144	0.100	96.5
Exclude if: $|\alpha|>1$	2.4	0.402	-0.8	0.116	0.083	96.6	2.4	0.405	-0.1	0.117	0.083	96.5	2.4	0.409	0.9	0.119	0.086	95.9
Exclude if: $|\alpha|>0.5$	14.8	0.400	-1.3	0.121	0.086	96.7	15.0	0.401	-1.0	0.122	0.086	96.5	15.0	0.404	-0.5	0.124	0.088	96.2
Exclude if: p<=0.1 & α>0.5	13.1	0.402	-1.0	0.121	0.086	96.8	13.3	0.403	-0.7	0.122	0.086	96.4	13.3	0.405	-0.1	0.124	0.088	96.2
Exclude if: p<=0.05 & α>0.5	11.7	0.402	-0.7	0.121	0.086	96.8	11.9	0.403	-0.5	0.122	0.086	96.4	11.9	0.406	0.1	0.124	0.088	96.2

b)

Strategy	Theta =0						Theta = 0.2						Theta = 0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.404	-0.5	0.115	0.083	96.5	0.0	0.435	7.3	0.115	0.086	95.6	0.0	0.465	14.8	0.117	0.098	92.8
Exclude if: p<=0.05	19.4	0.406	0.2	0.131	0.092	96.6	19.4	0.430	6.0	0.131	0.093	96.0	19.4	0.452	11.4	0.132	0.099	95.2
Exclude if: p<=0.1	26.7	0.406	0.2	0.140	0.096	96.7	26.7	0.429	5.9	0.139	0.098	96.1	26.7	0.450	10.9	0.141	0.103	95.7
Exclude if: $|\alpha|>1$	2.4	0.403	-0.6	0.116	0.083	96.4	2.5	0.433	6.7	0.116	0.085	95.8	2.4	0.462	13.8	0.117	0.096	93.3
Exclude if: $|\alpha|>0.5$	15.0	0.401	-1.1	0.121	0.086	96.2	15.1	0.425	4.9	0.121	0.087	95.9	15.0	0.449	10.7	0.121	0.093	94.7
Exclude if: p<=0.1 & α>0.5	13.3	0.402	-0.8	0.121	0.086	96.3	13.5	0.427	5.2	0.121	0.087	95.9	13.3	0.450	11.1	0.121	0.094	94.6
Exclude if: p<=0.05 & α>0.5	11.8	0.404	-0.5	0.121	0.086	96.3	12.0	0.428	5.5	0.120	0.087	95.9	11.8	0.452	11.5	0.121	0.094	94.4

Figure 4.10 – Box plots of the results for the different strategies obtained using the *main model* in the two scenarios; a) scenario 1; b) scenario 2. *1*: do nothing; *2*: exclu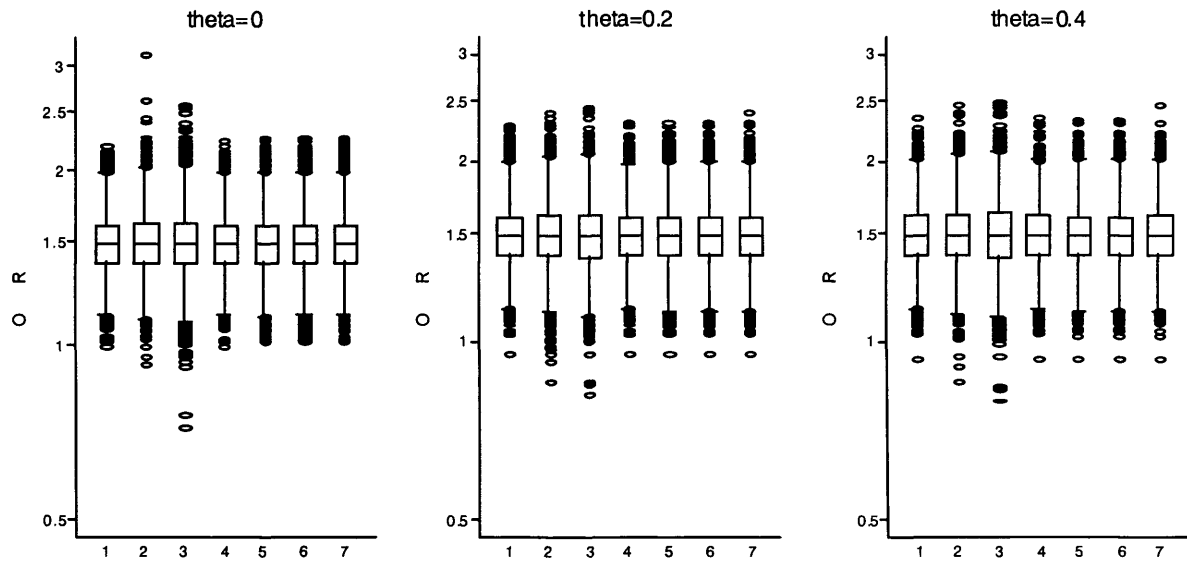de if $p<=0.05$; *3*: exclude if $p<=0.1$; *4*: exclude if |alpha|>1; *5*: exclude if |alpha|>0.5; *6*: exclude if |alpha|>0.5 & $p<=0.1$; *7*: exclude if |alpha|>0.5 & $p<=0.05$

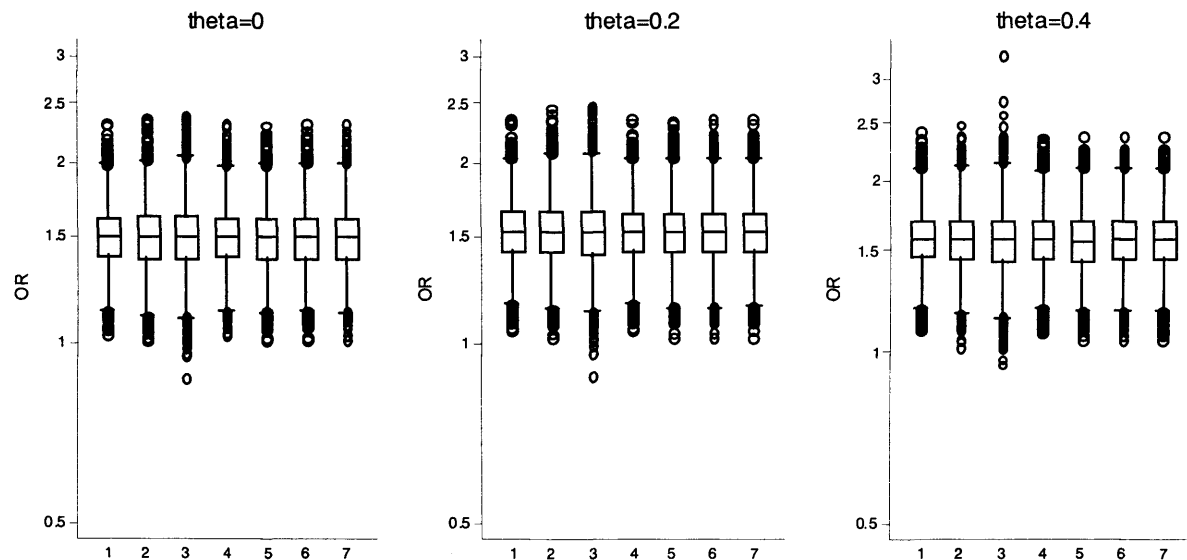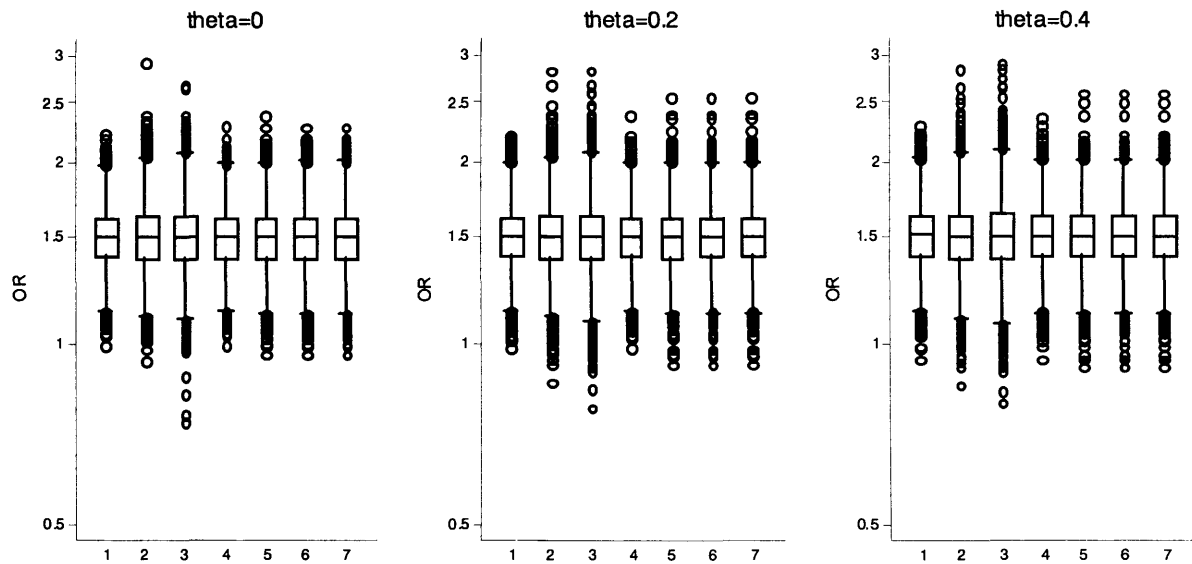**a)**

b)

Figure 4.11 – Box plots of the results for the different strategies obtained using the *mixture model* in the two scenarios; a) scenario 1; b) scenario 2. *1*: do nothing; *2*: exclude if $p<=0.05$; *3*: exclude if $p<=0.1$; *4*: exclude if |alpha|>1; *5*: exclude if |alpha|>0.5; *6*: exclude if |alpha|>0.5 & $p<=0.1$; *7*: exclude if |alpha|>0.5 & $p<=0.05$

a)

b)
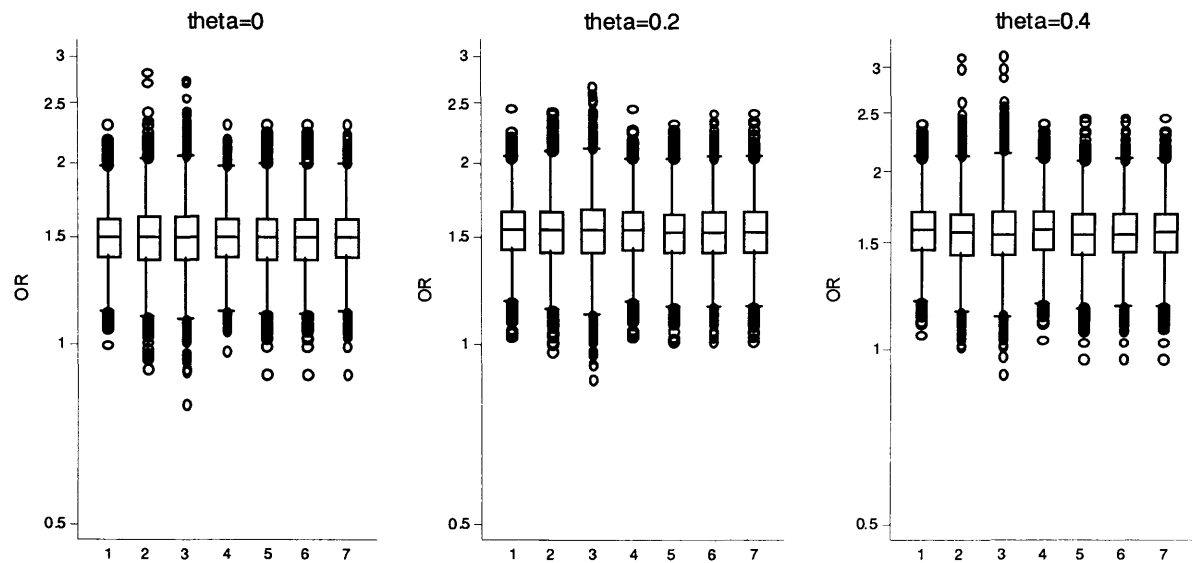
sensitivity of the OR_{GG} estimates to the choice of different strategies of handling HWE for any particular single dataset. In order to investigate what might happen in individual meta-analyses due to the play of chance, a graphical representation of the results of the meta-analyses on the 10,000 simulated datasets was also evaluated. Box plots of the pooled estimates of log OR_{GG} for each of the seven strategies under the different assumptions, i.e. different values of θ and different shape of the association between log OR_{GG} and alpha (scenario 1 and 2) were drawn for the main and the mixture model, as shown in figures 4.10 and 4.11 respectively. The second and third strategies are again those which perform worse for both the main and the mixture model and under all different assumptions on value of θ and shape of the association between alpha and estimate of log OR_{GG}. Adopting these two strategies, occasionally the pooled estimate of log OR_{GG} can be rather far from the true value, with this happening more frequently with the third strategy (p-value cut-off of 0.1). All the others perform similarly, with the possible exception of the first strategy which tends to perform better than all others under the mixture model. However, it has to be considered that the outlying estimates identified in these graphs are very rare, considering that overall 10,000 estimates are summarised by the scatter plots.

The results of the fixed effect meta-analysis models were similar to those of the random effects models in terms of differences between strategies for both the main and the mixture models. The tables with the measures of average performance as well as the graphs with the scatter plots for the fixed effect analyses are reported in Appendix 3 (tables A and B; figures A and B).

4.6   Discussion

4.6.1   What does evidence on departures from HWE suggest?

The review of 516 genetic associations included in 37 published meta-analyses presented in this chapter, not only confirms some of the aspects of HWE already highlighted by a number of authors, but also indicates issues that seem to have been overlooked. The problem of the lack of power when testing for HWE, even with the use of the unanimously suggested exact test, is in line with what previously indicated. In the 516 associations, more than half of which were published either in genetics journals or in

journals with a strong interest in genetics (e.g. *American Journal of Epidemiology*), the

percentage of studies showing statistically significant deviation from HWE was 10.5%,

which is within the range of figures reported by other authors. It particular, this

percentage was reported to vary from 10% to 35% in genetic association studies

published in non-genetics journals (Kocsis *et al.*, 2004a; Kocsis *et al.*, 2004b; Gyorffy,

Kocsis, and Vasarhelyi, 2004a; Gyorffy, Kocsis, and Vasarhelyi, 2004b; Bardoczy *et al.*,

2004; Nemeth *et al.*, 2004), where part of the variability was probably explained by

sampling error due to the relatively small number of associations considered in some of

the reports. The corresponding reported figure for studies published in three high-profile

genetics journals was 10%, that is on the low side of the range (Salanti *et al.*, 2005). As

pointed out by Salanti and colleagues, this might suggest the presence of publication bias,

whereby genetic association studies showing departures from HWE might be less likely

to be submitted for publication or, once submitted, to be published, in high-profile

genetic journals. In fact, editors and reviewers of such journals might be more prone to

consider the issue of HWE, and to assume poor methodology for a study with statistically

significant departure. In the results of the simulations performed in § 4.5, where

parameters were chosen to reflect those of the 516 associations, the percentage of

simulated studies reaching statistical significance for departure from HWE at the cut-off

level of 0.05 was higher than in the observed studies, with values of 16% and 19% for the

simulations under the main and the mixture model, respectively. Such finding seems to

support the possibility that publication bias might have been responsible for the low

percentage observed in the dataset of 516 studies included in the meta-analyses reviewed.

Another important point highlighted by the re-analysis of the 516 associations is the

dependence of the values of alpha on the size of the studies, with alpha increasing with

decreasing sample sizes. This is crucial since it means that not only large studies are more

likely to have the power needed to be able to detect departure from HWE, if such a

departure exists, but also that they are less likely to present major departures from HWE.

In fact, this finding is not at all surprising if departures from HWE are to be considered

proxy for poor quality, since in general the quality of large studies tend to be higher.

Therefore, the appropriateness of current recommendations on how to deal with HWE in

meta-analysis, which suggest using an exact test with *p*-value threshold of 0.05 to

identify studies with departures from HWE and then exclude them first in a sensitivity

analysis and eventually in the main analysis if they provide different estimates of the genetic effect, might be questioned.

In this chapter, the magnitude of departure from HWE was measured by the parameter alpha, which was estimated using a Bayesian approach with vague prior distributions. In fact, the inbreeding coefficient and disequilibrium parameter could also be used, but these parameters have sampling distributions that may be highly skew and are bounded by the requirement that all genotype frequencies be non-negative (Ayres and Balding, 1998). For this reason it is very difficult to specify vague prior distributions for these parameters.

4.6.2 Proposed strategy to detect and deal with departures from HWE

The simulation work presented in this chapter, and which is aimed at identifying the best strategy to be used for identifying and handling studies with departures from HWE, shows results which do not support the common belief that including studies with departures from HWE in a meta-analysis tends to bias the pooled estimate. In fact, the strategy of including all studies independently from departures from HWE is not associated with the presence of appreciably more bias than any other strategy under either of the models used for the simulations, and in any of the sensitivity analyses testing the impact of departures on the estimates of the genetic effect or shape of the relationship. Indeed, the strategy of "doing nothing" seems to always perform at least as well as the others. The second best choice is possibly that of excluding studies based on both the result of the exact test, with a p-value of 0.05, and the observed value of alpha, for example using thresholds of ±0.5. The worst strategies are those based on the exclusion of studies with statistically significant departures from HWE, due to the loss in precision of the pooled estimate caused by the exclusion of large studies. When adopting a cut-off significance level of 0.1 to compensate for the lack of power of the tests for HWE, the loss in precision gets worse due to the higher number of studies excluded.

In the simulations presented in this chapter, the assumptions on the values of the parameters required were mostly based on the values observed in the dataset of 516 published associations. However, further investigation is required to show whether the

conclusions hold under particular circumstances, such as a very low allele frequency or a specific genetic model.

Overall, the results of the simulation work presented here suggests that the real issue regarding current recommendations is not the lack of power of tests for HWE as previously suggested (Salanti *et al.*, 2005), but the failing to take into account the estimate of the magnitude of the departure, or, more radically, the fact that the belief that departure from HWE will bias the estimate of the genetic effect might be unjustified. This does not mean that the issue of HWE should be ignored at all in either primary studies or meta-analyses. Indeed, evidence of departures from HWE, which should be based on both the result of an exact test and the estimation of the magnitude of the departure, should be carefully considered as a warning signal for the possible presence of methodological problems requiring further investigation. In the context of a primary study, this may lead, for instance, to address the possibility of genotyping errors, try to measure the extent of the problem, and adjust the analysis for it. In the same way, evidence of departure from HWE in studies included in a meta-analysis should lead one to further investigate the possibility of methodological problems in those studies. However, decisions on whether such studies should be excluded, either in a sensitivity analysis or in the main analysis, should be taken based on the probability of bias associated with the specific methodological problem rather than on the evidence of departure from HWE *per se*.

4.6.3 Further work

In this chapter, hypothesis testing for HWE has been performed using a classical approach. However, although hypothesis testing is not part of the philosophy on which the Bayesian approach to statistical inference is based, a number of methods for testing HWE have been recently developed within a Bayesian framework to address some limitations of the frequentist approach. Bayesian methods can easily handle the problem of nuisance parameters, i.e. parameters that are not directly of interest but which need to be defined, by accounting for their different possible values (Spiegelhalter, Abrams, and Myles, 2004). Moreover, Bayesian methods can handle situations of sparse data for one or more genotypes, thus representing an alternative to exact tests (Zaykin, Zhivotovsky, and Weir, 1995). Another important advantage of Bayesian methods is that they allow

testing hypotheses other than the null. In fact, the presence of perfect HWE, as tested in the null hypothesis, is highly unlikely in any real population (§ 4.2.2). Thus, it would seem more appropriate to test whether the population is close enough to HWE for us to assume that this is the case. The investigators then have to face the problem of defining what "close enough" exactly means. Shoemaker and colleagues present an interesting application of this approach to the field of forensic science (Shoemaker, Painter, and Weir, 1998). They use the posterior distributions of a Bayesian analysis to find probabilities that the measures of departure from HWE adopted, the inbreeding coefficient and the disequilibrium parameter, lie inside or outside specific intervals. These intervals were based on the Untied States National Research Council report for use in the evaluation of forensic DNA evidence, where an inbreeding coefficient of 0.03 is suggested as the upper limit of departure from HWE in human populations (National Research Council, 1996). Finally, even if the investigator is indeed interested in testing the null hypothesis of exact HWE in the population, the p-value does not provide any quantifiable measure of strength of the evidence against the null hypothesis. This is a general problem of the frequentist approach, which is made worse by the fact that p-values are indeed often misinterpreted as measures of the evidence against the null, i.e. the p-value is confused with the probability of the population being in HWE. The Bayesian approach allows direct probability statements, which are not limited to the null hypothesis, but are applicable to any hypothesis of interest, such as what is the probability that the departure from HWE is greater than x (§ 1.5). As regards the estimation of the magnitude of the departure from HWE, Bayesian methods might be advantageous in providing more informative posterior distributions for the parameters of interest. Although not investigated in this chapter, previous knowledge about the parameter representing the magnitude of the departure or nuisance parameters, such as the allele frequencies, can be included in the model when estimating the measure of departure from HWE using a Bayesian approach. Previous information on the expected magnitude of the departure from HWE, for instance, might derive from some knowledge of the mating system operating in the population of interest.

The problem of HWE in meta-analysis of genetic association studies is in many ways similar to the problem of underlying risk of disease in meta-analysis of RCTs. In both cases, the effect of interest (genetic effect or treatment effect) in the different studies may be influenced by a factor (departure from HWE or underlying risk), which can vary from

study to study and reflect different characteristics of the study population or study design. For this reason, the methods recently developed to adjust for underlying risk of disease in meta-analyses of RCTs (Sharp and Thompson, 2000) might help in developing appropriate ways of adjusting for departure from HWE in meta-analyses of genetic association studies. Such an adjustment might be a more appropriate way of dealing with HWE, and could represent a compromise between the two opposite strategies of either excluding studies with statistically significant departures from HWE, as recommended by most authors, or including all studies by ignoring departures from HWE, as indicated by the simulation work presented in this chapter.

Another important issue which requires further investigation is whether, once assessed that there is no evidence against HWE in the studies included, the assumption of HWE should be incorporated in the meta-analysis model with the aim of increasing the efficiency in estimating the pooled genetic effect. Although the incorporation of the assumption of HWE could be done in either a frequentist or a Bayesian meta-analysis, the advantage of the Bayesian approach is that it allows modelling the uncertainty about the assumption, as illustrated by Cheng and Chen in individual case-control genetic association studies (Cheng and Chen, 2005). These authors define in their Bayesian model a variable representing departure from HWE, for which they use a "power prior" where the mean parameter is set to 0 while the precision parameter is chosen to reflect the uncertainty about HWE.

4.6.4 Conclusions

Departure from HWE is thought to be an indicator of methodological problems with the design and conduct of the study, in particular population stratification, bias in the selection of cases and controls, and genotyping error, and therefore needs to be investigated both in primary studies and meta-analyses of genetic association studies. However, the work presented in this chapter suggests that exclusion of studies from a meta-analysis based on evidence of departure from HWE is not justified, in particular when evidence of departure only consists of the result of a hypothesis test for HWE. It would appear more appropriate to use evidence of deviation from HWE, which should be based on both the result of the test and the estimation of the magnitude of the departure, as a warning signal for the possible presence of methodological problems requiring

further investigation. Decisions on whether studies with departures from HWE should be excluded from the meta-analysis, either in a sensitivity analysis or in the main analysis, should be taken based on the probability of bias associated with the specific problem rather than on the evidence of departure from HWE *per se*.

5   MENDELIAN RANDOMISATION

5.1   Chapter overview

This chapter addresses a promising application of genetic association studies as tools for assessing causality of associations between a risk factor and a disease, based on an approach known as "Mendelian randomisation". An introduction to the concept of Mendelian randomisation is presented in § 5.2, followed by a review of the literature on Mendelian randomisation and a discussion of the role of meta-analysis in this context in § 5.3. Section 5.4 presents an integrated meta-analytical approach to Mendelian randomisation, where a multivariate meta-analysis method is proposed and graphical ways of evaluating the assumptions underlying Mendelian randomisation are suggested. The meta-analysis models proposed are described in § 5.5, while in § 5.6 the integrated approach is applied to the example of *MTHFR* gene, homocysteine and Coronary Heart Disease (CHD). Finally, the issues raised by the use of Mendelian randomisation, the advantages of using an integrated approach to synthesise all evidence available, and the choice of the multivariate meta-analysis model are discussed in § 5.7.

5.2   Introduction

Traditional epidemiological studies that investigate the association between a biological risk factor (phenotype) and a disease are often biased because of confounding or reverse causation (Rothman and Greenland, 1998). Genetic association studies provide epidemiologists with a tool to obtain an unconfounded estimate of the association between a risk factor and a disease, when a certain genetic mutation exists which can alter the level of the risk factor. The approach, known as Mendelian randomisation, consists of carefully selecting a gene that can be used as an instrumental variable, provided that the associations between the gene and the phenotype and the gene and the disease can be accurately estimated (Davey Smith and Ebrahim, 2003). The necessary precision in these estimates will often be obtained only through the synthesis of all available evidence (Thompson, Tobin, and Minelli, 2003). Mendelian randomisation is likely to assume an important role in the near future, when greater knowledge of biological pathways will guide the choice of more suitable polymorphisms, and more

data on the genotype-disease and genotype-phenotype associations will become available from high quality studies. The potential of Mendelian randomisation, and the role of meta-analysis in this approach, have been recognised, as demonstrated by a rapidly growing interest in the topic. However, very limited methodological work has been done to date to investigate methods aimed not only at providing precise estimates, but also at checking the basic assumptions on which Mendelian randomisation is based (Minelli, Thompson, *et al.*, 2004 - included in Addenda).

5.3   Review of the literature

Katan was the first, in 1986, to suggest that data from genetic studies could be used to test for a relationship between a quantitative intermediate phenotype and a disease in a way that is not distorted by confounding or reverse causality, an idea later referred to as "Mendelian randomisation" (Katan, 1986). In fact, the term "Mendelian randomisation" simply means that, according to Mendel's laws of segregation and independence (Wijsman, 2002), a subject's genotype is determined by a seemingly random process at conception. Thus, epidemiological studies of the genotype-phenotype and genotype-disease associations show strong parallels with randomised trials and should not be affected by confounding or reverse causation in the way that makes classical epidemiological phenotype-disease studies so difficult to interpret (Clayton and McKeigue, 2001; Keavney, 2002; Davey Smith and Ebrahim, 2003). Although Mendelian randomisation is a fundamental assumption of *any* study evaluating the association between a genotype and an outcome, the term has been used to describe the whole process based on Katan's idea, that is to infer what is the effect of a phenotype on disease given information on the gene-disease and gene-phenotype associations. To avoid the ambiguity of this double meaning, the alternative name of "Mendelian deconfounding" has also been proposed to describe the approach (Tobin *et al.*, 2004 – included in Addenda).

5.3.1 From hypothesis testing to estimation

In his letter to the Lancet in 1986, Katan explained his idea using the example of cholesterol and cancer, where the hypothesis to test was whether low cholesterol favours tumour growth. Since differences in serum cholesterol levels in the population

are associated with different apolipoprotein E (*APOE*) genotypes, and the many environmental factors acting on cholesterol levels are equally distributed among the different genotypes, under the causal hypothesis we would expect to see a corresponding association between *APOE* and cancer. The absence of such a genetic association would suggest that the association between low cholesterol and cancer is spurious.

Youngman, Keavney and colleagues first used the term "Mendelian randomisation" when they applied the concept described by Katan to the evaluation of the association between fibrinogen and myocardial infarction based on both direct and indirect evidence, the latter obtained through Mendelian randomisation using the beta-fibrogen genotype (Youngman, 2000). They studied 4,700 cases and 6,000 controls from the ISIS-3 study and found evidence of an association between the beta-fibrogen genotype and fibrinogen levels. However, the non-significance of the association between the same genotype and myocardial infarction led the authors to reject the hypothesised causal pathway. They concluded that the observed association between fibrinogen and myocardial infarction (OR 1.20; 95% CI: 1.13 to 1.26) was probably a result of confounding or reverse causation.

The ultimate purpose of evaluating the causal association between a risk factor and a disease is to determine the magnitude of the effect of the causal factor. This can direct interventions aimed at preventing the disease. Thus, in order to inform decisions in health related research, estimation is of much greater utility than hypothesis testing (Rothman, 1978; Gardner and Altman, 1986). However, despite the growing interest in Mendelian randomisation, Katan's initial interest in hypothesis testing to confirm or refute the evidence for a particular phenotype-disease association found in observational studies, has not moved towards the estimation of the magnitude of the association, as one would have expected (Tobin *et al.*, 2004 – included in Addenda).

5.3.2   The assumptions of Mendelian randomisation

In 2003, Davey Smith and Ebrahim published a paper in which they thoroughly reviewed the literature and discussed the potentials of Mendelian randomisation for deriving unconfounded estimates of the association between a risk factor and a disease,

whilst emphasising the reliance of the approach on assumptions about the pathway from gene to disease (Davey Smith and Ebrahim, 2003). Among such assumptions, the most important is the one that excludes any pleiotropic effects of the genotype on the disease, that is the genotype influences the disease only through the phenotype considered. If the specific polymorphism also alters the risk of the disease via other pathways, then the estimate of the phenotype-disease association of interest might be seriously biased. Another assumption is that the locus under study is not in linkage disequilibrium with another locus associated with the disease, through either the same or a different phenotype, which means that the two loci segregate independently from one another and therefore their polymorphisms are not associated. The presence of linkage disequilibrium would result in the estimate of the effect of the polymorphism under investigation being confounded by the influence of the other polymorphism. On the other hand, if polymorphisms at more than one locus influence the same phenotype and the loci are not in linkage disequilibrium, then it may be possible to explore combinations of polymorphisms at different loci. The authors refer to this as "factorial Mendelian randomization", where the interest would be in evaluating those combinations of polymorphisms which produce the most extreme difference in the phenotype and thus the highest impact on the disease. Finally, Mendelian randomisation assumes the absence of "canalisation", a term indicating the situation where some developmental compensation has occurred to mitigate the effect of a polymorphism. This may happen when the effect of a polymorphism expressed during foetal development or post-natal growth influences the expression of other genes, which leads to permanent changes aimed at compensating for the influence of the polymorphism.

5.3.3   The role of meta-analysis in Mendelian randomisation

With the recent growth in knowledge about the human genome there has been a dramatic increase in the number of genetic epidemiological studies of the association between specific genes and diseases and between those genes and the risk factors or phenotypes that are thought to be intermediates on the causal pathway to disease. Studies of gene and disease tend to be more common than those on gene and intermediate phenotype (hereafter referred to simply as phenotype). The evidence for a genotype-phenotype association is often obtained as a spin-off from a study primarily

aimed at investigating the genotype-disease relationship, and this information is frequently obtained only on a subset of the subjects. As the number of genetic studies has grown, so meta-analyses have been produced to synthesise the evidence and overcome the limitations of power found in even moderately sized studies. The number of published meta-analyses of genetic association studies has increased at an incredible rate in this last decade; on the 15th of March 2005 the HuGE website archive listed 243 meta-analyses of genetic association studies published from 2000 (§ 4.4.1), while only 37 of such papers published between 1991 and 2000 had been previously identified (Attia, Thakkinstian and D'Este, 2003).

Where there is strong reason to suppose that the phenotype is intermediate on the causal pathway from gene to disease, it would be sensible to perform a meta-analysis that in some way integrates the evidence for all three relationships, genotype-phenotype, genotype-disease and phenotype-disease, where the size of the phenotype-disease association can be estimated from the other two. In fact, the use of Mendelian randomisation within individual studies where both genotype-phenotype and genotype-disease associations are evaluated is strongly limited by the uncertainty in the derived estimate of the phenotype-disease association, which can be very large as it depends on uncertainty in both the estimates of the genotype-phenotype and genotype-disease association (Thompson, Tobin, and Minelli, 2003). Thus, a sufficiently precise estimate of the phenotype-disease association is only likely to be obtained through a meta-analysis of all available evidence, which can also allow testing whether the phenotype is actually on the causal pathway to the disease.

Recently, Mendelian randomisation has been used in a meta-analysis context in three published examples, all evaluating the association of the *MTHFR* gene polymorphism with homocysteine levels but considering different diseases; CHD and myocardial infarction (Wald, Law and Morris, 2002; see also § 5.6); stroke (Casas *et al.*, 2005); venous trombosis (Den Heijer, Lewington and Clarke, 2005). Wald and colleagues estimated the phenotype-disease association using Mendelian randomisation by summarising the evidence available on genotype-phenotype and genotype-disease associations. The approach they used had some methodological problems; first, a meta-analysis was only carried out for the genotype-disease association, while a simple

average was used to summarise the genotype-phenotype association; second, the uncertainty in the genotype-phenotype association was not taken into account in the calculation of the confidence interval for the derived phenotype-disease association. Casas and colleagues addressed both these issues by performing two meta-analyses on genotype-phenotype and genotype-disease and by taking into account the uncertainty in both associations. By performing two separate meta-analyses, they assumed that the genotype-phenotype and genotype-disease estimates were independent and thus ignored the correlation induced by studies measuring both associations. In practice, that was not a problem in their case since there was only one study contributing to both estimates. They then compared this derived estimate with the direct estimate provided by a meta-analysis of classical observational studies, and formally tested the difference between the two using an interaction test. Den Heijer and colleagues followed a similar idea, although they did not attempt to formally derive the estimate the phenotype-disease association based on the evidence on genotype-phenotype and genotype-disease, but used such evidence to assess whether the association between homocysteine and venous thrombosis observed in their meta-analysis of classical observational studies was causal.

5.4 An integrated meta-analysis approach using Mendelian randomisation

5.4.1 Estimation of the phenotype-disease association

In order to use genetic studies to quantify the relationship between the phenotype and disease, the estimate of the genotype-disease association has to be combined with the estimate of the genotype-phenotype association (Figure 5.1). Suppose that a mutant genotype (GG) causes an increased risk of disease compared to the wild type (gg) and that this effect is measured by the odds ratio, $OR_{GG \text{ vs. } gg}$. Further suppose that GG compared to gg causes a mean difference, ΔP, in the level of the intermediate phenotype. Then, under the assumptions required for Mendelian randomisation and assuming linearity of the relationship between phenotype difference and log odds ratio for the disease, $OR_{GG \text{ vs. } gg}^{1/\Delta P}$ is an unconfounded estimate of the odds ratio of disease resulting from a unit change in the phenotype.

FIGURE 5.1 – Derivation of an unconfounded estimate of the effect of a change in phenotype on a disease based on the concept of Mendelian randomisation
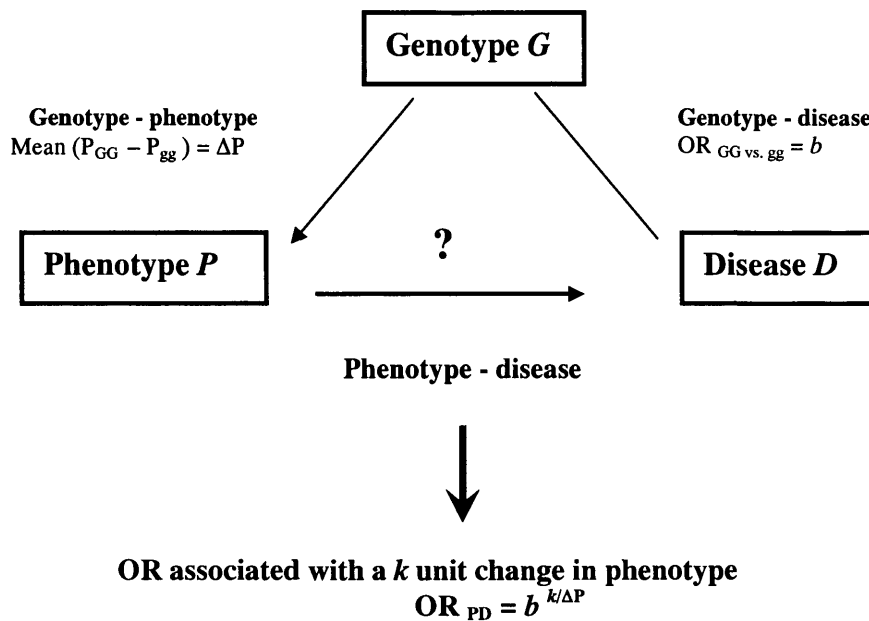
5.4.2 Sources of evidence

When searching for evidence on genotype-phenotype and genotype-disease associations, three different types of genetic studies are likely to be identified; those evaluating only the genotype-phenotype association, those evaluating only the genotype-disease association, and those evaluating both. As well as noting the usual estimates of effect and their precision, it is important to record when both associations are measured in the same study, in which case it is necessary to use a multivariate model in order to allow for the correlation in the genotype-phenotype and genotype-disease evidence arising from the studies that measure both associations (§ 5.5). It might be that studies classified as providing only genotype-phenotype information had, in fact, evaluated also genotype-disease association but using a different disease definition, so that this genotype-disease result cannot be pooled with those of other studies. When collecting data on genotype-phenotype it is important to note whether the information on the phenotype difference comes from cases, controls or a mixture of both. Whenever possible data from cases and controls should be analysed separately. If the disease itself affects the level of the phenotype in a way that is not simply linear,

then the data on the genotype-phenotype effect collected from cases may be less reliable due to reverse causation.

5.4.3 An integrated meta-analytical approach

If the genotype-phenotype and genotype-disease evidence all come from unrelated sources then separate meta-analyses will give estimates of the pooled effects that can, by appealing to Mendelian randomisation, be combined to estimate the size of the phenotype-disease association. Otherwise, it is necessary to use a multivariate model in order to allow for the correlation in the genotype-phenotype and genotype-disease evidence arising from the studies that measure both associations. In either case, the uncertainty in both estimates of genotype-phenotype and genotype-disease associations needs to be taken into account (Thompson, Tobin and Minelli, 2003). In practice there is likely to be a mixture of studies that measure the genotype-phenotype effect, those that measure genotype-disease and those that measure both.

Consider first a meta-analysis in which all available studies measure both genotype-phenotype and genotype-disease. We could proceed as before and separately pool the genotype-phenotype estimates and genotype-disease estimates before combining the pooled values in order to estimate the effect of phenotype on disease. However, the likely correlation in the sizes of pairs of estimates from the same study would affect the width of the confidence interval for the derived phenotype-disease effect and the validity of any hypothesis test. A more appropriate approach would be to combine the genotype-phenotype and genotype-disease estimates separately within each study to obtain study-specific estimates of the phenotype-disease effect. These study-specific estimates could be shown in their own forest plot and pooled to get an overall estimate.

In the more realistic situation in which some studies measure both genotype-disease and genotype-phenotype and some measure one or the other, then one needs to proceed with caution. The most important features of the data will be evident from a forest plot in two columns, one for genotype-disease and the other for genotype-phenotype, in which paired estimates from within the same study are aligned in the same row (see § 5.6). The forest plot is organized in three blocks representing the studies that measured both the odds ratio and the mean difference, those that measured only the odds ratio and

those that measured only the mean difference. Within blocks the studies are ordered by size of effect. Where both estimates are obtained in the same study, studies are ordered by size of the genotype-disease odds ratio. Having drawn the plot, the next stage should be to check that the genotype-disease estimates from studies that also report genotype-phenotype are consistent with the estimates from studies that do not report genotype-phenotype, suggesting that an assumption of exchangeability is appropriate. However, this is a subjective judgement (Spiegelhalter, Abrams and Myles, 2004). Similarly genotype-phenotype estimates should be compared between studies that also report genotype-disease and those that do not.

Funnel plots can be used to look for the presence of publication bias (Sutton *et al.*, 2000; Sterne, Egger and Smith, 2001). However, genetic studies may also be affected by a form of reporting bias in which both the odds ratio and mean difference are measured but only one is reported because the other contradicts the anticipated relationship. Moreover, it is possible that reporting both associations is a marker for a feature such as study quality. It is very difficult to detect this bias from the reported data themselves but a careful reading of the study methods may show if data were collected but not reported and it may be informative to distinguish studies that only reported one of the two associations using a different symbol on the funnel plots (see § 5.6).

When the phenotype does indeed lie on the causal pathway between gene and disease, then studies in populations with a large genotype-phenotype difference might be expected to show a large genotype-disease odds ratio. This can be investigated by plotting each study on a graph of log odds ratio of genotype-disease against average difference in phenotype with genotype (see § 5.6). A similar graphical approach has been used previously in the meta-analysis of randomised trials (Smith, Song and Sheldon, 1993; Midgley *et al.*, 1996). This graph would be expected to show a monotonic trend if the phenotype is intermediate on the aetiological pathway to disease and the line should pass through the origin. Lack of any correlation would cast doubt on whether the phenotype is truly intermediate. A line that does not pass through the origin might indicate that there is another intermediate phenotype through which the gene under study exerts its effect on disease (a special case of pleiotropy), or that the gene is

in linkage disequilibrium (associated at population level) with a gene which also affects the risk of disease, or that there is differential publication bias for the two associations (Sterne, Egger and Smith, 2001; Sutton, Abrams and Jones, 2002) . This graph will also show gross departures from linearity of the relationship between phenotype difference and log odds ratio of disease; approximate linearity being an assumption behind the averaging across studies to give an estimate of the pooled phenotype-disease association.

If it appears that genotype-phenotype and genotype-disease associations are consistent across all studies, then we may pool all genotype-phenotype estimates and all genotype-disease estimates before combining these overall estimates to derive a figure for the phenotype-disease association. The effect of the correlation on the confidence interval and hypothesis test will depend on the proportion of studies that report both genotype-disease and genotype-phenotype associations. The studies that provide both estimates can now be used as described above to provide a comparison with studies reporting one of the estimates and to investigate the consistency of the study-specific phenotype-disease estimates.

5.5 The multivariate meta-analytical approach

5.5.1 Multivariate meta-analysis models

The meta-analysis of genetic studies using Mendelian randomisation is a special case of a multivariate meta-analysis in which the synthesis is simultaneously performed on two correlated outcomes, namely the size of the genotype-phenotype difference and the genotype-disease log odds ratio. Multivariate models have previously been used in meta-analyses that collect data on correlated outcomes within the same study, such as in the synthesis of multiple-treatment studies, synthesis of multiple outcome (or endpoint) studies, and synthesis of studies with both multiple treatments and multiple outcomes (Raudenbush, Becker, and Kalaian, 1988; van Houwelingen, Zwinderman, and Stijnen, 1993; Berkey, Anderson, and Hoaglin, 1996; Arends, Voko, and Stijnen, 2003; Nam, Mengersen, and Garthwaite, 2003). However, unlike most of these other applications of multivariate meta-analysis, the outcome of real interest in Mendelian

randomisation is the single derived phenotype-disease association, calculated as a ratio of the two correlated outcomes.

In any multivariate meta-analysis it is important to allow for the correlation in the estimates from those studies that supply information on more than one outcome. If the ith study supplies correlated outcomes, x_i and y_i, which are assumed to be multivariate normally distributed (MVN) with variances v_{xi} and v_{yi}, then,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{xi} \\ \mu_{yi} \end{pmatrix}, \begin{pmatrix} v_{xi} & \psi\sqrt{v_{xi}v_{yi}} \\ \psi\sqrt{v_{xi}v_{yi}} & v_{yi} \end{pmatrix} \right) \tag{5.1}$$

By using random effects meta-analyses (§ 1.4), in the next level of the hierarchy we assume that different studies vary about the common means μ_x and μ_y with correlated between study variances τ_x^2 and τ_y^2,

$$\begin{pmatrix} \mu_{xi} \\ \mu_{yi} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \tau_x^2 & \rho\tau_x\tau_y \\ \rho\tau_x\tau_y & \tau_y^2 \end{pmatrix} \right) \tag{5.2}$$

The parameter ρ represents the *between-study correlation* across studies measuring both outcomes and ψ represents the *within-study correlation*. While the between-study correlation may be induced by the tendency of *studies* with large effect of genotype on phenotype to also have a large effect of genotype on disease, the within-study correlation may be induced by the tendency of *patients* with large effect of genotype on phenotype to also have a large effect of genotype on disease.

For those studies measuring only the first or second outcome we use the corresponding univariate normal distributions.

$$x_i \sim N(\mu_{xi}, v_{xi}) \quad \text{and} \quad \mu_{xi} \sim N(\mu_x, \tau_x^2)$$

and $\hfill (5.3)$

$$y_i \sim N(\mu_{yi}, v_{yi}) \quad \text{and} \quad \mu_{yi} \sim N(\mu_y, \tau_y^2)$$

In models for meta-analysis in Mendelian randomisation x will represent the log odds ratio of disease given genotype and y will represent the mean difference in phenotype. The object will be to estimate the ratio of x to y as this will give the log odds ratio of

the effect of phenotype on disease. The literature will provide the study estimates and their variances v_{x_i} and v_{y_i}, which may be reported implicitly in the form of standard errors or confidence intervals.

5.5.2   Within-study and between-study correlations

The *within-study correlation*, ψ, represents the possibility that when a particular study estimates one measure as being larger than its actual value then there may be a tendency for the other estimate also to be larger (or smaller) than its true value. This correlation can only be estimated from individual patient data, and is very rarely reported by primary studies. However, there are two good theoretical reasons to suppose that the within-study correlation will be negligible in most studies used in Mendelian randomisation. The study-specific odds ratio of genotype on disease is based on aggregated data and in most studies phenotype level is only measured in a sub-sample of the subjects.

To assess the likely size of the within-study correlation, Thompson (2005 – included in Addenda) performed a small simulation study, where parameters were chosen to reflect the values in studies of the *MTHFR*-homocysteine-CHD pathway. In all of the simulations the estimated correlation was within the range ±0.05. To further investigate the issue, a sensitivity analysis was performed using one of the models presented below (Model B, § 5.5.4), implemented using a maximum likelihood approach. In these analyses a small known amount of within-study correlation was allowed, and the result was that with ψ=-0.1 the OR of a 5. mol/l increase in homocysteine was 1.56 (95% CI: 1.18 to 2.05), while with ψ=+0.1 the OR was 1.53 (95% CI: 1.17 to 2.00). Thus, even if there is a small amount of within-study correlation this seems to have little effect on the final estimate.

The *between-study correlation*, ρ, represents the tendency for studies conducted in populations where the true effect of genotype on phenotype is large also to show a larger than average effect of genotype on disease. In contrast to the within-study correlation, for a meta-analysis of studies with a wide range of populations and designs this correlation may well be substantial, and can be adequately estimated when there

are a reasonable number of studies that report both outcomes.

In order to account for the between-study correlation, two models for the heterogeneities, or between-study variances, in the estimates of genotype-phenotype and genotype-disease associations have been developed. The first model, Model A (§ 5.5.3), is shown to give estimation problems even with large amounts of data. The second model, Model B (§ 5.5.4), overcomes this problem by making the extra assumption that the heterogeneity on the genotype-phenotype association is independent of the heterogeneity on the phenotype-disease association. Results of both models were obtained using a burn-in of 10,000 and a chain length of 50,000. Convergence was assessed via sensitivity analyses with respect to initial values, length of burn-in and length of sample, using both visual inspection of trace plots and by the Geweke, Heidelberger and Welch, and the Raftery and Lewis diagnostic tests (Cowles and Carlin, 1996) implemented in BOA (Smith, 2004). The WinBUGS codes for Model A and Model B are reported in Appendix 4.

5.5.3 Model A

The first model, which is represented schematically in Figure 5.2(a) and is based on the hierarchical bivariate normal distributions described in § 5.5.1, equations (5.1) to (5.3), models the heterogeneities on genotype-phenotype and genotype-disease, while the heterogeneity on phenotype-disease is derived from the other two. Without individual data it is not possible to estimate the within-study correlation, ψ, but for the reasons discussed in § 5.5.2, ψ will be very small and so ψ is assumed equal to 0. Denoting the log odds ratio of phenotype on disease by $\theta = \mu_x/\mu_y$, the marginal distribution becomes:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} \theta\mu_y \\ \mu_y \end{bmatrix} , \begin{bmatrix} v_{xi} + \tau_x^{\,2} & \rho\tau_x\tau_y \\ \rho\tau_x\tau_y & v_{yi} + \tau_y^{\,2} \end{bmatrix} \right) \tag{5.4}$$

When only one of the pair of estimates is reported, they are treated as univariate normal. Thus x_i is normally distributed with mean $\theta\mu_y$ and variance $v_{xi} + \tau_x^{\,2}$, or y_i is normally distributed with mean μ_y and variance $v_{yi} + \tau_y^{\,2}$.

For unknown mean parameters θ and μ_y, vague normal prior distributions, with mean 0 and variance 10,000, were used. A Wishart prior distribution was adopted for the inverse covariance matrix, in which degrees of freedom were chosen to be the rank of the covariance matrix in order to obtain a vague prior distribution (Carlin, 1996). *A priori* beliefs regarding the expectation of the covariance matrix were specified such that the variances of μ_{xi} and μ_{yi} were 0.35 and 7.75 respectively, and the corresponding prior correlation was 0.5, reflecting their observed value in the dataset. However, sensitivity analyses were performed to investigate the impact on the results of the model of the choice of different values for the variance and correlation parameters, while the degrees of freedom of the covariance matrix were kept the same to represent a vague Wishart prior distribution.

5.5.4 Model B

A second model, referred to as Model B, overcomes the limitations of Model A by modelling the heterogeneities of the associations in an alternative manner. The three stages in the triangle shown in Figure 5.1, namely genotype-phenotype, phenotype-disease and genotype-disease will all be subject to heterogeneity, but under the causal model implicit in that figure any one may be derived from the other two. Model B is parameterised in terms of the heterogeneities on the genotype-phenotype and phenotype-disease stages and is based on the critical assumption that they are independent as illustrated in Figure 5.2(b). It is based on the hierarchical bivariate normal distribution described (5.1), but now (5.2) is replaced by:

$$\mu_{y_i} \sim N\left(\mu_y, \tau^2_y\right) \quad , \quad \theta_i \sim N\left(\theta, \tau_\theta^2\right) \quad \text{and} \quad \mu_{x_i} = \mu_{y_i}\theta_i \qquad (5.5)$$

Even under this model, correlation will still be induced in the resultant heterogeneities on genotype-phenotype and genotype-disease. Independence implies that studies that report a large effect of genotype on phenotype will not tend to report relatively larger or smaller effects of the phenotype on disease. In fact, it is unlikely that the effect of a specific level of phenotype on disease would depend on the cause of that level (e.g. genotype rather than any other cause). This is almost certainly reasonable in the case of *MTHFR*, homocysteine and CHD, and is very likely to hold in most other cases.

FIGURE 5.2 - Modelling the heterogeneities of the associations in (a) Model A and (b) Model B

(a)

Model A
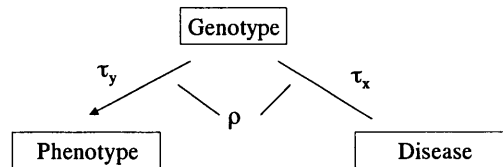
Heterogeneity in *phenotype-disease* association can be derived

(b)

Model B
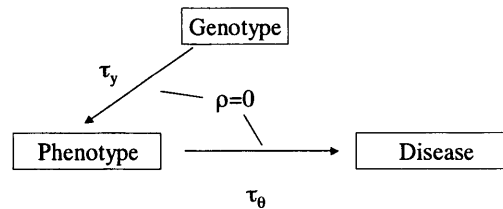
Heterogeneity in *genotype-disease* association can be derived

The specification of Model B, unlike Model A, does not need to approximate the distribution of the genotype-disease association by a normal distribution, and uses a modified product normal formulation (Spiegelhalter, 1998). Moreover, the specification of Model B uses a slightly different method to deal with those studies measuring only one association, either genotype-disease or genotype-phenotype. In these studies, the association that has not been evaluated is treated as missing at random, and missing values are sampled from the corresponding predictive distributions. Thus, all studies are modelled in a single step for both genotype-disease and genotype-phenotype associations.

Unknown mean parameters were assumed to have vague normal prior distributions, *Normal*(0, 10000). For the parameters representing the between-study heterogeneities a *Uniform*(0,2) on standard deviation was used, but sensitivity analyses were performed to investigate the impact of the choice of such prior distribution on the results of the models. In particular, a *Gamma*(0.001,0.001) on the precision parameter and a *Half-Normal*(0,1) on standard deviation were evaluated (see § 2.7.1).

5.6   Illustration

5.6.1   The example of MTHFR gene, homocysteine and CHD

A recent non-genetic meta-analysis on individual patient data from epidemiological studies showed a decrease of 11% in CHD for a 25% decrease of homocysteine levels, with an OR of 0.89 and 95% CI of 0.83 to 0.96 (Homocysteine Studies Collaboration, 2002). The meta-analysis showed heterogeneity between studies partly explained by study design. Retrospective studies yielded higher estimates of risk, perhaps due to reverse causation and/or unadjusted confounding. In particular, two major confounding factors were suggested; smoking and blood pressure. These are both strongly correlated with homocysteine and are known risk factors for CHD. The strong possibility of unadjusted confounding makes it very difficult to be sure that the relationship between homocysteine and CHD is causal.

A common polymorphism of the gene for the *MTHFR* enzyme leads to reduced enzyme activity, lower folate and consequently higher homocysteine levels (Bailey and Gregory, 1999). The polymorphism involves a C-to-T substitution at base 677, so the wild type homozygous genotype is referred to as CC and the mutant homozygous genotype as TT. This polymorphism can be used, together with the idea of Mendelian randomisation, to indirectly assess the effect of homocysteine on CHD.

A recent genetic meta-analysis of individual patient data has shown an increased risk of CHD of about 16% associated with genotype TT compared to CC, with an OR of 1.16 and 95% CI of 1.05 to 1.28 (Klerk et al., 2002). This result was similar to that of another meta-analysis published at the same time but carried out on aggregated data, which showed a pooled odds ratio of 1.21 for TT genotype, with 95% CI of 1.06 to

1.39 (Wald, Law and Morris, 2002). The later paper also mentioned those studies that evaluated the association between genotype and phenotype. They found an average difference of 2.7 μmol/l in homocysteine concentration (95% CI: 2.1 to 3.4) between TT and CC genotypes.

Combining the two meta-analyses by Wald *et al.* and Klerk *et al.*, a total of 66 genetic studies were identified. Classifying the studies that reported both the estimate and its precision, 32 evaluated only genotype-disease association, 16 only genotype-phenotype association, and 18 both. The definition of CHD used in the analysis presented is myocardial infarction or angiographically confirmed coronary artery occlusion (>50% of the luminal diameter). Genotype-disease associations were reported in an additional 13 studies in the original meta-analyses, but this information was not included because of either a different disease definition (e.g. atherosclerotic vascular disease), or a restricted study population (e.g. diabetic subjects). The data for the genotype-disease and/or genotype-phenotype associations for all the 66 studies are reported in Appendix 4, Table A, together with the study references.

Among the 18 studies evaluating both associations, 9 measured the mean difference in phenotype level with genotype in both cases and controls (4 reporting only combined means), 4 measured homocysteine only in cases, 3 only in controls, and 2 reports were unclear.

5.6.2   Results

Figure 5.3 shows the two-column forest plot with the first column representing the genotype-disease log odds ratio and the second the genotype-phenotype mean difference. It is clear that there is considerable variation between studies, with some even reporting average odds ratios less than one or mean differences in homocysteine in the opposite direction to that anticipated. Figures 5.4a and 5.4b show the funnel plots for the genotype-disease and genotype-phenotype associations. For the genotype-disease association, there seems to be no evidence of either publication bias, indicated by an overall lack of symmetry in the funnel plot, or reporting bias, suggested by a discrepancy in the shape of the funnel plot between studies reporting both and those reporting only one association. For the genotype-phenotype association, the funnel plot

FIGURE 5.3 – Two-column forest plot and pooled estimates for genotype-disease and genotype-phenotype associations. *A*=Asians; *E*=European; *F*=Female; *M*=Male. Full references are reported in Appendix 4, Table A
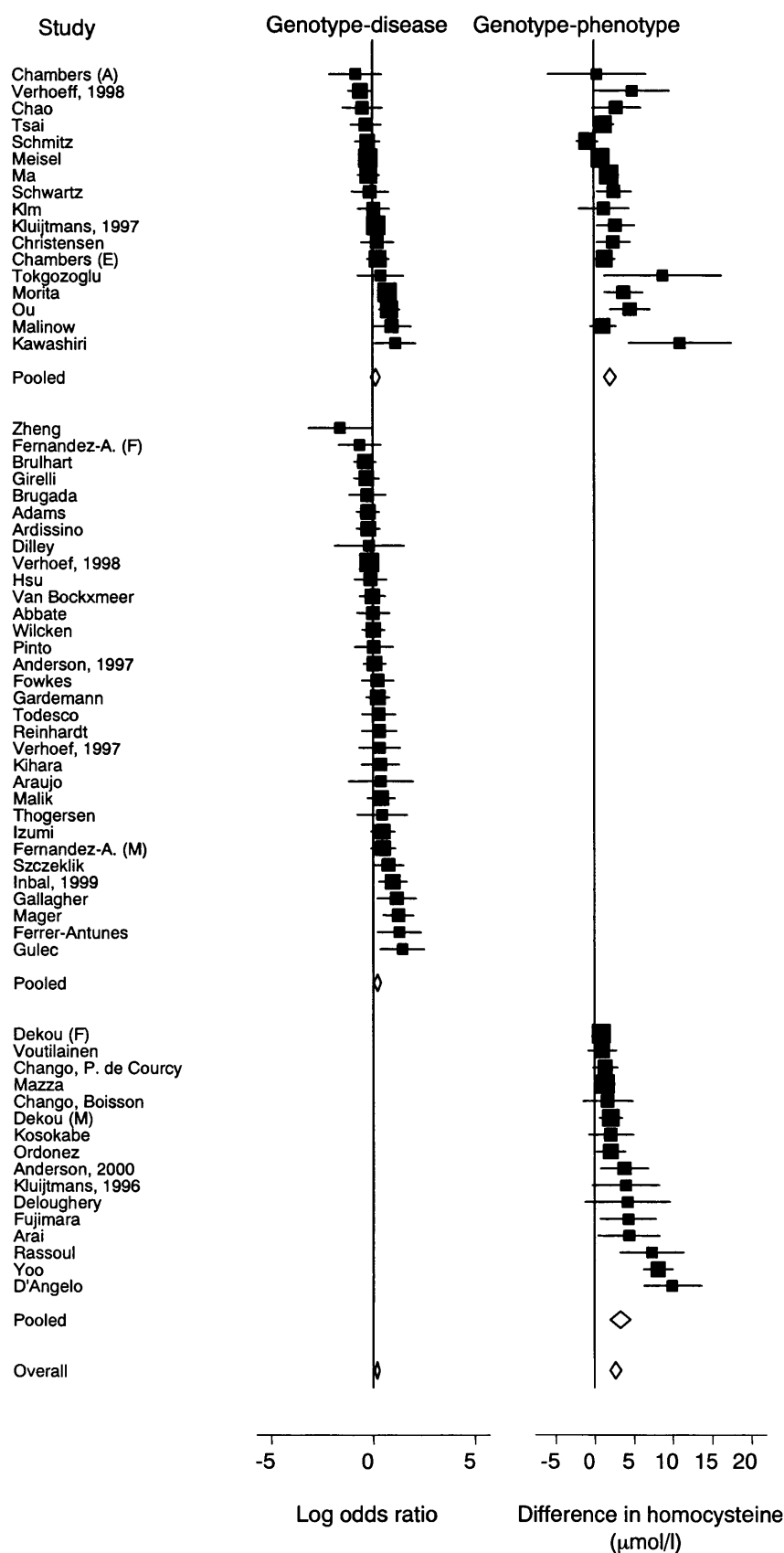
FIGURE 5.4 – Funnel plots for (a) genotype-disease and (b) genotype-phenotype associations. Different symbols used for those studies measuring both associations (Δ) and those measuring only the association of interest (o)
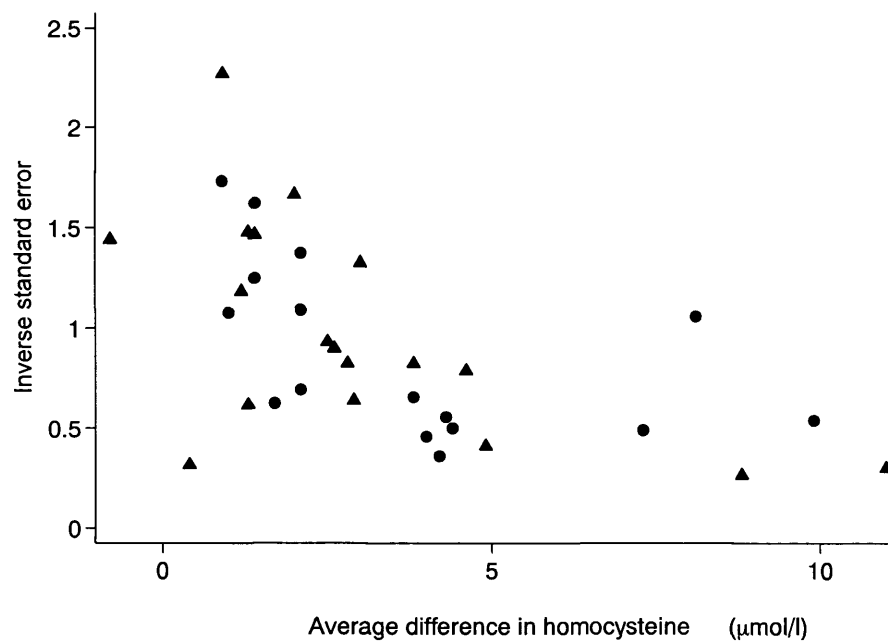
(a)

(b)

is suggestive of possible publication bias, while there appears to be little evidence of reporting bias.

The pooled estimate of the odds ratio of genotype on CHD based on studies that also reported the homocysteine change is 1.17 (95% CrI: 0.92 to 1.49) and where homocysteine was not reported the pooled odds ratio is 1.23 (95% CrI: 1.04 to 1.47). The difference is not statistically significant (p=0.75). Similarly the mean change in homocysteine in studies that also reported CHD is 2.06 μmol/l (95% CrI: 1.35 to 3.03) and in studies that did not report on CHD it is 3.35 μmol/l (95% CrI: 2.00 to 4.84). The direction of the difference is consistent with the presence of publication bias, whereby studies evaluating the genotype-phenotype association alone are only published if the effect size detected is large. However, this difference is not statistically significant (p=0.12). The lack of significant differences gives some confidence for pooling all odds ratio estimates to obtain 1.21 (95% CrI: 1.05 to 1.40) and all mean differences in homocysteine to obtain 2.75 μmol/l (95% CrI: 2.00 to 3.61).

The next step is to investigate more fully the 18 studies that report both an odds ratio and a mean difference. Figure 5.5 shows the forest plot of the study-specific estimates

FIGURE 5.5 – Forest plot for the derived study-specific estimates of the odds ratio of phenotype on disease (per 5 unit difference in homocysteine). *A*=Asians; *E*=European. Values have been truncated at ± 15
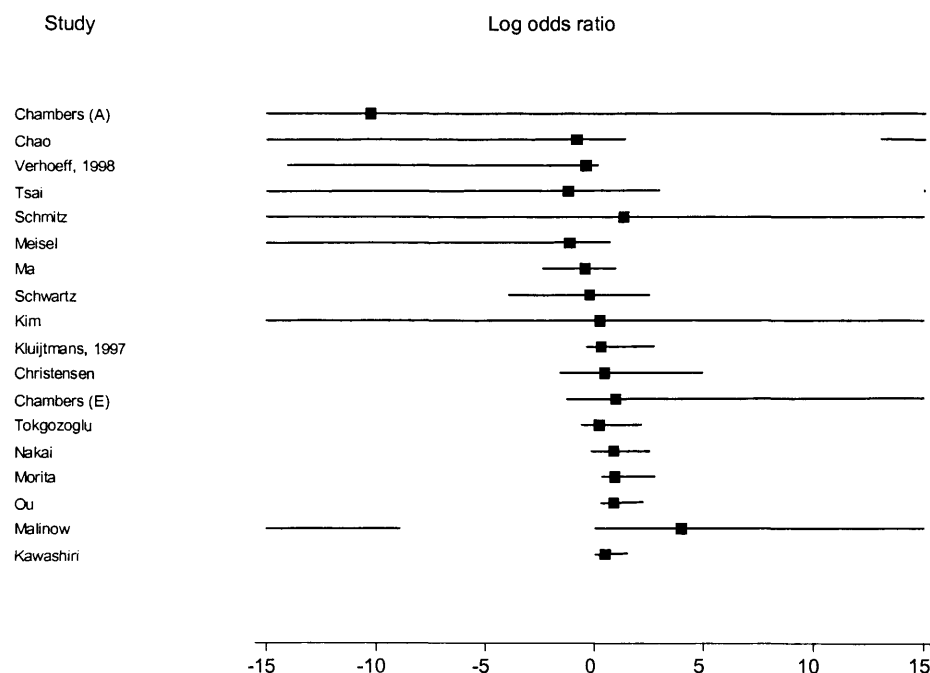
FIGURE 5.6 – Log odds ratio of *MTHFR* gene on coronary heart disease against mean difference in homocysteine between genotypes, separately for studies in which phenotype was measured on (a) cases and (b) controls. The axes of the ellipses are inversely proportional to the standard errors of the respective associations.

(a)
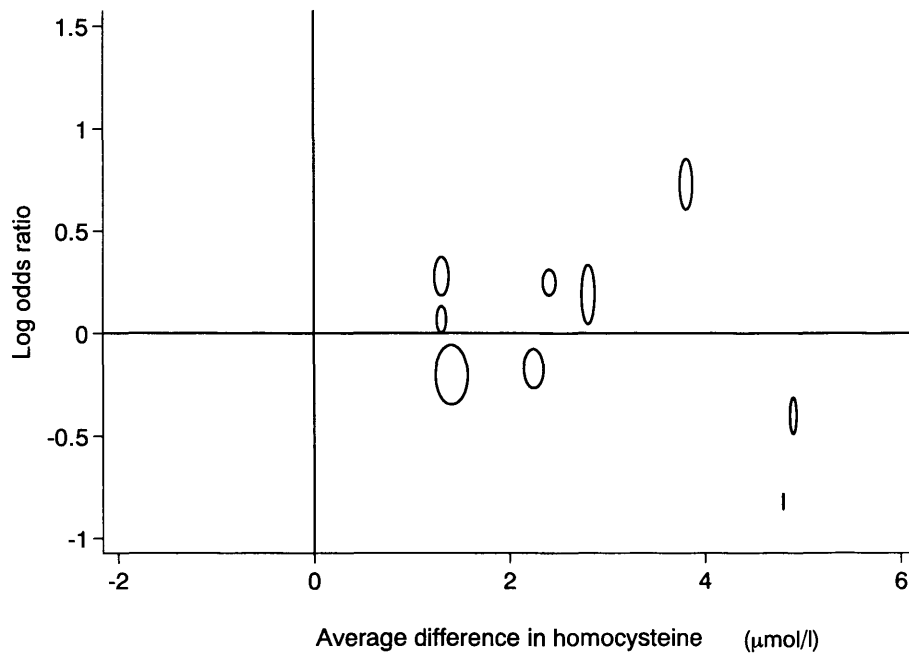
(b)
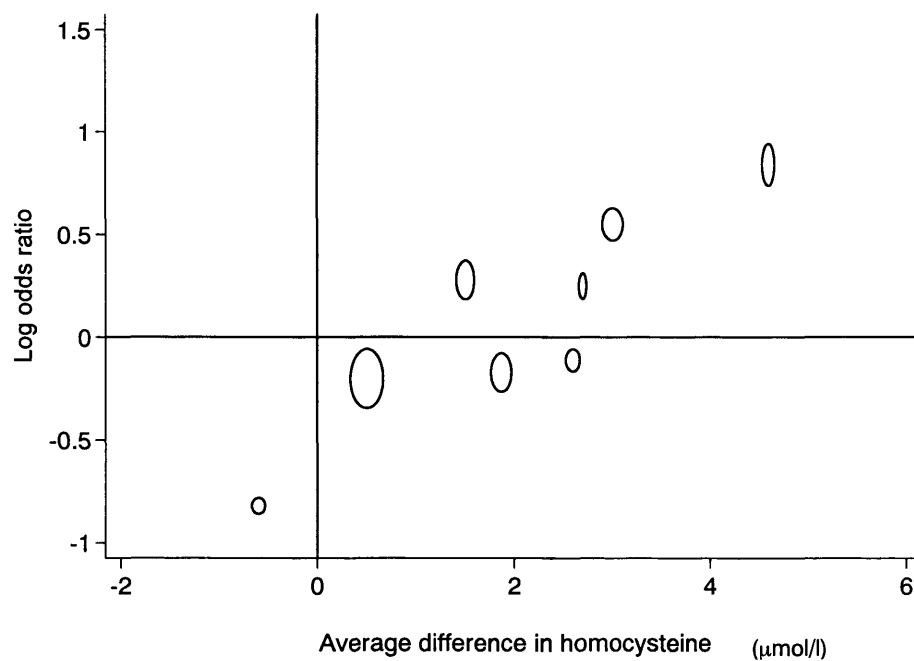
of the odds ratio of CHD associated with a 5μmol/l increase in homocysteine. Because, on a log odds ratio scale, the derived phenotype-disease association is obtained by division, when study results suggest that the homocysteine difference could reasonably be either side of zero, then the confidence interval for the ratio could stretch to plus or minus infinity (Fieller, 1954; Kendall and Stuart, 1973). In some cases this infinite range is accompanied by a gap, or clearing, in the forest plot associated with values that the ratio is very unlikely to take. The most important feature of Figure 5.5 is the poor precision in the derived estimate that can be obtained from most individual studies. Indeed, several derived estimates have infinite variance.

Figures 5.6a and 5.6b present plots of the log odds ratio of genotype-disease against the mean difference in homocysteine, separately for homocysteine measured in cases and controls. The two figures include all studies that either measured homocysteine only in cases or controls, or that measured it in both and reported separate estimates for the two groups. To allow for the large difference in precision of the different studies, the individual estimates are plotted as ellipses, with their axes inversely proportional to the standard error of the log odds ratio of genotype-disease and the mean change in homocysteine. Both figures show an approximately linear relationship, with the line passing close to the origin. As anticipated the pattern is somewhat clearer when control data are used. The unweighted correlations observed in Figures 5.6a and 5.6b are 0.37 ($p=0.29$) and 0.78 ($p=0.01$), respectively.

Table 5.1 reports the results of Model A and Model B for all parameters estimated. In particular, θ represents the log odds ratio of the phenotype-disease association, that is the increase in the log odds of CHD associated with a unit increment of homocysteine. It may be more informative to rescale this odds ratio for increments other than a unit increase in homocysteine. For instance, the odds ratio of CHD for a standard reference increment of 5 μmol/l, as used by Wald et al. (Wald, Law and Morris, 2002), is 1.48 (95% CrI: 1.15 to 1.99) and 1.52 (95% CrI: 1.17 to 2.06) for Model A and B, respectively, while for an increment of 3 μmol/l considered to reflect the possible size of a lowering homocysteine intervention with folic acid supplementation (Homocysteine Lowering Trialists' Collaboration, 1998), is 1.26 (95% CrI: 1.09 to 1.51) and 1.29 (95% CrI: 1.10 to 1.54) for Model A and B, respectively.

TABLE 5.1 – Results for Model A and Model B, expressed as median with 95% CrI

MODEL	PARAMETER				
	θ (95%CrI)	μ_y (95%CrI)	τ_x (95%CrI)	τ_y (95%CrI)	ρ (95%CrI)
Model A	0.078 (0.026 to 0.137)	2.673 (1.971 to 3.437)	0.101 (0.010 to 0.307)	3.178 (1.594 to 3.958)	0.689 (0.095 to 0.913)
Model B	0.082 (0.026 to 0.147)	2.640 (1.916 to 3.386)	0.120 (0.041 to 0.280)	3.014 (1.589 to 3.945)	Not in the model

Sensitivity analyses showed that the results of Model A are sensitive to the values specified for the hyperparameters chosen to represent *a priori* beliefs regarding the assumed variances and correlation in the Wishart prior distribution (Appendix 4, Table B). Indeed, it will usually be difficult to use Model A since there is unlikely to be sufficient information to estimate the between-study correlation accurately, so that any prior distribution will necessarily exert some influence on the final results. For both Model A and Model B, the results of sensitivity analyses considering different prior distributions for the heterogeneity terms, τ_x and τ_y, showed differences in the estimates of these parameters but minimal impact on the estimates of the parameters of interest, θ and μ_y (Appendix 4, Table C).

5.6.3 Effect of ignoring correlation

Performing two independent meta-analyses on genotype-disease and genotype-phenotype data and then estimating the phenotype-disease association based on the two pooled estimates means ignoring the correlation between the two outcomes, disease and phenotype, when measured in the same study. In general, multivariate models should be used whenever the outcomes of interest are correlated and measured within the same studies. However, the potential gain in terms of increased precision or reduced bias may be small, depending on the proportion of studies that measure both outcomes. When ignoring correlation in the example of the *MTHFR* gene, homocysteine and CHD, the OR for a 5μmol/l increase in homocysteine was 1.44 (95% CrI: 1.10 to 1.94), as compared with 1.48 (95% CrI: 1.15 to 1.99) and 1.52 (95% CrI: 1.17 to 2.06) of the multivariate meta-analysis when using Model A and Model B, respectively. Although in this example there is not much difference between the results of univariate and multivariate models, the impact of properly accounting for the correlation might be

greater in other situations.

5.7   Discussion

5.7.1   Mendelian randomisation: genotype as an instrumental variable

Mendelian randomisation is an example of the use of an instrumental variable; a technique that has been used in econometrics since the 1920s but only occasionally applied to health sciences to control for confounding and measurement error (Angrist, Imbens, and Rubin, 1996; Greenland, 2000). These methods derive an unconfounded estimate of the association between the exposure and outcome of interest from the observed relationship of an instrumental variable with both exposure and outcome. The use of instrumental variables, despite their potential role in epidemiology, has been mainly limited to measurement error (Hardin and Carroll, 2003) and to the field of randomised clinical trials. A typical application in randomised trials is to use the allocated treatment as an instrumental variable to control for the bias due to non-compliance as an alternative to the more usual analysis by intention-to-treat (Angrist, Imbens, and Rubin, 1996). A likely explanation for the limited use of these methods is that it is often difficult to find suitable instrumental variables, since the method requires that the variables have not only an unconfounded relation with both the exposure and the outcome of interest, but also an association with the outcome that is explained by the association with the exposure. In some situations Mendelian randomisation allows the subject's genotype to be used as an instrumental variable, with genetic studies providing information on the impact of a specific mutation on both the phenotype (risk factor) and disease of interest. If the gene is carefully chosen then both associations are unconfounded because the genotype is effectively randomly assigned, but the important assumption that the genotype is associated to the disease only through the phenotype of interest needs to be carefully assessed (Davey Smith and Ebrahim, 2003).

5.7.2   The proposed meta-analytical approach

Although genotype-disease associations are becoming better understood, it is only when we also have information about the causal pathway that we open up the possibility of preventive or therapeutic intervention. Thus, while the association

between *MTHFR* polymorphisms and CHD is scientifically interesting, disease prevention becomes possible only when we understand that this effect acts, at least in part, through homocysteine. Dietary folate supplementation is a relatively simple intervention that can be implemented at a population level to lower homocysteine and thereby reduce CHD. This intervention was introduced in the U.S.A. in the late 1990's with the fortification of cereals and grains (Rader, 2002). Classical epidemiological studies may provide evidence about the phenotype-disease association but it will almost certainly be affected by confounding and/or reverse causation. The use of Mendelian randomisation offers a novel way of deriving unconfounded estimates, although it should be remembered that Mendelian randomisation makes its own assumptions about the pathway from gene to disease (§ 5.3.2). The most crucial assumption is that the genotype influences the disease risk *only* through the modification of the specific phenotype. If the genetic polymorphism also alters the risk of the same disease via other pathways, then the estimate of a specific phenotype-disease association might be seriously biased. In the example of homocysteine and CHD this is probably not a problem, but for instance, polymorphisms of *APO-E* gene affect several different intermediate phenotypes related to lipid metabolism and atherosclerosis (Davey Smith and Ebrahim, 2003). Consequently it is advisable to limit the use of Mendelian randomisation to studies where there is good biological knowledge of the genotype-phenotype-disease pathway.

The approach proposed in this chapter stresses the need for meta-analyses to review simultaneously the stages in the genotype-phenotype-disease pathway and by implication advocates that, whenever possible, individual studies of genotype-disease associations should collect information on intermediate phenotypes. In fact, the analysis of studies which measure both associations allow an insight into the inter-relationships between genotype, phenotype and disease and gives the opportunity to check the assumptions of the analysis. In this respect, a meta-analysis of small studies might be more informative than a single large prospective study. In some meta-analyses inconsistencies across studies could result in departures from the linear trend seen in Figure 5.6. This might happen if study populations differ with respect to phenotype measurement, disease definition, gene-environment interactions, compensatory developmental processes (canalisation) or linkage disequilibrium with functional alleles

(Davey Smith and Ebrahim, 2003). Theoretical considerations (§ 5.4.2), and the example presented, suggest that it may be safer for primary researchers to measure the phenotype in controls in order to avoid any possibility of bias due to reverse causation. If the phenotype level is measured in both cases and controls, then two separate estimates should be reported. For meta-analyses, our recommendation is to perform sensitivity analyses, with phenotype data obtained on cases analyzed separately from those obtained on controls.

The need for an integrated meta-analytical approach to genetic studies when using Mendelian randomisation is particularly important. The uncertainty associated with the derived estimate of the phenotype-disease association can be large as it depends on uncertainty in both the estimate of the genotype-phenotype and genotype-disease association (Thompson, Tobin and Minelli, 2003). It is crucial in the use of Mendelian randomisation that both estimates are sufficiently precise, but especially that of the genotype-phenotype association. Such precision is only likely to be obtained through a meta-analysis of all available evidence. In fact, at present, almost all genetic studies are statistically underpowered to detect the relatively small effects of the many gene variants that underlie common, complex diseases (§ 1.6). Although massive reductions in genotyping costs offer the prospect of larger studies, study size remains limited by the cost of proper phenotyping (Thompson, 2002).

While meta-analyses can, in theory at least, partially alleviate the problem of inadequate statistical power, they cannot control the problems of publication and reporting bias (Sutton *et al.*, 2000; Sterne, Egger and Smith, 2001), that are thought to be particularly important in genetic epidemiology (Ioannidis *et al.*, 2001; Colhoun, McKeigue and Davey Smith, 2003). However, an integrated meta-analytical approach can start to address these issues by comparing the pooled estimates for genotype-phenotype and genotype-disease associations in studies reporting either only one or both associations, and by drawing the funnel plots in such a way that allows comparison between the two types of studies for each association.

The analysis of the correlation between the genotype-disease odds ratio and the genotype-phenotype difference, as typified by Figure 5.6, has to be interpreted with

care. The plot is based on data aggregated over studies and is analogous to an ecological study and potentially subject to the ecological fallacy, i.e. patterns seen in aggregate data do not necessarily translate to the individual (Greenland and Robins, 1994). Thus, when we see an increase in the risk of disease in studies that show an increased difference in phenotype it is probable, but not certain, that we would see a similar effect at the individual level. Equally a failure to see a pattern in aggregate data does not rule out the possibility of an individual level effect. Obviously an individual level causal effect is required for an intervention on the phenotype to have an impact on the risk of disease.

It is tempting to add non-genetic studies of the phenotype-disease association to the integrated approach presented here, if only to test whether they accord with the estimate derived from the application of Mendelian randomisation. Unfortunately the sample sizes required to establish equivalence of the measured and derived estimates are such that even a large meta-analysis may not suffice (Thompson, Tobin and Minelli, 2003). This clearly is an area that requires more work because the ultimate aim should be to produce an integrated meta-analysis that links together all relevant phenotypes, diseases and genotypes, including heterozygotes.

5.7.3 Choice of the multivariate model

When synthesising evidence of genetic studies for use in a Mendelian randomisation analysis, studies evaluating genotype-phenotype, genotype-disease or both associations together are likely to be encountered. Simulations suggest that in this situation the within-study correlation is likely to be very small, but it is still important to allow for the between-study correlation in the heterogeneities of studies that evaluate both genotype-phenotype and genotype-disease associations. Heterogeneities on the genotype-phenotype and genotype-disease associations may be highly correlated but a multivariate model that parameterises the heterogeneity directly (Model A) is difficult to fit because the correlation is poorly estimated. An alternative approach is therefore advocated, which treats the heterogeneities on genotype-phenotype and phenotype-disease as being independent, and implicitly defines the correlation between the heterogeneities on genotype-phenotype and genotype-disease (Model B).

A different method, which has not been investigated but could also be considered, is that of estimating the phenotype-disease association within each study evaluating both genotype-phenotype and genotype-disease associations. These estimates of the phenotype-disease association could be combined in the meta-analysis model with the indirect estimates of the phenotype-disease association obtained from those studies that only estimate either the genotype-phenotype or the genotype-disease association. Further work would be required in order to assess whether and when this method may be advantageous compared to the methods proposed.

The results of the two multivariate models presented in this chapter were obtained using a Bayesian approach with vague prior distributions for all parameters. However, the two models were also implemented by Thompson (2005 – included in Addenda) using a maximum likelihood approach. The use of Markov chain Monte Carlo (MCMC) methods for parameter estimation in the Bayesian approach avoided the requirement for Taylor series approximations of the maximum likelihood approach, which might have an impact on the results. Although simulation methods could also be used for maximum likelihood inference (Geyer, 1996), they are more difficult to implement, mainly because of the current lack of suitable software. However, despite the differences in the Bayesian and maximum likelihood approach in terms of approximations and structure, the results for the specific example considered were very similar. For the maximum likelihood as for the Bayesian approach, the paucity of information to accurately estimate the between-study correlation in Model A will usually make it difficult to use this model.

When using a Bayesian approach re-parameterisation may not be theoretically necessary as the posterior distribution of functions of the model parameters may be obtained directly from the MCMC samples. It can nevertheless, depending on the precise sampling algorithm used (Brooks, 1998), be desirable to re-parameterise in order to improve performance of the MCMC algorithm, especially in a hierarchical or non-linear model setting (Gelfand, Sahu, and Carlin, 1995). An important issue raised by the use of a Bayesian approach is the choice of vague distributions (O'Hagan, 1994; Spiegelhalter, Abrams, and Myles, 2004). In the meta-analysis context presented in this chapter, particularly important is not only the choice of the prior distribution used for

the between-study variance (§ 1.5.1, § 2.7 and § 2.8), but also of the values assumed for the parameters of the Wishart prior distribution when adopting Model A. Finally, although not considered here, a fully Bayesian approach with prior distributions based on data from other pertinent studies or expert opinion and other evidence could be adopted. However, elicitation of beliefs regarding the model parameters in such meta-analysis (e.g. correlations) is not straightforward (Gokhale and Press, 1982; Garthwaite and Dickey, 1988; Garthwaite, Kadane and O'Hagan, 2005).

5.7.4 Conclusions

Mendelian randomisation is an important tool for epidemiologists not only for testing the hypothesis of the existence of a particular phenotype-disease association found in classical observational studies - which are prone to confounding - but also for estimating the magnitude of the association. However, due to the great uncertainty in the estimates of phenotype-disease associations derived using Mendelian randomisation, evidence synthesis of genetic information on both genotype-disease and genotype-phenotype association is of crucial importance in obtaining estimates with sufficient precision. Such meta-analysis requires a multivariate approach and an adequate modelling of the underlying interrelated heterogeneities of the three associations. An integrated meta-analytical approach offers the possibility of checking some crucial assumptions underlying Mendelian randomisation. Should these assumptions not be satisfied, then the advocated "deconfounding" power of Mendelian randomisation would no longer be true.

6   GUIDELINES

6.1   Chapter overview

This chapter addresses the practical issue of how to provide comprehensive recommendations that might assist investigators undertaking a meta-analysis of genetic association studies. In § 6.2, an overview of the quality issues reported for published meta-analyses of genetic association studies is presented, and these findings are compared with those reported in other fields of medicine. Section 6.3 explains the need of comprehensive recommendations in this field, in the light of guidelines developed for general meta-analysis. In § 6.4, a set of guidelines are proposed which combine the findings presented in the previous chapters with the evidence already available in the literature. In particular, § 6.4.1 presents recommendations for dealing with general meta-analysis issues based on existing meta-analysis guidelines for RCTs and observational studies, which are adapted to consider characteristic problems of this field. Guidelines covering the specific issues of meta-analysis of genetic association studies are presented in § 6.4.2 at two levels of sophistication, one relatively simple although methodologically correct, the other more sophisticated but more efficient. In § 6.5 the recommendations are extended to consider the case of the meta-analysis of genetic association studies using Mendelian randomisation. Finally, § 6.6 discusses the relevance of guidelines in this field, the limitations of the guidelines proposed and the need for further work.

6.2   Evidence on the quality of published meta-analyses

The problem of poor methodological quality in the field of genetic association studies affects first of all primary studies. Indeed, poor quality has been considered one of the explanations for the lack of reproducibility of study results, which represents such a major problem of this field (Cardon and Bell, 2001). Meta-analysis is advocated as an important tool in making sense of conflicting results from different studies, but this can be achieved only if the evidence synthesis is performed using appropriate methodology. A few papers have recently evaluated different aspects of the quality of meta-analyses of genetic association studies, and their conclusion is unanimously that meta-analyses

of genetic association studies could be improved in many respects (Attia, Thakkinstian and D'Este, 2003; Munafo and Flint, 2004; Salanti, Sanderson and Higgins, 2005). The flaws in the conduct and reporting of meta-analysis highlighted in these reviews include both general meta-analysis issues and issues that are specific to this field.

Attia and colleagues reviewed meta-analyses of genetic association studies published between 1991 and 2000 (Attia, Thakkinstian and D'Este, 2003), with the aim of evaluating their quality and highlighting how traditional meta-analysis issues, as well as issues specific to the field, had been dealt with. Since no quality scale specific for meta-analyses of genetic association studies has been developed yet (§ 6.3), these authors used a set of criteria which included five general issues and two issues specific of this field. The five general items were: 1) description of the search strategy; 2) description of the inclusion/exclusion criteria; 3) whether the authors performed any test for heterogeneity and whether, in the presence of heterogeneity, they tried to identify the cause; 4) description of the statistical methods used to combine study results and whether these corresponded to accepted meta-analytical methods; 5) whether the authors explored the possibility of publication bias, and if so how. The two specific items were: 1) whether the authors checked Hardy-Weinberg equilibrium (HWE) and how; 2) what genetic model the authors chose to pool the data. Based on these criteria, they assessed the quality of the 37 genetic meta-analyses identified (see § 3.3), and concluded that meta-analyses of genetic association studies could be improved in virtually all respects.

Their results for the five general methodological issues were the following:

1. In almost all cases MEDLINE was the only source of studies used by the investigators. Twenty-eight studies (76%) did not report their search terms, and 13 (35%) did not even state their sources.

2. Nineteen studies (51%) did not describe either inclusion or exclusion criteria.

3. Nine studies (24%) did not assess statistically the presence of heterogeneity. Among the 13 studies which found and reported heterogeneity, 5 (38%) did not explore the reasons for that and 8 (62%) proceeded using a fixed effect model.

4. Among the 13 systematic reviews where a meta-analysis was performed, only 1 (8%) did not describe the statistical method used to combine the data. All of the other 12 meta-analyses used standard methods; in particular, 4 (33%) used a random effects model, 7 (58%) used a fixed effects model, and 1 (9%) used both.

5. Thirty (81%) did not check for publication bias (in those which did, visual inspection of a funnel plot was the sole method used).

In order to put the findings of Attia's review into context, a literature search was performed to identify similar reviews that evaluated the quality of published meta-analyses in different fields of medicine. The electronic search in MEDLINE, from 1966 to August 2005, proved to be particularly difficult due to a huge number of records obtained when trying to be comprehensive. After attempting different methods, the search strategy detailed in Figure 6.1 was used, where the search for keywords was limited to the title of the papers. This yielded 544 records, to which only one (Sacks, 1987) was added based on the cross-reference checking of all relevant papers. The retrieval of the articles was limited to those published in English and available in our library, and those which assessed more than 10 meta-analyses. In the reviews retrieved, different criteria were used by the authors to evaluate the quality of the meta-analyses, and not always these criteria were comparable with those adopted by Attia and colleagues. In particular, 7 reviews were excluded since they reported on less than two items out of the five described in Table 6.1. Details of the inclusion and exclusion of reviews are presented in Figure 6.2. Among the 13 reviews included, Oxman and Guyatt scale was adopted in about 60% of the cases (Oxman and Guyatt, 1991); without additional information provided by the authors, this quality scale contributed to three of the five items in Table 6.1. The results of the 13 reviews included, together with the field of research considered, number of meta-analyses included and years of publication of the meta-analyses, are shown in Table 6.1.

The findings reported in Table 6.1 are not comprehensive of what available in the literature and suffer from other limitations, including the fact that some meta-analyses might have been included in more than one review. Nonetheless, they provide a

Figure 6.1 – Search strategy used to identify in MEDLINE reviews that evaluated meta-analyses published in any field of medicine

#	Search History	Results
1	meta-analysis.m_titl.	5702
2	meta-analyses.m_titl.	380
3	systematic review.m_titl.	3435
4	systematic reviews.m_titl.	441
5	1 or 2 or 3 or 4	9577
6	quality.m_titl.	59819
7	evaluat$.m_titl.	221443
8	assessment.m_titl.	90177
9	survey.m_titl.	47373
10	appraisal.m_titl.	4960
11	methodolog$.m_titl.	15199
12	6 or 7 or 8 or 9 or 10 or 11	426684
13	5 and 12	**544**

Figure 6.2 – Flow chart of the inclusion and exclusion of reviews identified by MEDLINE search
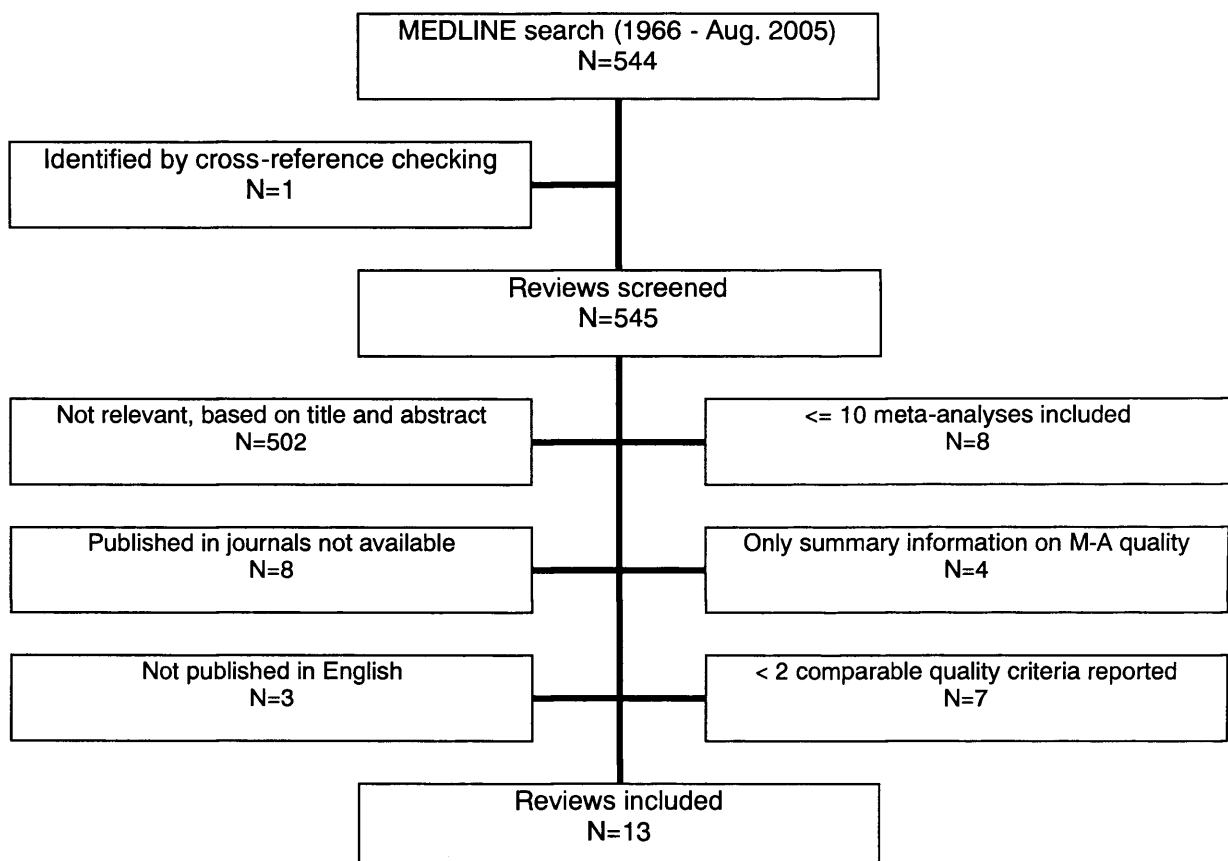
background which helps evaluating how the methodological quality of meta-analyses of genetic association studies compares with that observed in other areas of medical research. Table 6.1 clearly shows how the need for improvement of the quality of published meta-analyses is not limited to the field of genetic association, but indeed

Table 6.1 – General methodological issues reported in published meta-analyses of genetic association studies compared to published general meta-analyses

Author, year	Field of research	N. of Meta-analyses	Years of publication of meta-analyses	METHODOLOGICAL ISSUES (% of meta-analyses fulfilling the criteria)				
				Search strategy reported	Inclusion criteria reported	Pooling methods reported	Statistical heterogeneity evaluated	Publication bias evaluated
Attia, Thakkinstian, and D'Este, 2003	Genetic association studies	37	1991-2000	24	49	92	76	19
Dixon *et al.*, 2005	General surgical literature*	51	1997-2002	67	70	67	/	/
Shea *et al.*, 2002	Any field of medicine RCTs	52 Cochrane	1993-1996	31	74	98	29	8
		52 Journals	1990-1995	64	46	85	65	17
Moher *et al.*, 2002	Paediatric complementary and alternative medicine, and conventional therapy*	66	Not stated-2001	52	64	41	38	17
Bhandari *et al.*, 2001	Orthopaedic surgery RCTs and observational studies	40	1984-1999	83	78	70	/	/
Choi *et al.*, 2001	Anaesthesia*	82	1989-1999	73	81	82	35	5
Kelly *et al.*, 2001	Emergency medicine*	29	1990-1998	55	69	74	/	/
Fishbain *et al.*, 2000	Chronic pain treatment*	16	1988-1998	/	88	/	38	19
Jadad *et al.*, 2000	Asthma RCTs & observational studies	50	1988-1998	66	60	52	40	16
Jadad *et al.*, 1998	Any field of medicine RCTs	36 Cochrane	1995	/	90	/	47	/
		39 Journals			46		54	
Jadad and McQuay, 1996	Pain research RCTs & observational studies	80	1980-1993	61	73	71	/	/
Sacks *et al.*, 1996	Any field of medicine RCTs	58	1987-1990	69	67	78	47	41
Assendelft *et al.*, 1995	Spinal manipulation RCTs	51	1977-1993	27	35	/	/	/
Sacks *et al.*, 1987	Any field of medicine RCTs	86	1955-1986	35	44	66	23	2

* Not stated whether meta-analyses of RCTs, observational studies, or both.

represents a more general problem. This also applies to Cochrane meta-analyses published in The Cochrane database of Systematic Reviews, although they tend to show higher quality compared to meta-analyses published on peer-review journals (Jadad *et al.*, 1998; Olsen *et al.*, 2001; Shea, 2002). Overall, the variability of the figures reported for the five items in Table 6.1 appears to be more associated with differences across disciplines than years of publication of the meta-analyses. As regards the meta-analyses of genetic association studies, the comparisons in Table 6.1 shows how the description of the search strategy adopted by the authors is particularly flawed compared to other fields of medicine, with only 24% of the meta-analyses reporting this information. This is somewhat surprising, given that the importance of this aspect in systematic reviews and meta-analyses is well recognised (Hunt, Jaeschke and McKibbon, 2000; Robinson and Dickersin, 2002). The problem might be worse in meta-analyses of genetic association studies due to the lack of recommended strategies for the electronic literature search, which instead have been developed for the identification of RCTs by the Cochrane Collaboration (http://www.cochrane.dk/cochrane/handbook/hbook.htm). Poor compared with meta-analyses in other fields is also the reporting of inclusion criteria, particularly when comparing meta-analyses published in similar years. On the other side, the reporting of the statistical methods used to combine the data seems very good in meta-analyses of genetic association studies; as reported above, Attia's review showed how accepted meta-analysis methods were always used, although this does not imply that they were the most appropriate in each case. Particularly good in comparison to meta-analyses in other field of medicine seems to be the assessment of statistical heterogeneity based on the use of a heterogeneity test. This is probably explained by the general awareness to the extent of the problem of variability in study results in genetic association studies (§ 3.2.2). Finally, the assessment of the presence of publication bias based on funnel plots or other methods appears to be very poor, although this seems to reflect a general problem of published meta-analyses across all fields of medicine.

As regards the other two specific methodological items evaluated by Attia and colleagues, their findings were:

1. Seventy-six percent of the meta-analyses did not assess the presence of HWE, and among those which did that, 33% did not limit the evaluation of HWE to control subjects as should be done (see § 4.2.1).

2. The choice of how to pool data according to the underlying genetic model showed how the approaches used varied considerably, and justification for the choice of the approach was not provided in most cases (see § 2.3). In particular, 24% of the meta-analyses determined association solely by comparing allele frequency between cases and controls; 33% determined the effect by assuming one specific genetic model, and among these, 25% did not give any reason for the choice; finally, 43% performed multiple comparisons by applying different models (among which, only 13% adjusted for multiple testing).

6.3   Available guidelines and need for more

From what has been discussed in the previous section, it is reasonable to conclude that the quality of meta-analyses of genetic association studies do need to improve, and this concerns both the traditional meta-analysis issues and the methodological problems specific to this field. Such improvement is becoming an urgent need for the scientific community, due to the speed to which the number of meta-analyses of genetic association studies has increased in the last few years. While traditional meta-analysis issues encountered in the field of genetic association studies might benefit from borrowing and adapting recommendations developed for meta-analyses of RCTs and classical observational studies, specific issues require recommendations developed ad hoc.

In 1999, a conference for the Quality of Reporting of Meta-analyses (QUOROM) of RCTs was organised following the CONSORT initiative to improve the quality of reporting of RCTs (Begg, 1996), and its proceedings summarised and published (Moher *et al.*, 1999). In fact, from 1984 about 20 papers had already been published with checklists for meta-analyses (Shea, Dube' and Moher, 2001), and this included the guidelines for the conduct and reporting of meta-analysis of RCTs developed in 1995 by an international group of 20 scientists (Cook, Sackett and Spitzer, 1995). However, a consensus across disciplines did not seem to have developed at that time. The QUOROM statement was well received by the scientific community, and it soon became evident that a similar statement was also needed for the meta-analysis of observational studies, and in 2000 the MOOSE statement for the reporting of meta-analyses of observational studies was published (Stroup *et al.*, 2000).

In 1998, an international collaboration of individuals and organisations from different backgrounds called "Human Genome Epidemiology Network" (HuGE Net) was formed, which aimed at developing and disseminating population-based human genome epidemiological information (Centre for Disease Control and Prevention, 1999). Among the future objectives of this group, one is the development of guidelines for the evidence synthesis of genetic association studies, in recognition of the lack of consensus and need for more methodological work. This was the subject of a recent HuGE Net workshop, where the different methodological issues were discussed (Cambridge, 3-4 November 2004; report of the workshop available at *http://www.cgkp.org.uk/work/activities.html#syst_rev*). The workshop highlighted how, up to now, the methodological research carried out in the field of meta-analysis of genetic association studies seems to have focused more on the appraising of genetic association studies than on other aspects of the meta-analysis protocol, such as the more specific statistical issues. In 2002, based on the work of a previous HuGE Net methodological workshop, Little and colleagues published a comprehensive checklist for the appraising of primary genetic association studies (Little *et al.*, 2002). In 2003 Colhoun and colleagues proposed a shorter checklist that aimed at helping referees deciding whether a genetic association study had addressed the important methodological issues and might be suitable for publication (Colhoun, McKeigue and Davey Smith, 2003).

A number of authors have emphasised the need for quality improvement of meta-analyses of genetic association studies and given suggestions on how to achieve it (Attia, Thakkinstian and D'Este, 2003; Munafo and Flint, 2004; Salanti, Sanderson and Higgins, 2005), but no guidelines or comprehensive recommendations have been published yet. General recommendations of these authors are to; report the inclusion and exclusion criteria adopted; assess the presence of statistical heterogeneity and investigate possible causes; assess evidence of publication bias; for other general issues, Attia and colleagues suggest that meta-analyses of genetic association studies should adhere to general guidelines developed for the meta-analysis of traditional observational studies (Stroup *et al.*, 2000). Regarding the choice of how to pool data across genotypes, Attia and colleagues discourage the use of per-allele analysis, as well as the use of genotype model-based analyses unless there is *a priori* knowledge of what the genetic model might be. As for the issue of HWE, some authors recommend exclusion of studies showing statistical significant departures from HWE (Munafo and Flint, 2004), while others suggest performing sensitivity analyses to assess the

impact of studies showing departures from HWE on the results of the meta-analysis (Attia, Thakkinstian and D'Este, 2003; Salanti, Sanderson and Higgins, 2005).

The guidelines proposed in the following sections, although covering all methodological aspects of the meta-analysis of genetic association studies, will focus on those more statistical specific issues for which consensus and recommendations are particularly lacking. For such issues, guidelines are provided at two levels of statistical complexity, with the aim of addressing the needs of a broad spectrum of investigators involved in the evidence synthesis of genetic association studies. These guidelines are based on the work presented in the previous chapters and, although far from being complete and definitive, represent an attempt to improve the quality of the meta-analysis of genetic association studies in the light of what is known so far.

6.4   Proposed guidelines

6.4.1   General issues

In order to provide recommendations on how to deal with the general methodological issues of the meta-analysis of genetic association studies, the QUOROM and MOOSE statements for the reporting of meta-analysis of RCTs and observational studies, respectively, have been reviewed and those items that seemed relevant to the field of genetic association studies have been selected. Table 6.2 summarises these recommendations and their potential relevance to the field of genetic association. The choice of what items should be relevant to the meta-analysis of genetic association studies has been made after reviewing Attia's paper (Attia, Thakkinstian and D'Este, 2003) and the proceedings of the Huge Net methodological workshop, and in the light of the checklist proposed by Little and colleagues (2002) for the appraising of primary genetic association studies. Based on these same sources and on the work presented in the previous chapters of this thesis, other items have been added to those identified as relevant in Table 6.2, and are highlighted in red in Box 6.1, where the recommendations for the general issues of the meta-analysis of genetic association studies are presented. It can be noted in Table 6.2 how general recommendations developed for meta-analysis of RCTs apply to that of genetic association studies, while recommendations for meta-analysis of observational studies regarding the problem of confounding do not seem relevant to this field. In fact, although genetic association studies are observational in

Table 6.2 – Selection of items to inform guidelines on the general meta-analysis issues encountered in the meta-analysis of genetic association studies (x = used in guidelines)

Item	Meta-analysis of RCTs *(QUOROM)*	Meta-analysis of observational studies *(MOOSE)*	Relevance to meta-analysis of genetic association studies
Literature search			
Describe the information sources in detail	X	X	Yes
Search strategy — Described restrictions	X		Yes
Search strategy — Describe search strategy in detail		X	Yes
List of citations located & justification for exclusion		X	Yes
Description of addressing articles in other languages		X	Yes
Study selection			
Definition of inclusion and exclusion criteria	X	X	Yes
Study validity assessment			
Masked condition	X	X	Yes
Quality assessment	X	X	Yes
Data extraction			
Describe methods in detail	X	X	Yes
Describe study characteristics (clin. heterogeneity)	X		Yes
Methods for pooling			
Describe statistical methods for combining results	X	X	Yes
Investigation of heterogeneity			
Assessment of statistical heterogeneity	X	X	Yes
Sensitivity and subgroup analysis			
Describe sensitivity & subgroup analyses	X	X	Yes
Rationale for *a priori* sensitivity & subgroup analyses	X		Yes
Publication bias			
Assessment of publication bias	X	X	Yes
Reporting of results			
Provide summarising flow-chart of study selection	X		Yes
Present descriptive data for each study	X	X	Yes
Present data needed to calculate effect sizes and CIs	X		Yes
Other Issues			
Assessment of confounding		X	No
Adjusted analyses for possible confounders		X	No

nature, they tend to be more similar to RCTs in that the exposure to genotype is randomly determined at conception (§ 5.2). The only potential confounding of genetic associations is thought to be population stratification, which is caused by a mixture of different ethnic groups in the study population whenever the frequency of the polymorphism and the disease risk vary between ethnic groups and the study fails to match cases and controls for ethnicity (§ 4.2.2). Apart from being an indicator of quality for primary studies, this is not an issue when combining study results in a meta-analysis where the difference in ethnicity is *across* studies (or study groups), within which both cases and controls come from the same ethnic group (§ 3.7.1).

• Literature search

As in any meta-analysis, the search should be comprehensive. Although there is no direct evidence on how well MEDLINE covers genetic association studies, research in the field of RCTs shows how MEDLINE may miss a significant number of trials if used in isolation (Suarez-Almazor *et al.*, 2000). For this reason investigators should always use more than one electronic database, and they should also make use of grey-literature, reference lists and contacts with experts of the field. In general, database searches for gene-disease association studies should take the form: *("name of gene" OR "synonyms") AND ("name of disease" OR "synonyms")*. Unlike in meta-analysis of RCTs, database searches for genetic association studies should avoid the use of a methodological search filter, since there is currently inconsistent use of study design terms in papers and indexing. Such a filter should only be considered in situations when searches based on gene and disease yield an overwhelming number of citations. The search should not be restricted to English language, since, for example, much genetics research is undertaken in China and published in Chinese. Information sources and restrictions used for their search strategy should be described in detail.

• Study selection

Inclusion and exclusion criteria should be defined *a priori*. In general, inclusion criteria should try to be comprehensive regarding aspects of study designs and potential susceptibility to bias, and sample size, but sensitivity analyses should be planned to assess the robustness of the results and conclusions of the meta-analysis. In specific cases, it may be reasonable to exclude small studies, ideally using a pre-defined cut-point, in order to make the review more manageable; however this might introduce bias. The inclusion and exclusion criteria adopted should be described in detail.

• Study quality assessment

As in any meta-analysis, assessing the quality of primary studies is an important aspect of the meta-analysis of genetic association studies. Although incorporating study quality into study weighting in meta-analysis is discouraged by most authors, there is agreement that quality assessment of primary studies should be carried out routinely, possibly using scales or checklists developed for the specific field of application. This allows excluding studies with gross deficiencies and, more importantly, performing sensitivity analyses to evaluate

whether studies of lower quality tend to provide different results from studies of higher quality (Egger, Smith and Altman, 2001). Little and colleagues provide a comprehensive checklist for the appraising of genetic association studies (Little *et al.*, 2002), which, although originally not intended for use in isolation as a quality tool, is being used in practice.

• Data extraction

Information on the way that participants were selected should be recorded, in particular the selection of cases and controls in case-control designs. Whenever possible, results should be extracted for each genotype group; results based on unadjusted analyses should be used rather than results based on analyses that adjust for potential confounders using regression techniques. However, adjusted results might also be extracted when reported, for use in a sensitivity analysis. Important information that should be recorded for each study is the genotyping method adopted and, whenever available, the rate of genotyping error detected, which in the literature has been reported to vary widely, from 1 to 30%, depending on the technique used as well as other factors (§ 4.2.2). Evidence on the presence and magnitude of genotyping error in a study is important when interpreting an observed departure from HWE, since genotyping error might not only lead to loss in precision but also bias in the estimate of the genetic effect. Although population stratification is unlikely to be an important problem in well-designed studies, it might be present and bias the results of studies with lower quality. Thus, information on measures taken to prevent problems of population stratification should be extracted when reported.

• Investigation of heterogeneity

As in any meta-analysis, attempts should be made to assess the presence of heterogeneity between study results. Although graphically the forest plot used to present the results of a meta-analysis might help detecting heterogeneity, this should be formally evaluated on statistical grounds as well. Usually authors of meta-analysis assess the presence of statistical heterogeneity exclusively through statistical testing, but methodological research suggests that evaluating the statistical significance of a test for heterogeneity should not be used in isolation. The main problem of heterogeneity tests is that they have very low power, particularly when the number of studies included in the meta-analysis is small. Thus, estimation of the magnitude of heterogeneity should also be evaluated and reported (Higgins

and Thompson, 2002; Higgins *et al.*, 2003). Moreover, investigation of the possible causes of heterogeneity should always be performed using sensitivity analysis, subgroup analysis or meta-regression. The rationale for the selection of the variables on which to base such analyses should also be described.

• **General meta-analytical methods for pooling**

General meta-analysis methods for pooling study results can be classified in two categories, fixed effects models and random effects models (see § 1.4). The fixed effect model assumes no heterogeneity, that is all studies estimate the same true underlying effect size with the estimates differing only because of random fluctuation, and the combined effect is calculated as a weighted average of all estimates. In the random effects model, studies are assumed to estimate different underlying effect sizes, and a random term is included in the model to account for the between-study heterogeneity. Compared to fixed effect models, random effects models provide wider confidence intervals and give relatively more weight to smaller studies. Although there is no consensus about whether to use fixed effect or random effects models, since statistical tests for heterogeneity tend to have very low power, some authors have suggested that random effects models might be used in all cases; both models will provide similar results if heterogeneity is truly absent. However, random effects models might best be avoided when the number of studies is very small, due to the difficulty of estimating the between-study variance (§ 1.4). It is important to note that although between-study heterogeneity can be allowed for by using random effects models, the presence of important heterogeneity should prevent investigators from combining study results at all. The methods used for pooling should be always described in such way that the analyses could be replicated.

• **Publication and reporting bias**

As in any meta-analysis, the possible presence of publication bias should be assessed graphically by drawing a funnel plot and formally by evaluating the asymmetry, for example using Egger's regression test (Egger *et al.*, 1997). The possibility of selective reporting bias should also be considered, and this could be speculated by interpreting the results in the context of how many polymorphisms were studied.

Box 6.1 – General recommendations for the conduct of the meta-analysis of genetic association studies. In red are recommendations developed ad hoc for meta-analysis in this field

---

Literature search

➤ Use more than one database, grey-literature, reference lists and contacts with experts of the field

➤ Database search: *("name of gene" OR "synonyms") AND ("name of disease" OR "synonyms")*

➤ Don't use a methodological search filter (inconsistent use of study designs terms in papers and indexing)

➤ Search beyond English language (e.g. much genetics research in Chinese)

➤ Describe in detail the information sources and restrictions used

Study selection

➤ Define *a priori* inclusion and exclusion criteria

➤ Describe in detail inclusion and exclusion criteria used

Study quality assessment

➤ Assess the quality of primary studies (checklist for the appraising of genetic association studies available)

➤ Do not weight studies by quality, since single-number summaries of quality are unreliable

➤ Perform sensitivity analyses to assess the impact of studies susceptible to bias

➤ Describe the criteria used to assess the quality

Data extraction

➤ Information on selection of participants (e.g., cases and controls)

➤ Data on genotype frequencies in cases and controls rather than logORs

➤ Information on genotyping methods used

➤ When evaluated, information on degree of genotyping error

➤ Information on blinding of measurement of genotype to disease status

➤ Information on measures taken to prevent problems of population stratification

Investigation of heterogeneity

➤ Assess the presence of statistical heterogeneity using both testing and estimation

➤ Investigate possible causes of heterogeneity using sensitivity analysis/subgroup analysis/meta-regression

➤ Describe the rationale for sensitivity & subgroup analyses

General meta-analytical methods for pooling

➤ Use always random effects meta-analysis, unless very few studies included

➤ Describe methods used

Publication bias

➤ To assess presence of publication bias draw a funnel plot and assess asymmetry, e.g. using Egger's regression test

➤ Consider the possibility of selective reporting bias

Reporting of results

➤ Provide summarising flow chart of study selection

➤ Present descriptive data for each study

➤ Present data needed to calculate effect sizes and CIs.

In particular, provide a 2x3 table with genotype frequencies for cases and controls

---

- **Reporting of results**

Investigators should provide a summarising flow-chart of study selection, with number of citations located and justification for exclusion of studies, such as that proposed by the QUOROM statement (Moher *et al.*, 1999). They should also present descriptive data for each study. In particular, they should provide the data needed to calculate effect sizes and CIs for all studies included, for instance in the form of a 2x3 table with genotype frequencies for cases and controls for each study.

6.4.2 Specific issues

The following recommendations to deal with the specific issues of meta-analysis of genetic association studies are based on the limited methodological work which is available at the moment, and to which the work presented in this thesis contributes. Given the current state of knowledge, we are still far from having reached conclusions about the approaches to be adopted in dealing with such issues, and the recommendations presented in this section should be seen very much as necessarily being subject to future developments.

For each methodological issue considered, recommendations are presented at two different levels of statistical technicality. Level 1 addresses the needs of researchers with limited statistical skills and provides recommendations of what seem to be acceptable approaches and, more importantly, recommendations of what approaches should be avoided. Level 2 represents a step further, where more complex statistical models are suggested to improve statistical efficiency, but which require the presence of a statistician in the research team. The recommendations for level 1 and 2 are summarised in Box 6.2 and 6.3, respectively.

These recommendations consider the case of a bi-allelic polymorphism, with only three possible genotype groups (gg, Gg, GG), and a dichotomous disease outcome. In this situation there are two odds ratios expressing the genetic effect, the odds ratio of GG vs. gg (OR_{GG}) and the odds ratio of Gg vs. gg (OR_{Gg}), and the relationship between them define the genetic model, λ ($\lambda = \log OR_{Gg}/\log OR_{GG}$). However, the same concepts can be applied to more complex situations.

- **Pooling across genotypes**

The methods available for pooling the data across genotypes in a meta-analysis of genetic association studies have been presented and discussed in Chapter 2, where a novel method, the genetic model-free approach, has also been proposed.

In summary, all methods available for pooling the data in a meta-analysis of genetic association studies can be classified into two categories; those which use information from all genotype groups by assuming an underlying, known or unknown, genetic model (recessive, co-dominant, dominant) that establish a relationship between the odds ratios for the different genotypes; those which do not assume a genetic model but simply analyse the genotype groups two at a time (pairwise comparisons), either separately or in a bivariate meta-analysis. Although the second type of methods require no assumptions about the genetic model and thus might be appealing in some circumstances, they tend to be less efficient than the former. Efficiency is an important aspect of meta-analysis of genetic association studies, where the genetic effect size is usually small and the power to detect a gene-association often low (see § 1.6). For this reason, the position taken in these guidelines is that less efficient methods requiring no assumption about the genetic model should be used only when the others are not considered appropriate, and in the following sections ways of deciding about the appropriateness of the different methods will be suggested.

As regards the methods included in the first category, while genetic model-based methods to be used when the underlying genetic model is known are straightforward to implement, the genetic model-free approach suggested in this thesis for the common case when the genetic model is unknown and has to be estimated from the data, is relatively complex. For this reason, the genetic model-free approach will be recommended only in level 2 guidelines. As for the methods included in the second category, that is the pairwise comparisons, in theory they should always be performed using a bivariate meta-analysis, which takes into account the correlation between the two estimated odds ratios, OR_{GG} and OR_{Gg}, induced by the common baseline group (gg). Separate pairwise analyses ignore such correlation and are thus inefficient, as the estimates of the two odds ratios cannot "borrow strength" from one another. However, a joint pairwise analysis is difficult to perform, as opposed to the simple separate pairwise comparisons, and for this reason it is recommended only in level 2 guidelines.

In order to decide which method should be used to pool the data across genotypes, there are three questions that need to be addressed (Box 6.2 and 6.3):

1 - Can we assume that all studies included share the same underlying genetic model?

When combining information on all genotype groups across studies, we cannot assume a single, known or unknown, genetic model if such a model varies from study to study. In these circumstances we can only pool the data using pairwise comparisons. In Chapter 2 ways of investigating whether the assumption of common genetic model, λ, across studies might hold have been suggested. The simpler way is a graphical evaluation, which consists on first plotting the logOR$_{Gg}$ versus logOR$_{GG}$, where the slope of the association between the two log odds ratios represents λ. If λ is common across studies then the points should all lie along the straight line with slope λ, with deviation from this only due to sampling error. To check whether departures from linearity in the graph are consistent with sampling error, in a second graph the study-specific estimates of λ and their 95% CIs should be plotted. Ninety-five percent CI for λ can be either obtained by bootstrapping or calculated by using an approximation.

2 - Do we know what the underlying genetic model is?

Whenever prior information, in the form of evidence coming from studies not included in the meta-analysis and/or expert opinion, is available on what the underlying genetic model is, genetic model-based methods should be used. In these circumstances, such methods are the most efficient since they do not include uncertainty relatively to the genetic model. However, genetic model-based methods can lead to erroneous pooled estimates with deceptively high precision when the wrong genetic model is adopted (see § 2.5.2). Thus, whenever there is no prior knowledge on the genetic model, a genetic model-free approach should be used. If the statistical complexity of implementing such approach represents a problem, then the second best choice is to perform pairwise comparisons.

3 – Can we exclude heterosis?

An additional question after choosing the genetic model-free approach is whether the possibility of heterosis can be excluded or not. Heterosis, which means that the risk of the Gg group can be higher or lower than either of the homozygous groups, although rare, has been described. While the genetic model is usually included in the spectrum between

recessive and dominant (λ=0 and λ=1, respectively), so that λ can be modelled as bounded between 0 and 1, the presence of heterosis implies values of λ lower than 0 or higher than 1. Heterosis can be excluded based either on prior knowledge from studies not included in the meta-analysis and/or expert opinion, or on the data, for instance by plotting the study-specific estimates and 95% CIs of λ and checking whether there is evidence of λ being outside the range 0 and 1. The genetic model-free approach based on bounded λ (bounded analysis) should be used if heterosis can be excluded since it tends to be more efficient compared with the unbounded analysis.

Finally, even when performing pairwise comparisons, an attempt should be made to estimate the genetic model, λ, from the estimates of the two odds ratios, OR_{GG} and OR_{Gg}. Although the estimate of λ will not be very accurate in most cases, it might be informative in large meta-analyses (see § 2.5.2). The estimate of λ should always be reported together with its 95% CI, which could be obtained using methods such as bootstrapping or normal-based approximations.

- **Borrowing information across subgroups**

1 – Is the genetic effect the same across subgroups?

In the meta-analysis of genetic association studies, a difference in the genetic effect across subgroups of studies implies an interaction between the gene effect and some characteristics of the studies defining the subgroups. Thus, in theory, statistical tests for interaction could be used to evaluate whether the genetic effects vary across subgroups. However, the power to detect an interaction might be low even in large meta-analyses so that, in practice, the decision whether to use subgroups can rarely be based on statistical grounds (Altman and Bland, 2003). Regardless of which is the variable that defines the subgroups, be it ethnicity, gender, type of disease outcome considered, or any other, investigators are usually concerned whether the genetic effect is the same across subgroups. Indeed, although there are instances when subgroups are defined *a priori* with the aim of evaluating the genetic effect in the specific sub-populations, in most cases subgroup analyses are performed with the intent of explaining the observed heterogeneity in the genetic effect across studies. This situation is the same as in the meta-analysis of RCTs and classical observational studies, and conventional ways for evaluating whether there is evidence of the effect varying across

subgroups, such as subgroup analysis and diagnostic meta-regression, can be used (Sutton *et al.*, 2000). Moreover, prior information on whether the magnitude of the genetic effect varies across defined subgroups might be available from evidence external to the meta-analysis and could help investigators decide about the appropriateness of estimating separate genetic effects. If the genetic effect is found to vary across subgroups, then investigators usually perform separate meta-analyses on the different groups (e.g. in Black, Asian and Whites; males and females; etc.), while otherwise all subgroups are combined in one meta-analysis.

2 – Can we borrow information on secondary parameters across subgroups?

In the meta-analysis of genetic association studies, there are other "secondary" parameters that might or might not vary across subgroups, independently from differences in the genetic effect. These include the parameter representing the genetic model, λ, and the parameter indicating the amount of heterogeneity in the size of the genetic effect, τ. In Chapter 3, the rationale for assuming that secondary parameters are common across studies, in order to increase the efficiency of the meta-analysis, is presented and discussed. Where there is evidence that the genetic effect may vary across subgroups, an alternative to perform separate meta-analyses for each subgroup is to jointly perform the meta-analyses by modelling secondary parameters, such as λ and τ, as common across subgroups. This approach can be particularly beneficial for the meta-analysis of small subgroups, since the increased precision in estimating secondary parameters, which is due to the borrowing of information from the meta-analyses of other subgroups, is reflected in an increased precision in estimating the genetic effect of interest. The implementation of such approach, however, is suggested only for guidelines at level 2 (Box 6.3) since it requires the use of relatively complex models. These models, which are described in detail in § 3.4, should be applied only after checking that there is no evidence against the assumption of common λ and/or τ across subgroups. Again, prior knowledge might help decide about the similarity of λ and τ across subgroups, but such knowledge is unlikely to be available in many cases. Thus, decisions on whether these assumptions might hold will often have to be based on the data in the meta-analyses, by evaluating and comparing estimates and confidence intervals of the parameter in the different subgroups. Another secondary parameter in meta-analysis of genetic association studies is the allele frequency in controls. Although the models presented in § 3.4 could easily be extended to accommodate the assumption of common allele

Box 6.2 –Recommendations for specific issues of the meta-analysis of genetic association studies: level 1

Pooling across genotypes

    1 - *Question*: Is the genetic model common across studies?

      ➤ Graphical check: a) plot, for each study, $logOR_{Gg}$ vs. $logOR_{GG}$; the slope of the association between the two is λ

               b) plot the study-specific estimates of λ with their confidence intervals

      Are departures from linear trend in a) explained by random noise in b)?

        Answer: YES: Go to question 2

                NO: use separate pairwise comparisons

    2 - *Question*: Do we know what is the genetic model operating?

      ➤ Look for prior knowledge (other studies and/or expert opinion)

        Answer: YES: use genetic model-based analysis

                NO: use separate pairwise comparisons

    3 - If performing separate pairwise comparisons: estimate the genetic model from pooled ORs (report λ with CI)

Borrowing information across subgroups

    1 - *Question*: Is the genetic effect the same across subgroups?

    ➤ Look for prior knowledge (other studies and/or expert opinion)

    ➤ Compare results for subgroups

    ➤ Diagnostic meta-regression

      Answer: YES: Combine all subgroups in one meta-analysis

            NO: Perform separate meta-analyses

Dealing with Hardy-Weinberg equilibrium

    ➤ Evaluate departures from HWE in the studies included by:

        Testing for HWE. Use an exact test whenever in the presence of sparse data

        Estimating the magnitude of the departure. Report its CI

    ➤ Re-evaluate the quality of those studies showing departures from HWE

    ➤ Perform sensitivity analyses to exclude studies with departures from HWE only if evidence of specific problems

---

frequency across subgroups, in fact this parameter is very likely to vary between studies with different study populations.

    • **Dealing with Hardy-Weinberg equilibrium**

Since departures from HWE have been shown to be associated with methodological problems of genetic association studies, in particular genotyping error, population stratification and selection bias, this issue should always be addressed when performing a meta-analysis. Whenever raw data on genotype frequencies are reported, evidence on departure from HWE should be evaluated, based on both hypothesis testing and estimation, for all studies included in the meta-analysis. Testing for HWE should be performed using an exact test whenever in the presence of sparse data. For the estimation of the magnitude of

---

Box 6.3 – Recommendations for specific issues of the meta-analysis of genetic association studies: level 2

Pooling across genotypes

1 - *Question*: Is the genetic model common across studies?

➤ Graphical check: a) plot, for each study, $logOR_{Gg}$ vs. $logOR_{GG}$; the slope of the association between the two is λ

b) plot the study-specific estimates of λ with their confidence intervals

Are departures from linear trend in a) explained by random noise in b)?

➤ Statistical check: Is the model fit of model-free analysis equal/better than bivariate pairwise analysis?

Answer: YES: Go to question 2

NO: use bivariate pairwise analysis

2 - *Question*: Do we know what is the genetic model operating?

➤ Look for prior knowledge

Answer: YES: use genetic model-based analysis

NO: use genetic model-free analysis. Go to question 3

3 - *Question*: Can we exclude heterosis?

➤ Look for prior knowledge

➤ Plot the study-specific estimates of λ with their confidence intervals

Answer: YES: use bounded analysis ($0 \leq \lambda \leq 1$)

NO: use unbounded analysis

4 - If performing bivariate pairwise analysis: estimate the genetic model from pooled ORs (report λ with CI)

Borrowing information across subgroups

1 - *Question*: Is the genetic effect the same across subgroups?

➤ Look for prior knowledge

➤ Compare results for subgroups

➤ Diagnostic meta-regression

Answer: YES: Combine all subgroups in one meta-analysis

NO: Go to question 2

2 - *Question*: Can we borrow information across subgroups on secondary parameters?

➤ Look for prior knowledge

➤ Evaluate and compare estimates and CIs of the parameter(s) in the different subgroups

Answer: YES: Assume common parameter(s) across meta-analyses of different subgroups

NO: Perform separate meta-analyses

Dealing with Hardy-Weinberg equilibrium

As in level 1 - Box 6.2

---

the departure from HWE any of the available measures, such as the inbreeding coefficient, the disequilibrium parameter or the alpha parameter, can be used (§ 4.3.2), and such estimate should be reported together with its confidence interval. Once studies with departures from HWE have been identified, their quality should be carefully re-evaluated in the light of this finding, which represents a red-light signal for the possible presence of methodological problems requiring further investigation. Sensitivity analyses that exclude

studies with departures from HWE, and/or exclusion of these studies from the main analysis, should only be performed if there is any indication of specific methodological problems which may have caused the departure. In fact, the work presented in this thesis suggests that the exclusion of studies based on evidence of departure from HWE *per se* is not justified.

6.5   Extending the guidelines to Mendelian randomisation

The rationale for Mendelian randomisation, with its potentials and limitations, has been presented and discussed in Chapter 5, where an integrated approach to the meta-analysis of genetic association studies when using Mendelian randomisation has been proposed.

In summary, genetic association studies might be used to provide unconfounded estimates of the association between a risk factor (phenotype) and a disease using Mendelian randomisation, where the role of genetic information is that of an "instrument". Evidence from Genotype-Disease (G-D) association studies is combined with evidence from Genotype-Phenotype (G-P) association studies, and such genetic information is used to infer the Phenotype-Disease (P-D) association rather than being of interest in itself. Although in theory the concept of Mendelian randomisation could be applied within a single study measuring both G-D and G-P associations, in practice this is not possible since the derived P-D estimate would have extremely low precision even when the study is large. Thus, evidence synthesis of a number of studies evaluating G-D association on one side and G-P association on the other is needed. However, the validity and accuracy of the derived estimate of the P-D association relies not only on validity and accuracy of the evidence synthesis performed on both sides of the genetic information, but also on the appropriateness of the method used to combine the two sources of genetic data. While the recommendations presented in the previous sections address the first methodological aspect, the second involves specific issues based on the same principles regulating the use of instrumental variables, of which Mendelian randomisation represents an application, and are summarised below. These recommendations refer to the situation where only the differences in phenotype and disease risk between the two extreme genotypes (GG and gg) are considered for deriving the estimate of P-D association.

- **Data extraction**

For data extraction on G-D association, all recommendations presented in § 6.4.1 apply. As for G-P association, in addition to extracting the estimate of the mean phenotype difference between genotypes and its precision, information on whether the data on the phenotype difference come from cases, from controls, or from a mixture of both, should also be extracted. The reason is that data on G-P collected from cases may be less reliable than data from controls if the disease itself affects the phenotype level (reverse causation) in a way that is not linear (see § 5.6.2 and 5.7.2). Moreover, information of whether both associations are measured in the same study should also be recorded, for the reasons explained in the sections below.

- **Checking the assumption on which Mendelian randomisation is based**

The assumptions underlying the triangulation of G-D, G-P and P-D associations, on which the use of Mendelian randomisation is based, need to be checked. If such assumptions do not hold, serious biases may arise when applying the methods described in Chapter 5 to derive an estimate of P-D association.

When synthesising the evidence on genetic information, a mixture of studies evaluating G-D association, G-P association, or both is likely to be encountered. The presence of studies measuring both associations provides an important opportunity for evaluating whether the assumptions underlying Mendelian randomisation might hold. This can be assessed using graphs which, although not capable to show minor deviations, have the advantage of being simple and thus should always be drawn and evaluated. These graphs, which are described in detail in Chapter 5, are the more informative the larger is the number of studies measuring both G-D and G-P associations.

First, one should draw a forest plot with two columns, one for G-D and the other for G-P, where paired estimates from within the same study are aligned in the same row. The forest plot will thus be organised in three blocks, two blocks with studies measuring either G-D or G-P, and the third block with studies measuring both associations. Estimates of G-D from studies measuring only G-D association should be compared with those from studies measuring also G-P association, and the same should be done for estimates of G-P. If it appears that G-D and G-P associations are consistent across all studies, then all G-D

estimates and all G-P estimates can be pooled before combining these overall estimates to derive a figure for the P-D association. A possible reason for which, for example, G-P estimates from studies measuring only G-P could be different from studies measuring also G-D is publication bias, whereby results from studies evaluating G-P alone would be reported only if the effect size detected is large. The issue of publication bias should be further explored by drawing funnel plots for G-D and G-P, where asymmetry in the plots suggests presence of publication bias. If a different symbol on the funnel plots is used to distinguish studies that reported data on only one of the two associations from studies reporting on both, such funnel plots could also help detect reporting bias. Reporting bias means that both the odds ratio and the mean phenotype difference are measured but only one is reported, possibly because the other contradicts the anticipated relation. This situation may be suggested by a discrepancy in the shape of the funnel plot between studies reporting on both associations and those reporting on only one association.

The most crucial assumption underlying Mendelian randomisation is that the genotype influences the disease risk only through modification of the specific phenotype, and thus investigators should limit the use of Mendelian randomisation to the case where there is good biological knowledge of the genotype-phenotype-disease pathway. Whether this assumption might hold should always be checked based on the evidence from those studies that measured both G-D and G-P associations. When the phenotype does indeed lie on the causal pathway between gene and disease, studies carried out in populations with a large difference in phenotype with genotype (G-P) might be expected to show a large G-D odds ratio. This can be investigated by plotting the findings from each study on a graph of G-D log odds ratio against G-P difference, which is expected to show a monotonic trend if the phenotype is intermediate on the causal pathway, with the line passing through the origin. Lack of any correlation would cast doubt on whether the phenotype is truly intermediate, while a line not passing trough the origin might indicate that there is another phenotype through which the gene affects the disease risk, or that the gene is in linkage disequilibrium with a gene which also affects the disease risk, or that there is differential publication bias for the two associations. This graph might also show gross departures from linearity of the relation between G-D and G-P, in which case Mendelian randomisation can still be used but the methods described in § 5.5 do not apply, since they are based on the assumption of approximate linearity.

- **Deriving the estimate for the phenotype-disease association**

Are ALL studies measuring both G-D and G-P associations?

In the situation where all available studies measured both G-D and G-P associations, the correlation in the sizes of pairs of estimates from the same study affects the size of the confidence interval for the derived P-D estimate and the result of the hypothesis test. A simple way to account for such correlation is to combine the G-D and G-P estimates separately within each study to obtain study-specific estimates of P-D. These P-D estimates should be graphed in their own forest plot and pooled to obtain an overall estimate, using standard meta-analysis methods.

Are there ANY studies measuring both G-D and G-P associations?

In the more realistic situation in which some studies measured both G-D and G-P associations and some measured one or the other, bivariate meta-analysis models should be used to account for the correlation induced by those studies measuring both associations. Since these models, which are described in detail in § 5.5, are rather complex, they are suggested only for guidelines at level 2. The impact of ignoring the correlation induced by studies that measured both associations depends on the proportion of such studies.

If the evidence on G-D and G-P came from two separate sources, then independent meta-analyses on G-D and G-P can be performed and the P-D estimate calculated from the two pooled estimates. The confidence interval for P-D must be calculated by taking into account the uncertainty of both G-D and G-P estimates. This can be done using the following formula, which is based on the formula developed by Kendall and Stuart for the confidence interval of the ratio of two normal variates. Here the original formula is adapted for large samples in which the two estimates used in the ratio are independent, that is come from separate meta-analyses (Thompson, Tobin, and Minelli, 2003). Denoting the log odds ratio for G-D with ξ and the mean phenotype difference by δ, the $(100-\alpha)\%$ confidence interval

for the ratio has limits:
$$\frac{\hat{\xi}}{\hat{\delta}} \frac{1 \pm z_{\alpha/2} \sqrt{\left[\dfrac{s_{\delta}^2}{\hat{\delta}^2} + \dfrac{s_{\xi}^2}{\hat{\xi}^2}\right] - z_{\alpha/2}^2 \dfrac{s_{\delta}^2}{\hat{\delta}^2} \dfrac{s_{\xi}^2}{\hat{\xi}^2}}}{\left[1 - z_{\alpha/2}^2 \dfrac{s_{\delta}^2}{\hat{\delta}^2}\right]}$$

where $z_{\alpha/2}$ is the $(100-\alpha/2)\%$ value from a standard normal distribution, for instance 1.96 for a 95% confidence interval, and s denotes the standard error of the corresponding variable.

Box 6.4 – Specific recommendations for the conduct of the meta-analysis of genetic association studies when using Mendelian randomisation: level 1

Data extraction

    1 - Information on whether G-D and G-P associations are measured in the same study

    2 - Information on whether G-P data are collected from cases, from controls, or from a mixture of both

    3 - See Data extraction in Box 6.1

Checking the assumption on which Mendelian randomisation is based

    1 - If there are studies measuring both associations: draw a two-column forest plot with G-D and G-P data

    2 - For studies measuring both associations: plot logOR for G-D against mean difference for G-P for each study

    3 - Draw funnel plots for G-D and G-P, with different symbols for studies measuring only one or both associations

Deriving the estimate for the phenotype-disease association

    Question: Are *all* studies measuring both G-D and G-P associations?

    Answer: YES: 1 – Combine G-D and G-P estimates to derive a P-D estimate *within* each study

                2 – Draw forest plot of study specific P-D estimates; pool them using standard meta-analysis methods

       NO: 1 - Perform separate meta-analyses on G-D and G-P; derive P-D from the two pooled estimates

                2 - Calculate the CI for P-D estimate based on uncertainty of both G-D and G-P estimates*

* See § 6.5 "Deriving the estimate for the phenotype-disease association" for formula to calculate CI for P-D estimate.

Box 6.5 – Specific recommendations for the conduct of the meta-analysis of genetic association studies when using Mendelian randomisation: level 2

Data extraction

    1 - Information on whether G-D and G-P associations are measured in the same study

    2 - Information on whether G-P data are collected from cases, from controls, or from a mixture of both

    3 - See Data extraction in Box 6.1

Checking the assumption on which Mendelian randomisation is based

    1 - If there are studies measuring both associations: draw a two-column forest plot with G-D and G-P data

    2 - For studies measuring both associations: plot logOR for G-D against mean difference for G-P for each study

    3 - Draw funnel plots for G-D and G-P, with different symbols for studies measuring only one or both associations

Deriving the estimate for the phenotype-disease association

    1 - *Question*: Are *all* studies measuring both G-D and G-P associations?

    Answer: YES: 1 – Combine G-D and G-P estimates to derive a P-D estimate *within* each study

                 2 – Draw forest plot of study specific P-D estimates; pool them using standard meta-analysis methods

       NO: Go to question 2

    2 - *Question*: Are there *any* studies measuring both G-D and G-P associations?

    Answer: YES: Use bivariate meta-analysis model and calculate P-D within the model

            NO: 1 - Perform separate meta-analyses on G-D and G-P; derive P-D from the two pooled estimates

                2 - Calculate the CI for P-D estimate based on uncertainty of both G-D and G-P estimates*

* See § 6.5 "Deriving the estimate for the phenotype-disease association" for formula to calculate CI for P-D estimate.

6.6   Discussion

The increasingly important role that evidence synthesis has acquired in medical decision-making (Harbour and Miller, 2001; Atkins *et al.*, 2004) and the recognition of methodological weaknesses of published meta-analyses in all fields of medical research (§ 6.2) have led to a number of initiatives aimed at developing guidelines that could assist researchers in producing high-quality evidence synthesis. Important steps in this direction have been the publishing of the QUOROM statement for the reporting of meta-analyses of RCTs in 1999, followed by the MOOSE statement for the reporting of meta-analyses of observational studies in 2000. Following these initiatives and in recognition of the fact that evidence synthesis in specific areas might require more specific recommendations, a number of groups have developed "specialised" guidelines such as the STARD (Standards for Reporting of Diagnostic Accuracy) statement for diagnostic studies (Bossuyt *et al.*, 2003a; Bossuyt *et al.*, 2003b) and the TREND (Transparent Reporting of Evaluations with Non-randomized Designs) statement for behavioural and public health intervention studies (Des Jarlais *et al.*, 2004). These initiatives are now recognised as high research priorities, as demonstrated by the publication in June 2005 of a whole issue of the *Annals of Internal Medicine*, one of the leading journals in medicine, entitled "Challenges of Summarizing Better Information for Better Health: The Evidence-based Practice Center Experience" (*Annals of Internal Medicine*, 21 June 2005; volume 142, number 12, part 2). This issue includes a series of 9 articles targeting different areas of medical research, with the aim of aiding researchers in preparing high-quality systematic reviews and meta-analyses.

The Cochrane Collaboration, with its primary aim of generating and disseminating high-quality systematic reviews of health care interventions, has significantly contributed to the development of comprehensive and practical guidelines for good-standard meta-analysis in the field of RCTs (Bero and Rennie, 1995). In the field of genetic association studies, the HuGE Net group has recently started working in the same direction as the Cochrane Collaboration, by organising and promoting international methodology workshops on the issues of meta-analysis of genetic association studies, with the aim of reaching a consensus and developing standards for conduct and reporting of such meta-analyses. However, there is still much methodological work that needs to be done before a consensus can be reached.

The number of published meta-analyses of genetic association studies has increased at an incredible rate in this last decade. While Attia and colleagues could identify 37 papers published between 1991 and 2000 (Attia, Thakkinstian and D'Este, 2003), when I accessed the HuGE website archive on the 15th of March 2005 there were 243 papers published from 2000 (§ 4.4.1). Given the number of meta-analyses of genetic association studies being produced and the specific methodological issues that characterise this field, it might seem surprising that no guidelines for such an important area of medical research have yet been published. The lack of recommendations for good quality evidence synthesis of genetic association studies reflects the lack of consensus and indeed the paucity of methodological work carried out to investigate the methodological issues of this field.

The guidelines proposed in this chapter have been developed with the aim of assisting the conduct of meta-analyses of genetic association studies rather than suggesting the reporting of such meta-analyses. Although often used interchangeably, as it appears to be the case for the QUOROM statement, these two types of guidelines are intrinsically different. It is true that guidelines for reporting of meta-analysis, although structured in a different way from guidelines for conducting a meta-analysis, tend to serve the same purpose in that they indicate the methodological issues which need to be reported, and thus necessarily addressed, in high-quality evidence synthesis. However, guidelines for reporting might not necessarily provide recommendations on which method should be adopted for dealing with a particular issue when more options are available. This seems to be particularly relevant in the field of meta-analysis of genetic association studies where the approaches adopted vary substantially and little attempt has been done to compare the different methods and provide advice on how to choose between them.

In this chapter the recommendations have also been extended to address the case of meta-analysis of genetic association studies which use Mendelian randomisation, in recognition of the importance of this novel application of genetic epidemiology as a potentially valuable tool for deriving unconfounded estimates of the effect of a risk factor (phenotype) on a disease risk. The work presented in Chapter 5 shows how the near future of Mendelian randomisation relies on the possibility of synthesising all evidence available, since no individual study, unless extremely large, will have sufficient data to provide a precise estimate of the phenotype-disease association of interest. The meta-analysis of genetic association studies that uses Mendelian randomisation has specific methodological issues that need to be considered, and which have been addressed in this chapter. Moreover, the

meta-analytical approach potentially provides a valuable tool for assessing some basic assumptions on which the use of Mendelian randomisation is based, and simple ways of investigating such assumptions have also been presented.

6.6.1 Further work

The recommendations proposed here for the meta-analysis of genetic association studies consider only the case of a bi-allelic polymorphism with three possible genotypes, gg, Gg, GG, and a dichotomous disease outcome. However, the same concepts hold for the more complex situations where the polymorphism involves more than two alleles, where the relationships between the odds ratios for the different genotypes that define the mode of inheritance become more difficult to specify or evaluate. Moreover, apart from studies evaluating the association between a gene and a disease, there are studies which evaluate the association between the gene and an intermediate phenotype usually measured as a quantitative outcome variable, as discussed for Mendelian randomisation. Although adapting the framework of the genetic model-free approach presented in Chapter 2 to accommodate the case of a continuous outcome should be relatively straightforward, this route has not been taken yet and it represents one of the issues which will follow the work summarised in this thesis. In the meanwhile, it would appear sensible to suggest that a bivariate meta-analysis should be carried out whenever there is no prior information about the genetic model. As regards the recommendations proposed for integrating the evidence on genotype-disease and genotype-phenotype associations when using Mendelian randomisation, the methods proposed in this thesis consider the simplest scenario of a single gene, a single phenotype and a single disease in the triangulation. This might not be realistic in some situations where more genes may influence the same phenotype, or one gene may influence more phenotypes acting on the disease of interest. Such situation might require much more complicated models, and again represent a value area of future research if Mendelian randomisation has to be used for the study of diseases with complex causal pathways.

Finally, it is important to remember that reaching a consensus and developing guidelines, although crucially important, will not automatically lead to improved methodological quality of published meta-analyses of genetic association studies. A number of authors have shown how poor quality has affected meta-analyses of RCTs through the years (Table 6.1) although guidelines have been published starting from 1984 (Shea, Dube' and Moher, 2001). It is

likely that other ways can help promoting good standards. Journal editors, for instance, can influence the way current research is published by imposing stricter methodological control during the review process, and they could be beneficially involved in the battle to improve the quality of research in this field. Moreover, it has to be considered that hardly any positive feedback will follow the dissemination of any recommendation for sound methodology if the methods proposed are out of reach for most end users. In the field of meta-analysis of RCTs, the Cochrane Collaboration has contributed to improve the general quality of published meta-analyses not only by disseminating guidelines, but also by releasing software user-friendly (Revman software, available at *http://www.cc-ims.net/RevMan*) to help researchers with limited statistical skills implement standard meta-analytical models. Indeed, the choice of providing two different levels of recommendations in this chapter has been made in recognition of the importance of reaching a wide range of researchers with different statistical skills. While an increase in the statistical complexity of the model is often associated with an increase in the efficiency of the method, still in practice complex models are likely to be used by a minority of the investigators carrying out meta-analyses of genetic association studies. Ideally, sophisticated models could be made accessible to all researchers by developing appropriate software by which these can be implemented using a user-friendly interface. However, it might be argued that user-friendly software might encourage researchers to use methods which they do not understand, unless graphical and formal investigations of the underlying assumptions for each method are built into the program and explained in such a way that they guarantee its appropriate use. This represents a stimulating idea for further work.

6.6.2 Conclusions

The guidelines proposed in this chapter have been developed to advise researchers considering a meta-analysis of genetic association studies as to which methods should be used for the collection, synthesis, analysis and interpretation of the evidence. Particular emphasis has been given to the statistical methods which address the specific meta-analysis issues of this field, where there seems to be complete lack of consistency, justification and transparency in the published literature.

Possibly the most important message for researchers carrying out a meta-analysis of genetic association studies is the importance of understanding which are the assumptions behind

each method available to synthesise the data, since it is only based on such knowledge that appropriate choices can be made. An attempt should always be made to investigate whether such assumptions might hold, and judgement should ideally be based on both prior knowledge and evidence from the data. Graphical and statistical ways of investigating whether these assumptions might hold have been presented. Unfortunately, the data may be not sufficient to provide evidence for or against such assumptions, and prior knowledge thus become valuable. In the absence of any prior knowledge, still investigators have to make a rational choice and have the responsibility to write reports that are as transparent as possible, where the choice is discussed and justified, so that readers can critically evaluate their work.

7 DISCUSSION & CONCLUSIONS

7.1 Summary

Meta-analysis is a recognised powerful tool for evidence-based decision making, since it can critically appraise, summarise and attempt to reconcile the results from a number of studies. Many institutional bodies which are involved in the development of evidence-based clinical guidelines consider meta-analysis as the strongest source of evidence, and indeed its role in medical decision-making is reflected by the enormous increase of this type of study in medical research over the last 15 years (Harbour and Miller, 2001; Atkins *et al.*, 2004). This is particularly true in the field of genetic epidemiology, where the recent completion of the sequencing of the human genome, which is considered an important milestone in the history of medicine, has led many scientists to believe in an imminent integration of gene discoveries into medical practice. Some scientists have forecast that by the year 2010 genetic tests will be available for prediction of as many as a dozen common conditions, with individuals being able to learn their individual susceptibilities and take steps to reduce those risks (Collins and McKusick, 2001). Moreover, in a premature way and with worrying rapidity, commercial marketing of genetic testing for disease prevention has already been implemented, both in Europe and the United States (Khoury, Little and Burke, 2004).

With the increased number of published meta-analyses in all fields of medical research, evidence has been reported on how meta-analyses carried out on the same topic can produce discordant results (Jadad, Cook and Browman, 1997), and also produce discordant results compared to subsequent large RCTs (LeLorier *et al.*, 1997). As LeLorier and colleagues point out, the appealing idea of a meta-analysis as a tool for "summarizing all the information contained in a set of trials into a single odds ratio may greatly oversimplify an extremely complex issue", and lead to inappropriate conclusions. Genetic association studies have shown a worrying failure to replicate and validate postulated associations, to the point that some scientists have raised concerns on the actual value of human genome discoveries to health care (Khoury, Little and Burke, 2004). This situation, while providing a very strong case for evidence synthesis, makes the task of combining study results particularly difficult. Unfortunately, as discussed in Chapter 6, relatively little methodological work has been carried out in this field and, as a

consequence, a lack of consensus and standardisation of the methods to be adopted is apparent in the literature. Particularly neglected are the specific statistical issues of the meta-analysis of genetic association studies, which has been the motivating reason for the work presented in this thesis. These include the choice of how to pool the data across genotypes based on assumptions about the underlying genetic model, the appropriate use of subgroup analysis, and the issue of how to deal with departures from Hardy-Weinberg equilibrium of primary studies.

One recurring issue encountered when reviewing the way investigators deal with these aspects of meta-analysis of genetic association studies is the lack of a coherent and transparent approach to the choice of a specific method among those available. In particular, a general problem is the lack of adequate consideration, and sometimes understanding, of what are the assumptions on which each method relies. This seems to lead to two opposite behaviours; one is to use the method which is most likely to provide narrower confidence intervals, or even statistical significance, without any consideration of the underlying assumptions; the other is to "play safe" and avoid making any assumptions at all. Indeed, avoiding assumptions tends to lead to a loss of power in the analysis, and often turns out to be an illusion anyway. An example of this is the way investigators deal with the issue of pooling data across genotype groups. Many investigators combine the data by assuming a specific genetic model, common to all studies included, even when there is no *a priori* knowledge of what such a model might be. However, the examples presented in Chapter 2 show how assuming an inappropriate model can lead to biased estimates of the genetic effects, often accompanied by deceptively high precision. On the other hand, some investigators choose to avoid the need to make *any* assumption about the underlying genetic model by performing independent pairwise comparisons, where the two, or more, odds ratios are estimated in separate meta-analyses. This approach tends to be inefficient since it ignores the correlation between the odds ratios, and thus the possibility for one of the odds ratios to provide additional information on the other. Moreover, avoiding assumptions on the genetic model indeed implies assuming that the two odds ratios are independent from one another. The philosophical approach adopted in this thesis is closer to the usual way of thinking of the statistician, who accepts the idea of assumptions as part of the interpretation of the evidence, and who is used to assessing which assumptions might hold by applying different models and choosing the simplest that can explain the data. In

choosing the approach to combine data across genotypes, for instance, each assumption is critically evaluated. While the assumption of a specific genetic model in the absence of prior knowledge is considered unjustifiable and discarded, the assumption that the same unknown genetic model is shared by all studies in the meta-analysis is viewed as reasonable. However, a graphical assessment of whether there is any evidence in the data against this assumption is performed to support theoretical considerations. In general, the assumptions underlying the method adopted should be always discussed and an effort should be made to evaluate whether they might hold, based on evidence from the data and external information. Whenever in doubt, sensitivity analyses should be performed to evaluate the impact of different assumptions on the result of the meta-analysis.

The same philosophical approach is adopted when addressing the issue of subgroup analysis. Here the investigator might have reasons to believe that the assumption that the parameter of interest, i.e. the genetic effect, is common across subgroups is unjustified. However, the assumptions that other secondary parameters in the meta-analysis model are shared by all subgroups might well hold. The review of published meta-analysis presented in Chapter 3 suggests that investigators deal with subgroups either by combining them in a single meta-analysis, thus assuming that they share *all* parameters in the model, or by performing separate meta-analyses, thus assuming that they do not share *any* parameter. The work presented in this thesis suggests that assuming that subgroups share secondary parameters, such as the genetic model and between-study heterogeneity of the genetic effect, might be reasonable and might provide a potential gain in the precision of the genetic effect. Again, sensitivity analyses should be performed whenever in doubt.

The issue of HWE in the meta-analysis of genetic association studies, where departures from HWE are considered a proxy for poor quality of primary studies, highlights an interesting example of how adopting a conservative strategy might not necessarily represent the best strategy. Although there is agreement on the fact that departure from HWE should be evaluated, there is little evidence supporting the current recommendation to exclude studies not in HWE. In fact, there is no straightforward way to define what is meant by departure from HWE and how it can be best identified, with possible options being; the use of a statistical test, estimation of the magnitude of departures, or a combination of both. Indeed, the only option which is currently used is the identification

of departure based on the result of a test. The risk of a conservative strategy of excluding studies which show statistically significant departure is that large studies are those more likely to be excluded, although they are not necessarily those showing the larger departures. In fact, the review of more than 500 genetic association studies presented in Chapter 4 suggests an inverse correlation between sample size of the study and magnitude of the observed departure from HWE. The work presented in this thesis suggests that the presence of a departure from HWE does not, *per se*, justify exclusion of a study from the meta-analysis, but it should be considered as a warning signal for the possible presence of methodological problems in the study. The decision whether to exclude the study from the main analysis or in sensitivity analyses should rather be based on the detection of a specific problem and consideration of its possible impact on the results of the meta-analysis.

The meta-analytical approach to Mendelian randomisation presented in Chapter 5 illustrates how attention to the underlying assumptions of a method can greatly help assessing its validity. Mendelian randomisation is a very promising application of genetic association studies, whereby genetic data are used to derive an estimate of the association between a risk factor (phenotype) and a disease which is free of the confounding and reverse causation typical of classical epidemiology. However, the validity of its use strongly depends on the fulfilment of the assumptions on which this method is based (Davey Smith and Ebrahim, 2003). An integrated meta-analytical approach can not only allow for the phenotype-disease association to be estimated with sufficient precision, but also provides a valuable tool for assessing whether the crucial assumptions behind the triangulation genotype-phenotype-disease might hold.

In order to put the work of this thesis into context, the findings presented and methods developed have been integrated with the evidence already available in the literature to provide an overview of all issues that require consideration in the meta-analysis of genetic association studies. This has led to the development of a set of recommendations for the conduct of meta-analysis in this field. In recognition of the complexity of some of the methods proposed, an effort has been made to identify simpler, but yet methodologically correct, approaches, and in Chapter 6 guidelines are presented at two different levels of statistical sophistication. Indeed, while an increase in the statistical complexity of the meta-analysis model is often associated with an increase in its

efficiency, in practice complex models are likely to be used only by a minority of the investigators carrying out meta-analyses of genetic association studies. The key message is to not use *incorrect* methods. The guidelines proposed, although far from being definitive, might assist either a statistician new to the field as to what are the specific challenges of the meta-analysis of genetic association studies, or a researcher with experience in the field but limited statistical skills, as to what assumptions should be evaluated before choosing to apply a specific method.

7.2 Discussion and further work

For all issues investigated in this thesis, specific aspects have been identified which require further consideration and work. In the choice of how to combine data across genotype groups, the proposed genetic model-free approach overcomes a number of limitations of currently used methods, but the problem of how to assess the assumption of constancy of the genetic model across studies has not been completely resolved. The proposed graphical ways of evaluating whether the data suggest that the differences in the estimates of the genetic model within individual studies might be explained by sampling error, are limited by the fact that such estimates are often very imprecise. When addressing the issue of how to handle subgroups, no satisfactory statistical measure could be found to compare models based on different assumptions on the similarity of secondary parameters across subgroups within the Bayesian framework. Both these issues might greatly benefit from the use of a "full" Bayesian approach, where prior knowledge based on external data or experts' opinion is formally incorporated in the analysis (Spiegelhalter, Abrams, and Myles, 2004). This interesting, but yet problematic aspect, which represents the major advantage of the Bayesian approach compared to the classical approach, has not been investigated in this thesis, although its potentials have been highlighted. In chapters 2 and 3, the possibility of including prior knowledge not necessarily on the parameter of interest but also on secondary parameters included in the model, has been discussed. The incorporation of prior information on *any* of the parameters in the meta-analysis model increases, to some extent, the efficiency of the approach in terms of precision of the estimate of interest.

In addressing the issue of HWE, the work presented in Chapter 4 has concentrated on how to assess departures from HWE in individual studies and how to handle studies showing departures. However, there are a number of other related questions which require further investigation. One is whether it is possible to develop an overall assessment of the presence and magnitude of departures from HWE in the whole meta-analysis, rather than in individual studies. Another important question is whether, once having assessed that there is no evidence of departure, HWE can indeed be assumed in the model in order to increase the efficiency in estimating the pooled genetic effect.

Given the important role of meta-analysis in the use of Mendelian randomisation, this is another field where further research needs to be carried out in order to extend the use of such a promising approach to a range of complex but more realistic situations. The possibility of modelling genetic and environmental factors which interfere with the triangulation gene-phenotype-disease would greatly increase its applicability. Moreover, the statistical efficiency of the approach would benefit from the use of information from all genotype groups for both the genotype-phenotype and genotype-disease associations. Although the genetic model-free approach presented in Chapter 2 could be used, further work is needed to extend the method to the case of a continuous outcome, such as the phenotype level.

In general, in order to improve the standard of published meta-analysis, there are two separate aspects that need consideration; the first is whether there is consensus on what are the most appropriate methods to deal with each methodological issue, which can lead to a standardised approach to evidence synthesis; the second is how these methods can be disseminated in such a way that they can be implemented by all researchers carrying out meta-analysis. Chapter 6 addresses the first question and provides guidelines to assist investigators considering a meta-analysis of genetic association studies. These are based on current knowledge, to which this thesis contributes and which is still far from being conclusive. More methodological work is needed to extend these recommendations not only to cover a wider range of situations that might be encountered, such as multiple-allele polymorphisms and continuous disease outcomes, but also to improve the efficiency of the methods described, such as use of information from all genotypes for Mendelian randomisation. The question of how to encourage the implementation of methodological recommendations represents an important further step in the research

agenda. In this thesis the importance of considering the feasibility of application of the methods proposed has been recognised and has led to the development of guidelines at two different levels of statistical complexity. However, the ideal way to address this issue might be to develop user-friendly software packages, which can allow the implementation of complex models by investigators with limited statistical skills. The counter argument for this is that oversimplification of sophisticated methods might lead to their inappropriate use, due to lack of understanding of what their underlying assumptions are.

7.3   Conclusions

Empirical evidence on the quality of published meta-analyses of genetic association studies suggests that the standard is relatively low, and the methodology needs to be improved in virtually all respects. This thesis highlights what are the main methodological pitfalls specific to this field and proposes new approaches specifically developed for dealing with them. The findings presented are combined with evidence available in the literature to provide guidelines on how to deal with all issues which require consideration in the meta-analysis of genetic association studies. In order to be of potential use to investigators with varying statistical skills, such guidelines are developed at two levels of sophistication, one relatively simple although methodologically correct, the other more sophisticated but more efficient. The key message is, do not use *incorrect* methods, and perhaps the best way to achieve this is to understand that each method adopted is based on assumptions, which need to be carefully considered. These assumptions should be always made explicit so that the reader can critically evaluate them, and an effort should be made to assess whether these assumptions hold, based on prior knowledge and evidence from the data. As has been suggested for meta-analysis generally (Sutton *et al.*, 2000), whenever in doubt sensitivity analyses should be performed to evaluate the impact of different assumptions on the results of the meta-analysis.

<div style="border: 1px solid black; padding: 10px;">

APPENDIX 1

</div>

CHAPTER 2

The choice of genetic model

Model-free approach

a) WinBUGS code for the models based on prospective likelihood

Model

```
{

        for( i in 1:n) {

    r_gg[i] ~ dbin(p_gg[i], n_gg[i])
    r_Gg[i] ~ dbin(p_Gg[i], n_Gg[i])
    r_GG[i] ~ dbin(p_GG[i], n_GG[i])
    logit(p_gg[i]) <- mu[i] - delta[i]/2
    logit(p_GG[i]) <- mu[i] + delta[i]/2
    logit(p_Gg[i]) <- mu[i] + lambda*delta[i] - delta[i]/2
    mu[i] ~ dnorm(0.0,0.0001)
    delta[i] ~ dnorm(d, prec)


                        }

        d ~ dnorm(0.0,0.0001)

# Bounded analysis
        lambda ~ dbeta(0.5,0.5)

# Unbounded analysis
        lambda ~ dnorm(0.5,0.1)

        prec <- 1/var
        var <- pow(sd,2)
        sd ~ dunif(0,2)
        OR_GG <- exp(d)
        OR_Gg <- exp(d*lambda)

}
```

b) WinBUGS code for the models based on retrospective likelihood

Model

```
{
        for( i in 1:n) {

case[i,1:3] ~ dmulti(p.case[i,], N.case[i])
control[i,1:3] ~ dmulti(p.ctrl[i,], N.ctrl[i])
sum[i] <- p.ctrl[i,1] + exp(lambda*delta[i])*p.ctrl[i,2]+exp(delta[i])*p.ctrl[i,3]
p.case[i,1] <- p.ctrl[i,1]/sum[i]
p.case[i,2] <- exp(lambda*delta[i])*p.ctrl[i,2]/sum[i]
p.case[i,3] <- exp(delta[i])*p.ctrl[i,3]/sum[i]
delta[i] ~ dnorm(d, prec)
p.ctrl[i,1] <- exp(a[i])/(1 + exp(a[i])) * exp(b[i])/(1 + exp(b[i]))
p.ctrl[i,2] <- exp(a[i])/(1 + exp(a[i])) - p.ctrl[i,1]
p.ctrl[i,3] <- 1- p.ctrl[i,1] - p.ctrl[i,2]
a[i] ~ dnorm(0.0, 0.0001)
b[i] ~ dnorm(0.0, 0.0001)

                }

        d ~ dnorm(0.0, 0.0001)

# Bounded analysis
        lambda ~ dbeta(0.5,0.5)

# Unbounded analysis
        lambda ~ dnorm(0.5,0.1)

        prec <- 1/var
        var <- pow(sd,2)
        sd ~ dunif(0,2)
        OR_GG <- exp(d)
        OR_Gg <- exp(d*lambda)

}
```

Joint pairwise comparisons

WinBUGS code (prospective likelihood)

Model

```
{
        for( i in 1:n) {

  r_GG[i] ~ dbin(p_GG[i], n_GG[i])
  r_Gg[i] ~ dbin(p_ Gg [i], n_ Gg [i])
  r_gg[i] ~ dbin(p_ gg[i], n_ gg[i])
  logit(p_ gg[i]) <- mu[i]
  logit(p_ GG[i]) <- mu[i] + delta_ GG[i]
  logit(p_ Gg[i]) <- mu[i] + delta_ Gg[i]
  mu[i] ~ dnorm(0.0, 0.0001)
  delta_GG[i] ~ dnorm(d.GG, prec.GG)
  delta_ Gg[i] ~ dnorm(d.Gg, prec.Gg)

        }

        d.GG ~ dnorm(0.0, 0.0001)
        d.Gg ~ dnorm(0.0, 0.0001)
        prec.GG <- 1/ var.GG
        var.GG <- pow(sd.GG,2)
        sd.GG ~ dunif(0,2)
        prec.Gg <- 1/ var.Gg
        var.Gg <- pow(sd.Gg,2)
        sd.Gg ~ dunif(0,2)
        OR.GG<- exp(d.GG)
        OR.Gg<- exp(d.Gg)

}
```

Figure A – Bland-Altman plots of the difference in the estimates for Hani meta-analysis obtained by models with different prior against their average value for; a) OR_{Gg}; b) OR_{GG}; c) λ; d) τ.

a)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

b)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

c)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

d)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)
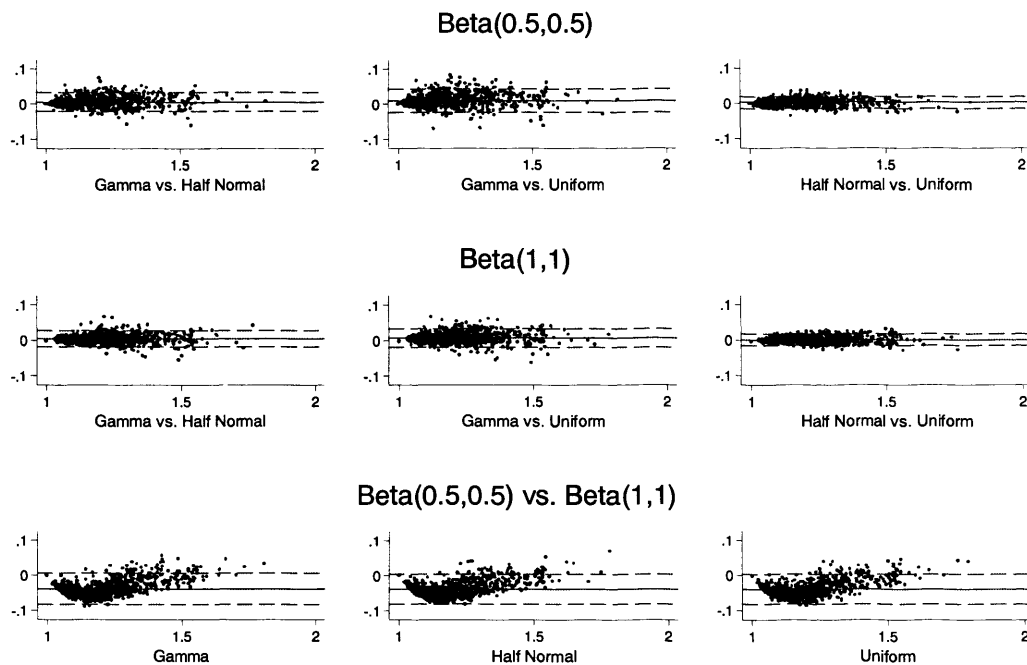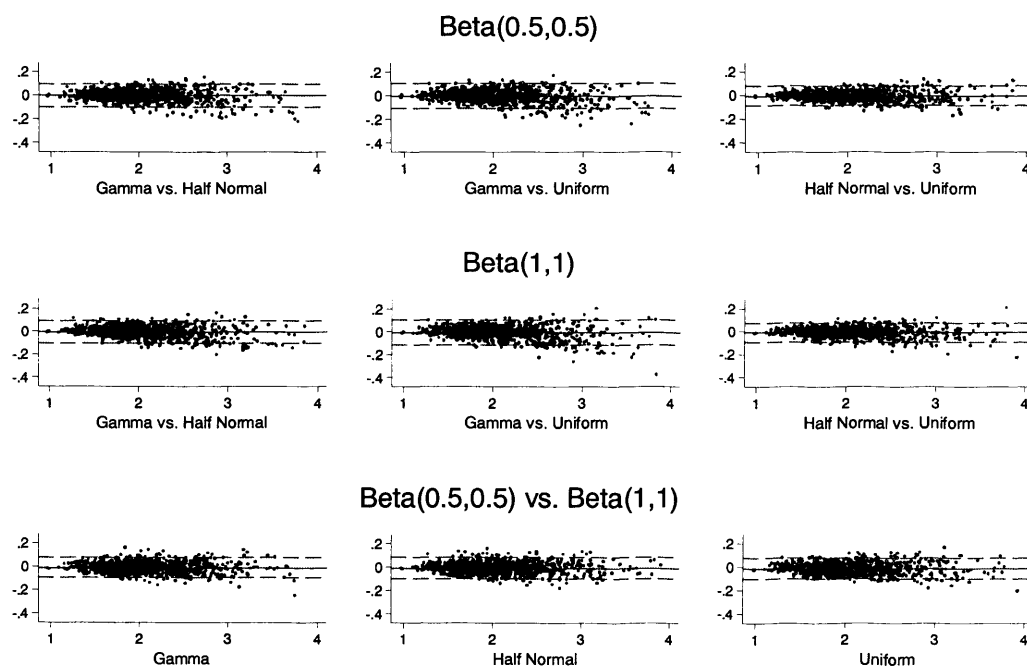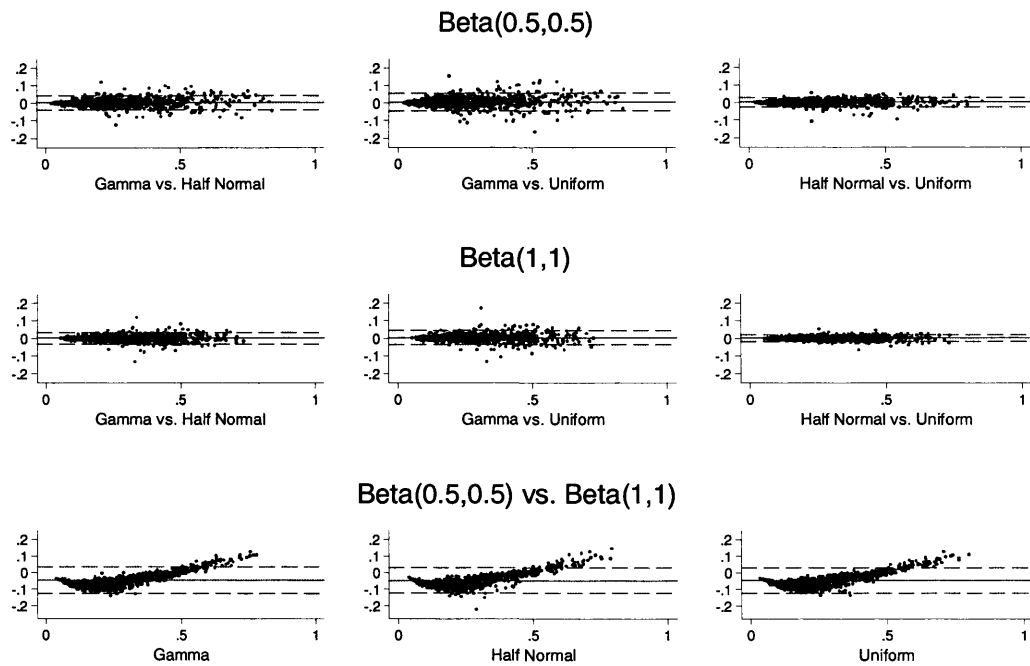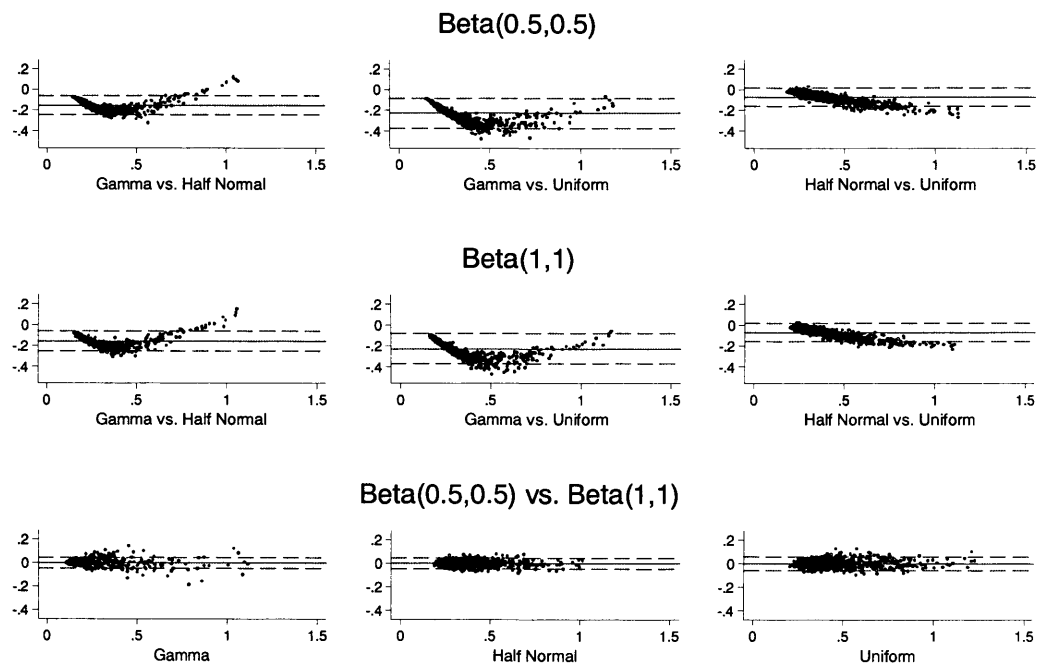
Figure B – Bland-Altman plots of the difference in the estimates for Wheeler meta-analysis obtained by models with different prior against their average value for; a) OR_{Gg}; b) OR_{GG}; c) λ; d) τ.

a)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

b)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

c)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)

d)

Beta(0.5,0.5)

Beta(1,1)

Beta(0.5,0.5) vs. Beta(1,1)
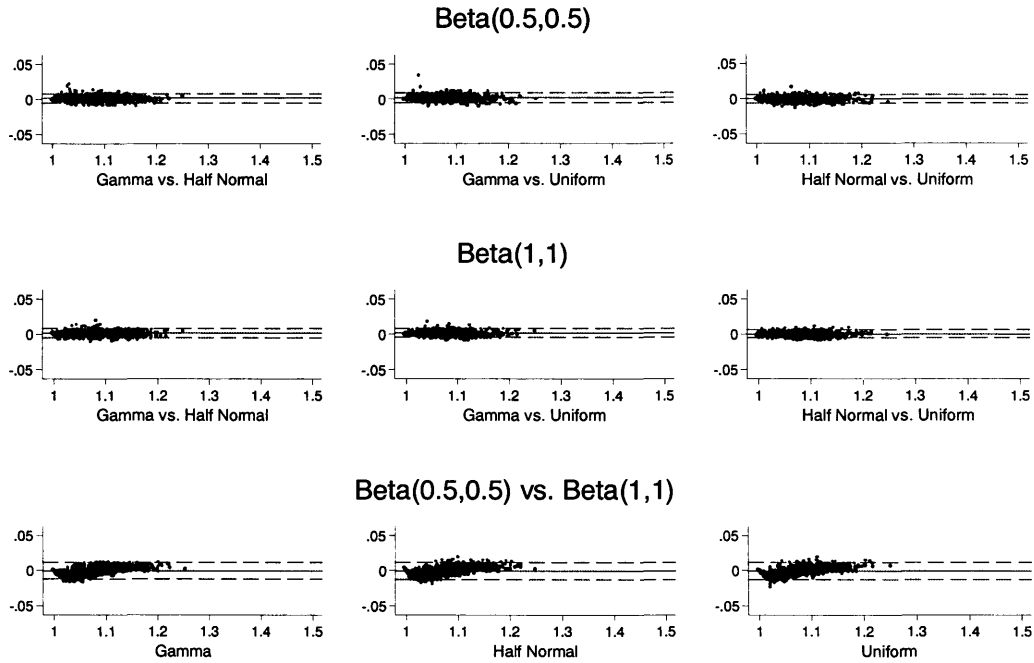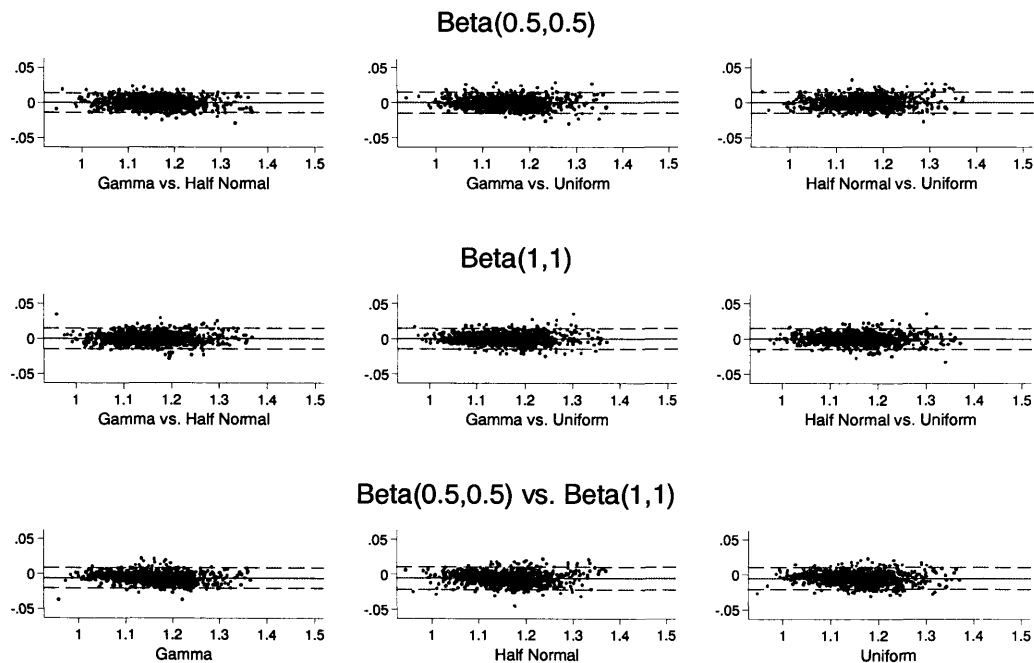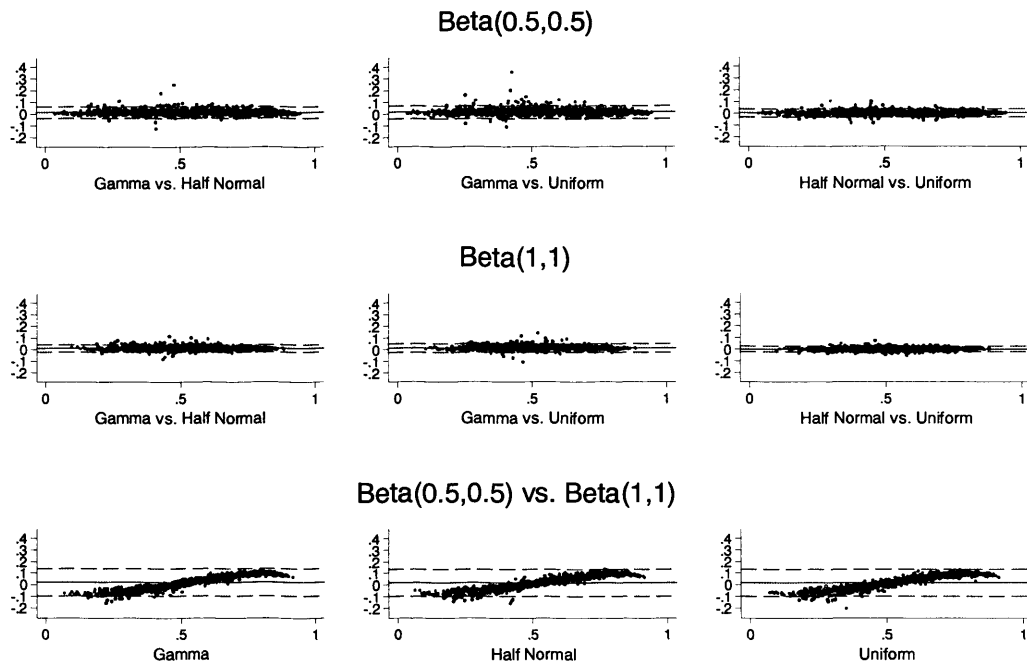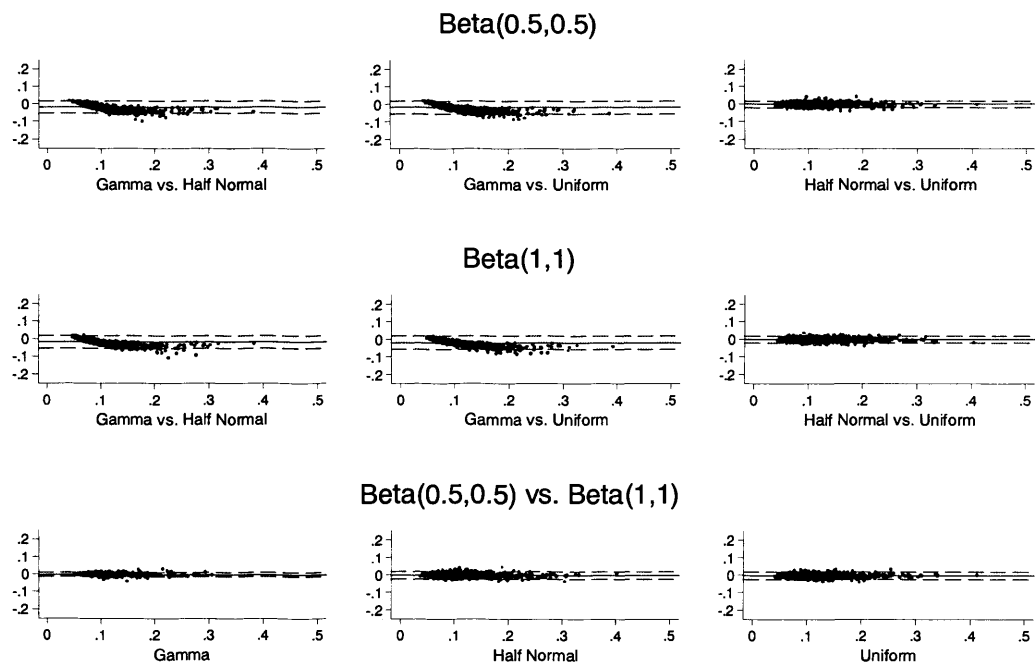
APPENDIX 2

CHAPTER 3

Subgroup analysis

References for Table 3.1 (§ 3.3)

Allison, D. B., Heo, M., Faith, M. S., and Pietrobelli, A. (1998) Meta-analysis of the Trp64Arg polymorphism in the b3 adrenergic receptor with body mass index. *International Journal of Obesity*; **22**:559-66.

Arranz, M. J., Munro, J., Sham, P., et al. (1998) Meta-analysis of studies on genetic variation in 5-HT2A receptors and clozapine response. *Schizophrenia Research*; **32**:93-9.

Breen, G., Brown, J., Maude, S., et al. (1999) -141 C del/ins polymorphism of the dopamine receptor 2 gene is associated with schizophrenia in a British population. *American Journal of Medical Genetics*; **88**:407-10.

Carlquist, J. F., Menlove, R. L., Murray, M. B., O'Connell, J. B., and Anderson, J. L. (1991) HLA class II (DR and DQ) antigen associations in idiopathic dilated cardiomyopathy. Validation study and meta-analysis of published HLA association studies. *Circulation*; **83**:515-22.

Christensen, P. M., Gotzsche, P. C., and Brosen, K. (1997) The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of lung cancer: a meta-analysis. *European Journal of Clinical Pharmacology*; **51**:389-93.

Dubertret, C., Gorwood, P., Ades, J., Feingold, J., Schwartz, J. C., and Sokoloff, P. (1998) Meta-analysis of DRD3 gene and schizophrenia: ethnic heterogeneity and significant association in Caucasians. *American Journal of Medical Genetics*; **81**:318-22.

Fujisawa, T., Ikegami, H., Kawaguchi, Y., et al. (1998) Meta-analysis of association of insertion/deletion polymorphism of angiotensin I-converting enzyme gene with diabetic nephropathy and retinopathy. *Diabetologia*; **41**:47-53.

Fujisawa, T., Ikegami, H., Kawaguchi, Y., and Ogihara, T. (1998) Meta-analysis of the association of Trp64Arg polymorphism of beta 3-adrenergic receptor gene with body mass index. *Journal of Clinical Endocrinology & Metabolism*; **83**:2441-4.

Furlong, R. A., Ho, L., Rubinsztein, J. S., Walsh, C., Paykel, E. S., and Rubinsztein, D. C. (1999) Analysis of the monoamine oxidase A (MAOA) gene in bipolar affective disorder by association studies, meta-analyses, and sequencing of the promoter. *American Journal of Medical Genetics*; **88**:398-406.

Furlong, R. A., Ho, L., Walsh, C., et al. (1998) Analysis and meta-analysis of two serotonin transporter gene polymorphisms in bipolar and unipolar affective disorders. *American Journal of Medical Genetics*; **81**:58-63.

Furlong, R. A., Rubinsztein, J. S., Ho, L., et al. (1999) Analysis and meta-analysis of two polymorphisms within the tyrosine hydroxylase gene in bipolar and unipolar affective disorders. *American Journal of Medical Genetics*; **88**:88-94.

Hani, E. H., Boutin, P., Durand, E., et al. (1998) Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia*; **41**:1511-5.

Hokanson, J. E. (1997) Lipoprotein lipase gene variants and risk of coronary disease: a quantitative analysis of population-based studies. *International Journal of Clinical & Laboratory Research*; 27:24-34.

Houlston, R. S. (2000) CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. *Pharmacogenetics*; 10:105-14.

Houlston, R. S. (1999) Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention*; 8:675-82.

Iacoviello, L., Burzotta, F., Di Castelnuovo, A., Zito, F., Marchioli, R., and Donati, M. B. (1998) The 4G/5G polymorphism of PAI-1 promoter gene and the risk of myocardial infarction: a meta-analysis. *Thrombosis & Haemostasis*; 80:1029-30.

Jenkins, D., Penny, M. A., Fletcher, J. A., et al. (1992) HLA class II gene polymorphism contributes little to Hashimoto's thyroiditis. *Clinical Endocrinology*; 37:141-5.

Joost, O., Taylor, C. A., Thomas, C. A., et al. (1999) Absence of effect of seven functional mutations in the CYP2D6 gene in Parkinson's disease. *Movement Disorders*; 14:590-5.

Juo, S. H., Wyszynski, D. F., Beaty, T. H., Huang, H. Y., and Bailey-Wilson, J. E. (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. *American Journal of Medical Genetics*; 82:235-41.

Kato, N., Sugiyama, T., Morita, H., Kurihara, H., Yamori, Y., and Yazaki, Y. (1999) Angiotensinogen gene and essential hypertension in the Japanese: extensive association study and meta-analysis on six reported studies. *Journal of Hypertension*; 17:757-63.

Keavney, B., McKenzie, C., Parish, S., et al. (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*; 355:434-42.

Kunz, R., Kreutz, R., Beige, J., Distler, A., and Sharma, A. (1997) M. Association between the angiotensinogen 235T-variant and essential hypertension in whites: a systematic review and methodological appraisal. *Hypertension*; 30:1331-7.

McCarron, M. O., Delong, D., and Alberts, M. J. (1999) APOE genotype as a risk factor for ischemic cerebrovascular disease: a meta-analysis. *Neurology*; 53:1308-11.

McWilliams, J. E., Sanderson, B. J., Harris, E. L., Richert-Boe, K. E., and Henner, W. D. (1995) Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. *Cancer Epidemiology Biomarkers & Prevention*; 4:589-94.

Ordovas, J. M., Lopez-Miranda, J., Perez-Jimenez, F., et al. (1995) Effect of apolipoprotein E and A-IV phenotypes on the low density lipoprotein response to HMG CoA reductase inhibitor therapy. *Atherosclerosis*; 113:157-66.

Pato, C. N., Macciardi, F., Pato, M. T., Verga, M., and Kennedy, J. L. (1993) Review of the putative association of dopamine D2 receptor and alcoholism: a meta-analysis. *American Journal of Medical Genetics*; 48:78-82.

Prasher, V. P., Chowdhury, T. A., Rowe, B. R., and Bain, S. C. (1997) ApoE genotype and Alzheimer's disease in adults with Down syndrome: meta-analysis. *American Journal of Mental Retardation*; **102**:103-10.

Preisig, M., Bellivier, F., Fenton, B. T., et al. (2000) Association between bipolar disorder and monoamine oxidase A gene polymorphisms: results of a multicenter study. *American Journal of Psychiatry*; **157**:948-55.

Rantala, M., Rantala, T. T., Savolainen, M. J., Friedlander, Y., and Kesaniemi, Y. A. (2000) Apolipoprotein B gene polymorphisms and serum lipids: meta-analysis of the role of genetic variation in responsiveness to diet. *American Journal of Clinical Nutrition*; **71**:713-24.

Rostami-Hodjegan, A., Lennard, M. S., Woods, H. F., and Tucker, G. T. (1998) Meta-analysis of studies of the CYP2D6 polymorphism in relation to lung cancer and Parkinson's disease. *Pharmacogenetics*; **8**:227-38.

Shaikh, S., Collier, D. A., Sham, P. C., et al. (1996) Allelic association between a Ser-9-Gly polymorphism in the dopamine D3 receptor gene and schizophrenia. *Human Genetics*; **97**:714-9.

Sharma, P. (1998) Meta-analysis of the ACE gene in ischaemic stroke. *Journal of Neurology, Neurosurgery & Psychiatry*; **64**:227-30.

Staessen, J. A., Wang, J. G., Ginocchio, G., et al. (1997) The deletion/insertion polymorphism of the angiotensin converting enzyme gene and cardiovascular-renal risk. *Journal of Hypertension*; **15**:1579-92.

Staessen, J. A., Kuznetsova, T., Wang, J. G., Emelianov, D., Vlietinck, R., and Fagard, R. (1999) M235T angiotensinogen gene polymorphism and cardiovascular renal risk. *Journal of Hypertension*; **17**:9-17.

Weston, A. and Godbold, J. H. (1997) Polymorphisms of H-ras-1 and p53 in breast cancer and lung cancer: a meta-analysis. *Environmental Health Perspectives*; **105**(Suppl.4):919-26.

Williams, J., Spurlock, G., Holmans, P., et al. (1998) Meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia. *Molecular Psychiatry*; **3**:141-9 [erratum appears in Molecula Psychiatry 1998; 3:458].

Wong, N. A., Rae, F., Simpson, K. J., Murray, G. D., and Harrison, D. J. (2000) Genetic polymorphisms of cytochrome p4502E1 and susceptibility to alcoholic liver disease and hepatocellular carcinoma in a white population: a study and literature review, including meta-analysis. *Molecular Pathology*; **53**:88-93.

Table A - Sensitivity analyses for λ [ε_j= logit(λ_j) – logit(λ_1), with j=2,3; $\varepsilon_j \sim N(0,\tau_\varepsilon^{\,2})$], when τ is not assumed common across subgroups

META-ANALYSIS APPROACH	White population (n=22)				Asian population (n=12)				Black population (n=6)			
	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)
τ_ε=0	1.31 (1.09 to 1.57)	1.12 (1.03 to 1.24)	0.45 (0.28 to 0.60)	OR$_{GG}$: 0.30 (0.15 to 0.51) OR$_{Gg}$: 0.13 (0.06 to 0.24)	1.81 (1.17 to 3.03)	1.30 (1.06 to 1.76)	0.45 (0.28 to 0.60)	OR$_{GG}$: 0.62 (0.29 to 1.23) OR$_{Gg}$: 0.27 (0.11 to 0.61)	1.31 (0.76 to 2.34)	1.12 (0.88 to 1.52)	0.45 (0.28 to 0.60)	OR$_{GG}$: 0.29 (0.01 to 1.11) OR$_{Gg}$: 0.10 (0.00 to 0.53)
τ_ε=0.43 (λ: 0.3 to 0.7)	1.31 (1.09 to 1.56)	1.13 (1.04 to 1.25)	0.46 (0.29 to 0.63)	OR$_{GG}$: 0.30 (0.15 to 0.51) OR$_{Gg}$: 0.13 (0.06 to 0.25)	1.78 (1.16 to 2.99)	1.26 (1.05 to 1.74)	0.42 (0.22 to 0.62)	OR$_{GG}$: 0.61 (0.29 to 1.21) OR$_{Gg}$: 0.25 (0.09 to 0.60)	1.31 (0.74 to 2.39)	1.12 (0.86 to 1.67)	0.48 (0.23 to 0.73)	OR$_{GG}$: 0.24 (0.01 to 1.11) OR$_{Gg}$: 0.11 (0.00 to 0.61)
τ_ε=1.12 (λ: 0.1 to 0.9)	1.30 (1.09 to 1.56)	1.13 (1.04 to 1.26)	0.47 (0.28 to 0.68)	OR$_{GG}$: 0.58 (0.27 to 1.17) OR$_{Gg}$: 0.14 (0.06 to 0.25)	1.70 (1.13 to 2.86)	1.19 (1.02 to 1.69)	0.35 (0.09 to 0.61)	OR$_{GG}$: 0.24 (0.01 to 1.24) OR$_{Gg}$: 0.20 (0.04 to 0.56)	1.29 (0.66 to 2.60)	1.11 (0.74 to 2.00)	0.56 (0.11 to 0.91)	OR$_{GG}$: 0.29 (0.15 to 0.50) OR$_{Gg}$: 0.11 (0.00 to 0.89)
τ_ε=+∞	1.30 (1.09 to 1.56)	1.13 (1.04 to 1.26)	0.48 (0.28 to 0.72)	OR$_{GG}$: 0.29 (0.15 to 0.50) OR$_{Gg}$: 0.14 (0.06 to 0.26)	1.63 (1.12 to 2.71)	1.12 (1.00 to 1.63)	0.26 (0.00 to 0.59)	OR$_{GG}$: 0.53 (0.24 to 1.10) OR$_{Gg}$: 0.13 (0.00 to 0.52)	1.26 (0.58 to 2.72)	1.09 (0.62 to 2.31)	0.71 (0.01 to 1.00)	OR$_{GG}$: 0.26 (0.01 to 1.33) OR$_{Gg}$: 0.14 (0.00 to 1.12)

Table B - Sensitivity analyses for $\tau\,[\phi_j = \log(\tau_j) - \log(\tau_1),\ \text{with } j=2,3;\ \phi_j \sim N(0,\tau_\phi^{\,2})]$, when λ is not assumed common across subgroups

META-ANALYSIS APPROACH	White population (n=22)				Asian population (n=12)				Black population (n=6)			
	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95% CrI)	Heterogeneity (SD) (95% CrI)	OR$_{GG}$ (95% CrI)	OR$_{Gg}$ (95% CrI)	λ (95%CrI)	Heterogeneity (SD) (95% CrI)
τ_ϕ=0	1.30 (1.08 to 1.58)	1.13 (1.03 to 1.27)	0.48 (0.28 to 0.69)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.16 (0.08 to 0.26)	1.52 (1.17 to 2.14)	1.06 (1.00 to 1.42)	0.15 (0.00 to 0.55)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.05 (0.00 to 0.21)	1.26 (0.63 to 2.50)	1.11 (0.66 to 2.14)	0.72 (0.01 to 1.00)	OR$_{GG}$: 0.33 (0.20 to 0.50) OR$_{Gg}$: 0.23 (0.00 to 0.45)
τ_ϕ=0.56 (τ: 0.20 to 0.35)	1.30 (1.08 to 1.56)	1.13 (1.03 to 1.26)	0.48 (0.28 to 0.71)	OR$_{GG}$: 0.30 (0.17 to 0.49) OR$_{Gg}$: 0.14 (0.07 to 0.25)	1.56 (1.16 to 2.35)	1.08 (1.00 to 1.49)	0.20 (0.01 to 0.57)	OR$_{GG}$: 0.40 (0.20 to 0.76) OR$_{Gg}$: 0.08 (0.00 to 0.33)	1.26 (0.65 to 2.45)	1.10 (0.68 to 2.12)	0.69 (0.01 to 1.00)	OR$_{GG}$: 0.24 (0.08 to 0.66) OR$_{Gg}$: 0.14 (0.00 to 0.54)
τ_ϕ=1.01 (τ: 0.20 to 0.55)	1.30 (1.09 to 1.55)	1.13 (1.04 to 1.26)	0.48 (0.27 to 0.71)	OR$_{GG}$: 0.29 (0.15 to 0.48) OR$_{Gg}$: 0.14 (0.06 to 0.25)	1.58 (1.15 to 2.46)	1.09 (1.00 to 1.53)	0.22 (0.00 to 0.57)	OR$_{GG}$: 0.44 (0.20 to 0.86) OR$_{Gg}$: 0.09 (0.00 to 0.39)	1.24 (0.62 to 2.38)	1.08 (0.65 to 2.05)	0.66 (0.01 to 1.00)	OR$_{GG}$: 0.17 (0.03 to 0.73) OR$_{Gg}$: 0.09 (0.00 to 0.60)
τ_ϕ=+∞	1.30 (1.09 to 1.56)	1.13 (1.04 to 1.26)	0.48 (0.28 to 0.72)	OR$_{GG}$: 0.29 (0.15 to 0.50) OR$_{Gg}$: 0.14 (0.06 to 0.26)	1.63 (1.12 to 2.71)	1.12 (1.00 to 1.63)	0.26 (0.00 to 0.59)	OR$_{GG}$: 0.53 (0.24 to 1.10) OR$_{Gg}$: 0.13 (0.00 to 0.52)	1.26 (0.58 to 2.72)	1.09 (0.62 to 2.31)	0.71 (0.01 to 1.00)	OR$_{GG}$: 0.26 (0.01 to 1.33) OR$_{Gg}$: 0.14 (0.00 to 1.12)

APPENDIX 3

CHAPTER 4

Hardy-Weinberg equilibrium

References for Table 4.1 (§ 4.4.1)

Anguelova, M., Benkelfat, C., and Turecki, G. (2003) A systematic review of association studies investigating genes coding for serotonin receptors and the serotonin transporter: II. Suicidal behavior. *Molecular Psychiatry*; **8**:646-53.

Botto, L. D. and Yang, Q. (2000) 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *American Journal of Epidemiology*; **151**:862-77.

Brennan, P., Lewis, S., Hashibe, M., et al. (2004) Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: a HuGE review. *American Journal of Epidemiology*; **159**:1-16.

Camp, N. J. and Tavtigian, S. V. (2002) Meta-analysis of associations of the Ser217Leu and Ala541Thr variants in ELAC2 (HPC2) and prostate cancer. *American Journal of Human Genetics*; **71**:1475-8.

Di Castelnuovo, A., de Gaetano, G., Donati, M. B., and Iacoviello, L. (2001) Platelet glycoprotein receptor IIIa polymorphism PLA1/PLA2 and coronary risk: a meta-analysis. *Thrombosis and Haemostasis*; **85**:626-33.

Benhamou, S. and Sarasin, A. (2005) ERCC2 /XPD gene polymorphisms and lung cancer: a HuGE review. *American Journal of Epidemiology*; **161**:1-14.

Efstathiadou, Z., Tsatsoulis, A., and Ioannidis, J. P. (2001) Association of collagen I 1 Sp1 polymorphism with the risk of prevalent fractures: a meta-analysis. *Journal of Bone & Mineral Research*; **16**:1586-92.

Ertekin-Taner N, Ronald J, Feuk L, et al.(2005) Elevated amyloid . protein (A. 42) and late onset Alzheimer's disease are associated with single nucleotide polymorphisms in the urokinase-type plasminogen activator gene. *Human Molecular Genetics*; **14**:447-60.

Feyler, A., Voho, A., Bouchardy, C., et al. (2002) Point: myeloperoxidase -463G --> a polymorphism and lung cancer risk. *Cancer Epidemiology, Biomarkers and Prevention*; **11**:1550-4.

Ioannidis, J. P., Stavrou, I., Trikalinos, T. A., et al. (2002) Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis. *Journal of Bone and Mineral Research*; **17**:2048-60.

Jee, S. H., Beaty, T. H., Suh, I., Yoon, Y., and Appel, L. J. (2000) The methylenetetrahydrofolate reductase gene is associated with increased cardiovascular risk in Japan, but not in other populations. *Atherosclerosis*; **153**:161-8.

Kaklamani, V. G., Hou, N., Bian, Y., et al. (2003) TGFBR1*6A and cancer risk: a meta-analysis of seven case-control studies. *Journal of Clinical Oncology*; **21**:3236-43.

Kavvoura, F. K. and Ioannidis, J. P. (2005) Ala45Thr polymorphism of the NEUROD1 gene and diabetes susceptibility: a meta-analysis. *Human Genetics*; **116**:192-9.

Kehoe, P. G., Katzov, H., Feuk, L., et al. (2003) Haplotypes extending across ACE are associated with Alzheimer's disease. *Human Molecular Genetics*; **12**:859-67.

Lawlor, D. A., Day, I. N., Gaunt, T. R., et al. (2004) The association of the PON1 Q192R polymorphism with coronary heart disease: findings from the British Women's Heart and Health cohort study and a meta-analysis. *BMC Genetics*; **5**:17.

Lee, W. J., Brennan, P., Boffetta, P., et al. (2002) Microsomal epoxide hydrolase polymorphisms and lung cancer risk: a quantitative review. *Biomarkers*; **7**:230-41.

Matakidou, A., Eisen, T., and Houlston, R. S. (2003) TP53 polymorphisms and lung cancer risk: a systematic review and meta-analysis. *Mutagenesis*; **18**:377-85.

Lin, P. Y. and Tsai, G. (2004) Association between serotonin transporter gene promoter polymorphism and suicide: results of a meta-analysis. *Biological Psychiatry*; **55**:1023-30.

Noble, E. P. (2003) D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*; **116B**:103-25.

Ntais, C., Polycarpou, A., and Ioannidis, J. P. (2003) Association of the CYP17 gene polymorphism with the risk of prostate cancer: a meta-analysis. *Cancer Epidemiology, Biomarkers and Prevention*; **12**:120-6.

Ntais, C., Polycarpou, A., and Ioannidis, J. P. (2003) SRD5A2 gene polymorphisms and the risk of prostate cancer: a meta-analysis. *Cancer Epidemiology, Biomarkers and Prevention*; **12**:618-24.

Ntais, C., Polycarpou, A., and Ioannidis, J. P. (2003) Vitamin D receptor gene polymorphisms and risk of prostate cancer: a meta-analysis. *Cancer Epidemiology, Biomarkers and Prevention*; **12**:1395-402.

Ntais, C., Polycarpou, A., and Ioannidis, J. P. (2004) Meta-analysis of the association of the cathepsin D Ala224Val gene polymorphism with the risk of Alzheimer's disease: a HuGE gene-disease association review. *American Journal of Epidemiology*; **159**:527-36.

Ray, J. G., Shmorgun, D., and Chan, W. S. (2002) Common C677T polymorphism of the methylenetetrahydrofolate reductase gene and the risk of venous thromboembolism: meta-analysis of 31 studies. *Pathophysiology of Haemostasis and Thrombosis*; **32**:51-8.

Rujescu, D., Giegling, I., Sato, T., Hartmann, A. M., and Moller, H. J. (2003) Genetic variations in tryptophan hydroxylase in suicidal behavior: analysis and meta-analysis. *Biological Psychiatry*; **54**:465-73.

Schena, F. P., D'Altri, C., Cerullo, G., Manno, C., and Gesualdo, L. (2001) ACE gene polymorphism and IgA nephropathy: an ethnically homogeneous study and a meta-analysis. *Kidney International*; **60**:732-40.

Sethi, A. A., Nordestgaard, B. G., and Tybjaerg-Hansen, A. (2003) Angiotensinogen gene polymorphism, plasma angiotensinogen, and risk of hypertension and ischemic heart disease: a meta-analysis. *Arteriosclerosis, Thrombosis and Vascular Biology*; **23**:1269-75.

Ye, Z. and Parry, J. M. (2002) The CYP17 MspA1 polymorphism and breast cancer risk: a meta-analysis. *Mutagenesis*; **17**:119-26.

Zintzaras, E. and Stefanidis, I. (2005) Association between the GLUT1 gene polymorphism and the risk of diabetic nephropathy: a meta-analysis. *Journal of Human Genetics*; **50**:84-91.

Zhu, M. M., Weedon, J., and Clark, L. T. (2000) Meta-analysis of the association of platelet glycoprotein IIIa PlA1/A2 polymorphism with myocardial infarction. *American Journal of Cardiology*; **86**:1000-5.

Table A – Fixed effects analysis: Results for the logOR$_{GG}$ under the different strategies obtained using the *main model* in the two scenarios; a) scenario 1; b) scenario 2.

a)

Strategy	Theta=0						Theta=0.2						Theta=0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.405	0.0	0.102	0.081	95.1	0.0	0.407	0.4	0.102	0.081	94.7	0.0	0.412	1.7	0.102	0.083	94.9
Exclude if: p<=0.05	16.3	0.407	0.4	0.111	0.087	95.3	16.1	0.408	0.5	0.111	0.088	94.9	16.2	0.411	1.4	0.111	0.089	95.2
Exclude if: p<=0.1	23.6	0.407	0.4	0.118	0.091	95.3	23.5	0.407	0.4	0.118	0.092	95.0	23.6	0.411	1.3	0.118	0.093	95.0
Exclude if: |α|>1	1.9	0.405	-0.2	0.102	0.081	95.0	1.9	0.406	0.2	0.102	0.082	94.7	1.9	0.411	1.4	0.102	0.083	94.9
Exclude if: |α|>0.5	12.5	0.402	-0.7	0.105	0.083	95.1	12.4	0.404	-0.4	0.104	0.083	94.9	12.5	0.408	0.5	0.105	0.084	94.9
Exclude if: p<=0.1 & α>0.5	10.8	0.404	-0.5	0.104	0.083	95.1	10.7	0.405	-0.2	0.104	0.083	95.0	10.8	0.409	0.8	0.104	0.084	94.9
Exclude if: p<=0.05 & α>0.5	9.4	0.404	-0.3	0.104	0.083	95.1	9.4	0.406	0.0	0.104	0.083	94.9	9.4	0.410	1.0	0.104	0.084	94.9

b)

Strategy	Theta=0						Theta=0.2						Theta=0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.407	0.3	0.102	0.081	95.0	0.0	0.430	6.0	0.102	0.083	94.1	0.0	0.453	11.8	0.101	0.090	92.5
Exclude if: p<=0.05	16.5	0.407	0.5	0.111	0.087	95.5	16.3	0.428	5.6	0.111	0.089	94.8	16.5	0.447	10.3	0.111	0.093	93.6
Exclude if: p<=0.1	23.9	0.408	0.7	0.118	0.091	95.3	23.6	0.428	5.5	0.117	0.092	95.1	23.9	0.446	10.1	0.118	0.097	93.8
Exclude if: |α|>1	1.9	0.406	0.1	0.102	0.081	95.1	1.8	0.429	5.7	0.102	0.083	94.2	1.9	0.452	11.4	0.102	0.089	92.7
Exclude if: |α|>0.5	12.6	0.404	-0.4	0.105	0.082	95.4	12.5	0.424	4.6	0.104	0.083	94.7	12.6	0.444	9.6	0.104	0.088	93.4
Exclude if: p<=0.1 & α>0.5	11.0	0.405	-0.2	0.105	0.082	95.4	10.8	0.426	5.0	0.104	0.083	94.5	11.0	0.446	9.9	0.104	0.088	93.4
Exclude if: p<=0.05 & α>0.5	9.6	0.406	0.0	0.104	0.082	95.4	9.4	0.427	5.2	0.104	0.084	94.5	9.6	0.447	10.2	0.104	0.089	93.3

Table B – Fixed effects analysis: Results for the logOR_GG under the different strategies obtained using the *mixture model* in the two scenarios; a) scenario 1; b) scenario 2.

a)

Strategy	Theta=0						Theta=0.2						Theta=0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.405	-0.1	0.102	0.081	95.0	0.0	0.410	1.2	0.102	0.084	94.5	0.0	0.417	2.9	0.102	0.085	93.6
Exclude if: p<=0.05	19.4	0.408	0.7	0.115	0.091	95.3	19.4	0.409	0.9	0.116	0.093	94.9	19.6	0.413	1.8	0.115	0.093	94.7
Exclude if: p<=0.1	26.6	0.408	0.6	0.122	0.095	95.3	26.8	0.409	0.8	0.123	0.098	95.1	26.8	0.413	1.8	0.122	0.098	94.6
Exclude if: |α|>1	2.5	0.404	-0.3	0.102	0.081	95.1	2.4	0.409	0.9	0.102	0.084	94.5	2.5	0.415	2.4	0.102	0.085	93.8
Exclude if: |α|>0.5	14.9	0.403	-0.7	0.107	0.085	95.2	14.9	0.405	0.0	0.107	0.087	95.0	15.0	0.410	1.0	0.107	0.087	94.3
Exclude if: p<=0.1 & α>0.5	13.1	0.404	-0.4	0.107	0.085	95.2	13.2	0.406	0.2	0.107	0.087	95.0	13.3	0.411	1.3	0.106	0.087	94.3
Exclude if: p<=0.05 & α>0.5	11.8	0.405	-0.1	0.106	0.085	95.2	11.8	0.407	0.5	0.106	0.086	94.9	11.9	0.412	1.6	0.106	0.087	94.4

b)

Strategy	Theta=0						Theta=0.2						Theta=0.4					
	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage	Studies excl. (%)	Mean	% bias	Aver. SE	RMSE	Coverage
Do nothing	0.0	0.406	0.1	0.102	0.080	95.3	0.0	0.435	7.2	0.102	0.084	94.3	0.0	0.464	14.4	0.101	0.095	90.5
Exclude if: p<=0.05	19.2	0.408	0.7	0.115	0.088	95.7	19.3	0.430	6.1	0.115	0.091	95.2	19.4	0.450	11.0	0.115	0.096	93.7
Exclude if: p<=0.1	26.7	0.409	0.9	0.122	0.093	95.8	26.5	0.430	6.0	0.122	0.096	95.2	26.8	0.449	10.7	0.122	0.100	94.2
Exclude if: |α|>1	2.5	0.405	-0.1	0.102	0.081	95.3	2.4	0.432	6.6	0.102	0.084	94.3	2.5	0.460	13.4	0.102	0.094	91.3
Exclude if: |α|>0.5	14.9	0.403	-0.6	0.107	0.083	95.5	14.9	0.425	4.9	0.107	0.085	95.1	15.0	0.447	10.3	0.106	0.091	93.3
Exclude if: p<=0.1 & α>0.5	13.3	0.404	-0.3	0.106	0.083	95.5	13.2	0.427	5.2	0.106	0.085	94.9	13.4	0.449	10.7	0.106	0.091	93.1
Exclude if: p<=0.05 & α>0.5	11.8	0.405	-0.1	0.106	0.083	95.4	11.8	0.428	5.5	0.106	0.085	94.9	11.9	0.450	11.0	0.106	0.092	92.9

Figure A – Fixed effects analysis: Box plots of the results for the different strategies obtained using the *main model* in the two scenarios; a) scenario 1; b) scenario 2. *1*: do nothing; *2*: exclu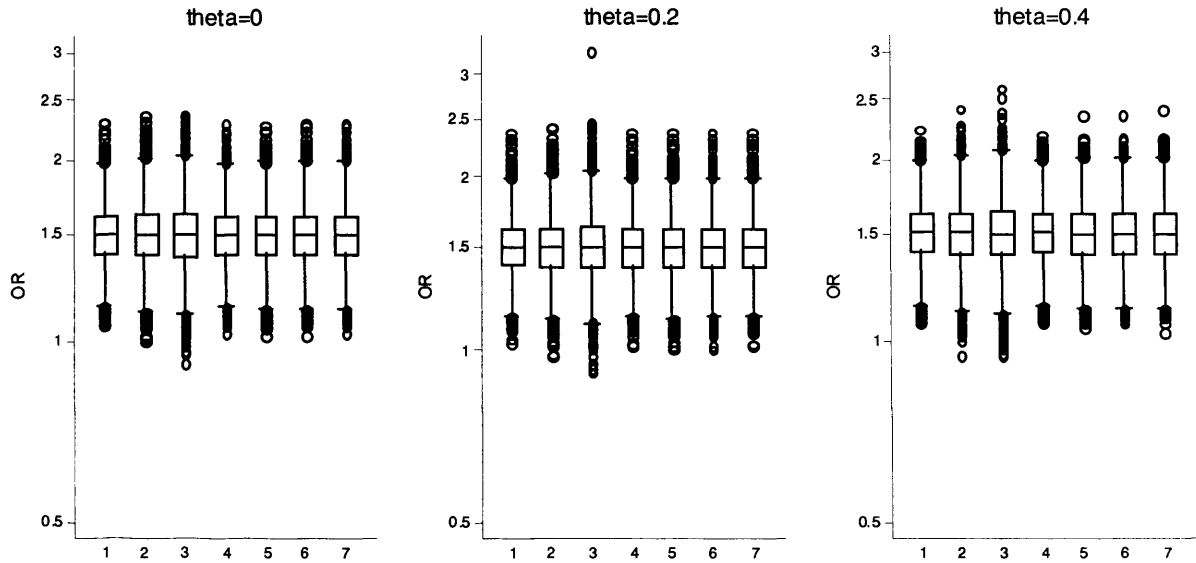de if p<=0.05; *3*: exclude if p<=0.1; *4*: exclude if |alpha|>1; *5*: exclude if |alpha|>0.5; *6*: exclude if |alpha|>0.5 & p<=0.1; *7*: exclude if |alpha|>0.5 & p<=0.05

a)

b)
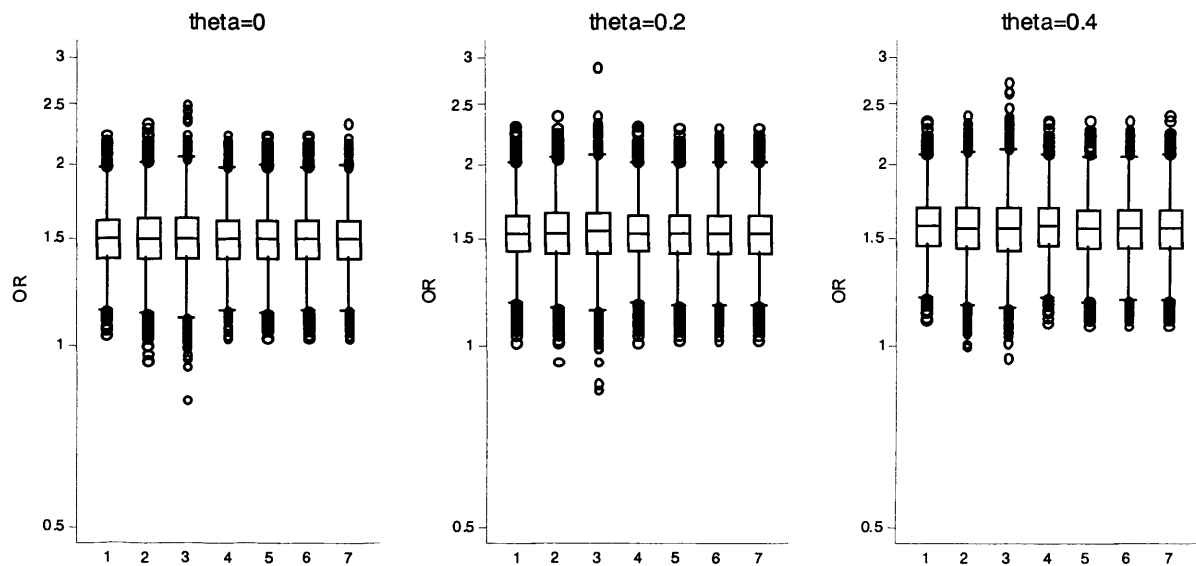
Figure B – Fixed effects analysis: Box plots of the results for the different strategies obtained using the *mixture model* in the two scenarios; a) scenario 1; b) scenario 2. *1*: do nothing; *2*: exclude if p<=0.05; *3*: exclude if p<=0.1; *4*: exclude if |alpha|>1; *5*: exclude if |alpha|>0.5; *6*: exclude if |alpha|>0.5 & p<=0.1; *7*: exclude if |alpha|>0.5 & p<=0.05

a)
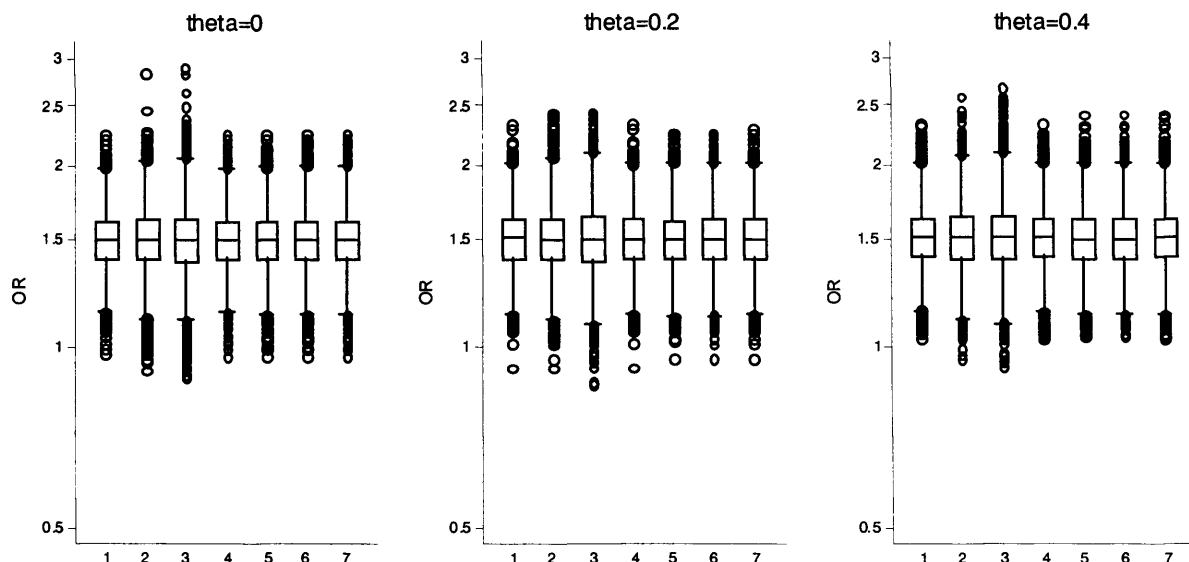
b)
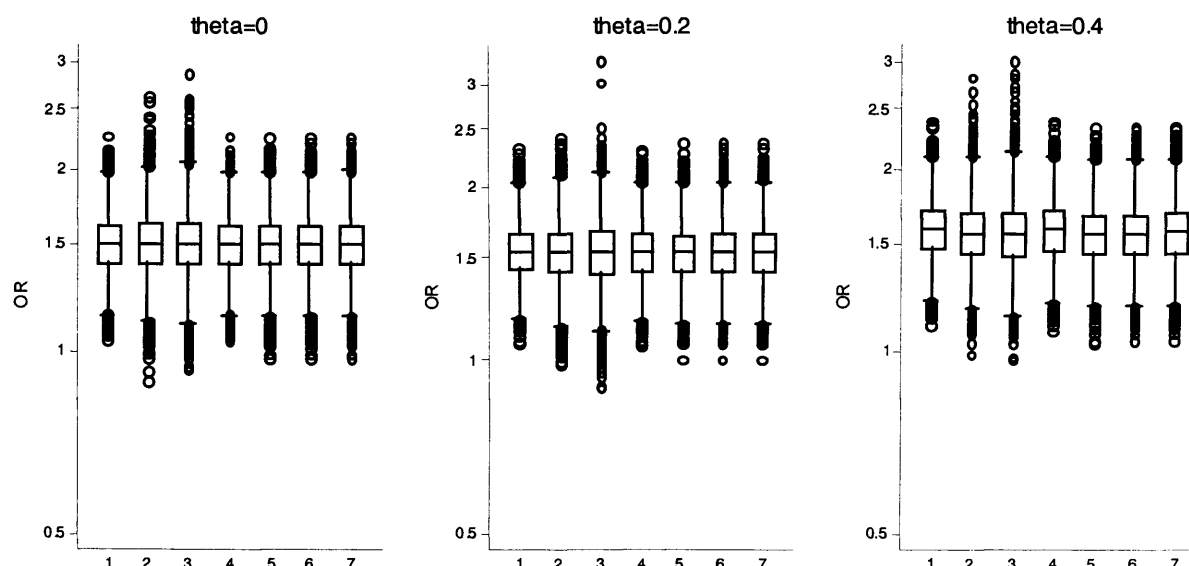
Simulation work

Stata code for the ado.file for simulating the 10,000 datasets

```
program sim_meta_analysis
    syntax newvarlist(min = 6 max = 6 generate) , ///
       AF(varlist) ///
       NCAses(varlist) ///
       NCOntrols(varlist) ///
       OR(varlist) ///
       alpha(varlist)

    quietly {
         tokenize `varlist'
         local ncase0 = "`1'"
         local ncase1 = "`2'"
         local ncase2 = "`3'"
         local ncontrol0 = "`4'"
         local ncontrol1 = "`5'"
         local ncontrol2 = "`6'"

         tokenize `or'
         local or1 = "`1'"
         local or2 = "`2'"
         local nstudy = _N


         forvalues i = 1 / `nstudy' {

local a1 = exp(-`alpha'[`i']+`af'[`i']+log(2))
local a2 = exp(2*`af'[`i'])
              local p1 = 1/(1+`a1'+`a2')
              local p2 = `a1'/(1+`a1'+`a2')
              local p3 = `a2'/(1+`a1'+`a2')


*=======================================
* numbers of controls
*=======================================
              local nC = `ncontrols'[`i']
              preserve
              drop _all
              set obs `nC'
              tempvar u g
              gen `u' = uniform()
              gen `g' = (`u' < `p3') + (`u' < (`p2' + `p3'))
              forvalues gt = 0 / 2 {
                   count if `g' == `gt'
                   local ct`gt' = r(N)
                   }
              restore
              forvalues gt = 0 / 2 {
                   replace `ncontrol`gt'' = `ct`gt'' in `i'
                   }
*=======================================
* numbers of cases
*=======================================
              local q1 = `p1'
              local q2 = `p2' * `or1'[`i']
              local q3 = `p3' * `or2'[`i']
```

```
        local sq = `q1' + `q2' + `q3'
        local q1 = `q1' / `sq'
        local q2 = `q2' / `sq'
        local q3 = `q3' / `sq'
        local nC = `ncases'[`i']
        preserve
        drop _all
        set obs `nC'
        tempvar u g
        gen `u' = uniform()
        gen `g' = (`u' < `q3') + (`u' < (`q2' + `q3'))
        forvalues gt = 0 / 2 {
            count if `g' == `gt'
            local ct`gt' = r(N)
            }
        restore
        forvalues gt = 0 / 2 {
            replace `ncase`gt'' = `ct`gt'' in `i'
            }
        }
    }
end
```

a) Stata code for the main model

```
*=======================================
* Fixed parameters
*=======================================

input ncase
40
75
95
115
135
165
200
250
345
695
end

gen ncontrol=ncase
gen af=-0.75

*gen theta=0
*gen theta=0.2
gen theta=0.4


*=======================================
* Parameters generated at each simulation
*=======================================

quietly{
forvalues x = 1 / 10000 {

        keep ncase ncontrol af theta
        set obs 10

        gen rep = `x'
        gen alpha=invnorm(uniform())* 40/(ncase+40)
        *gen delta = log(1.5)+ theta*alpha
        gen delta = log(1.5)+ theta*abs(alpha)

        gen lambda=0.5
        gen or1 = exp(lambda * delta)
        gen or2 = exp(delta)

sim_meta_analysis case1 case2 case3 control1 control2 control3 , af(af)
ncas(ncase) ncon(ncontrol) or(or1 or2) alpha(alpha)

    di " "
    di "SIMULATION `x'"
    gen a = 0.5*log(4*control1*control3/control2^2)

                        }
            }
```

b) Stata code for the mixture model

```
*=====================================
* Fixed parameters
*=====================================

input ncase
40
75
95
115
135
165
200
250
345
695
end

gen ncontrol=ncase
gen af=-0.75

*gen theta=0
*gen theta=0.2
gen theta=0.4


*=====================================
* Parameters generated at each simulation
*=====================================

quietly{
forvalues x = 1 / 10000 {

        keep ncase ncontrol af theta
        set obs 10

        gen rep = `x'
        gen a_1=invnorm(uniform())* 40/(ncase+40)
        gen a_2=invnorm(uniform())* 0.5
        gen u=uniform()
        gen alpha=a_1*(u<0.9) + a_2*(u>0.9)
        *gen delta = log(1.5)+ theta*alpha
        gen delta = log(1.5)+ theta*abs(alpha)
        gen lambda=0.5
        gen or1 = exp(lambda * delta)
        gen or2 = exp(delta)

sim_meta_analysis case1 case2 case3 control1 control2 control3 , af(af)
ncas(ncase) ncon(ncontrol) or(or1 or2) alpha(alpha)

    di " "
    di "SIMULATION `x'"
    gen a = 0.5*log(4*control1*control3/control2^2)


                    }
        }
```

APPENDIX 4

CHAPTER 5

Mendelian randomisation

a)   WinBUGS code for the Model A

Model

```
{
# Studies evaluating G-P and G-D associations (n=18)

for( i in 1:18) {
        delta[i,1:2] ~ dmnorm(mu[],T[ , ])
        Diff_GP[i] ~ dnorm(delta[i,1], Weight_GP[i])
        LogOR_GD[i] ~ dnorm(delta[i,2], Weight_GD[i])
            }

# Studies with only G-D association (n=32)

for (i in 19:50) {
        LogOR_GD[i] ~ dnorm(delta[i,2], Weight_GD[i])
        delta[i,2] ~ dnorm(mu[2],tau_GD)
            }

# Studies with only G-P association (n=16)

for (i in 51:66) {
        Diff_GP[i] ~ dnorm(delta[i,1], Weight_GP[i])
        delta[i,1] ~ dnorm(mu[1],tau_GP)
            }

mu[1] ~ dnorm(0.0,0.0001)
mu[2] ~ dnorm(0.0,0.0001)

tau_GP ~ dgamma(0.001,0.001)
var2_GP <- 1/tau_GP
sd_GD <- sd(delta_GD[])
var_GD <-pow(sd_GD,2)
tau_PD ~ dgamma(0.001,0.001)
var2_PD <- 1/tau_PD

# Sensitivity analysis for tau_GP and tau_GD - half normal distribution

#tau_GP <- 1/var2_GP
#var2_GP <- pow(sd_GP,2)
#sd_GP ~ dnorm(0,1)I(0,)
#tau_PD <- 1/var2_PD
#var2_PD <- pow(sd_PD,2)
#sd_PD ~ dnorm(0,1)I(0,)

# Sensitivity analysis for tau_GP and tau_GD - uniform distribution

#tau_GP <- 1/var2_GP
#var2_GP <- pow(sd_GP,2)
#sd_GP ~ dunif(0,2)

#tau_GD <- 1/var2_GD
#var2_GD <- pow(sd_GD,2)
#sd_GD ~ dunif(0,2)

T[1:2,1:2] ~ dwish(R[ , ],2)
var_GP <- 1/T[1,1]
var_GD <- 1/T[2,2]
cov <- 1/T[1,2]
for (i in 1:2) {
```

```
        for (j in 1:2) {
        covmat[i,j] <- inverse(T[ , ], i, j)
                        }
                }
corr <- covmat[1,2]/sqrt(covmat[1,1]*covmat[2,2])
logOR_PD <- mu[2]*5/mu[1]
pOR_GD <- exp(mu[2])
pDiff_GP <- mu[1]
OR_PD <- exp(logOR_PD)


        }
```

Data

R[,1]	R[,2]
7.75	0.82
0.82	0.35

Assumed values for Wishart prior distribution in sensitivity analysis:

1) variances halved, same correlation (0.5)

R[,1]	R[,2]
3.88	0.41
0.41	0.18

2) variances doubled, same correlation (0.5)

R[,1]	R[,2]
15.50	1.64
1.64	0.70

3) same variances, correlation of 0

R[,1]	R[,2]
7.75	0
0	0.35

4) same variances, correlation of 0.99

R[,1]	R[,2]
7.75	1.63
1.63	0.35

LogOR_GD[]	Weight_GD[]	Diff_GP[]	Weight_GP[]
-0.82	2.37	0.4	0.1
-0.58	10.53	4.9	0.17
-0.48	4	2.9	0.41
-0.31	6.93	1.3	2.19
-0.22	10.41	-0.8	2.08
-0.2	39.06	0.9	5.15
-0.17	13.72	2	2.78
-0.11	4.73	2.6	0.81
0.07	6.57	1.3	0.38
0.19	34.6	2.8	0.68
0.25	6.25	2.5	0.87
0.28	13.72	1.4	2.15
0.41	3.08	8.8	0.07
0.55	10.41	3	1.76
0.73	25	3.8	0.68
0.84	16	4.6	0.62
0.96	4.34	1.2	1.4
1.14	3.84	11	0.09
-1.61	1.65	NA	NA
-0.63	3.51	NA	NA
-0.37	13.49	NA	NA

-0.29	10.01	NA	NA
-0.26	4.74	NA	NA
-0.22	11.95	NA	NA
-0.21	11.59	NA	NA
-0.17	1.3	NA	NA
-0.17	25.63	NA	NA
-0.11	6.25	NA	NA
-0.01	9.75	NA	NA
0.03	6.1	NA	NA
0.04	12.94	NA	NA
0.06	4.43	NA	NA
0.1	13.26	NA	NA
0.25	6.43	NA	NA
0.26	11.87	NA	NA
0.3	5.89	NA	NA
0.32	5.38	NA	NA
0.35	3.86	NA	NA
0.38	4.54	NA	NA
0.4	1.52	NA	NA
0.42	8.62	NA	NA
0.46	2.58	NA	NA
0.49	11.82	NA	NA
0.5	11	NA	NA
0.77	7.14	NA	NA
0.98	8.35	NA	NA
1.18	4.22	NA	NA
1.26	6.87	NA	NA
1.3	3.29	NA	NA
1.46	3.26	NA	NA
NA	NA	0.9	3
NA	NA	1	1.16
NA	NA	1.4	1.56
NA	NA	1.4	2.63
NA	NA	1.7	0.39
NA	NA	2.1	1.89
NA	NA	2.1	0.48
NA	NA	2.1	1.19
NA	NA	3.8	0.43
NA	NA	4	0.21
NA	NA	4.2	0.13
NA	NA	4.3	0.31
NA	NA	4.4	0.25
NA	NA	7.3	0.24
NA	NA	8.1	1.12
NA	NA	9.9	0.29

END

b) WinBUGS code for the Model B

Model

```
{
# All studies evaluating either G-P, G-D or both associations (n=66)
for( i in 1:66) {
        Diff_GP[i] ~ dnorm(delta_GP[i], Weight_GP[i])
        LogOR_GD[i] ~ dnorm(delta_GD[i], Weight_GD[i])
        delta_GD[i] <- delta_GP[i]*delta_PD[i]/5
        delta_GP[i] ~ dnorm(d_GP, tau_GP)
        delta_PD[i] ~ dnorm(d_PD, tau_PD)
        Weight_GP[i] ~ dlnorm(mu_WGP,psi_GP)
        Weight_GD[i] ~ dlnorm(mu_WGD,psi_GD)
        PD[i] <- exp(delta_PD[i])
            }

m_GD <- mean(delta_GD[])
mu_WGP ~ dnorm(0.0,0.0001)
mu_WGD ~ dnorm(0.0,0.0001)
d_GP ~ dnorm(0.0,0.0001)
d_PD ~ dnorm(0.0,0.0001)

tau_GP ~ dgamma(0.001,0.001)
var_GP <- 1/tau_GP
tau_PD ~ dgamma(0.001,0.001)
var_PD <- 1/tau_PD
psi_GP ~ dgamma(0.001,0.001)
psi_GD ~ dgamma(0.001,0.001)

# Sensitivity analysis for tau_GP and tau_GD - half normal distribution

#tau_GP <- 1/var_GP
#var_GP <- pow(sd_GP,2)
#sd_GP ~ dnorm(0,1)I(0,)
#tau_PD <- 1/var_PD
#var_PD <- pow(sd_PD,2)
#sd_PD ~ dnorm(0,1)I(0,)
#psi_GP <- 1/var_psi_GP
#var_psi_GP <- pow(sd_psi_GP,2)
#sd_psi_GP ~ dnorm(0,1)I(0,)
#psi_GD <- 1/var_psi_GD
#var_psi_GD <- pow(sd_psi_GD,2)
#sd_psi_GD ~ dnorm(0,1)I(0,)

# Sensitivity analysis for tau_GP and tau_GD – uniform distribution

#tau_GP <- 1/var_GP
#var_GP <- pow(sd_GP,2)
#sd_GP ~ dunif(0,2)
#tau_PD <- 1/var_PD
#var_PD <- pow(sd_PD,2)
#sd_PD ~ dunif(0,2)
#psi_GP <- 1/var_psi_GP
#var_psi_GP <- pow(sd_psi_GP,2)
#sd_psi_GP ~ dunif(0,2)
#psi_GD <- 1/var_psi_GD
#var_psi_GD <- pow(sd_psi_GD,2)
```

#sd_psi_GD ~ dunif(0,2)

pDiff_GP <- d_GP
pOR_GD <- exp(m_GD)
OR_PD <- exp(d_PD)

     }

Data

LogOR_GD[]   Weight_GD[]   Diff_GP[]   Weight_GP[]

Same as for Model A

Study	GENOTYPE-DISEASE		GENOTYPE-PHENOTYPE	
	log OR	Variance logOR	Diff. Hcy	Variance Diff. Hcy
Chambers (Asians)	-0.82	0.42	0.4	10.00
Verhoeff BJ	-0.58	0.10	4.9	5.90
Chao	-0.48	0.25	2.9	2.44
Tsai	-0.31	0.14	1.3	0.46
Schmitz	-0.22	0.10	-0.8	0.48
Meisel	-0.20	0.03	0.9	0.19
Ma	-0.17	0.07	2.0	0.36
Schwartz	-0.11	0.21	2.6	1.24
Kim	0.07	0.15	1.3	2.63
Kluijtmans (1997)	0.19	0.03	2.8	1.47
Christensen	0.25	0.16	2.5	1.15
Chambers (European)	0.28	0.07	1.4	0.47
Tokgozoglu	0.41	0.33	8.8	14.29
Nakai	0.55	0.10	3.0	0.57
Morita	0.73	0.04	3.8	1.47
Ou	0.84	0.06	4.6	1.61
Malinow	0.96	0.23	1.2	0.71
Kawashiri	1.14	0.26	11.0	11.11
Zheng	-1.61	0.61	/	/
Fernandez-A. (Females)	-0.63	0.29	/	/
Brulhart	-0.37	0.07	/	/
Girelli	-0.29	0.10	/	/
Brugada	-0.26	0.21	/	/
Adams	-0.22	0.08	/	/
Ardissino	-0.21	0.09	/	/
Dilley	-0.17	0.77	/	/
Verhoef (1998)	-0.17	0.04	/	/
Hsu	-0.11	0.16	/	/
Van Bockxmeer	-0.01	0.10	/	/
Abbate	0.03	0.16	/	/
Wilcken	0.04	0.08	/	/
Pinto	0.06	0.23	/	/
Anderson (1997)	0.1	0.08	/	/
Fowkes	0.25	0.16	/	/
Gardemann	0.26	0.08	/	/
Todesco	0.3	0.17	/	/
Reinhardt	0.32	0.19	/	/
Verhoef (1997)	0.35	0.26	/	/
Kihara	0.38	0.22	/	/
Araujo	0.4	0.66	/	/
Malik	0.42	0.12	/	/
Thogersen	0.46	0.39	/	/
Izumi	0.49	0.09	/	/
Fernandez-A. (Males)	0.5	0.09	/	/
Szczeklik	0.77	0.14	/	/
Inbal	0.98	0.12	/	/
Gallagher	1.18	0.24	/	/
Mager	1.26	0.15	/	/
Ferrer-Antunes	1.30	0.30	/	/
Gulec	1.46	0.31	/	/
Dekou (Females)	/	/	0.9	0.33
Voutilainen	/	/	1.0	0.86
Chango (a)	/	/	1.4	0.64
Mazza	/	/	1.4	0.38
Chango (b)	/	/	1.7	2.56
Dekou (Males)	/	/	2.1	0.53
Kosokabe	/	/	2.1	2.08
Gonzalez Ordonez	/	/	2.1	0.84
Anderson (2000)	/	/	3.8	2.33
Kluijtmans (1996)	/	/	4.0	4.76
Deloughery	/	/	4.2	7.69
Fujimara	/	/	4.3	3.23
Arai	/	/	4.4	4.00
Rassoul	/	/	7.3	4.17
Yoo	/	/	8.1	0.89
D'Angelo	/	/	9.9	3.45

Table A – All 66 studies included in the meta-analyses on *MTHFR*, homocysteine (Hcy) & CHD

References for Table A

Abbate, R., Sardi, I., Pepe, G., et al. (1998) The high prevalence of thermolabile 5-10 methylenetetrahydrofolate reductase (MTHFR) in Italians is not associated to an increased risk for coronary artery disease (CAD). *Thrombosis and Haemostasis*; **79**:727-30.

Adams, M., Smith, P. D., Martin, D., Thompson, J.R., Lodwick, D., and Samani, N. J. (1996) Genetic analysis of thermolabile methylenetetrahydrofolate reductase as a risk factor for myocardial infarction. *QJM: monthly journal of the Association of Physicians*; **89**:437-44.

Anderson, J. L., King, G. J., Thomson, M. J., et al. (1997) A mutation in the methylenetetrahydrofolate reductase gene is not associated with increased risk for coronary artery disease or myocardial infarction. *Journal of the American College of Cardiology*; **30**:1206-11.

Anderson, J. L., Muhlestein, J. B., Horne, B. D., et al. (2000) Plasma homocysteine predicts mortality independently of traditional risk factors and C-reactive protein in patients with angiographically defined coronary artery disease. *Circulation*; **102**:1227-32.

Arai, K., Yamasaki, Y., Kajimoto, Y., et al. (1997) Association of methylenetetrahydrofolate reductase gene polymorphism with carotid arterial wall thickening and myocardial infarction risk in NIDDM. *Diabetes*; **46**:2102-4.

Araujo, F., Lopes, M., Goncalves, L., Maciel, M. J., and Cunha-Ribeiro, L. M. (2000) Hyperhomocysteinemia, MTHFR C677T genotype and low folate levels: a risk combination for acute coronary disease in a Portuguese population. *Thrombosis and Haemostasis*; **83**:517-8.

Ardissino, D., Mannucci, P. M., Merlini, P. A., et al. (1999) Prothrombotic genetic risk factors in young survivors of myocardial infarction. *Blood*; **94**:46-51.

Brugada, R., and Marian, A. J. (1997) A common mutation in methylenetetrahydrofolate reductase gene is not a major risk of coronary artery disease or myocardial infarction. *Atherosclerosis*; **128**:107-12.

Brulhart, M. C., Dussoix, P., Ruiz, J., Passa, P., Froguel, P., and James, R. W. (1997) The (Ala-Val) Mutation of methylenetetrahydrofolate reductase as a genetic risk factor for vascular disease in non-insulin dependent diabetic patients. *American Journal of Human Genetics*; **60**:228-9.

Chambers, J. C., Ireland, H., Thompson, E., et al. (2000) Methylenetetrahydrofolate reductase 677 C-->T mutation and coronary heart disease risk in UK Indian Asians. *Arteriosclerosis, Thrombosis & Vascular Biology*; **20**:2448-52.

Chango, A., Boisson, F., Barbe, F., et al. (2000) The effect of 677C-->T and 1298A-->C mutations on plasma homocysteine and 5,10-methylenetetrahydrofolate reductase activity in healthy subjects. *British Journal of Nutrition*; **83**:593-6.

Chango, A., Potier De Courcy, G., Boisson, F., et al. (2000) 5,10-methylenetetrahydrofolate reductase common mutations, folate status and plasma homocysteine in healthy French adults of the Supplementation en Vitamines et Mineraux Antioxydants (SU.VI.MAX) cohort. *British Journal of Nutrition*; **84**:891-6.

Chao, C. L., Tsai, H. H., Lee, C. M., et al. (1999) The graded effect of hyperhomocysteinemia on the severity and extent of coronary atherosclerosis. *Atherosclerosis*; **147**:379-86.

Christensen, B., Frosst, P., Lussier-Cacan, S., et al. (1997) Correlation of a common mutation in

the methylenetetrahydrofolate reductase gene with plasma homocysteine in patients with premature coronary artery disease. *Arteriosclerosis, Thrombosis & Vascular Biology*; 17:569-73.

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

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

Deloughery, T. G., Evans, A., Sadeghi, A., et al. (1996) Common mutation in methylenetetrahydrofolate reductase. Correlation with homocysteine metabolism and late-onset vascular disease. *Circulation*; 94:3074-8.

Dilley, A., Hooper, W. C., El-Jamil, M., Renshaw, M., Wenger, N. K., and Evatt, B. L. (2001) Mutations in the genes regulating methylene tetrahydrofolate reductase (MTHFR C-->T677) and cystathione beta synthase (CBS G-->A919, CBS T-- >c833) are not associated with myocardial infarction in African Americans. *Thrombosis Research*; 103:109-15.

Fernandez-Arcas, N., Dieguez-Lucena, J. L., Munoz-Moran, E., et al. (1999) The genotype interactions of methylenetetrahydrofolate reductase and rennin angiotensin system genes are associated with myocardial infarction. *Atherosclerosis*; 145:293-300.

Ferrer-Antunes, C., Palmeiro, A., Morais, J., Lourenco, M., Freitas, M., and Providencia, L. (1998) The mutation C677T in the methylene tetrahydrofolate reductase gene as a risk factor for myocardial infarction in the Portuguese population. *Thrombosis and Haemostasis*; 80:521-2.

Fowkes, F. G., Lee, A. J., Hau, C. M., Cooke, A., Connor, J. M., and Lowe, G. D. (2000) Methylene tetrahydrofolate reductase (MTHFR) and nitric oxide synthase (ecNOS) genes and risks of peripheral arterial disease and coronary heart disease: Edinburgh Artery Study. *Atherosclerosis*; 150:179-85.

Fujimura, H., Kawasaki, T., Sakata, T., et al. (2000) Common C677T polymorphism in the methylenetetrahydrofolate reductase gene increases the risk for deep vein thrombosis in patients with predisposition of thrombophilia. *Thrombosis Research*; 98:1-8.

Gallagher, P. M., Meleady, R., Shields, D. C., et al. (1996) Homocysteine and risk of premature coronary heart disease. Evidence for a common gene mutation. *Circulation*; 94:2154-8.

Gardemann, A., Weidemann, H., Philipp, M., et al. (1999) The TT genotype of the methylenetetrahydrofolate reductase C677T gene polymorphism is associated with the extent of coronary atherosclerosis in patients at high risk for coronary artery disease. *European Heart Journal*; 20:584-92.

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

Gonzalez Ordonez, A. J., Fernandez Alvarez, C. R., Rodriguez, J. M., Garcia, E. C., and Alvarez, M. V. (1999) Genetic polymorphism of methylenetetrahydrofolate reductase and venous thromboembolism: a case-control study. *Haematologica*; 84:190-1.

Gulec, S., Aras, O., Akar, E., et al. (2001) Methylenetetrahydrofolate reductase gene polymorphism and risk of premature myocardial infarction. *Clinical Cardiology*; **24**:281-4.

Hsu, L. A., Ko, Y. L., Wang, S. M., et al. (2001) The C677T mutation of the methylenetetrahydrofolate reductase gene is not associated with the risk of coronary artery disease or venous thrombosis among Chinese in Taiwan. *Human Heredity*; **51**:41-5.

Inbal, A., Freimark, D., Modan, B., et al. (1999) Synergistic effects of prothrombotic polymorphisms and atherogenic factors on the risk of myocardial infarction in young males. *Blood*; **93**:2186-90.

Izumi, M., Iwai, N., Ohmichi, N., Nakamura, Y., Shimoike, H., and Kinoshita, M. (1996) Molecular variant of 5,10-methylenetetrahydrofolate reductase is a risk factor of ischemic heart disease in the Japanese population. *Atherosclerosis*; **121**:293-4.

Kawashiri, M., Kajinami, K., Nohara, A., et al. (2000) Effect of common methylenetetrahydrofolate reductase gene mutation on coronary artery disease in familial hypercholesterolemia. *American Journal of Cardiology*; **86**:840-5.

Kihara, T., Abe, S., Saigo M, et al. (1997) Methylenetetrahydrofolate reductase gene polymorphism and premature myocardial infarction. *Circulation*; **96**(8S) Supplement:101-I.

Kim, C. H., Hwang, K. Y., Choi, T. M., Shin, W. Y., and Hong, S. Y. (2001) The methylenetetrahydrofolate reductase gene polymorphism in Koreans with coronary artery disease. *International Journal of Cardiology*; **78**:13-7.

Kluijtmans, L. A., Kastelein, J. J., Lindemans, J., et al. (1997) Thermolabile methylenetetrahydrofolate reductase in coronary artery disease. *Circulation*; **96**:2573-7.

Kluijtmans LA, van den Heuvel LP, Boers GH, et al. (1996) Molecular genetic analysis in mild hyperhomocysteinemia: a common mutation in the methylenetetrahydrofolate reductase gene is a genetic risk factor for cardiovascular disease. *American Journal of Human Genetics*; **58**:35-41.

Kosokabe, T., Okumura, K., Sone, T., et al. (2001) Relation of a common methylenetetrahydrofolate reductase mutation and plasma homocysteine with intimal hyperplasia after coronary stenting. *Circulation*; **103**:2048-54.

Ma, J., Stampfer, M. J., Hennekens, C. H., *et al.* (1996) Methylenetetrahydrofolate reductase polymorphism, plasma folate, homocysteine, and risk of myocardial infarction in US physicians. *Circulation*; **94**:2410-6.

Mager, A., Lalezari, S., Shohat, T., et al. (1999) Methylenetetrahydrofolate reductase genotypes and early-onset coronary artery disease. *Circulation*; **100**:2406-10.

Malik, N. M., Syrris, P., Schwartzman, R., et al. (1998) Methylenetetrahydrofolate reductase polymorphism (C-677T) and coronary artery disease. *Clinical Science (London, England: 1979)*; **95**:311-5.

Malinow, M. R., Nieto, F. J., Kruger, W. D., et al. (1997) The effects of folic acid supplementation on plasma total homocysteine are modulated by multivitamin use and methylenetetrahydrofolate reductase genotypes. *Arteriosclerosis, Thrombosis & Vascular Biology*; **17**:1157-62.

Mazza, A., Motti, C., Nulli, A., et al. (2000) Lack of association between carotid intima-media thickness and methylenetetrahydrofolate reductase gene polymorphism or serum homocysteine in

non-insulin-dependent diabetes mellitus. *Metabolism*; **49**:718-23.

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

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

Nakai, K., Fusazaki, T., Suzuki, T., et al. (2000) Genetic polymorphism of 5,10-methylenetetrahydrofolate increases risk of myocardial infarction and is correlated to elevated levels of homocysteine in the Japanese general population. *Coronary Artery Disease*; **11**:47-51.

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

Pinto, X., Vilaseca, M. A., Garcia-Giralt, N., et al. (2001) Homocysteine and the MTHFR 677C-->T allele in premature coronary artery disease. Case control and family studies. *European Journal of Clinical Investigation*; **31**:24-30.

Rassoul, F., Richter, V., Janke, C., et al. (2000) Plasma homocysteine and lipoprotein profile in patients with peripheral arterial occlusive disease. *Angiology*; **51**:189-96.

Reinhardt, D., Sigusch, H. H., Vogt, S. F., Farker, K., Muller, S., and Hoffmann, A. (1998) Absence of association between a common mutation in the methylenetetrahydrofolate reductase gene and the risk of coronary artery disease. *European Journal of Clinical Investigation*; **28**:20-3.

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

Schwartz, S. M., Siscovick, D. S., Malinow, M. R., et al. (1997) Myocardial infarction in young women in relation to plasma total homocysteine, folate, and a common variant in the methylenetetrahydrofolate reductase gene. *Circulation*; **96**:412-7.

Szczeklik, A., Sanak, M., Jankowski, M., et al. (2001) Mutation A1298C of methylenetetrahydrofolate reductase: risk for early coronary disease not associated with hyperhomocysteinemia. *American Journal of Medical Genetics*; **101**:36-9.

Thogersen, A. M., Nilsson, T. K., Dahlen, G., et al. (2001) Homozygosity for the C677-->T mutation of 5,10- methylenetetrahydrofolate reductase and total plasma homocyst(e) ine are not associated with greater than normal risk of a first myocardial infarction in northern Sweden. *Coronary Artery Disease*; **12**:85-90.

Todesco, L., Angst, C., Litynski, P., Loehrer, F., Fowler, B., and Haefeli, W. E. (1999) Methylenetetrahydrofolate reductase polymorphism, plasma homocysteine and age. *European Journal of Clinical Investigation*; **29**:1003-9.

Tokgozoglu, S. L., Alikasifoglu, M., Unsal, I., et al. (1999) Methylene tetrahydrofolate reductase genotype and the risk and extent of coronary artery disease in a population with low plasma folate. *Heart*; **81**:518-22.

Tsai, M. Y., Welge, B. G., Hanson, N. Q., et al. (1999) Genetic causes of mild hyperhomocysteinemia in patients with premature occlusive coronary artery diseases. *Atherosclerosis*; **143**:163-70.

van Bockxmeer, F. M., Mamotte, C. D., Vasikaran, S. D., Taylor, R. R. (1997) Methylenetetrahydrofolate reductase gene and coronary artery disease. *Circulation*; **95**:21-3.

Verhoef, P., Kok, F. J., Kluijtmans, L. A., et al. (1997) The 677C-->T mutation in the methylenetetrahydrofolate reductase gene: associations with plasma total homocysteine levels and risk of coronary atherosclerotic disease. *Atherosclerosis*; **132**:105-13.

Verhoef, P., Rimm, E. B., Hunter, D. J., et al. (1998) A common mutation in the methylenetetrahydrofolate reductase gene and risk of coronary heart disease: results among U.S. men. *Journal of the American College of Cardiology*; **32**:353-9.

Verhoeff, B. J., Trip, M. D., Prins, M. H., Kastelein, J. J., and Reitsma, P. H. (1998) The effect of a common methylenetetrahydrofolate reductase mutation on levels of homocysteine, folate, vitamin B12 and on the risk of premature atherosclerosis. *Atherosclerosis*; **141**:161-6.

Voutilainen, S., Lakka, T. A., Hamelahti, P., Lehtimaki, T., Poulsen, H. E., and Salonen, J. T. (2000) Plasma total homocysteine concentration and the risk of acute coronary events: the Kuopio Ischaemic Heart Disease Risk Factor Study. *Journal of Internal Medicine*; **248**:217-22.

Wilcken, D. E., Wang, X. L., Sim, A. S., and McCredie, R. M. (1996) Distribution in healthy and coronary populations of the methylenetetrahydrofolate reductase (MTHFR) C677T mutation. *Arteriosclerosis, Thrombosis & Vascular Biology*; **16**:878-82.

Yoo, J-H. (2001) A thermolabile variant of methylenetetrahydrofolate reductase is a determinant of hyperhomocyst(e)inemia in the elderly. *Annals of the New York Academy of Science*; **928**:344.

Zheng, Y. Z., Tong, J., Do, X. P., Pu, X. Q., and Zhou, B. T. (2000) Prevalence of methylenetetrahydrofolate reductase C677T and its association with arterial and venous thrombosis in the Chinese population. *British Journal of Haematology*; **109**:870-4.

Table B – Results of sensitivity analyses for Model A, performed by specifying different sets of values for the covariance matrix of the Wishart prior distribution (see WinBUGS code for Model A). Results are expressed as median with 95% CrI and compared with those of the main analysis

MODEL		PARAMETER				
		θ (95%CrI)	μ_y (95%CrI)	τ_x (95%CrI)	τ_y (95%CrI)	ρ (95%CrI)
Main analysis		0.078 (0.029 to 0.138)	2.607 (1.850 to 3.446)	0.070 (0.002 to 0.251)	5.272 (1.922 to 14.730)	0.691 (0.105 to 0.914)
Sensitivity analyses	Variances halved, same correlation (0.5)	0.080 (0.031 to 0.139)	2.537 (1.813 to 3.348)	0.070 (0.002 to 0.250)	5.329 (1.930 to 14.950)	0.762 (0.171 to 0.943)
	Variances doubled, same correlation (0.5)	0.075 (0.026 to 0.136)	2.702 (1.891 to 3.579)	0.070 (0.002 to 0.252)	5.207 (1.918 to 14.490)	0.613 (0.051 to 0.875)
	Same variances, correlation of 0	0.076 (0.026 to 0.139)	2.628 (1.849 to 3.494)	0.069 (0.002 to 0.251)	5.260 (1.924 to 14.690)	0.499 (-0.161 to 0.842)
	Same variances, correlation of 0.99	0.084 (0.036 to 0.132)	2.555 (1.868 to 3.300)	0.071 (0.002 to 0.252)	5.284 (1.917 to 14.790)	0.984 (0.829 to 0.997)

Table C – Results of sensitivity analyses for Model A and Model B, performed by specifying different vague prior distributions for the heterogeneity terms (see WinBUGS codes). Results are expressed as median with 95% CrI and compared with those of the main analysis

MODEL	Prior distributions for the heterogeneities	PARAMETER				
		θ (95%CrI)	μ_y (95%CrI)	τ_x (95%CrI)	τ_y (95%CrI)	ρ (95%CrI)
Model A	Gamma(0.001,0.001)	0.078 (0.029 to 0.138)	2.607 (1.850 to 3.446)	0.070 (0.002 to 0.251)	5.272 (1.922 to 14.73)	0.691 (0.105 to 0.914)
	Half-Normal(0,1) on standard deviation	0.079 (0.027 to 0.138)	2.653 (1.934 to 3.449)	0.101 (0.010 to 0.301)	3.647 (1.493 to 7.955)	0.689 (0.098 to 0.911)
	Uniform(0,2) on standard deviation	0.078 (0.026 to 0.137)	2.673 (1.971 to 3.437)	0.101 (0.010 to 0.307)	3.178 (1.594 to 3.958)	0.689 (0.095 to 0.913)
Model B	Gamma(0.001,0.001)	0.083 (0.030 to 0.148)	2.682 (1.914 to 3.533)	0.102 (0.032 to 0.253)	3.610 (1.686 to 7.426)	*Not in the model*
	Half-Normal(0,1) on standard deviation	0.083 (0.027 to 0.146)	2.650 (1.923 to 3.426)	0.116 (0.039 to 0.269)	3.027 (1.508 to 5.822)	*Not in the model*
	Uniform(0,2) on standard deviation	0.082 (0.026 to 0.147)	2.640 (1.916 to 3.386)	0.120 (0.041 to 0.280)	3.014 (1.589 to 3.945)	*Not in the model*

BIBLIOGRAPHY

Akey, J. M., Zhang, K., Xiong, M., Doris, P., Jin, L. (2001) The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *American Journal of Human Genetics*; **68**:1447-56.

Altman, D. G. and Bland, J. M. (2003) Interaction revisited: the difference between two estimates. *British Medical Journal*; **326**:219.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*; **91**:444-55.

Arends, L. R., Voko, Z., and Stijnen T. (2003) Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine*; **22**:1335-53.

Assendelft, W. J., Koes, B. W., Knipschild, P. G., and Bouter, L. M. (1995) The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA*; **274**:1942-8.

Assmann, S. F., Pocock, S. J., Enos, L. E., and Kasten, L. E. (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*; **355**:1064-9.

Atkins, D., Best, D., Briss, P. A., et al. - for the GRADE Working Group (2004) Grading quality of evidence and strength of recommendations. *British Medical Journal*; **328**:1490-1497.

Attia, J., Thakkinstian, A., and D'Este, C. (2003) Meta-analyses of molecular association studies: Methodologic lessons for genetic epidemiology. *Journal of Clinical Epidemiology*; **56**:297-303.

Ayres, K. L. and Balding, D. J. (1998) Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity*; **80**:769-77.

Bailey, L. B. and Gregory, J. F. (1999) Polymorphisms of methylenetetrahydrofolate reductase and other enzymes: metabolic significance, risks and impact on folate requirement. *Journal of Nutrition*; **129**:919-22.

Bardoczy, Z., Gyorffy, B., Kocsis, I., and Vasarhelyi B. (2004) Re-calculated Hardy-Weinberg values in papers published in Atherosclerosis between 1995 and 2003. *Atherosclerosis*; **173**:141-143.

Begg, C., Cho, M., Eastwood, S., et al. (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*; **276**:637-9.

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*. Second edition. New York, U.S.A.: Springer-Verlag.

Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian theory*. Chichester, U. K.: John Wiley & Sons Ltd.

Berkey, C. S., Anderson, J. J., and Hoaglin, D. C. (1996) Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine*; **15**:537-57.

Bero, L. and Rennie, D. (1995) The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA*; **274**:1935-8.

Best, N. G., Spiegelhalter, D. J., Thomas, A., and Brayne, C. E. G. (1996) Bayesian Analysis of Realistically Complex Models. *Journal of the Royal Statistical Society*, Series A;**159**: 323-42.

Bland, J. M. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*; **1**:307-10.

Bogardus, S. T. Jr., Concato, J., and Feinstein, A. R. (1999) Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *JAMA*; **281**:1919-26.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., et al. (2003a) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Annals of Internal Medicine*; **138**:40-4.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., et al. (2003b) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine*; **138**:W1-12.

Box, G. E. P. and Tiao, G. C. (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., and Peters, T. J. (2004) Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology*; **57**:229-36.

Brooks, S. P. (1998) Markov chain Monte Carlo method and its applications. *Journal of the Royal Statistical Society*, Series D; **47**:69-100.

Cardon, L. R. and Bell, J. I. (2001) Association study designs for complex diseases. *Nature Review Genetics*; **2**:91-9.

Cardon, L. R. and Palmer, L. J. (2003) Population stratification and spurious allelic association. *Lancet*; **361**:598-604.

Carlin, B. P. (1996) *Hierarchical longitudinal modelling*. In Markov Chain Monte Carlo In Practice, Gilks WR, Richardson S, Spiegelhalter DJ (eds), London: Chapman & Hall; 303-19.

Casas, J. P., Bautista, L. E., Smeeth, L., Sharma, P., and Hingorani, A. D. (2005) Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet*; **365**:224-32.

Centres for Disease Control and Prevention (1999) *The Human Genome Epidemiology Network (HuGE Net)*; http://www.cdc.gov/genetics/hugenet/default.htm.

Clayton, D. and McKeigue, P. M. (2001) Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet*; **358**:1356-60.

Cheng, K. F and Chen, J. H. (2005) Bayesian models for population-based case-control studies when the population is in Hardy-Weinberg equilibrium. *Genetic Epidemiology*; **28**:183-92.

Colhoun, H. M., McKeigue, P. M., and Davey Smith, G. (2003) Problems of reporting genetic associations with complex outcomes. *Lancet*; **361**:865-72.

Collins, F. S. and McKusick, V. A. (2001) Implications of the Human Genome Project for medical science. *JAMA*; **285**:540-4.

Cook, D. J., Sackett, D. L., and Spitzer, W. O. (1995) Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *Journal of Clinical Epidemiology*; 48:167-71.

Cowles, M. K. and Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*; **91**:883-904.

Cui, J. (2000) Hardy-Weinberg equilibrium test in case-control studies. *Stata Technical Bulletin*; **57**:17-9.

Davey Smith, G. and Ebrahim, S. (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*; **32**:1-22.

Day, I. N., Gu, D., Ganderton, R. H., Spanakis, E., and Ye, S. (2001) Epidemiology and the genetic basis of disease. *International Journal of Epidemiology*; **30**:661-7.

Den Heijer, M., Lewington, S., and Clarke, R. (2005) Homocysteine, *MTHFR* and risk of venous thrombosis: a meta-analysis of published epidemiological studies. *Journal of Thrombosis and Haemostasis*; **3**:292-9.

Devlin, B., Risch, N. and Roeder, S. (1991) Forensic DNA tests and Hardy-Weinberg equilibrium. *Science*; **253**: 1039-41.

Des Jarlais, D. C., Lyles, C., Crepaz, N., and TREND Group. (2004) Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health*; **94**:361-6.

Dixon, D. O. and Simon, R. (1991) Bayesian subset analysis. *Biometrics*; **47**: 871-81.

Dixon, D. O. and Simon, R. (1992) Bayesian subset analysis in a colorectal cancer clinical trial. *Statistics in Medicine*; **11**:13-22.

Donner, A. (1982) A Bayesian approach to the interpretation of subgroup results in clinical trials. *Journal of Chronic Diseases*; **35**:429-35.

Drews, J. and Ryser, S. (1997) The role of innovation in drug development. *Nature Biotechnology*; **15**: 1318-9.

Edland, S. D., Slager, S., and Farrer, M. (2004) Genetic association studies in Alzheimer's disease research: challenges and opportunities. *Statistics in Medicine*; **23**:169-78.

Egger, M., Smith, G. D., and Altman, D. G. (2001) *Systematic Reviews in Health Care - Meta-analysis in context*. London, U. K.: BMJ Books.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*; 315:629-34.

Elston, R. C. and Forthofer, R. (1977) Testing for Hardy-Weinberg equilibrium in small samples. *Biometrics*; 33:536-42

Emigh, T. H. (1980) A comparison of tests for Hardy-Weinberg Equilibrium. *Biometrics*; 36:627-642.

Fieller, E. (1954) Some problems in interval estimation. *Journal of the Royal Statistical Society, B*; 16:175.

Fujisawa, T., Ikegami, H., Kawaguchi, Y., et al. (1998) Meta-analysis of association of insertion/deletion polymorphism of angiotensin I-converting enzyme gene with diabetic nephropathy and retinopathy. *Diabetologia*; 41:47-53.

Freedman, M. L., Reich, D., Penney, K. L., et al. (2004) Assessing the impact of population stratification on genetic association studies. *Nature Genetics*; 36:388-93.

Gambaro, G., Anglani, F., and D'Angelo, A. (2000) Association studies of genetic polymorphisms and complex disease. *Lancet*; 355:308-11.

Gardner, M. J. and Altman, D. G. (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal Clinical Research Edition*; 292:746-50.

Garthwaite, P. H. and Dickey, J. M. (1988) Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society*, Series B; 50:462-74.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005) Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*; **100**:680-701.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*; **85**:398-409.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995) Efficient parameterisations for normal linear mixed models. *Biometrika*; **82**:479-88.

Gelman, A. (2004) *Prior distributions for variance parameters in hierarchical models.* Technical report. Department of Statistics, Columbia University. Available at http://www.stat.columbia.edu/~gelman/research/unpublished/tau7.pdf

Geyer, C. J. (1996) *Estimation and optimization of functions.* In: Markov Chain Monte Carlo In Practice, Gilks WR, Richardson S, Spiegelhalter DJ (eds), London: Chapman & Hall; 241-58.

Gilks, W. R., Richardson, S. , and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice.* London, U. K.: Chapman and Hall.

Gokhale, D. V. and Press, S. J. (1982) Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Journal of the Royal Statistical Society*, Series A; **145**:237-49.

Gomes, I., Collins, A., Lonjou, C., et al. (1999) Hardy-Weinberg quality control. *Annals of Human Genetics*; **63**: 535-8.

Gordon, D., Finch, S. J., Nothnagel, M., and Ott, J. (2002) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Human Heredity*; **54**:22-33

Gould, W. and Sribney, W. (1999) *Maximum likelihood estimation with Stata*. College Station, Texas: Stata Press.

Gorroochurn, P., Hodge, S. E., Heiman, G., and Greenberg, D. A. (2004) Effect of population stratification on case-control association studies. II. False-positive rates and their limiting behaviour as number of subpopulations increases. *Human Heredity*; **58**:40-8.

Greenland, S. (2000) An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*; **29**:722-9.

Greenland, S. and Robins, J. (1994) Invited commentary: ecologic studies – biases, misconceptions and counterexamples. *American Journal of Epidemiology*; **139**:747-760.

Gyorffy, B., Kocsis, I., and Vasarhelyi, B. (2004a) Biallelic genotype distributions in papers published in Gut between 1998 and 2003: altered conclusions after recalculating the Hardy-Weinberg equilibrium. *Gut*; **53**:614-615.

Gyorffy, B., Kocsis, I., and Vasarhelyi, B. (2004b) Missed calculations and new conclusions: re-calculation of genotype distribution data published in Journal of Investigative Dermatology, 1998-2003. *Journal of Investigative Dermatology*; **122**:644-646.

Hani, E. H., Boutin, P., Durand, E., et al. (1998) Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia*; **41**:1511-5.

Harbour, R. and Miller, J. (2001) A new system for grading recommendations in evidence based guidelines. *British Medical Journal*; **323**:334-6.

Hardin, J. W. and Carroll, R. J. (2003) Variance estimation for the instrumental variables approach to measurement error in generalized linear models. *The Stata Journal*; 3:342-50.

Hardy, R. J. and Thompson, S. G. (1996) A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*; 15:619-29.

Hartl, D. L. and Clark, A. G. (1997) *Principles of population genetics*. Third edition. Sunderland, MA, U.S.A.: Sinauer Associates.

Hedges, L. V. and Vevea, J. L. (1998) Fixed- and random-effects models in meta-analysis. *Psychological Methods*; 3:486-504.

Heiman, G. A., Hodge, S. E., Gorroochurn, P., Zhang, J., and Greenberg, D. A. (2004) Effect of population stratification on case-control association studies. I. Elevation in false positive rates and comparison to confounding risk ratios (a simulation study). *Human Heredity*; 58:30-9.

Hernandez, J. L. and Weir, B. S. (1989) A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics*; 45:53-70.

Higgins, J. P. and Thompson, S. G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21:1539-58.

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal*; 327:557-60.

Higgins, J. P. and Whitehead, A. (1996) Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*; 15:2733-49.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002) A comprehensive review of genetic association studies. *Genetics in Medicine*; 4:45-61.

Hosking, L., Lumsden, S., Lewis, K., et al. (2004) Detection of genotyping error by Hardy-Weinberg equilibrium testing. *European Journal of Human Genetics*; **12**:395-9.

Homocysteine Lowering Trialists' Collaboration (1998) Lowering blood homocysteine with folic acid based supplements: meta- analysis of randomised trials. *British Medical Journal*; **316**:894-8.

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

Hunt, D. L., Jaeschke, R., and McKibbon, K. A. (2000) Users' guides to the medical literature: XXI. Using electronic health information resources in evidence-based practice. Evidence-Based Medicine Working Group. *JAMA*; **283**:1875-9.

Ioannidis, J. P. A. (2003) Genetic associations: false or true? *Trends in Molecular Medicine*; **9**:135-8.

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., and Contopoulos-Ioannidis, D. G. (2001) Replication validity of genetic association studies. *Nature Genetics*; **29**:306-9.

Ioannidis, J. P., Ntzani, E. E., and Trikalinos, T. A. (2004) "Racial" differences in genetic effects for complex diseases. *Nature Genetics*; **36**:1312-8.

Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., and Contopoulos-Ioannidis, D. G. (2003) Genetic associations in large versus small studies: an empirical assessment. *Lancet*; **361**:567-71.

Irony, T. Z. and Singpurwalla, N. D. (1997) Noninformative priors do not exist: A discussion with Jose M. Bernardo. *Journal of Statistical Inference and Planning*; **65**:159-89.

Jadad, A. R., Cook, D. J., and Browman, G. P. (1997) A guide to interpreting discordant systematic reviews. *Canadian Medical Association Journal*; **156**:1411-6.

Jadad, A. R., Cook, D. J., Jones, A., et al. (1998) Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA*; **280**:278-80.

Jadad, A. R. and McQuay, H. J. (1996) Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *Journal of Clinical Epidemiology*; **49**:235-43.

Jadad, A. R., Moher, M., Browman, G. P., et al. (2000) Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ*; **320**:537-40 [erratum in BMJ 2000; **320**:984].

Juengel, J. L., Quirke, L. D., Tisdall, D. J., Smith, P., Hudson, N. L., McNatty, K. P. (2000) Gene expression in abnormal ovarian structures of ewes homozygous for the inverdale prolificacy gene. *Biology of Reproduction*; **62**:1467-78.

Kang, S. J., Gordon, D., and Finch, S. J. (2004) What SNP genotyping errors are most costly for genetic association studies? *Genetic Epidemiology*; **26**:132-41.

Kass, R. E. and Wasserman, L. (1996) The selection of prior distributions by formal rules. *Journal of the American Statistical Association*; **91**:1343-70.

Katan, M. B. (1986) Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*; **1**:507-8.

Kato, N., Sugiyama, T., Morita, H., Kurihara, H., Yamori, Y., and Yazaki, Y. (1999) Angiotensinogen gene and essential hypertension in the Japanese: extensive association study and meta-analysis on six reported studies. *Journal of Hypertension*; **17**:757-63.

Keavney, B. (2000) Genetic association studies in complex diseases. *Journal of Human Hypertension*; **14**:361-7.

Keavney, B. (2002) Genetic epidemiological studies of coronary heart disease. *International Journal of Epidemiology*; **31**:730-6.

Keavney, B., Youngman, L., Palmer, A., et al. (2000) Large-Scale Test of Hypothesised Associations between Polymorphisms of Lipid-Related Genes and Myocardial Infarction in About 5000 Cases and 6000 Controls. *Circulation*; **102**(suppl II):852.

Kendall, M., and Stuart A. (1973 a) *Distribution theory*. In: The advanced theory of statistics. Volume 1. London, U. K.: Griffin.

Kendall, M. and Stuart A. (1973 b) *Inference and relationship*. In: The advanced theory of statistics. Volume 2. London, U. K.: Griffin.

Khoury, M. J., Beaty, T. H., and Cohen, B. H. (1993) *Fundamentals of genetic epidemiology*. New York, USA: Oxford University Press.

Khoury, M. J., Little, J., and Burke, W. (2004) *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease*. New York, USA: Oxford University Press.

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

Kocsis, I., Gyorffy, B., Nemeth, E., and Vasarhelyi, B. (2004a) Examination of Hardy-Weinberg equilibrium in papers of Kidney International: an underused tool. *Kidney International*; **65**:1956-1958.

Kocsis, I., Vasarhelyi, B., Gyorffy, A., and Gyorffy, B. (2004b) Reanalysis of genotype distributions published in Neurology between 1999 and 2002. *Neurology*; **63**:357-8.

Lambert, P. C., Sutton, A. J., Abrams, K. R., and Jones, D. R. (2002) A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*; **55**:86-94.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., Jones, D. R. (2005) How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*; **24**:2401-28.

Lee, W. C. (2003) Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibrium in a gene bank of affected individuals. *American Journal of Epidemiology*; **158**:397-400.

LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., and Derderian, F. (1997) Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine*; **337**:536-42.

Lindley, D. (1988) Statistical inference concerning Hardy-Weinberg equilibrium. *Bayesian Statistics*; **3**:307-26.

Little, J., Bradley, L., Bray, M. S., et al. (2002) Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *American Journal of Epidemiology*; **156**:300-10.

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., and Hirschhorn, J. N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*; **33**:177-82.

Makni, H., Franco, E. L., Kaiano, J., et al. (2000) P53 polymorphism in codon 72 and risk of human papillomavirus-induced cervical cancer: effect of inter-laboratory variation. *International Journal of Cancer*; **87**:528-33.

Marsh, D. G. and Huang, S. K. (1991) Molecular genetics of human immune responsiveness to pollen allergens. *Clinical & Experimental Allergy*; **21**(Suppl 1):168-72.

Midgley, J. P., Matthew, A. G., Greenwood, C. M., et al. (1996) Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. *JAMA*; **275**:1590-7.

Minelli, C.*, Thompson, J. R.*, Abrams, K. R., Lambert, P.C. (2005a) Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Statistics in Medicine*; **24**:3845-61. (*joint first authors)

Minelli, C.*, Thompson, J. R.*, Abrams, K. R., Thakkinstian, A., and Attia, J. (2005b) The choice of genetic model in the meta-analysis of molecular association studies. *International Journal of Epidemiology*; **34**:1319-28. (*joint first authors)

Minelli, C.*, Thompson, J. R*, Tobin, M. D., and Abrams, K. R. (2004) An integrated approach to the meta-analysis of genetic association studies using Mendelian Randomization. *American Journal of Epidemiology*; **160**:445-52. (* joint first authors)

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F. (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet*; **354**:1896-1900.

Montoya-Delgado, L. E., Irony, T. Z., de B. Pereira, C. A., and Whittle, M. R. (2001) An unconditional exact test for the Hardy-Weinberg equilibrium law: sample-space ordering using the Bayes factor. *Genetics*; **158**:875-83.

Munafo, M. R. and Flint, J. (2004) Meta-analysis of genetic association studies. *Trends in Genetics*; **20**:439-44.

Munafo, MR, Clark, T and Flint, J. (2005) Promise and pitfalls in the meta-analysis of genetic association studies: a response to Sen and Schinka. *Molecular Psychiatry*; June 28 [Epub ahead of print].

Nam, I. S., Mengersen, K., and Garthwaite, P. (2003) Multivariate meta-analysis. *Statistics in Medicine*; **22**:2309-33.

National Research Council (1996) *The Evaluation of forensic DNA evidence*. Washington, DC, U.S.A.: National Academy Press.

Nemeth, E., Vasarhelyi, B., Gyorffy, B., and Kocsis, I. (2004) Unreported deviations of genotype distributions from Hardy-Weinberg equilibrium in articles published in Critical Care Medicine between 1999 and 2003. *Critical Care Medicine*; **32**:1431-1433.

O'Hagan, A. (1994) *Bayesian Inference*. In: Kendall's advanced theory of statistics. Volume 2B. London: Hodder Arnold.

Oxman, A. D. and Guyatt, G. H. (1992) A consumer's guide to subgroup analyses. *Annals of Internal Medicine*; **116**:78-84.

Perneger, T. V. (1998) What's wrong with Bonferroni adjustments. *British Medical Journal*; **316**:1236-8.

Pocock, S. J. and Hughes, M. D. (1990) Estimation issues in clinical trials and overviews. *Statistics in Medicine*; **9**:657-71.

Prentice, R. L. and Pyke, R. (1979) Logistic Disease Incidence Models and Case-Control Studies. *Biometrika*; **66**:403-11.

Rader, J. I. (2002) Folic acid fortification, folate status and plasma homocysteine. *Journal of Nutrition*; **132**:2466S-70S.

Raudenbush, S. W., Becker, B. J., and Kalaian, H. (1988) Modeling multivariate effect sizes. *Psychological Bulletin*; **103**:111-20.

Risch, N. J. (2000) Searching for genetic determinants in the new millennium. *Nature*; **405**:847-56.

Robinson, K. A. and Dickersin, K. (2002) Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *International Journal of Epidemiology*; **31**:150-3.

Rogatko, A, Slifker, M. J., and Babb, J. S. (2002) Hardy-Weinberg equilibrium diagnostics. *Theoretical Population Biology*; **62**:251-257.

Rothman, K. J. (1978) A show of confidence. *New England Journal of Medicine*; **299**:1362-3.

Rothman, K. J. and Greenland, S. (1998) *Modern epidemiology*. Second edition. Philadelphia, U.S.A.: Lippincott Williams & Wilkins.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., and Chalmers, T. C. (1987) Meta-analyses of randomized controlled trials. *New England Journal of Medicine*; **316**: 450-5.

Sacks, H. S., Reitman, D., Pagano, D., and Kupelnick, B. (1996) Meta-analysis: an update. *Mount Sinai Journal of Medicine* **63**:216-24.

Sasazuki, T., Nishimura, Y., Muto, M., and Ohta, N. (1983) HLA-linked genes controlling immune response and disease susceptibility. *Immunological Reviews*; **70**:51-75.

Salanti, G., Amountza, G., Ntzani, E. E., and Ioannidis, J. P.A. (2005) Hardy-Weinberg Equilibrium in Genetic Association Studies: An Empirical Evaluation of Reporting, Deviations, and Power. *European Journal of Human Genetics*; **13**:840-8.

Salanti, G., Sanderson, S., and Higgins, J. P. (2005) Obstacles and opportunities in meta-analysis of genetic association studies. *Genetics in Medicine* 2005; **7**:13-20.

Sasieni, P. D. (1997) From genotypes to genes: doubling the sample size. *Biometrics*; **53**: 1253-61.

Seaman, S. R. and Richardson, S. (2004) Equivalence of prospective and retrospective models in the Bayesian analysis of case-control studies. *Biometrika*; **91**:15-25.

Sethi, A. A., Nordestgaard, B. G., and Tybjaerg-Hansen, A. (2003) Angiotensinogen gene polymorphism, plasma angiotensinogen, and risk of hypertension and ischemic heart disease: a meta-analysis. *Arteriosclerosis, Thrombosis & Vascular Biology*; **23**:1269-75.

Schaid, D. J. and Jacobsen, S. J. (1999) Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. *American Journal of Epidemiology*; **149**:706-11.

Sham, P. (2001) *Statistics in Human Genetics*. London, U. K.: Arnold Publishers.

Sharp, S. J. and Thompson, S. G. (2000) Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine*; **19**:3251-74

Shea, B., Dube', C., and Moher, D. (2001) Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context, pp 122-39. London: BMJ Publishing Group.

Shea, B., Moher, D., Graham, I., Pham, B., and Tugwell, P. (2002) A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. *Evaluation & the Health Professions*; **25**:116-29.

Shoemaker, J., Painter, I., and Weir, B. S. (1998) A Bayesian characterization of Hardy-Weinberg disequilibrium. *Genetics*; **149**:2079-2088.

Smith, B. J. (2004) *Bayesian Output Analysis Program* (BOA). Version 1.1 user's manual.

Smith, G. D., Song, F., and Sheldon, T. A. (1993) Cholesterol lowering and mortality: the importance of considering initial level of risk. *British Medical Journal*; **306**:1367-73.

Snedecor, G. W. and Cochran, W. G. (1989) *Statistical Methods*. Eighth edition. Ames, Iowa, U.S.A: Iowa State Press.

Spiegelhalter, D. J. (1998) Bayesian graphical modelling: a case study in monitoring health outcomes. *Journal of the Royal Statistical Society, Series C*; **47**:115-33.

Spiegelhalter, D. J. (2001) Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine*; **20**:435-52.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, U. K.: John Wiley & Sons Ltd.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*; **64**:583-639.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R., and Abrams, K. R. (1999) Methods in health service research. An introduction to bayesian methods in health technology assessment. *British Medical Journal*; **319**:508-12.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2004) *WinBUGS User Manual*. Version 1.4.1. MRC Biostatistics Unit; Cambridge.

Sterne, J. A., Egger, M., and Smith, G. D. (2001) Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*; **323**:101-5.

Stewart, L. A. and Tierney, J. F. (2002) To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions*; **25**:76-97.

Strachan, T. and Read, A. P. (1999) *Human molecular genetics*. Second edition. Oxford, U. K.: BIOS Scientific Publishers Ltd.

Stroup, D. F., Berlin, J. A., Morton, S. C., et al. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*; **283**:2008-12.

Suarez-Almazor, M. E., Belseck, E., Homik, J., Dorgan, M., and Ramos-Remus, C. (2000) Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials*; **21**:476-87.

Sutton, A. J. and Abrams, K. R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*; **10**:277-303.

Sutton, A. J., Abrams, K. R., and Jones, D. R. (2002) Generalized synthesis of evidence and the threat of dissemination bias. The example of electronic fetal heart rate monitoring (EFM). *Journal of Clinical Epidemiology*; **55**:1013-24.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song F. (2000) *Methods for meta-analysis in medical research*. Chichester, U. K.: John Wiley & Sons Ltd.

Sweeting, M. J., Sutton, A. J., and Lambert, P. C. (2004) What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of rare events. *Statistics in Medicine*; **23**:1351-75.

Thakkinstian, A., McEvoy, M., Minelli, C., et al. (2005) Systematic review and meta-analysis of the association between . 2-adrenoceptor polymorphisms and asthma: a HuGE review. *American Journal of Epidemiology*; **162**:201-11.

The Lancet (2003) In search of genetic precision. *Lancet*; **361**:357.

Thomas, D. C. (2004) *Statistical methods in genetic epidemiology*. New York, USA: Oxford University Press.

Thomas, D. C. and Witte, J. S. (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiology, Biomarkers & Prevention*; **11**:505-12.

Thompson, D., Witte, J. S., Slattery, M., and Goldgar, D. (2004) Increased power for case-control studies of single nucleotide polymorphisms through incorporation of family history and genetic constraints. *Genetic Epidemiology*; **27**:215-24.

Thompson, E. A. (2002) *Human Genetics, Overview*. In: Palmer L, ed. Biostatistical Genetics and Genetic Epidemiology. Chichester, U. K.: John Wiley & Sons Ltd.: 386-90.

Thompson, J. R., Minelli, C., Abrams, K. R., Tobin, M. D., and Riley, R. D. (2005) Meta-analysis of genetic studies using Mendelian randomization – a multivariate approach. *Statistics in Medicine*; **24**:2241-54.

Thompson, J. R., Tobin, M. D., Minelli, C. (2003) On the accuracy of estimates of the effect of phenotype on disease derived from Mendelian randomisation studies. Technical Report 2003_GE1 - available at http://www.prw.le.ac.uk/research/HCG/getechrep.html.

Thompson, S. G. (1994) Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*; **309**:1351-5.

Thompson, S. G. and Pocock, S. J. (1991) Can meta-analyses be trusted? *Lancet*; **338**:1127-30.

Tobin, M. D., Minelli, C., Burton, P. R., Thompson, J. R. (2004) The development of Mendelian randomisation: from hypothesis test to "Mendelian deconfounding". *International Journal of Epidemiology*; **33**:26-9.

Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001) Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine*; **20**: 453-72.

van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*; **21**:589-624.

van Houwelingen, H. C., Zwinderman, K. H., and Stijnen T. (1993) A bivariate approach to meta-analysis. *Statistics in Medicine*; **12**:2273-84.

Wacholder, S., Rothman, N., and Caporaso, N. (2002) Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiology, Biomarkers & Prevention; **11**:513-20.

Wald, D. S., Law, M., and Morris, J. K. (2002) Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *British Medical Journal*; **325**:1202.

Weir, B. S. (1996) *Genetic data analysis II: methods for discrete population genetic data*. Second edition. Sunderland, MA, U.S.A.: Sinauer Associates.

Wittke-Thompson, J. K., Pluzhnikov, A., Cox, N.J. (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. *American Journal of Human Genetics*; **76**:967-86.

Wheeler, J. G., Keavney, B. D., Watkins, H., Collins, R., and Danesh, J. (2004) Four paraoxonase gene polymorphisms in 11212 cases of coronary heart disease and 12786 controls: meta-analysis of 43 studies. *Lancet*; **363**:689-95.

Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., and Kleijnen, J. (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*; **3**:25.

Wijsman, E. M. (2002) *Mendel's Laws*. In: Elston R, Olson JM, Palmer LJ, editors. Biostatistical Genetics and Genetic Epidemiology. Chichester, U. K.: John Wiley & Sons Ltd.: 527-9.

Williams, J., Spurlock, G., Holmans, P., et al. (1998) A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia. *Molecular Psychiatry*; **3**:141-9.

Wittke-Thompson, J. K., Pluzhnikov, A., and Cox, N. J. (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. *American Journal of Human Genetics*; **76**:967-86.

Xu, J., Turner, A., Little, J., Bleecker, E. R., and Meyers, D. A. (2002) Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Human Genetics*; **111**:573-4.

Youngman, L., Keavney, B., Palmer, A., et al. (2000) Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by "Mendelian randomisation". *Circulation*; **102**(suppl II):31-32.

Zaykin, D., Zhivotovsky, L., and Weir, B. S. (1995) Exact tests for association between alleles at arbitrary numbers of loci. *Genetica*; **96**:169-78.

ADDENDA

Published papers to which I contributed during my PhD

1. **Minelli, C.***, Thompson, J. R.*, Tobin, M. D., Abrams, K. R. (2004) An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology*; **160**:445-52 (*joint first authors).

   I undertook the first-draft of the paper and participated in the planning of the analyses, interpretation of the results and revising of the paper.

2. Tobin, M. D., **Minelli, C.**, Burton, P. R., Thompson, J. R. (2004) Commentary: Development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'. *International Journal of Epidemiology*; **33**: 26-29.

   I participated in the planning and revising of the paper.

3. Thompson, J. R., **Minelli, C.**, Abrams, K. R., Tobin, M. D., Riley, R. D. (2005) Meta-analysis of genetic studies using Mendelian randomization – a multivariate approach. *Statistics in Medicine*; **24**:2241-54.

   I participated in the planning of the analyses, interpretation of the results, and drafting and revising of the paper.

4. **Minelli, C.***, Thompson, J. R.*, Abrams, K. R., Lambert, P.C. (2005a) Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Statistics in Medicine*; **24**:3845-61 (*joint first authors)

   I participated in the planning and performing of the analyses, interpretation of the results, drafting and revising of the paper.

5. **Minelli, C.***, Thompson, J. R.*, Abrams, K. R., Thakkinstian, A., Attia, J. (2005b) The choice of genetic model in the meta-analysis of molecular association studies. *International Journal of Epidemiology*; **34**:1319-28 (*joint first authors)

   I undertook the first-draft of the paper and participated in the planning of the analyses, interpretation of the results and revising of the paper.

6. Thakkinstian, A., McEvoy, M., **Minelli, C.,** et al. (2005) Systematic review and meta-analysis of the association between . 2-adrenoceptor polymorphisms and asthma: a HuGE review. *American Journal of Epidemiology*; **162**:201-11.

I gave advice on the analyses and participated in the interpretation of the results and revising of the paper.

ORIGINAL CONTRIBUTIONS

An Integrated Approach to the Meta-Analysis of Genetic Association Studies using Mendelian Randomization

Cosetta Minelli,, John R. Thompson, Martin D. Tobin, and Keith R. Abrams

From the Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, Leicester Medical School, University of Leicester, Leicester, United Kingdom.

A natural randomization process, sometimes called Mendelian randomization, occurs at conception to determine a person's genotype. By combining information from genotype-disease and genotype-phenotype studies, it is possible to use this Mendelian randomization to derive an estimate of the association between phenotype (risk factor) and disease that is free of the confounding and reverse causation typical of classical epidemiology. When one is synthesizing evidence, studies evaluating genotype-phenotype associations, studies evaluating genotype-disease associations, and studies evaluating both are encountered, and methods should be used that allow for this structure. Plotting the log odds ratio of genotype-disease against the mean genotype-phenotype difference may help investigators detect departures from the assumptions underlying Mendelian randomization. Testing for differences between studies reporting on only the genotype-phenotype or genotype-disease association and those reporting on both associations may help in detecting reporting bias. This integrated approach to the meta-analysis of genotype-phenotype and genotype-disease studies is illustrated here using the example of the methylenetetrahydrofolate reductase (*MTHFR*) gene, homocysteine level, and coronary heart disease. An integrated meta-analytical approach may increase the precision of this estimate and provide information on the assumptions underlying Mendelian randomization. Serious biases may arise if the assumptions behind the analysis based on Mendelian randomization are not met.

epidemiologic methods; genetic epidemiology; genetics; genotype; meta-analysis; phenotype

Abbreviations: CI, confidence interval; MTHFR, methylenetetrahydrofolate reductase.

With the recent growth in knowledge about the human genome, there has been a dramatic increase in the number of genetic epidemiologic studies of the association between specific genes and diseases and between those genes and the risk factors or phenotypes that are thought to be intermediates on the causal pathway to disease. In many instances, these studies have supplemented preexisting research on the association between the phenotype and the disease. For instance, many recent studies have evaluated the associations between polymorphisms in the methylenetetrahydrofolate reductase (*MTHFR*) gene and coronary heart disease and between the *MTHFR* gene and homocysteine level. These studies have been motivated, in part, by the preexisting evidence of an association between homocysteine level and coronary heart disease. Similarly, there have been many studies of polymorphisms in the apolipoprotein E gene and coronary heart disease or stroke and many studies of the apolipoprotein E gene and lipid levels, stemming from epidemiologic evidence of an association between lipids and coronary heart disease or stroke.

As the number of genetic studies has grown, so have meta-analyses been produced to synthesize the evidence and over-
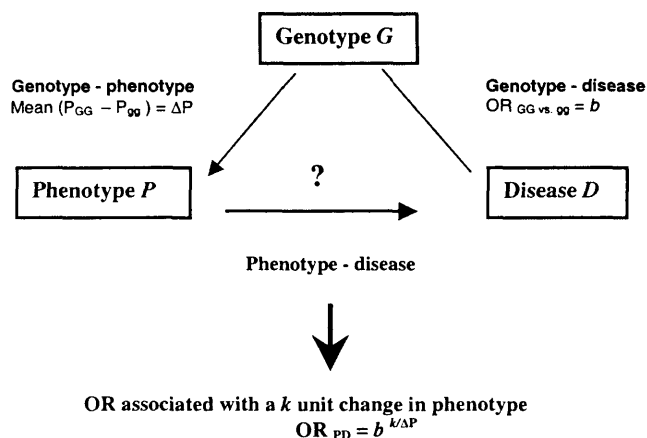
FIGURE 1. Calculation of an unconfounded estimate of the effect of a change in phenotype on a disease based on the concept of Mendelian randomization. OR, odds ratio.

come the limitations of power found in even moderate-sized studies (1). Two factors are evident from reviewing these meta-analyses: first, studies of the relation between a gene and an intermediate phenotype (hereafter referred to simply as phenotype) tend to be less common than studies of a gene and a disease; and second, evidence for a genotype-phenotype association is often obtained as a spin-off of a study aimed primarily at investigating the genotype-disease relation, and thus this information is often obtained only on a subset of the subjects. Where there is a strong reason to suppose that the phenotype is intermediate on the causal pathway from gene to disease, it would be sensible to perform a meta-analysis that in some way integrates the evidence for all three relations: genotype-phenotype, genotype-disease, and phenotype-disease. The logic behind this approach is greatly strengthened by an appeal to Mendelian randomization—that is, the fact that one's genes are inherited at conception through a seemingly random process. Accordingly, epidemiologic studies of genotype-phenotype and genotype-disease associations show strong parallels with randomized trials and should not be affected by confounding or reverse causation in the way that makes classical epidemiologic phenotype-disease studies so difficult to interpret (2–4). In theory, the genotype acts as an instrumental variable, and by combining the information obtained from genotype-disease and genotype-phenotype studies, it should be possible to derive an unconfounded estimate of the phenotype-disease association. Integrated meta-analyses may be able to take advantage of Mendelian randomization in order to test whether the phenotype is actually on the causal pathway and to obtain an unconfounded estimate of the size of the effect of phenotype on disease.

In this paper, we present a framework for an integrated meta-analytical approach to the study of genotype-disease and genotype-phenotype associations that takes advantage of the benefits of Mendelian randomization to derive an indirect estimate of the phenotype-disease relation. To illustrate the methods involved, we use the example of an integrated

meta-analysis of studies on *MTHFR* and coronary heart disease and *MTHFR* and homocysteine level.

METHODS

Mendelian randomization

In order to use genetic studies to quantify the relation between phenotype and disease, the estimate of the genotype-disease association must be combined with the estimate of the genotype-phenotype association (figure 1). Suppose that a mutant genotype (GG) causes an increased risk of disease in comparison with the wild type (gg) and that this effect is measured by the odds ratio, $OR_{GG \text{ vs. } gg}$. Further suppose that GG compared with gg causes a mean difference, ΔP, in the level of the intermediate phenotype. Then, under the assumptions required for Mendelian randomization and assuming linearity of the relation between phenotype difference and log odds ratio for the disease, $OR_{GG \text{ vs. } gg}^{1/\Delta P}$ is an unconfounded estimate of the odds ratio of disease resulting from a unit change in the phenotype.

Sources of evidence

When one is searching for evidence on genotype-phenotype and genotype-disease associations, three different types of genetic studies are likely to be identified: those evaluating only the genotype-phenotype association, those evaluating only the genotype-disease association, and those evaluating both. In addition to noting the usual estimates of effect and their precision, it is important to record when both associations are measured in the same study. It might be that studies classified as providing only genotype-phenotype information in fact also evaluated the genotype-disease association but used a different disease definition, so that this genotype-disease result cannot be pooled with the results of other studies. When collecting data on genotype-phenotype, it is important to note whether the information on the phenotype difference comes from cases, from controls, or from a mixture of both. Whenever possible, data from cases and data from controls should be analyzed separately. If the disease itself affects the level of the phenotype in a way that is not simply linear, the data on the genotype-phenotype effect collected from cases may be less reliable because of reverse causation.

Meta-analytical approaches

If the genotype-phenotype evidence and the genotype-disease evidence come from unrelated sources, separate meta-analyses will provide estimates of the pooled effects that can, by using Mendelian randomization, be combined to estimate the size of the phenotype-disease association. In practice, there is likely to be a mixture of studies that measured the genotype-phenotype effect, studies that measured the genotype-disease effect, and studies that measured both. Studies that measured both associations need to be modeled correctly in order to properly account for the correlation in their estimates of the genotype-phenotype and genotype-disease associations (5).

Consider first a meta-analysis in which all available studies measured both genotype-phenotype and genotype-disease. We could proceed as before and separately pool the genotype-phenotype estimates and genotype-disease estimates before combining the pooled values in order to estimate the effect of phenotype on disease. However, the likely correlation in the sizes of pairs of estimates from the same study would affect the size of the confidence interval for the derived phenotype-disease effect and the validity of any hypothesis test. A better procedure would be to combine the genotype-phenotype and genotype-disease estimates separately within each study to obtain study-specific estimates of the phenotype-disease effect. These study-specific estimates could be graphed in their own forest plot and pooled to obtain an overall estimate.

In the more realistic situation in which some studies measured both genotype-disease and genotype-phenotype and some measured one or the other, we need to proceed with caution. Most important features of the data will be evident from a forest plot with two columns, one for genotype-disease and the other for genotype-phenotype, in which paired estimates from within the same study are aligned in the same row (see figure 2). The forest plot is organized in three blocks, representing the studies that measured both the odds ratio and the mean difference, the studies that measured only the odds ratio, and the studies that measured only the mean difference. Within blocks, the studies are ordered by size of effect. Where both estimates were obtained in the same study, the studies are ordered by the size of the genotype-disease odds ratio. Having drawn the plot, the next stage should be to check that the genotype-disease estimates from studies that also reported data on genotype-phenotype are consistent with the estimates from studies that did not report data on genotype-phenotype. Similarly, genotype-phenotype estimates should be compared between studies that also reported data on genotype-disease and those that did not.

Funnel plots can be used to search for the presence of publication bias (6, 7). However, genetic studies may also be affected by a form of reporting bias in which both the odds ratio and the mean difference are measured but only one is reported because the other contradicts the anticipated relation. Moreover, it is possible that the reporting of data on both associations is a marker for some feature such as study quality. It is very difficult to detect this bias from the reported data; but a careful reading of the study's methods section may show whether data were collected but not reported, and it may be informative to distinguish studies that reported data on only one of the two associations using a different symbol on the funnel plots.

When the phenotype does indeed lie on the causal pathway between gene and disease, studies carried out in populations with a large genotype-phenotype difference might be expected to show a large genotype-disease odds ratio. This can be investigated by plotting the findings from each study on a graph of the log odds ratio of genotype-disease against the average difference in phenotype with genotype. A similar graphic approach was used previously in the meta-analysis of randomized trials (8, 9). This graph would be expected to show a monotonic trend if the phenotype is inter-

mediate on the etiologic pathway to disease, and the line should pass through the origin. Lack of any correlation would cast doubt on whether the phenotype is truly intermediate. A line that does not pass through the origin might indicate that there is another intermediate phenotype through which the gene under study exerts its effect on disease (a special case of pleiotropy) or that the gene is in linkage disequilibrium (associated at population level) with a gene which also affects the risk of disease, or that there is differential publication bias for the two associations (7, 10). This graph will also show gross departures from linearity of the relation between phenotype difference and log odds ratio of disease, as approximate linearity is an assumption behind the averaging across studies to obtain an estimate of the pooled phenotype-disease association.

If it appears that genotype-phenotype and genotype-disease associations are consistent across all studies, we may pool all genotype-phenotype estimates and all genotype-disease estimates before combining these overall estimates to derive a figure for the phenotype-disease association. The effect of the correlation on the confidence interval and hypothesis test will depend on the proportion of studies that reported on both genotype-disease and genotype-phenotype associations. The studies that provided both estimates can now be used as described above to provide a comparison with studies reporting one of the estimates and to investigate the consistency of the study-specific phenotype-disease estimates. Methods for the adjustment of confidence intervals and hypothesis tests that allow for between-study correlation have been described elsewhere (5).

EXAMPLE

MTHFR, homocysteine, and coronary heart disease

A recent nongenetic meta-analysis of individual patient data from epidemiologic studies showed a decrease of 11 percent in coronary heart disease for a 25 percent decrease in homocysteine level (odds ratio = 0.89, 95 percent confidence interval (CI): 0.83, 0.96) (11). The meta-analysis showed that heterogeneity between studies was partly explained by study design. Retrospective studies yielded higher estimates of risk, perhaps because of reverse causation and/or unadjusted confounding. In particular, two major confounding factors were suggested: smoking and blood pressure. These factors are both strongly correlated with homocysteine level and are known risk factors for coronary heart disease. The strong possibility of unadjusted confounding makes it very difficult to be sure that the relation between homocysteine and coronary heart disease is causal.

A common polymorphism in the gene for the *MTHFR* enzyme leads to reduced enzyme activity, a lower folate level, and consequently a higher homocysteine level (12). The polymorphism involves a C→T base pair substitution at nucleotide 677, so the wild-type homozygous genotype is referred to as CC and the mutant homozygous genotype as TT. This polymorphism can be used, together with the idea of Mendelian randomization, to indirectly assess the effect of homocysteine on coronary heart disease.
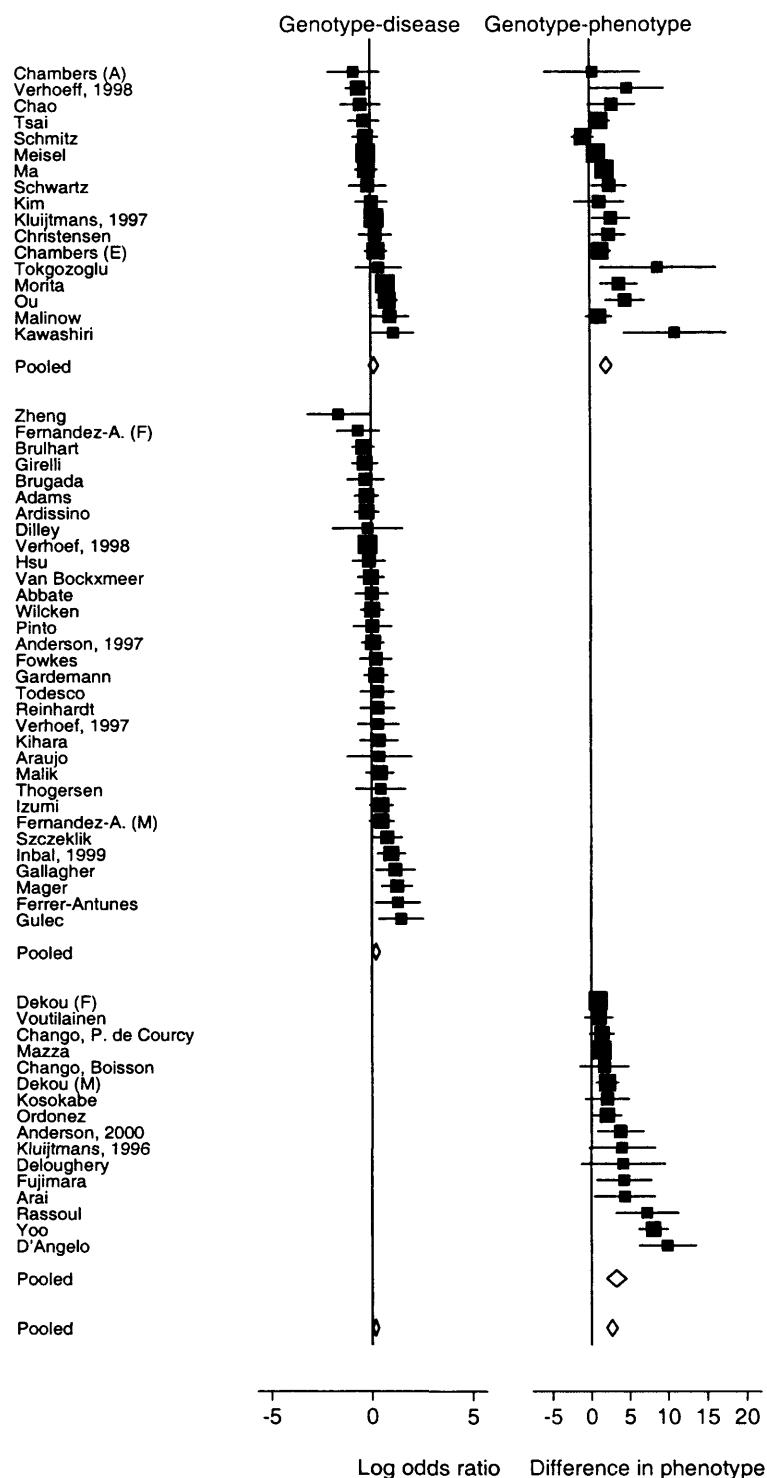
FIGURE 2. Two-column forest plot and pooled estimates for genotype-disease and genotype-phenotype associations. *A*, Asians; *E*, Europeans; *F*, females; *M*, males. Horizontal bars, 95% confidence interval. (For references, see the technical report by Minelli et al. (5).)

A recent genetic meta-analysis of individual patient data has shown an increased risk of coronary heart disease of approximately 16 percent associated with genotype TT as compared with CC (odds ratio = 1.16, 95 percent CI: 1.05,

1.28) (13). This result was similar to that of another meta-analysis published at the same time but carried out on aggregated data, which showed a pooled odds ratio of 1.21 for the TT genotype (95 percent CI: 1.06, 1.39) (14). The latter

paper also mentioned those studies that evaluated the association between genotype and phenotype. Wald et al. (14) found an average difference of 2.7 μmol/liter in homocysteine concentration (95 percent CI: 2.1, 3.4) between the TT and CC genotypes.

Sources of evidence

When the two meta-analyses by Wald et al. (14) and Klerk et al. (13) were combined, a total of 66 genetic studies were identified. Classifying the studies that reported both the estimate and its precision, 32 evaluated only the genotype-disease association, 16 evaluated only the genotype-phenotype association, and 18 evaluated both. The definition of coronary heart disease used in our analysis was myocardial infarction or angiographically confirmed coronary artery occlusion (>50 percent of the luminal diameter). Genotype-disease associations were reported in an additional 13 studies in the original meta-analyses, but this information was not included because of either a different disease definition or a restricted study population (5).

Among the 18 studies that evaluated both associations, nine measured the mean difference in phenotype level with genotype in both cases and controls (four reporting only combined means); four studies measured homocysteine only in cases and three only in controls, and two reports were unclear.

RESULTS

Figure 2 shows the two-column forest plot, with the first column representing the genotype-disease log odds ratio and the second representing the genotype-phenotype mean difference. It is clear that there is considerable variation between studies, with some even reporting average odds ratios less than 1 or mean homocysteine differences in the direction opposite that anticipated.

Parts *a* and *b* of figure 3 show the funnel plots for the genotype-disease and genotype-phenotype associations, respectively. For the genotype-disease association, there seems to be no evidence of either publication bias, indicated by an overall lack of symmetry in the funnel plot, or reporting bias, suggested by a discrepancy in the shape of the funnel plot between studies reporting on both associations and those reporting on only one association. For the genotype-phenotype association, the funnel plot is suggestive of possible publication bias, while there appears to be little evidence of reporting bias.

All of the following meta-analyses were based on random-effects models, which take into account between-study heterogeneity (6). The pooled odds ratio estimate for the effect of genotype on coronary heart disease based on studies that also reported on the homocysteine change was 1.17 (95 percent CI: 0.93, 1.48), and where data on homocysteine were not reported, the pooled odds ratio was 1.24 (95 percent CI: 1.04, 1.48). The difference was not statistically significant (*p* = 0.68). Similarly, the mean change in homocysteine level in studies that also reported on coronary heart disease was 2.14 μmol/liter (95 percent CI: 1.37, 2.91), and in studies that did not report data on coronary heart disease it
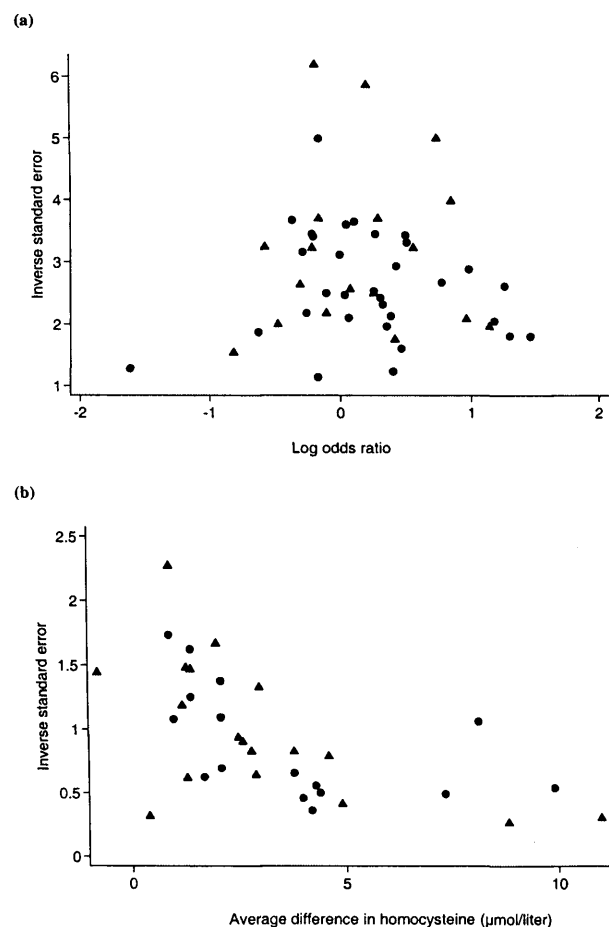

(a)

(b)

FIGURE 3. Funnel plots for genotype-disease (*a*) and genotype-phenotype (*b*) associations. Different symbols are used for those studies that measured both associations (triangles) and those that measured only the association of interest (circles).

was 3.34 μmol/liter (95 percent CI: 2.10, 4.59). The direction of the difference is consistent with the presence of publication bias, whereby results from studies evaluating the genotype-phenotype association alone are published only if the effect size detected is large. However, this difference was not statistically significant (*p* = 0.11). The lack of significant differences justified pooling all of the odds ratio estimates to obtain an odds ratio of 1.21 (95 percent CI: 1.06, 1.40) and all mean differences in homocysteine to obtain a difference of 2.71 μmol/liter (95 percent CI: 2.02, 3.41). Combining the two estimates, we obtain an estimate of the odds ratio for coronary heart disease per unit (1.0-μmol/liter) change in homocysteine level of 1.07. If, for the moment, we ignore the correlation between the studies that measured both, we obtain the 95 percent confidence interval 1.02, 1.14. It may be more informative to rescale this odds ratio for increments other than a unit increase in homocysteine level. For instance, the odds ratio for coronary heart disease for a standard reference increment of 5 μmol/liter, as used by Wald et al. (14), is 1.43 (95 percent CI: 1.10, 1.95), while for an increment of 3 μmol/liter, considered to reflect the possible
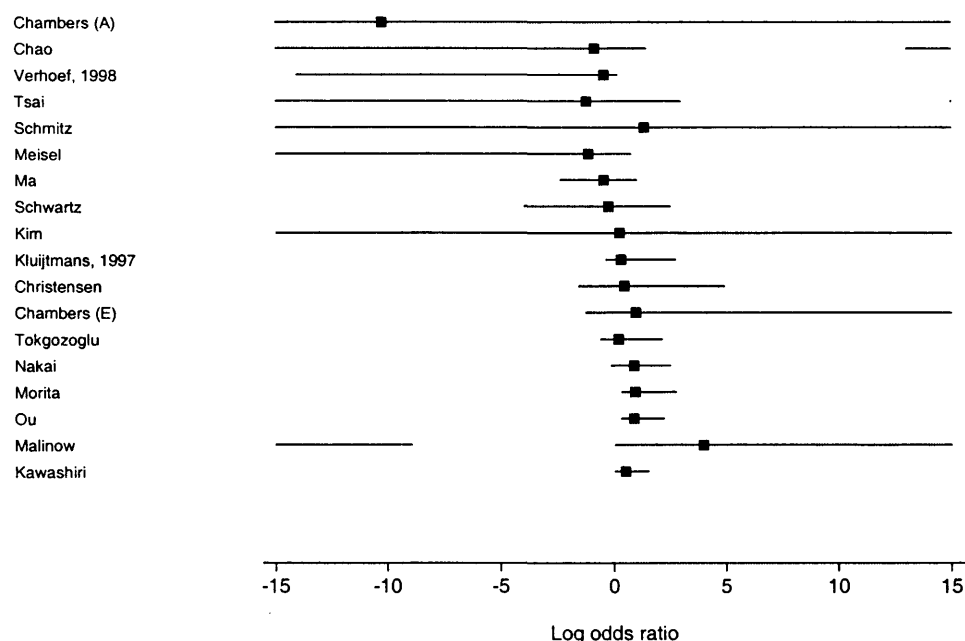
FIGURE 4. Forest plot of the derived study-specific estimates of the odds ratio for the effect of phenotype on disease (per 5-μmol/liter difference in homocysteine level). *A*, Asians; *E*, Europeans. Horizontal bars, 95% confidence interval. (For references, see the technical report by Minelli et al. (5).)

size of a homocysteine-lowering intervention with folic acid supplementation (15), the odds ratio is 1.24 (95 percent CI: 1.06, 1.49).

The next step is to investigate more fully the findings of the 18 studies that reported both an odds ratio and a mean difference. Figure 4 shows the forest plot of the study-specific estimates of the coronary heart disease odds ratio associated with a 5-μmol/liter increase in homocysteine level. Because, on a log odds ratio scale, the derived phenotype-disease association is obtained by division, when study results suggest that the homocysteine difference could reasonably be on either side of zero, the confidence interval for the ratio could stretch to plus or minus infinity (16, 17). In some cases, this infinite range is accompanied by a gap or clearing in the forest plot associated with values that the ratio is very unlikely to take. The most important feature of figure 4 is the poor precision in the derived estimate that can be obtained from most individual studies. Indeed, several derived estimates have infinite variance.

Parts *a* and *b* of figure 5 show the log odds ratio of genotype-disease plotted against the mean difference in homocysteine level, separately for homocysteine measured in cases and in controls. The two parts of the figure include all studies that either measured homocysteine only in cases or controls or measured it in both and reported separate estimates for the two groups. To allow for the large differences in the precision of the different studies, the individual estimates are plotted as ellipses, with their axes inversely proportional to the standard error of the log odds ratio of genotype-disease and the mean change in homocysteine level. Both figures show an approximately linear relation,

with the line passing close to the origin. As anticipated, the pattern is somewhat clearer when control data are used.

The unweighted correlations observed in parts *a* and *b* of figure 5 are 0.37 (*p* = 0.29) and 0.78 (*p* = 0.01), respectively. Adjusting for the correlation in the 18 studies reporting data on both measures would alter our odds ratio estimate and confidence interval for the effect of a 5-μmol/liter increase in homocysteine level on coronary heart disease risk from 1.43 (95 percent CI: 1.10, 1.95) to 1.54 (95 percent CI: 1.17, 2.06) (5).

DISCUSSION

Although genotype-disease associations are becoming better understood, it is only when we also have information about the causal pathway that we open up the possibility of preventive or therapeutic intervention. Thus, while the association between *MTHFR* polymorphisms and coronary heart disease is scientifically interesting, disease prevention becomes possible when we understand that this effect acts, at least in part, through homocysteine. Dietary folate supplementation is a relatively simple intervention that can be implemented at a population level to lower homocysteine level and thereby reduce coronary heart disease risk. This intervention was introduced in the United States in the late 1990s with the fortification of cereals and grains (18). Classical epidemiologic studies may provide evidence about the phenotype-disease association, but it will almost certainly be affected by confounding and/or reverse causation. The use of Mendelian randomization offers a novel way of deriving unconfounded estimates, although Mendelian randomization makes its own assumptions about the pathway from gene to
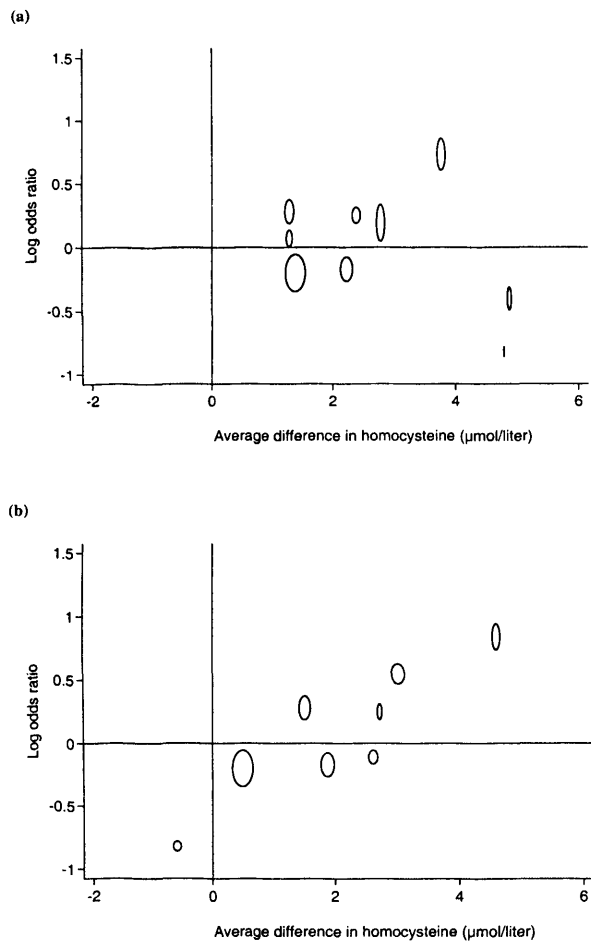
(a)

Average difference in homocysteine (µmol/liter)

(b)

Average difference in homocysteine (µmol/liter)

FIGURE 5. Plot of the log odds ratio for the effect of the methylene-tetrahydrofolate reductase gene on coronary heart disease risk against the mean difference in homocysteine level between genotypes, separately for studies in which phenotype was measured among cases (a) and controls (b). The axes of the ellipses are inversely proportional to the standard errors of the respective associations.

disease (4). The most crucial assumption is that the genotype influences disease risk only through modification of the specific phenotype. If the genetic polymorphism also alters the risk of the same disease via other pathways, the estimate of a specific phenotype-disease association might be seriously biased. In the example of homocysteine and coronary heart disease, this is probably not a problem; but, for instance, polymorphisms of the apolipoprotein E gene affect several different intermediate phenotypes related to lipid metabolism and atherosclerosis (4). Consequently, it is advisable to limit the use of Mendelian randomization to studies where there is good biologic knowledge of the genotype-phenotype-disease pathway.

The approach proposed in this paper stresses the need for investigators conducting meta-analyses to review simultaneously the stages in the genotype-phenotype-disease pathway, and it implies that individual studies of genotype-disease associations should collect information on interme-

diate phenotypes whenever possible. In fact, the analysis of studies that measured both associations allows insight into the interrelations between genotype, phenotype, and disease and gives researchers an opportunity to check the assumptions of the analysis. In this respect, a meta-analysis of small studies might be more informative than a single large prospective study. In some meta-analyses, inconsistencies across studies could result in departures from the linear trend seen in figure 5. This might happen if study populations differed with respect to phenotype measurement, disease definition, gene-environment interactions, compensatory developmental processes (canalization), or linkage disequilibrium with functional alleles (4). Theoretical considerations and our example suggest that it may be safer for primary researchers to measure the phenotype in controls in order to avoid any possibility of bias due to reverse causation. If the phenotype level is measured in both cases and controls, two separate estimates should be reported. For meta-analyses, our recommendation is to perform sensitivity analyses, with phenotype data obtained on cases analyzed separately from phenotype data obtained on controls.

The need for an integrated meta-analytical approach to genetic studies when using Mendelian randomization is particularly important. The uncertainty associated with the derived estimate of the phenotype-disease association can be large, since it depends on uncertainty in both the estimate of the genotype-phenotype association and the estimate of the genotype-disease association (19). It is crucial to the use of Mendelian randomization that both estimates are sufficiently precise, but especially that of the genotype-phenotype association. Such precision is only likely to be obtained through a meta-analysis of all available evidence. In fact, at present, almost all genetic studies lack the statistical power to detect the relatively small effects of the many gene variants that underlie common, complex diseases (20). Although massive reductions in genotyping costs offer the prospect of larger studies, study size remains limited by the cost of proper phenotyping (21).

While meta-analyses can (in theory, at least) partially alleviate the problem of inadequate statistical power, they cannot control the problems of publication and reporting bias (6, 7) that are thought to be particularly important in genetic epidemiology (1, 22). However, using an integrated meta-analytical approach, investigators can start to address these issues by comparing the pooled estimates for genotype-phenotype and genotype-disease associations in studies reporting on either one association only or both associations, and by drawing the funnel plots in a way that allows comparison between the two types of studies for each association.

Analysis of the correlation between the genotype-disease odds ratio and the genotype-phenotype difference, as typified by figure 5, must be done with care. The plot is based on data aggregated over studies and is analogous to an ecologic study that is potentially subject to the ecological fallacy—that is, patterns seen in aggregate data do not necessarily translate to the individual. Thus, when we see an increase in the risk of disease in studies that show an increased difference in phenotype, it is probable but not certain that we would see a similar effect at the individual level. Equally, failure to see a pattern in aggregate data does not rule out the

possibility of an individual-level effect. Obviously, an individual-level causal effect is required for an intervention on the phenotype to have an impact on the risk of disease.

It is tempting to add nongenetic studies of the phenotype-disease association to our integrated approach, if only to test whether they accord with the estimate derived from the application of Mendelian randomization. Unfortunately, the sample sizes required to establish equivalence of the measured and derived estimates are such that even a large meta-analysis may not suffice (19). This is clearly an area that requires more work, because our ultimate aim should be to produce an integrated meta-analysis that links together all relevant phenotypes, diseases, and genotypes, including the heterozygous group.

ACKNOWLEDGMENTS

REFERENCES

1. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. J Clin Epidemiol 2003;56:297–303.
2. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 2001;358:1356–60.
3. Keavney B. Genetic epidemiological studies of coronary heart disease. Int J Epidemiol 2002;31:730–6.
4. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
5. Minelli C, Thompson JR, Tobin MD, et al. Meta-analytical methods for the synthesis of genetic studies using Mendelian randomisation. (Genetic Epidemiology Technical Report no. 2003/GE2). Leicester, United Kingdom: Centre for Biostatistics, Department of Health Sciences, University of Leicester, 2003. (World Wide Web URL: http://www.prw.le.ac.uk/research/HCG/getechrep.html).
6. Sutton AJ, Abrams KR, Jones DR, et al. Methods for meta-analysis in medical research. Chichester, United Kingdom: John Wiley and Sons Ltd, 2000.
7. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. BMJ 2001;323:101–5.
8. Midgley JP, Matthew AG, Greenwood CM, et al. Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. JAMA 1996;275:1590–7.
9. Smith GD, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. BMJ 1993;306:1367–73.
10. Sutton AJ, Abrams KR, Jones DR. Generalized synthesis of evidence and the threat of dissemination bias: the example of electronic fetal heart rate monitoring (EFM). J Clin Epidemiol 2002;55:1013–24.
11. Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. JAMA 2002;288:2015–22.
12. Bailey LB, Gregory JF. Polymorphisms of methylenetetrahydrofolate reductase and other enzymes: metabolic significance, risks and impact on folate requirement. J Nutr 1999;129:919–22.
13. Klerk M, Verhoef P, Clarke R, et al. *MTHFR* 677C→T polymorphism and risk of coronary heart disease: a meta-analysis. JAMA 2002;288:2023–31.
14. Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. BMJ 2002;325:1202.
15. Homocysteine Lowering Trialists' Collaboration. Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomised trials. BMJ 1998;316:894–8.
16. Fieller EC. Some problems in interval estimation. J R Stat Soc B 1954;16:175–85.
17. Kendall MG, Stuart A. The advanced theory of statistics. Vol 2. Inference and relationship. 3rd ed. London, United Kingdom: Charles Griffin and Company Ltd, 1973.
18. Rader JI. Folic acid fortification, folate status and plasma homocysteine. J Nutr 2002;132(suppl):2466S–70S.
19. Thompson JR, Tobin MD, Minelli C. On the accuracy of estimates of the effect of phenotype on disease derived from Mendelian randomisation studies. (Genetic Epidemiology Technical Report no. 2003/GE1). Leicester, United Kingdom: Centre for Biostatistics, Department of Health Sciences, University of Leicester, 2003. (World Wide Web URL: http://www.prw.le.ac.uk/research/HCG/getechrep.html).
20. Hirschhorn JN, Lohmueller K, Byrne E, et al. A comprehensive review of genetic association studies. Genet Med 2002;4:45–61.
21. Thompson EA. Human genetics: overview. In: Palmer L, ed. Biostatistical genetics and genetic epidemiology. Chichester, United Kingdom: John Wiley and Sons Ltd, 2002:386–90.
22. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet 2003;361:865–72.

Commentary: Development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'

Martin D Tobin, Cosetta Minelli, Paul R Burton and John R Thompson

In his letter to the *Lancet* in 1986, reprinted in this issue of the *International Journal of Epidemiology* (*IJE*), Katan described the idea of using data from genetic studies to test for a relationship between a quantitative intermediate phenotype and a disease in a way that is not distorted by confounding or reverse causality.[1] Following the application of these ideas by other authors[2–5] interest in the concept has grown, although it is still not widely understood. This important and novel method has the potential to improve the way that the quantitative phenotypes that underlie common diseases are investigated, so better informing public health interventions that alter the level of the phenotype in order to reduce the risk of disease.[4]

Katan described how evidence of the effect of the apolipoprotein E (APOE) genotype on cancer risk could be used to test the hypothesis that 'a naturally low cholesterol favours tumour growth.'[1] Given that 'the gradient in serum cholesterol levels in the population is associated with a gradient in APOE [genotype]', under the causal hypothesis we would expect to see a corresponding association between APOE and cancer. The absence of such a genetic association 'would suggest that the association between low cholesterol and cancer is spurious'. Katan emphasized that APOE genotype is present since birth and is not disturbed by disease, so unlike conventional epidemiological methods the genetic test is not influenced by reverse causation or confounding.

The term 'Mendelian randomization' has been used by a number of researchers when applying this idea to investigate phenotype–disease associations[2,4,5] but not always in exactly the same way. The term was first used in a completely different context to describe a method of pseudo-randomization in a particular clinical trial for which the randomization of treatment was not otherwise possible[6] (see Wheatley and Gray's Commentary in this issue of *IJE*)[7]. At its most basic, 'Mendelian randomization' simply means that, according to Mendel's laws of segregation and independence,[8] a subject's genotype is determined by an apparently random process at conception. So 'Mendelian randomization' is a fundamental biological process that should reasonably underpin the appropriate interpretation of *any* study in which genotype is related to an outcome. However, by common usage, the term 'Mendelian randomization' has also become attached to the epidemiological method that appears to be based upon Katan's

ideas that generates *indirect*, and unconfounded, inferences about the association between a phenotype and a disease given *direct* information on the gene–disease and gene–phenotype associations.[4] The confusion and ambiguity that this double meaning engenders is impeding the transmission of ideas about the value of this important epidemiological approach. In our view it would be better to give the epidemiological method an alternative name, such as 'Mendelian deconfounding', and to reserve the term 'Mendelian randomization' for the more fundamental biological process.

In applying the concept described by Katan, the emphasis originally was on hypothesis testing to confirm or refute the evidence for particular phenotype–disease associations found in observational studies. However, the method can be developed to provide an estimate of the *size* of the unconfounded effect of a phenotype on disease together with a measure of its uncertainty. In this commentary we describe how these ideas have been developed since Katan's paper and in particular we emphasize the benefits of estimating the size of the effect of phenotype on disease over simple hypothesis testing. Finally we consider possible future developments particularly in regards to meta-analysis.

Applications of Mendelian randomization to learn about phenotype–disease relationships

Youngman *et al.* and Keavney *et al.* were the first authors to use the term 'Mendelian randomization' in a similar epidemiological context to that described by Katan.[2,3] Youngman *et al.* studied fibrinogen levels and beta-fibrogen genotype in premature myocardial infarction (MI) cases and related controls. From their dataset, three associations were observed. These are shown in Figure 1, where G represents the genotype (beta-fibrinogen
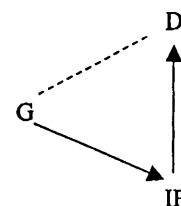

Figure 1 A pictorial representation of the model used to test for a causal association between intermediate phenotype and disease (IP-D) or to derive an unconfounded estimate of the size of that effect

University of Leicester, Department of Epidemiology and Public Health, 22–28 Princess Road West, Leicester LE1 6TP, UK. E-mail: mt47@leicester.ac.uk

HindIII), IP the intermediate phenotype (fibrinogen), and D the disease (MI). The association between G and D (shown as a broken line) is induced only through the causal effects of G on IP and of IP on D. The model assumes that there is no other pathway through which the gene exerts its effect on MI. Under the assumption of Mendelian randomization, the measurements of the G–D and G–IP associations are unconfounded, while the measured IP–D association is likely to be confounded and subject to reverse causation.

The information available enabled Youngman *et al.* to assess whether fibrinogen (IP) had a causal link with MI (D). The authors obtained unconfounded estimates of the G–IP (0.12 g/l per allele, standard error 0.018, $P < 0.00001$) and G–D, odds ratio = 1.03 (95% CI: 0.96, 1.10) associations. As the G–IP association is clearly established, a causal IP–D relationship would have resulted in an observed G–D association. As this G–D link is not seen, their observed IP–D odds ratio of 1.20 (95% CI: 1.13, 1.26) is probably a result of confounding or reverse causation.

The non-significance of the G–D association does lead us to doubt the hypothesized causal pathway. However, this conclusion is subject to all of the reservations we would have about using *P*-values. An alternative interpretation would be that there were insufficient data to rule out a 10% increase in MI risk for a (modest) 0.12-g/l change in the level of fibrinogen. Thought of in this way, the analysis based on the assumption of Mendelian randomization suggests that any causal IP–D association is not large but it does not rule it out completely. The use of the data to estimate the potential size of the IP–D association is likely to be much more informative than relying on hypothesis tests of G–IP and G–D to rule in or rule out a causal link.

To estimate the size of the unconfounded IP–D odds ratio associated with a specified change, K, in the intermediate phenotype when the measured G–IP difference is ΔIP per allele and the odds ratio (per allele) of G–D is OR, we may calculate

$$OR^{K/\Delta IP}$$

If this estimation approach is used then it is important that uncertainty in the derived IP–D odds ratio accurately reflects the uncertainties in *both* the G–IP and G–D estimates.[9] Some analyses have ignored the uncertainty in G–IP when calculating CI for the derived odds ratios, which may be very misleading when G–IP is inaccurately assessed.

Youngman and colleagues also used the assumption of Mendelian randomization to study the relationship between plasma apolipoproteins A1 (IP_1) and B (IP_2) and MI (D) using apolipoprotein E (G_1) and cholesteryl ester transfer protein (G_2) genotypes as illustrated in Figure 2.[3]
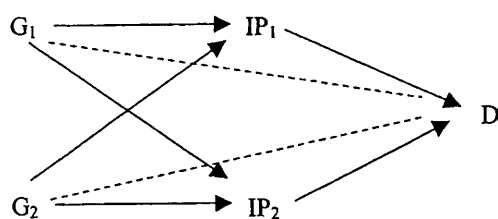

Figure 2 Modelling the effect of intermediate phenotypes on a complex disease

This model has the additional complications of more than one gene affecting each intermediate phenotype (genetic, specifically locus, heterogeneity) and each gene exerting its effect on more than one intermediate phenotype, a special case of pleiotropy. The development of approaches to derive unconfounded estimates of the effect of IP on D in such situations have not yet been fully developed but would be of enormous benefit, given that the phenotypes of greatest interest in public health terms are those that underlie common disorders where such complexity is the norm.[10] Further examples of the use of Mendelian randomization can be found in the extensive review by Davey Smith and Ebrahim.[4]

Developing the concept: from hypothesis testing to Mendelian deconfounding

The key research question from a public health perspective is: what is the unconfounded effect of IP on D? Suitably designed genetic studies provide epidemiologists with a tool to derive unconfounded estimates of the size of the effect of the IP on D together with a measure of its uncertainty.[9] As with most research involving human subjects, the purpose of such studies will usually be to determine the magnitude of the effect of a causal factor or an intervention aimed at preventing disease.[11] In these situations, estimation (of the magnitude of the IP–D association) rather than hypothesis testing (of whether observational epidemiology studies have been subject to confounding) will be of greater utility. There is a substantial literature that stresses the advantages of estimation over hypothesis testing to inform decisions in health-related research.[11,12] Clinically important effects may be statistically non-significant if the sample size is inadequate. On the other hand, clinically irrelevant effects may sometimes be statistically significant.

Unconfounded estimates of the IP–D association can readily be adjusted for a realistic reduction in the level of IP that one could expect from a public health intervention aimed at reducing D. Misleading inferences could result, however, if the intervention exerted its effect on D via causal pathways other than IP. This is the case with interventions aimed at reducing MI risk by lowering fibrinogen, which also appear to affect MI risk via other pathways.[13–15]

In order to produce tight CI for the IP–D odds ratio we need accurate estimates of both G–D and G–IP. Such information will often only be available from meta-analyses. Minelli *et al.* have described meta-analytical approaches for Mendelian deconfounding, addressing the important issue that G–IP and G–D associations may be correlated when both estimates are obtained from the same study and describing methods to allow for such correlation.[16]

The future

Precise estimates are needed to indirectly estimate the effect of IP on D from Mendelian randomized studies. As almost all current genetic studies are statistically underpowered to detect the relatively small effects of the frequent gene variants that underlie common, complex diseases,[17] there has been an increasing emphasis on evidence synthesis and meta-analysis in genetic epidemiology. The Human Genome Epidemiology

Network (HuGENet) now co-ordinates a series of reviews that integrate evidence from genetic association studies (http://www.cdc.gov/genomics/hugenet/default.htm).[18] However, only 2 of the 20 reviews published by April 2003 actually employed meta-analysis.[19-21] We are optimistic about the possibility of larger studies in the near future because of the substantial reductions in genotyping costs, but study size still remains limited by the cost of proper phenotyping and Mendelian randomization is likely to be based on evidence from meta-analyses for the foreseeable future.

When undertaking meta-analyses of genetic studies to derive unconfounded estimates of IP–D association, researchers will need to be mindful of the limitations of Mendelian randomization studies described by Davey Smith and Ebrahim[4] and the limitations of standard meta-analytical problems, such as publication bias.[22] Furthermore, reporting and publication bias are so pervasive in genetic association studies[17] that especial caution may be needed. Funnel plots should be visually inspected and the sensitivity of the results to methods which 'adjust' for the presence of publication bias could be tested.[23] In addition, approaches that account for the correlation between G–D and G–IP associations may be required where these estimates are obtained from the same studies.[16]

The ambitious drive to understand aetiological pathways that underlie the so-called complex diseases, such as asthma and coronary heart disease, has gathered pace with the plethora of biological knowledge and data that have arisen from the human genome project.[24] Given that such diseases are common, improved understanding of these pathways will probably be necessary for significant improvements in public health. However, complex diseases are also characterized by their multifactorial nature, an uncertain disease definition, pleiotropy, phenocopies, and genetic heterogeneity.[10] In short, individual genetic effects are modest, difficult to detect, and likely to be strongly influenced by environment. The diseases where 'Mendelian deconfounding' has the greatest potential also throw up the greatest challenges, since the assumptions that relate to Mendelian randomization are least likely to hold for complex diseases. The most important assumption is the absence of an alternative pathway through which the gene exerts its effect on disease (a special case of pleiotropy), which will affect the validity of a phenotype–disease association derived by from genetic studies. Furthermore, the findings may be less generalizable where there is linkage disequilibrium to, or interaction with, a gene with functional effects on the phenotype and/or disease under study, or gene–environment interaction.

However, given that direct estimates of the effect of an intermediate phenotype on disease in traditional observational studies are highly prone to confounding and reverse causation,[4] the derivation of phenotype–disease associations from genetic studies should be considered as a valuable alternative to observational studies and efforts should be directed towards developing methods to appropriately model this complexity. These methods have the potential to be particularly useful in the future as knowledge of biological pathways improves, more suitable polymorphisms can be used, high quality data from large genetic association studies become available, and methods to derive estimates from Mendelian randomised studies are refined.

Acknowledgements

References

[1] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986;i:507–08. (Reprinted *Int J Epidemiol* 2004;33:9.)

[2] Youngman L, Keavney B, Palmer A *et al.* Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomisation'. *Circulation* 2000;102(Suppl.II):31–32.

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

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

[5] Verhoef P, Klerk M, Katan MB. Use of 'Mendelian randomisation' for testing causality. *J Inherit Metab Dis* 2003;26(Suppl.1):39.

[6] Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 1991;7(Suppl.3):9–12.

[7] Wheatley K, Gray R. Commentary: Mendelian randomization—an update on its use to evaluate allogenic stem cell transplantation in leukaemia. *Int J Epidemiol* 2004;33:15–17.

[8] Wijsman EM. Mendel's Laws. In: Elston R, Olson JM, Palmer LJ (eds). *Biostatistical Genetics and Genetic Epidemiology.* Chichester: Wiley, 2002, pp. 527–29.

[9] Thompson JR, Tobin MD, Minelli C. *On the Accuracy of Estimates of Phenotype on Disease Derived from Mendelian Randomised Studies.* Technical report 2003/GE1. Leicester: University of Leicester, 2003 (http://www.prw.le.ac.uk/research/HCG/getechrep.html).

[10] Palmer LJ. Complex Diseases. In: Elston R, Olson JM, Palmer LJ (eds). *Biostatistical Genetics and Genetic Epidemiology.* Chichester: Wiley, 2002, pp. 206–17.

[11] Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ (Clin Res Ed)* 1986;292:746–50.

[12] Rothman KJ. A show of confidence. *N Engl J Med* 1978;299: 1362–63.

[13] Rosenson RS, Tangney CC. Antiatherothrombotic properties of statins: implications for cardiovascular event reduction. [comment] *JAMA* 1998;279:1643–50.

[14] Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* 1998; 279:1477–82.

[15] Lowe GD. Why do smokers have higher plasma fibrinogen levels than non-smokers? *Clin Sci (Lond)* 2001;101:209–10.

[16] Minelli C, Thompson JR, Tobin MD. *Meta-Analytical Methods for the Synthesis of Genetic Studies Using Mendelian Randomisation.* Technical report 2003/GE2. Leicester: University of Leicester, 2003 (http://www.prw.le.ac.uk/research/HCG/getechrep.html).

[17] Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91–99.

[18] Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol* 2000;151:2–3.

[19] Little J, Khoury MJ, Bradley L *et al.* The human genome project is complete. How do we develop a handle for the pump? *Am J Epidemiol* 2003;157:667–73.

[20] Botto LD, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am J Epidemiol* 2000;**151**:862–77.

[21] Engel LS, Taioli E, Pfeiffer R *et al*. Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. *Am J Epidemiol* 2002;**156**:95–109.

[22] Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context. 2nd Edn.* London: BMJ Books, 2001, pp.189–211.

[23] Sutton AJ, Abrams KR, Jones DR. Generalized synthesis of evidence and the threat of dissemination bias. The example of electronic fetal heart rate monitoring (EFM). *J Clin Epidemiol* 2002;**55**:1013–24.

[24] Burton P, Tobin MD. Epidemiology and Genetic Epidemiology. In: Balding DJ, Bishop M, Cannings C (eds). *Handbook of Statistical Genetics. 2nd Edn.* Chichester: Wiley, 2003.

Meta-analysis of genetic studies using Mendelian randomization—a multivariate approach

John R. Thompson*,†, Cosetta Minelli, Keith R. Abrams, Martin D. Tobin
and Richard D. Riley

*Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, University of Leicester,
22-28 Princess Rd West, Leicester, LE1 6TP, U.K.*

SUMMARY

In traditional epidemiological studies the association between phenotype (risk factor) and disease is often biased by confounding and reverse causation. As a person's genotype is assigned by a seemingly random process, genes are potentially useful instrumental variables for adjusting for such bias. This type of adjustment combines information on the genotype–disease association and the genotype–phenotype association to estimate the phenotype–disease association and has become known as Mendelian randomization. The information on genotype–disease and genotype–phenotype may well come from a meta-analysis. In such a synthesis, a multivariate approach needs to be used whenever some studies provide evidence on both the genotype–phenotype and genotype–disease associations. This paper presents two multivariate meta-analytical models, which differ in their treatment of the heterogeneities (between-study variances). Heterogeneities on the genotype–phenotype and genotype–disease associations may be highly correlated, but a multivariate model that parameterizes the heterogeneity directly is difficult to fit because that correlation is poorly estimated. We advocate an alternative model that treats the heterogeneities on genotype–phenotype and phenotype–disease as being independent. This model fits readily and implicitly defines the correlation between the heterogeneities on genotype–phenotype and genotype–disease. We show how either maximum likelihood or a Bayesian approach with vague prior distributions can be used to fit the alternative model. Copyright © 2005 John Wiley & Sons, Ltd.

1. INTRODUCTION

Traditional epidemiological studies that investigate the association between phenotypes (biological risk factors) and diseases are often biased because of confounding or reverse

---
*Correspondence to: John R. Thompson, Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, University of Leicester, 22-28 Princess Rd West, Leicester, LE1 6TP, U.K.
†E-mail: john.thompson@le.ac.uk

causation. A useful technique for adjusting for such bias is to employ a carefully selected gene as an instrumental variable, an approach that has recently become known as Mendelian randomization. This method is effective provided that the gene is well chosen [1] and that the associations between the genotype and the phenotype and between the genotype and the disease can be accurately estimated. To obtain the necessary precision in these estimates will often require a meta-analysis. In such a synthesis, a multivariate approach needs to be used whenever some studies provide evidence on both the genotype–phenotype and genotype–disease associations, in order to allow for the correlation in those outcomes. This paper develops some models for the meta-analysis of genetic studies that are used for Mendelian randomization.

1.1. Meta-analysis of genetic studies using Mendelian randomization

With the recent growth in knowledge about the human genome there has been a dramatic increase in the number of genetic epidemiological studies of the association between specific genes and diseases and between those genes and the risk factors, or phenotypes, that are thought to be intermediates on the causal pathway to disease. As the number of these studies has grown, so meta-analyses have been produced to synthesise the evidence and overcome the limitations of precision found in even moderately sized studies [2, 3].

A new use of this genetic epidemiological evidence has recently been advocated and is based on the concept of Mendelian randomization, that is, the fact that one's genes are inherited before birth by a seemingly random process analogous to treatment allocation in a randomized trial. Because of the nature of genetic inheritance, neither the genotype–disease nor the genotype–phenotype relationship can be affected by confounding or reverse causation, hence combining them provides an unconfounded estimate of the phenotype–disease association [1, 4, 5]. Suppose that a mutant genotype (GG) is associated with an increased risk of disease compared to the wildtype (gg) and that this effect is measured by its odd ratio ($OR_{GG\,vs\,gg}$). Further suppose that GG is associated with a mean difference, ΔP, in the level of the intermediate phenotype compared with gg. Then, assuming linearity of the relationship between phenotype and the log OR of disease, $OR_{GG\,vs\,gg}^{1/\Delta P}$ is an unconfounded estimate of the OR of disease resulting from a unit change in the phenotype. Equivalently the log OR of a unit change in phenotype on disease is estimated by the ratio $\log OR_{GG\,vs\,gg}/\Delta P$. It may be more informative to rescale this OR for increments other than a unit change in phenotype. For an increment of k units the formula becomes $OR_{GG\,vs\,gg}^{k/\Delta P}$. In a meta-analysis the assumption of linearity should be assessed by plotting on a graph, for each study, log OR of genotype–disease against average difference in phenotype with genotype [6].

In a meta-analysis of genetic studies for Mendelian randomization, there is likely to be a mixture of studies that measure the genotype–phenotype effect, those that measure genotype–disease and those that measure both. If the evidence on genotype–phenotype and genotype–disease comes from unrelated sources then separate meta-analyses would give estimates of the pooled effects that could, by appealing to Mendelian randomization, be combined to estimate the size of the phenotype–disease association. Otherwise, it is necessary to use a multivariate model in order to allow for the correlation in the genotype–phenotype and genotype–disease evidence arising from the studies that measure both associations. In either case, the uncertainty in both estimates of genotype–phenotype and genotype–disease associations needs to be taken into account [6].

The meta-analysis of genetic studies using Mendelian randomization is a special case of a multivariate meta-analysis in which the synthesis is simultaneously performed on two correlated outcomes, namely the size of the genotype–phenotype difference and genotype–disease log OR. Multivariate models have previously been used in meta-analyses that collect data on correlated outcomes within the same study, such as in the synthesis of multiple-treatment studies, synthesis of multiple outcome (or endpoint) studies, and synthesis of studies with both multiple treatments and multiple outcomes [7–11]. However, unlike most of these other applications of multivariate meta-analysis, the outcome of real interest in Mendelian randomization is the single derived phenotype–disease association, calculated as a ratio of the two correlated outcomes.

1.2. Overview of the paper

In this paper we consider two models for the heterogeneities, or between-study variances, in the estimates of genotype–phenotype and genotype–disease associations. The first model is shown to give estimation problems even with large amounts of data. The second model overcomes this problem by making the extra assumption that the heterogeneity on the genotype–phenotype association is independent of the heterogeneity on the phenotype–disease association. In Section 2, we introduce the example that will be used to illustrate the models. Section 3 introduces the two models and uses them to analyse the example.

2. ILLUSTRATIVE EXAMPLE: *MTHFR* GENE, HOMOCYSTEINE AND CHD

To illustrate the models we consider the use of the methylene tetrahydrofolate reductase (*MTHFR*) gene as an instrumental variable in the estimation of the effect of homocysteine on coronary heart disease (CHD). Many traditional observational studies have suggested that an increase in blood homocysteine level is associated with an increased risk of CHD. However, important potential confounders, such as smoking and blood pressure, may be responsible for some or all of the observed association. This confounding may explain the heterogeneity between studies that was found in a recent meta-analysis of individual patient data [12]. Moreover, unknown (and therefore unmeasured) confounders are also thought to influence this relationship, as is reverse causation, that is, elevations in blood homocysteine may result from atherosclerosis and CHD. The likely presence of unadjusted confounding makes it very difficult to be sure that the relationship between homocysteine and CHD is causal. The *MTHFR* gene has a common polymorphism that has been shown to affect the level of homocysteine in the blood, which makes it a possible choice as an instrumental variable. The polymorphism in question involves a C-to-T substitution at base 677, so the wildtype homozygous genotype is referred to as CC and the mutant homozygous genotype as TT. The TT polymorphism leads to reduced enzyme activity and consequently higher homocysteine levels [13], so if homocysteine is causally related to CHD we would expect the TT polymorphism to be associated with CHD.

A recent genetic meta-analysis of individual patient data has shown an increased risk of CHD of about 16 per cent associated with genotype TT compared to CC (OR: 1.16; 95 per cent CI: 1.05–1.28) [14]. This result is similar to that of another meta-analysis based on aggregated data published by Wald *et al.* [15], which showed an OR of 1.21 (95 per cent CI: 1.06–1.39). The latter paper also considered the evidence on the association between
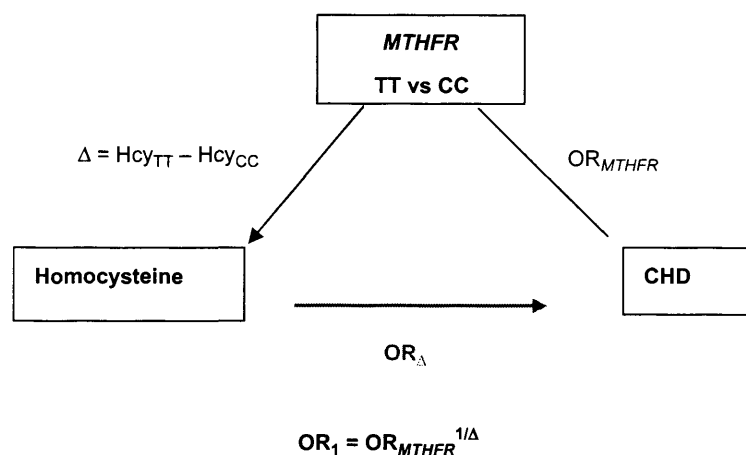
Figure 1. Calculation of an unconfounded estimate of the effect of an increase in homocysteine (Hcy) of 1 µmol/l on the risk of CHD, based on the concept of Mendelian randomization.

genotype and phenotype and reported a simple average mean difference of 2.7 (95 per cent CI: 2.1–3.4), µmol/l in homocysteine concentration between TT and CC genotypes (Figure 1).

Combining the two genetic meta-analyses, a total of 66 genetic studies were identified (data set available at http://www.prw.le.ac.uk/research/HCG/MTHFRdataSet.html). Of these, 32 evaluated only the genotype–disease association, 16 only the genotype–phenotype association, and 18 reported both [6]. The two-column forest plot in Figure 2 illustrates the results of the studies.

3. MODELS FOR MULTIVARIATE META-ANALYSIS

In any multivariate meta-analysis it is important to allow for the correlation in the estimates from those studies that supply information on more than one outcome. If the ith study supplies correlated outcomes, x_i and y_i, which are assumed to be multivariate normally (MVN) distributed with variances v_{xi} and v_{yi}, then,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} \mu_{xi} \\ \mu_{yi} \end{pmatrix}, \begin{pmatrix} v_{xi} & \psi\sqrt{v_{xi}v_{yi}} \\ \psi\sqrt{v_{xi}v_{yi}} & v_{yi} \end{pmatrix} \right)$$

In the next level of the hierarchy we assume that different studies vary about the common means μ_x and μ_y with correlated heterogeneities τ_x and τ_y.

$$\begin{pmatrix} \mu_{xi} \\ \mu_{yi} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \tau_x & \rho\sqrt{\tau_x\tau_y} \\ \rho\sqrt{\tau_x\tau_y} & \tau_y \end{pmatrix} \right)$$

The parameter ρ represents the between-study correlation across studies measuring both outcomes and ψ represents the within-study correlation. For those studies measuring only the

Figure 2. Two-column forest plot and pooled estimates for genotype–disease and genotype–phenotype associations (for references see data set available at http://www.prw.le.ac.uk/research/HCG/MTHFRdataSet.html).

first or second outcome we use the corresponding univariate normal distributions.

$$x_i \sim \mathrm{N}(\mu_{xi}, v_{xi}) \quad \text{and} \quad \mu_{xi} \sim N(\mu_x, \tau_x)$$

and

$$y_i \sim \mathrm{N}(\mu_{yi}, v_{yi}) \quad \text{and} \quad \mu_{yi} \sim \mathrm{N}(\mu_y, \tau_y)$$

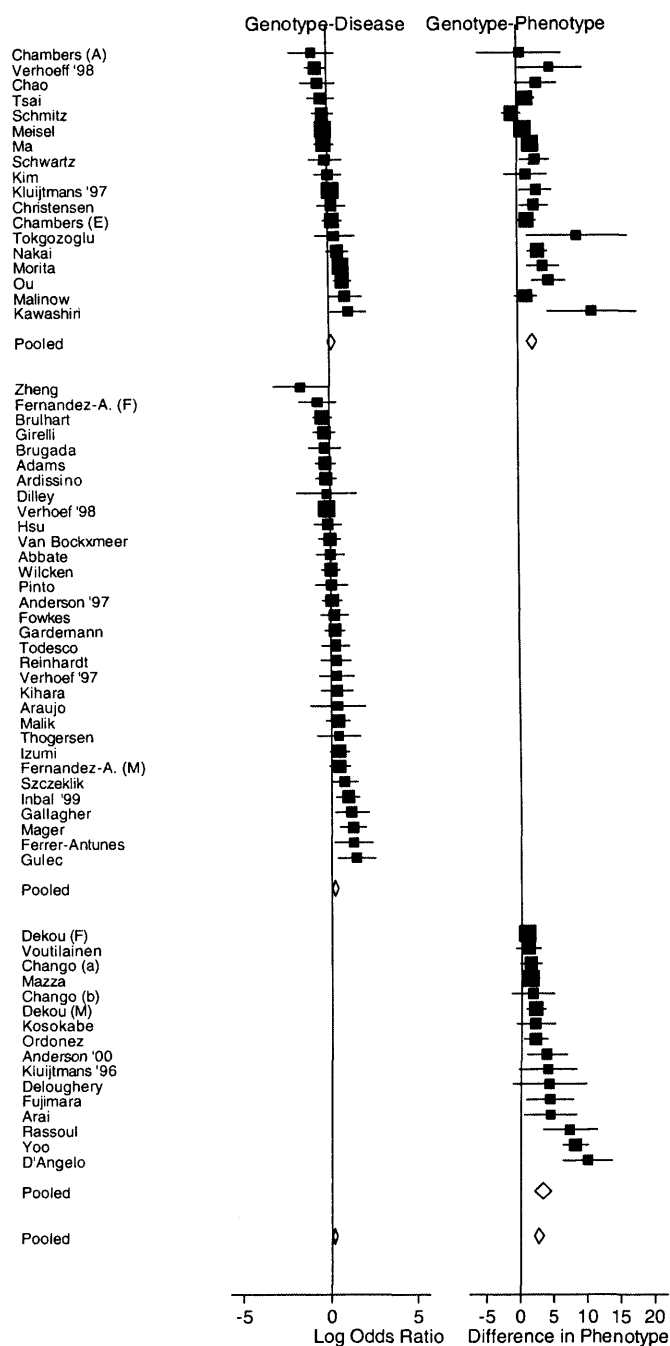In models for meta-analysis in Mendelian randomization x will represent the log OR of disease given genotype and y will represent the mean difference in phenotype. The object will be to estimate the ratio of x to y as this will give the log OR of the effect of phenotype on disease. The literature will provide the study estimates and their variances v_{xi} and v_{yi}, which may be reported implicitly in the form of standard errors or confidence intervals. We will assume that these variances are known without error.

3.1. Within-study and between-study correlations

The within-study correlation, ψ, represents the possibility that when a particular study estimates one measure as being larger than its actual value then there may be a tendency for the other estimate also to be larger (or smaller) than its true value. This correlation can only be estimated from individual patient data, and is never reported by primary studies. However there are two good theoretical reasons to suppose that the within-study correlation will be negligible in most studies used in a Mendelian randomization. The study-specific OR of genotype on disease is based on aggregated data and in most studies phenotype level is only measured in a subsample of the subjects.

A small simulation study was performed to assess the likely size of the within-study correlation. This simulated study consisted of the same number of cases and controls ($n = 100$). The probability of disease given the phenotype, was assumed to be controlled by a logistic function of the phenotype with parameters α and β. The distribution of the phenotype over the population was assumed to be normal with a fixed genotype effect on phenotype. Repeated studies were simulated under identical conditions and the correlation between the estimated the log OR for the genotype–disease association and the mean difference in phenotype was calculated. The parameters were chosen to reflect the values in studies of the *MTHFR*–homocysteine–CHD pathway. Initially the following parameters were considered; an OR per unit change in phenotype on disease of 1.1 [$\beta = \log(1.1) = 0.095$] and a difference in phenotype due to genotype of 4 units, which give an OR of genotype on disease of 1.46; a frequency of mutant genotype of 12 per cent; a between-subject variation in levels of phenotype that is normally distributed with parameters $\mathrm{N}(8,2)$ or $\mathrm{N}(12,2)$ depending on genotype; a baseline risk of disease (α) with mean of -4 and standard deviation of 0.2, reflecting unmeasured covariates. Under these conditions 3.9 per cent of people with the wildtype genotype and 5.6 per cent of those with the mutant genotype develop the disease. As recommended [6] the effect of genotype on phenotype was estimated from the controls only.

From 5000 repeated simulations the correlation between the estimated log OR of genotype on disease and the mean difference in phenotype was 0.015. Using further sets of 5000 simulations, the basic situation was altered by changing one parameter (sample size 50 or 100, phenotype difference 2 or 6, OR 1.05 or 1.2, baseline risk of disease -3) and keeping the others fixed at the values given above. In all of these simulations the estimated correlation was within the range ± 0.05.

The between-study correlation, ρ, represents the tendency for studies conducted in populations where the true effect of genotype on phenotype is large also to show a larger than average effect of genotype on disease. In contrast to the within-study correlation, for a meta-analysis of studies with a wide range of populations and designs this correlation may well be substantial.

3.2. Model A—heterogeneity on genotype–phenotype and genotype–disease

In the first model, represented schematically in Figure 3(a), we adopt the hierarchical bivariate normal distributions described at the start of Section 3. Without individual data it is not possible to estimate the within-study correlation, ψ, but for the reasons set out above ψ will be very small and so we assume $\psi = 0$. Denoting the log OR of phenotype on disease by $\theta = \mu_x/\mu_y$, the marginal distribution becomes:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \theta\mu_y \\ \mu_y \end{bmatrix}, \begin{bmatrix} v_{xi} + \tau_x & \rho\sqrt{\tau_x\tau_y} \\ \rho\sqrt{\tau_x\tau_y} & v_{yi} + \tau_y \end{bmatrix} \right)$$

When only one of the pair of estimates is reported we treat them as univariate normal. Thus x_i is normally distributed with mean $\theta\mu_y$ and variance $v_{xi} + \tau_x$, or y_i is normally distributed with mean μ_y and variance $v_{yi} + \tau_y$.

The model has five parameters $(\theta, \mu_y, \tau_x, \tau_y, \rho)$, and numerical maximization of the log likelihood using the ml command in Stata [16] is based on the transformed parameter values $(\theta, \mu_y, \log(\tau_x), \log(\tau_y), \log[(1 + \rho)/(1 - \rho)])$. This avoids the need to check the natural range constraints on the parameters, such as τ_x and τ_y which need to be greater than 0 if not log-transformed. Starting values were calculated from the raw data as

$$\theta = \frac{\bar{x}}{\bar{y}}, \quad \mu_y = \bar{y}, \quad \tau_x = \text{Var}(x), \quad \tau_y = \text{Var}(y), \quad \rho = \text{Corr}(x, y)$$

For the *MTHFR* data maximization was started at $\theta = 0.06$, $\mu_y = 3.33$, $\tau_x = 0.35$, $\tau_y = 7.75$, $\rho = 0.51$ and converged to the values reported in Table I. This solution corresponds to an OR of 1.53 (95 per cent CI: 1.21–1.93) for the effect of a 5 µmol/l increase in homocysteine on CHD. The between-study correlation is poorly estimated, and the log likelihood increases with ρ with a maximum at $\rho = 1$.

Under Model A the final estimate of the OR depends on the between-study correlation, which is poorly estimated. This was demonstrated by sensitivity analyses that fitted the model with a range of different known values for the correlation. The results varied from an OR
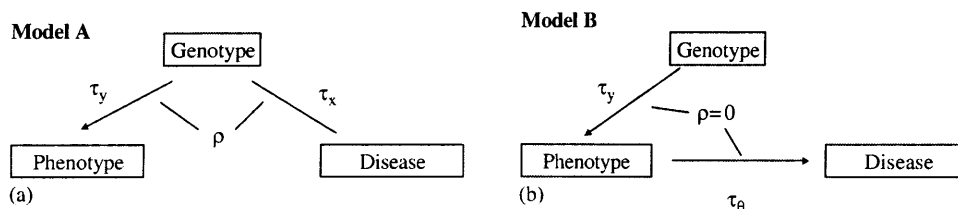

Figure 3. (a) Modelling the heterogeneities of the associations in Model A; and (b) modelling the heterogeneities of the associations in Model B.

Table I. Results of Model A and B fitted to the *MTHFR* data, using the maximum-likelihood approach.

Model	Parameter	Estimate	95 per cent CI
A	θ	0.085	0.038–0.132
	μ_y	2.678	1.989–3.366
	τ_x	0.127	0.056–0.287
	τ_y	3.083	1.513–6.281
	ρ^*	1.000	—
B	θ	0.087	0.033–0.141
	μ_y	2.714	1.932–3.495
	τ_x	0.009	0.002–0.035
	τ_y	3.370	1.589–7.148

* The estimate is at the extreme of its range and the 95 per cent CI cannot be calculated.

of 1.42 (95 per cent CI: 1.09–1.85) obtained when the between-study correlation was 0 to 1.53 (95 per cent CI: 1.21–1.93) when the correlation was 1. It will usually be difficult to use Model A because the results depend on the between-study correlation and there is unlikely to be sufficient information to estimate the correlation accurately.

3.3. Model B—heterogeneity on genotype–phenotype and phenotype–disease

A second model, referred to as Model B, overcomes the limitations of Model A by modelling the heterogeneities of the associations in an alternative manner. The three stages in the triangle shown in Figure 1, namely genotype–phenotype, phenotype–disease and genotype–disease will all be subject to heterogeneity but under the causal model implicit in that figure we may derive any one from the other two. In Model B we parameterize the heterogeneities on the genotype–phenotype and phenotype–disease stages and critically assume that they are independent as illustrated in Figure 3(b). Even under this model, correlation will still be induced in the resultant heterogeneities on genotype–phenotype and genotype–disease. Independence implies that studies that find a large effect of genotype on phenotype will not tend to find relatively larger or smaller effects of the phenotype on disease. In fact, it is unlikely that the effect of a specific level of phenotype on disease would depend on the cause of that level (e.g. genotype rather than any other cause). This is almost certainly reasonable in the case of *MTHFR*, homocysteine and CHD, and is very likely to hold in most other cases.

Assuming bivariate normal distributions we have a hierarchical model with, at the lower level, means μ_{xi} and μ_{yi}, variances v_{xi} and v_{yi}, and 0 within-study correlation. Suppose that θ_i and μ_{yi} represent the log OR of phenotype on disease and the average change in phenotype within the ith study and that the corresponding population means are θ and μ_y. If the between-study variances are τ_θ and τ_y, respectively, and there is zero between-study correlation, then since $\mu_{xi} = \theta_i \mu_{yi}$, we can use a Taylor series approximation (delta method) [17] to derive the mean and variances of the two observed measures; genotype–phenotype and genotype–disease. If we approximate their distribution by a bivariate normal we obtain:

$$\begin{bmatrix} \mu_{xi} \\ \mu_{yi} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \theta\mu_y \\ \mu_y \end{bmatrix}, \begin{bmatrix} \mu_y^2\tau_\theta + \theta^2\tau_y & \theta\tau_y \\ \theta\tau_y & \tau_y \end{bmatrix} \right)$$

so that:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \theta\mu_y \\ \mu_y \end{bmatrix}, \begin{bmatrix} v_{xi} + \mu_y^2\tau_\theta + \theta^2\tau_y & \theta\tau_y \\ \theta\tau_y & v_{yi} + \tau_y \end{bmatrix} \right)$$

As before, when only one of the pair is observed we treat them as univariate normal; thus x_i is normally distributed with mean $\theta\mu_y$ and variance $v_{xi} + \mu_y^2\tau_\theta + \theta^2\tau_y$, or y_1 is normally distributed with mean μ_y and variance $v_{yi} + \tau_y$.

This new model has four parameters $(\theta, \mu_y, \tau_0, \tau_y)$, and numerical maximization of the log likelihood is based on transformed parameter values θ, $\mu_y, \log(\tau_\theta), \log(\tau_y)$. To obtain initial values we take:

$$\theta = \frac{\bar{x}}{\bar{y}}, \quad \mu_y = \bar{y}, \quad \tau_\theta = \text{Var}\left(\frac{x}{y}\right), \quad \tau_y = \text{Var}(y)$$

For the *MTHFR*–homocysteine–CHD example, maximization starts at $\theta = 0.06$, $\mu_y = 3.33$, $\tau_\theta = 0.31$, $\tau_y = 7.75$ and converges to the values reported in Table I. These figures lead to an estimate of the induced correlation between true study-specific means of genotype–phenotype and genotype–disease of $\rho = 0.52$. The result of Model B expressed as the effect of a 5 μmol/l increase in homocysteine on CHD is an OR of 1.54 (95 per cent CI: 1.18–2.02).

A sensitivity analysis allowing a small known amount of within-study correlation found that with $\psi = -0.1$ the OR of a 5 μmol/l increase in homocysteine was 1.56 (95 per cent CI: 1.18–2.05) while with $\psi = +0.1$ the OR was 1.53 (95 per cent CI: 1.17–2.00). Thus even if there is a small amount of within-study correlation this has little effect on the final estimate.

The main difference between Models A and B is that B adds the assumption that the heterogeneity in the mean difference in phenotype with genotype is independent of the heterogeneity in the log OR of phenotype on disease. Under Model B, a plot of the true log OR of genotype on disease for each study, $\mu_{xi} = \theta_i\mu_{yi}$, against the true average difference in phenotype μ_{yi}, would produce points that would lie about a straight line passing through the origin with slope θ, and the correlation would be approximately $\theta\sqrt{(\tau_y/\tau_x)}$. Using the parameter estimates from the fitting of Model A (Table I) this correlation should be about $0.085 * \sqrt{(3.083/0.127)} = 0.42$, similar to the implied correlation when Model B is fitted. The fact that Model A goes on to find an implausible solution with a correlation of one is partly a reflection of the lack of data to estimate that correlation in the absence of extra assumptions, and partly due to the detection of some non-linearity. The maximized log likelihood for Model A is −37.3 and for Model B is −39.2. As there is one extra parameter in Model A the choice of model is not clear cut but we prefer Model B since it accords more closely with the pattern of between-study variation that we expected to see.

3.4. Bayesian approach to parameter estimation

Equivalent models have also been developed within a Bayesian framework, using MCMC methods as implemented by WinBUGS software 1.4 [18], with 'vague' prior distributions adopted for all model parameters, such that they contain relatively little information in comparison to the likelihood [19]. Convergence was assessed via sensitivity analyses with respect to initial values, length of 'burn-in' and length of sample, using both visual inspection of trace plots and by the Geweke, Heidelberger and Welch, and the Raftery and Lewis diagnostic tests [20] implemented in BOA [21]. The WinBUGS code is available on our website at http://www.prw.le.ac.uk/research/HCG/gebugs.html

3.4.1. Model A. When adopting a Bayesian approach, the same structure as for the maximum likelihood approach was followed. For unknown mean parameters vague normal prior distributions, with mean 0 and variance 1000, were used. A Wishart prior distribution was adopted for the inverse covariance matrix, in which degrees of freedom were chosen to be the rank of the covariance matrix in order to obtain a vague prior distribution [22]. *A priori* beliefs regarding the expectation of the covariance matrix were specified such that the variances of μ_{xi} and μ_{yi} were 0.35 and 7.75, respectively, and the corresponding prior correlation was 0.5, all of which were considered *a priori* plausible values. The result of the Bayesian model expressed as the effect of a 5 μmol/l increase in homocysteine on CHD is similar to that of the maximum likelihood approach, with an OR of 1.50 (95 per cent CrI: 1.15–1.99). However, a problem similar to that encountered in the maximum likelihood approach for the correlation estimation was present, with the results of the model being sensitive to the values specified for the hyperparameters chosen to represent *a priori* beliefs regarding the expected covariance matrix in the Wishart prior distribution.

3.4.2. Model B. When using a Bayesian approach, the specification of Model B, unlike Model A, is slightly different from that used for the maximum likelihood approach, and uses a modified product normal formulation [23]. Here we do not need to approximate the distribution of the genotype–disease association by a normal distribution, as we did in the maximum likelihood approach in order to use the delta method. Indeed, strictly speaking, this distribution cannot be normal being the product of two distributions, which we assumed to be normal (for the genotype–phenotype and phenotype–disease associations). Unknown mean parameters have vague normal prior distributions with mean 0 and variance 1000, whilst unknown variance parameters have inverse gamma prior distributions with parameters 0.001 and 0.001. Moreover, the Bayesian specification of Model B uses a slightly different method to deal with those studies measuring only one association (either genotype–disease or genotype–phenotype). In these studies, the association that has not been evaluated is treated as missing at random, and missing values are sampled from the corresponding predictive distributions. Thus, all 66 studies are modelled in a single step for both genotype–disease and genotype–phenotype associations.

The result for the Bayesian approach expressed as the effect of a 5 μmol/l increase in homocysteine on CHD is an OR of 1.54 (95 per cent CrI: 1.17–2.06). Despite the differences in the two approaches in terms of approximations and structure, the results are very similar.

3.5. Effect of ignoring correlation

Multivariate models should be used whenever the outcomes of interest are correlated and measured within the same studies. However, the potential gain in terms of increased precision or reduced bias may be small, depending on the proportion of studies that measure both outcomes. When ignoring correlation in the example of the *MTHFR* gene, homocysteine and CHD, thus carrying out two independent meta-analyses on genotype–disease and genotype–phenotype data and then estimating the phenotype–disease association based on the two pooled estimates, the OR for a 5 μmol/l increase in homocysteine was 1.43 (95 per cent CI: 1.10–1.95) and 1.44 (95 per cent CrI: 1.10–1.94), for the classical and Bayesian meta-analyses, respectively, as compared with 1.54 (95 per cent CI: 1.18–2.02) and 1.54 (95 per cent CrI: 1.17–2.06) of the multivariate meta-analysis. The confidence interval for the classical

approach when ignoring correlation is calculated based on the uncertainty in the estimates of both genotype–disease and genotype–phenotype associations [24]. In our example, the gain in precision between univariate and multivariate models is minimal, however in other examples the gain might be greater.

The hierarchical nature of the model could be extended if it were thought that the assumption of partial exchangeability was more appropriate [19]: i.e. conditional upon study-level covariates, for example, year of study.

4. DISCUSSION

If the assumptions of Mendelian randomization are met [1], then it is possible to derive an unconfounded estimate of the size of the phenotype–disease association. The most important assumption behind the triangulation shown in Figure 1 for our case-study is that the *MTHFR* polymorphism only influences the risk of CHD through its impact on homocysteine and not by any other pathway. We also assume that *MTHFR* is not in linkage disequilibrium with any gene that influences CHD or homocysteine and that the relationship between the ratio of log $OR_{GG\,vs\,gg}$ and the mean difference in phenotype, ΔP, is constant across studies even though the size of the phenotype difference may vary. In our case-study these assumptions are reasonable [6] and therefore the estimated association between homocysteine and CHD is believed to be causal. Clearly it is advisable to limit the use of Mendelian randomization to examples where there is a good understanding of the biological pathway from gene to phenotype to disease.

When synthesising evidence of genetic studies for use in a Mendelian randomization analysis, studies evaluating genotype–phenotype, genotype–disease or both associations together are likely to be encountered. Simulations suggest that in this situation the within-study correlation is likely to be very small, but it is still important to allow for the between-study correlation in the heterogeneities of studies that evaluate both genotype–phenotype and genotype–disease associations. Heterogeneities on the genotype–phenotype and genotype–disease associations may be highly correlated but a multivariate model that parameterizes the heterogeneity directly (Model A) is difficult to fit because the correlation is poorly estimated. We advocate an alternative approach that treats the heterogeneities on genotype–phenotype and phenotype–disease as being independent (Model B). This model fits readily and implicitly defines the correlation between the heterogeneities on genotype–phenotype and genotype–disease.

In our paper we adopted both a maximum likelihood and a Bayesian approach with vague prior distributions for all parameters. Although no differences were shown in any of the results, there are differences in the statistical paradigms and in the subsequent interpretation of the results. MCMC was used for parameter estimation in the Bayesian approach due to the analytical intractability. The use of MCMC avoided the requirement for Taylor series approximations, which might have an impact on the results, although they made no difference in our example. Although MCMC methods could also be used for maximum likelihood inference [25], we found good performance using the Newton–Raphson algorithm together with approximate variances. When using a Bayesian approach re-parameterization may not be theoretically necessary as the posterior distribution of functions of the model parameters may be obtained directly from the MCMC samples. It can nevertheless, depending on the precise sampling algorithm used [26], be desirable to re-parameterize in order to improve performance of the

MCMC algorithm, especially in a hierarchical or non-linear model setting [27]. An additional issue raised by the use of a Bayesian approach is the choice of 'vague' or non-informative prior distributions [19, 28]. This is particularly important for the prior on the between-study variance in a meta-analysis context, to which the results may be sensitive [19, 29–31]. The difference in the interpretation of the results of the two approaches is reflected by the different meaning of the confidence/credibility intervals, and, more importantly, by the possibility offered by Bayesian methods to allow direct probability statements regarding the estimate of the effect of interest. Finally, although we did not consider it, a fully Bayesian approach with prior distributions based on data from other pertinent studies [32–34] or expert opinion and other evidence could be adopted [19]. However, elicitation of beliefs regarding the model parameters in such meta-analysis (e.g. correlations) is not straightforward [35, 36].

Mendelian randomization is an example of the use of an instrumental variable; a technique that has been used in econometrics since the 1920s but only occasionally applied to health sciences to control for confounding and measurement error [37, 38]. These methods derive an unconfounded estimate of the association between the exposure and outcome of interest from the observed relationship of an instrumental variable with both exposure and outcome. The use of instrumental variables, despite their potential role in epidemiology, has been mainly limited to measurement error [39] and to the field of randomized clinical trials. A typical application is to use the allocated treatment as an instrumental variable to control for the bias due to non-compliance as an alternative to the more usual analysis by intention-to-treat [38]. A likely explanation for the limited use of these methods is that it is often difficult to find suitable instrumental variables, since the method requires the variables have not only an unconfounded relation with both the exposure and the outcome of interest, but also an association with the outcome that is explained by the association with the exposure. In some situations Mendelian randomization allows the subject's genotype to be used as an instrumental variable, with genetic studies providing information on the impact of a specific mutation on both the phenotype (risk factor) and disease of interest. If the gene is carefully chosen then both associations are unconfounded because the genotype is effectively randomly assigned, but the important assumption that the genotype is associated to the disease only through the phenotype of interest needs to be carefully assessed [1].

There are many other issues regarding the synthesis of genetic studies using Mendelian randomization that might be considered and developed as extensions of the model presented. An important issue for any meta-analysis that is particularly relevant in genetic epidemiology [2, 40], is that of dissemination bias. Not only do we need to be concerned about the possibility of publication bias but also of reporting bias, due to researchers choosing not to report either a log OR or a phenotype difference when both are measured but one is not consistent with some accepted theory. An integrated meta-analytical approach can start addressing these issues by comparing the pooled estimates and the funnel plots for genotype–phenotype and genotype–disease associations in studies reporting either only one or both associations [6]. It is important to note that, whilst all meta-analyses are based on the assumption of missing at random for those studies which might not have been included (no publication bias), in the case of Mendelian randomization we also assume that in those studies reporting only one estimate, the other is also missing at random (no reporting bias). A valuable extension would be to model explicitly the reporting and publication bias mechanism in the meta-analysis [41, 42]. In practise, however, explicit information about these mechanisms is usually lacking and strong assumptions about reporting and publication mechanisms operating are likely to be required.

Another important extension would be to include the information on genotype–disease association obtained from the analysis of subjects heterozygous (rather than only homozygous) for the polymorphism. The use of Mendelian randomization could also be extended to more complex situations where several genotypes influence the disease through the same phenotype or where a single genotype influences the disease through different phenotypes.

ACKNOWLEDGEMENTS

REFERENCES

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**:1–22.
2. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *Journal of Clinical Epidemiology* 2003; **56**:297–303.
3. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genetics in Medicine* 2002; **4**:45–61.
4. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358**:1356–1360.
5. Keavney B. Genetic epidemiological studies of coronary heart disease. *International Journal of Epidemiology* 2002; **31**:730–736.
6. Minelli C, Thompson JR, Tobin MD, Abrams KR. An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology* 2004; **160**:445–452.
7. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**:2309–2333.
8. Arends LR, Voko Z, Stijnen T. Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* 2003; **22**:1335–1353.
9. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**:537–557.
10. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993; **12**:2273–2284.
11. Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. *Psychological Bulletin* 1988; **103**:111–120.
12. Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *Journal of the American Medical Association* 2002; **288**:2015–2022.
13. Bailey LB, Gregory III JF. Polymorphisms of methylenetetrahydrofolate reductase and other enzymes: metabolic significance, risks and impact on folate requirement. *Journal of Nutrition* 1999; **129**:919–922.
14. Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, Schouten EG. MTHFR 677C→T polymorphism and risk of coronary heart disease: a meta-analysis. *Journal of the American Medical Association* 2002; **288**:2023–2031.
15. Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *British Medical Journal* 2002; **325**:1202.
16. Gould W, Sribney W. *Maximum Likelihood Estimation with Stata*. Stata Press: College Station, TX, 1999.
17. Kendall M, Stuart A. *The Advanced Theory of Statistics: Distribution Theory*, vol. 1. Griffin: London, 1973.
18. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS User Manual, Version 1.4*. MRC Biostatistics Unit, 2001.
19. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Wiley: Chichester, U.K., 2004.
20. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 1996; **91**:883–904.
21. Smith BJ. Bayesian output analysis program (BOA). *Version 1.1 User's Manual*, 2004.
22. Carlin BP. Hierarchical longitudinal modelling. In *Markov Chain Monte Carlo In Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London, 1996; 303–319.
23. Spiegelhalter DJ. Bayesian graphical modelling: a case study in monitoring health outcomes. *Journal of the Royal Statistical Society, Series C* 1998; **47**:115–133.
24. Thompson JR, Tobin MD, Minelli C. On the accuracy of estimates of the effect of phenotype on disease derived from Mendelian randomization studies. *Technical Report* 2003_GE1. Available at http://www.prw.le.ac.uk/research/HCG/getechrep.html

25. Geyer CJ. Estimation and optimization of functions. In *Markov Chain Monte Carlo In Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London, 1996; 241–258.
26. Brooks SP. Markov chain Monte Carlo method and its applications. *Journal of the Royal Statistical Society, Series D* 1998; **47**:69–100.
27. Gelfand AE, Sahu SK, Carlin BP. Efficient parameterizations for normal linear mixed models. *Biometrika* 1995; **82**:479–488.
28. O'Hagan A. Bayesian inference. In *Kendall's Advanced Theory of Statistics*, vol. 2B. Hodder Arnold: London, 1994.
29. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**:277–303.
30. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study to investigate the use of 'vague' prior distributions in MCMC using WinBUGS. *Statistics in Medicine* (in press).
31. Gelman A. Prior distributions for variance parameters in hierarchical models. *Technical Report*, Department of Statistics, Columbia University, 2004. Available at http://www.stat.columbia.edu/~gelman/research/unpublished/tau7.pdf
32. Abrams KR, Lambert PC, Sanso B, Shaw C, Marteau T. Meta-analysis of heterogeneously reported study results—a Bayesian approach. In *Meta-analysis in Medicine and Health Policy*, Chapter 2, Berry D, Stangl D (eds). Marcel Dekker: New York, 2000; 29–63.
33. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**:2733–2749.
34. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
35. Garthwaite PH, Dickey JM. Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B* 1988; **50**:462–474.
36. Gokhale DV, Press SJ. Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Journal of the Royal Statistical Society, Series A* 1982; **145**:237–249.
37. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**:722–729.
38. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–455.
39. Hardin JW, Carroll RJ. Variance estimation for the instrumental variables approach to measurement error in generalized linear models. *The Stata Journal* 2003; **3**:342–350.
40. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; **361**:865–872.
41. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: Chichester, U.K., 2000.
42. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal* 2001; **323**:101–105.

Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies

Cosetta Minelli[*,†,‖] John R. Thompson[‡,‖] Keith R. Abrams[§] and Paul C. Lambert[¶]

Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, U.K.

SUMMARY

A genetic model-free method for the meta-analysis of genetic association studies is described that estimates the mode of inheritance from the data rather than assuming that it is known. For a bi-allelic polymorphism, with G as risk allele and g as wild-type, the genetic model depends on the ratio of the two log odds ratios, $\lambda = \log OR_{Gg} / \log OR_{GG}$, where OR_{GG} compares GG with gg and OR_{Gg} compares Gg with gg. Modelling $\log OR_{GG}$ as a random effect creates a hierarchical model that can be implemented within a Bayesian framework.

In Bayesian modelling, vague prior distributions have to be specified for all unknown parameters when no external information is available. When the data are sparse even supposedly vague prior distributions may have an influence on the posterior estimates. We investigate the impact of different vague prior distributions for the between-study standard deviation of $\log OR_{GG}$ and for λ, by considering three published meta-analyses and associated simulations. Our results show that depending on the characteristics of the meta-analysis the results may indeed be sensitive to the choice of vague prior distribution for either parameter.

Genetic association studies usually use a case-control design that should be analysed by the corresponding retrospective likelihood. However, under some circumstances the prospective likelihood has been shown to produce identical results and it is usually preferred for its simplicity. In our meta-analyses the two likelihoods give very similar results. Copyright © 2005 John Wiley & Sons, Ltd.

---

[*]Correspondence to: Cosetta Minelli, Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LE1 6TP, U.K.
[†]E-mail: cm109@le.ac.uk
[‡]E-mail: trj@le.ac.uk
[§]E-mail: kral@le.ac.uk
[¶]E-mail: pl4@le.ac.uk
[‖]Joint First Authors.

1. INTRODUCTION

1.1. Bayesian models for meta-analysis

Meta-analysis is the quantitative synthesis of results from multiple studies [1]. If the results for the individual studies are similar they can be pooled using a fixed effects model, but where the studies show heterogeneity it is important to use a random effects model that allows for the between-study variability [1]. Ideally the reasons for the heterogeneity should then be investigated [2]. A random effects meta-analysis is an example of a two-level hierarchical (or multilevel) model. By assuming exchangeability between studies, each can 'borrow strength' from the others. This results in the estimated effects for the individual studies being shrunk towards the overall mean and usually gives increased precision. More importantly, the precision of the estimate of that overall mean will allow for the variability between studies [3, 4].

The Bayesian approach to hierarchical modelling has been advocated for theoretical and practical reasons [4]. Bayesian analysis allows researchers to use external information either in the form of subjective beliefs or based on other data sources, and enables them to produce subjective probability statements about the model parameters [4]. Fitting Bayesian models by Markov chain Monte Carlo (MCMC) methods is particularly flexible and makes it practical to use relatively complex hierarchical models while allowing for uncertainty in all parameters [5, 6]. Although possible, inclusion of uncertainty for variance parameters in an equivalent likelihood analysis is not straightforward and this source of variability is often ignored [7].

The main practical problem in undertaking a Bayesian meta-analysis is to specify appropriate prior distributions for the model parameters based on external information. When expert opinion is available it can be difficult to use this to derive probability distributions, especially for parameters such as variances or correlations [8, 9]. When there is no external information we are left with the equally difficult problem of trying to specify non-informative prior distributions for all of the model parameters, including the hyperparameters if the model is hierarchical [3]. Although a number of such prior distributions have been proposed, and routinely used, strictly speaking 'non-informative' prior distributions, i.e. prior distributions that formally represent ignorance, and thus do not favour any particular parameter values, do not exist [10, 11]. In fact, *any* prior distribution exerts some influence on the shape of the posterior distribution, the more so in the presence of sparse data. The real aim is to identify a prior distribution that has minimal effect on the final inference relative to the data [12]. For this reason, the term non-informative prior distribution is better replaced by 'vague' prior distribution [13, 14], which indicates a density with high spread that gives similar prior probability to a wide range of parameter values. The problem of choosing vague prior distributions has been demonstrated to be particularly critical for hierarchical variance parameters, since prior distributions proposed as vague might in fact influence the analysis due to limited data [13, 15–17]. Thus, sensitivity analyses that examine the robustness of the choice of prior distributions are an essential part of a Bayesian hierarchical analysis.

1.2. Meta-analysis of genetic association studies: a genetic model-free approach

The meta-analysis of genetic association studies introduces specific methodological problems [18], among which the most characteristic is the presence of at least three possible

genotypes as exposure groups and the fact that these are related by the underlying mode of inheritance. In the simplest case of a bi-allelic genetic polymorphism, with a wild-type allele, g, and a mutant allele, G, thought to be associated with the disease of interest, association studies will collect information on the relative frequency of disease in subjects with each of the three genotypes (gg, Gg and GG). There are thus two relative risks or odds ratios to be estimated, GG and Gg, each compared with the wild genotype gg. The relationship between these two relative risks is dependent on the mode of inheritance, also called the genetic model.

Methods currently used for the meta-analysis of such studies usually reduce the three groups to two by assuming a specific genetic model and thus combining the groups accordingly (e.g. assuming a recessive model to justify combining Gg and gg) or assigning to the heterozygous group Gg half the effect, on the log scale, of the GG group (co-dominant model or 'per-allele' analysis) [18].

Since the underlying genetic model is usually not known, we propose a method that avoids the assumption of a specific genetic model, but which takes into account the correlation between the two estimates of the odds ratios [19]. The model treats the log odds ratio of Gg versus gg ($\log OR_{Gg}$) as an unknown proportion, λ, of the log odds ratio of GG versus gg ($\log OR_{GG}$), i.e. $\lambda = \log OR_{Gg}/\log OR_{GG}$, and thus $OR_{Gg} = [OR_{GG}]^{\lambda}$. Under this model the ratio, λ, is assumed constant across studies. We refer to this approach as 'genetic model-free', a term already in use in genetic epidemiology to indicate that no underlying *genetic* model is assumed, though the analyses are still based on an assumed *statistical* model. Values of λ equal to $0, 0.5$ and 1 correspond to the recessive, co-dominant and dominant genetic model, respectively, but we allow λ to take any value between 0 and 1. In very rare situations a gene may be over-dominant, that is the risk of the Gg group can be higher or lower than either of the homozygous groups [20]. This would be characterized by values of λ higher than 1 or lower than 0. This rare situation is not considered in this particular investigation.

1.3. Overview of the paper

In this paper we consider the statistical aspects of a Bayesian implementation of the genetic model-free approach by applying the model to three previously published meta-analyses and to simulations based on those three scenarios.

When adopting a Bayesian approach to a hierarchical model that allows heterogeneity in OR_{GG}, unless there is prior knowledge, we have the problem of specifying vague prior distributions for the between-study variance of $\log OR_{GG}$ and for the parameter λ. Since we rule out the over-dominant case in our examples, we need vague prior distributions for λ which are constrained to cover the range between 0 and 1. The sensitivity of the analysis to the choice of vague prior distributions is investigated for the three published meta-analyses introduced in Section 2 and then for the simulated data in Section 3.

By far the majority of genetic association studies use a case–control design that requires a retrospective likelihood based on the probability of exposure given disease. Prentice and Pyke [21] showed that a maximum likelihood analysis based on the corresponding prospective likelihood gives the same results as an analysis of the retrospective likelihood for a single study. Because the form of the prospective likelihood is simpler, it is very widely used. Equivalence within the Bayesian framework does not generally exist and has only been established for very particular choices of prior distributions [22]. Although not exactly equivalent, the results of Prentice and Pyke would suggest that with vague prior distributions the

retrospective and prospective Bayesian analyses should give similar answers. This issue is investigated in the context of meta-analysis in Section 4.

2. ILLUSTRATIVE EXAMPLES

2.1. Illustrative meta-analyses

For illustrative purposes, the genetic model-free method and all sensitivity analyses are applied to three previously published meta-analyses. In all cases, the polymorphism is bi-allelic, and we will call the two alleles G and g, where G is the one thought to be associated with the disease. The examples are

(a) *AGT* gene and essential hypertension, reported by Kato *et al.* [23]. The meta-analysis includes 7 case–control studies, with an average number of cases and controls per study of 191 and 175, respectively, and an average frequency of the G allele of 0.75.

(b) *KIR6.2* gene and type II diabetes, reported by Hani *et al.* [24]. This meta-analysis includes 4 case-control studies, with an average number of cases and controls per study of 130 and 92, respectively, and an average frequency of the G allele of 0.34.

(c) *PON1* Q192R polymorphism and myocardial infarction, reported by Wheeler *et al.* [25]. This is a meta-analysis including 19 studies, 17 of which are case–control and 2 cohort studies. The average number of cases and controls per study is 301 and 424, respectively, and the average frequency of the G allele is 0.33.

The variation between the three meta-analyses in terms of the number and sizes of the studies and the frequency of the allele of interest allows us to assess whether the sensitivity to the choice of prior distributions varies according to these characteristics.

2.2. The meta-analytical model based on retrospective likelihood

The meta-analysis is based on a retrospective likelihood that mirrors the method of sampling in case–control studies. Subjects are selected dependent on their disease status and then their exposure status is ascertained.

Denoting by y_{0j} and y_{1j} the number of controls and cases, respectively, in genotype group j, with $j = 1,2,3$ (corresponding to gg, Gg and GG), the retrospective likelihood (L_R) for each study included in the meta-analysis is derived from a pair of multinomial distributions

$$y_{0j} \sim \text{Multinomial}(n_0, p_{0j}) \quad y_{1j} \sim \text{Multinomial}(n_1, p_{1j})$$

where n_0 and n_1 are the total number of controls and cases, respectively

$$p_{dj} = \frac{\beta_j \exp(d\delta_j)}{\sum_{k=1}^{3} \beta_k \exp(d\delta_k)}, \quad j = 1,2,3$$

and d is an indicator of the disease status, taking the value of 0 for controls and 1 for cases. The probability that a control has exposure j is $\beta_j / \sum_{k=1}^{3} \beta_k$, with $\beta_1 = 1$. The log odds ratios of disease for the exposure groups Gg and GG compared to no exposure (gg) are represented by δ_2 and δ_3, respectively, while δ_1 is zero by definition. The likelihood for each study will

thus take the form

$$L_{\mathrm{R}}(\beta, \delta; y) = \prod_{d=0}^{1} \prod_{j=1}^{3} \left\{ \frac{\beta_j \exp(d\delta_j)}{\sum_{k=1}^{3} \beta_k \exp(d\delta_k)} \right\}^{y_{dj}} \tag{1}$$

In the meta-analysis, the full likelihood is then obtained as the product of likelihoods (1) over the i studies, under the assumption of independence of the studies. The study-specific log odds ratios for GG *versus* gg, δ_{3i}, are modelled as normally distributed random effects parameters, which vary about an overall mean, θ, with variance, τ^2

$$\delta_{3i} \sim \mathrm{N}(\theta, \tau^2)$$

The study-specific log odds ratios for Gg *versus* gg, δ_{2i}, are equal to the product of δ_{3i} and λ, i.e. $\lambda = \delta_{2i}/\delta_{3i}$, and the mode of inheritance, λ, is assumed constant across studies and thus modelled as a fixed effect parameter. It is usually not possible to model both δ_{3i} and λ as random effects because, without extra information, it is very difficult to simultaneously estimate the heterogeneity of the two parameters. However, if there are reasons to believe that λ differs across populations, the model could be generalized to include subgroups of studies within which λ is constant.

Prior distributions have to be specified for the unknown model parameters θ, τ and λ. While a diffuse normal distribution is used for θ in all models, i.e. $\theta \sim \mathrm{Normal}(0, 10\,000)$, prior distributions for τ and λ are discussed in the following paragraphs. Corresponding posterior distributions are obtained using MCMC methods implemented using WinBUGS 1.4.1 [26], and details on the WinBUGS code for fitting this model can be found on our website, www.hs.le.ac.uk/research/HCG/AppendixSiM2005.doc. The number of simulations was varied and the traces were inspected for evidence of non-convergence before deciding on a 'burn-in' of 5000 iterations followed by chains of length 10 000.

2.2.1. Prior distributions for the heterogeneity term.

Three prior distributions were considered for the between-study standard deviation, τ. Figure 1(a) shows the densities for the three prior distributions, all presented on the standard deviation scale.

The first prior distribution is a gamma distribution for the precision parameter (the inverse of the variance),

$$\frac{1}{\tau^2} \sim \mathrm{Gamma}(0.001, 0.001)$$

This corresponds to an inverse-gamma distribution on the between-study variance, and is approximately uniform apart from a 'spike' of probability mass close to zero. Although this is perhaps the most commonly used vague prior distribution for the heterogeneity parameter, it has been recently criticized and prior distributions on the standard deviation parameter have been recommended, as they are more directly interpretable [15, 26].

The second prior distribution for the standard deviation, τ, is a standardized half-normal distribution truncated at zero,

$$\tau \sim \mathrm{Half\text{-}Normal}(0, 1) \quad \tau > 0$$

This prior distribution gives a low probability to values greater than 2.
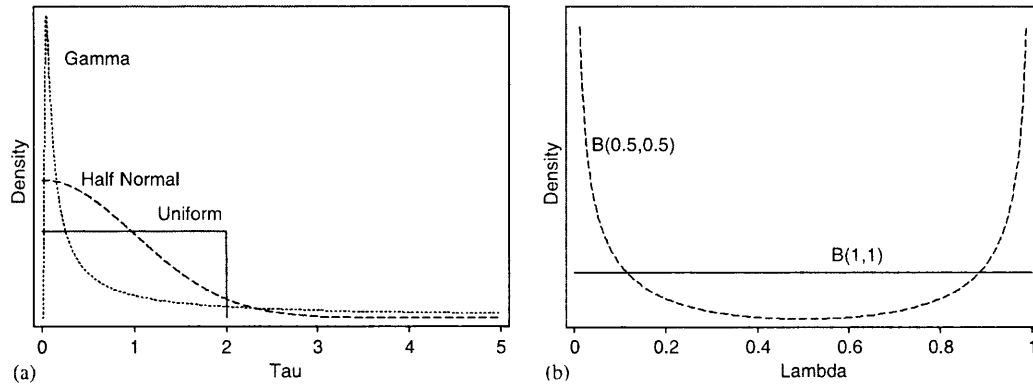
Figure 1. Density plots for the prior distributions considered for: (a) τ; and (b) λ. In (a) the densities for the three prior distributions are all presented on the standard deviation scale; the gamma distribution has been re-scaled in order to show the shape of the distribution.

Finally, the third prior distribution considered is a uniform distribution over the range 0 to 2 and excludes the possibility that the standard deviation can be over 2

$$\tau \sim \text{Uniform}(0,2)$$

2.2.2. Prior distributions for λ. We consider two beta prior distributions for the parameter, λ. Both are constrained to cover the range between 0 and 1, and have been used for modelling vague prior beliefs about proportions [4]. Figure 1(b) shows the densities for the two prior distributions.

The first prior is a beta distribution with both parameters equal to one

$$\lambda \sim \text{Beta}(1,1)$$

This distribution is uniform between 0 and 1. However, when parameters have values very close to the extremes, i.e. 0 or 1, and the data are sparse, this prior distribution will tend to pull the posterior estimates towards 0.5. For instance, for a near recessive model where the true value of λ is very close to 0, this prior distribution will tend to distort the posterior estimates because it gives 90 per cent prior probability to values greater than 0.1.

The second prior is a beta distribution with both parameters equal to 0.5

$$\lambda \sim \text{Beta}(0.5,0.5)$$

and corresponds to a Jeffreys' prior distribution for a binomial likelihood. This distribution gives greater prior probability to values of λ close to the extremes [27], i.e. to models which are close to recessive or dominant. However, if the genetic model is actually close to co-dominant, i.e. $\lambda = 0.5$, and the data are sparse, this distribution may tend to inflate the uncertainty surrounding λ.

2.3. Results

The results for the 6 combinations of prior distributions for τ and λ for the meta-analyses of Kato, Hani and Wheeler are illustrated in Figures 2(a), (b) and (c), respectively. Point
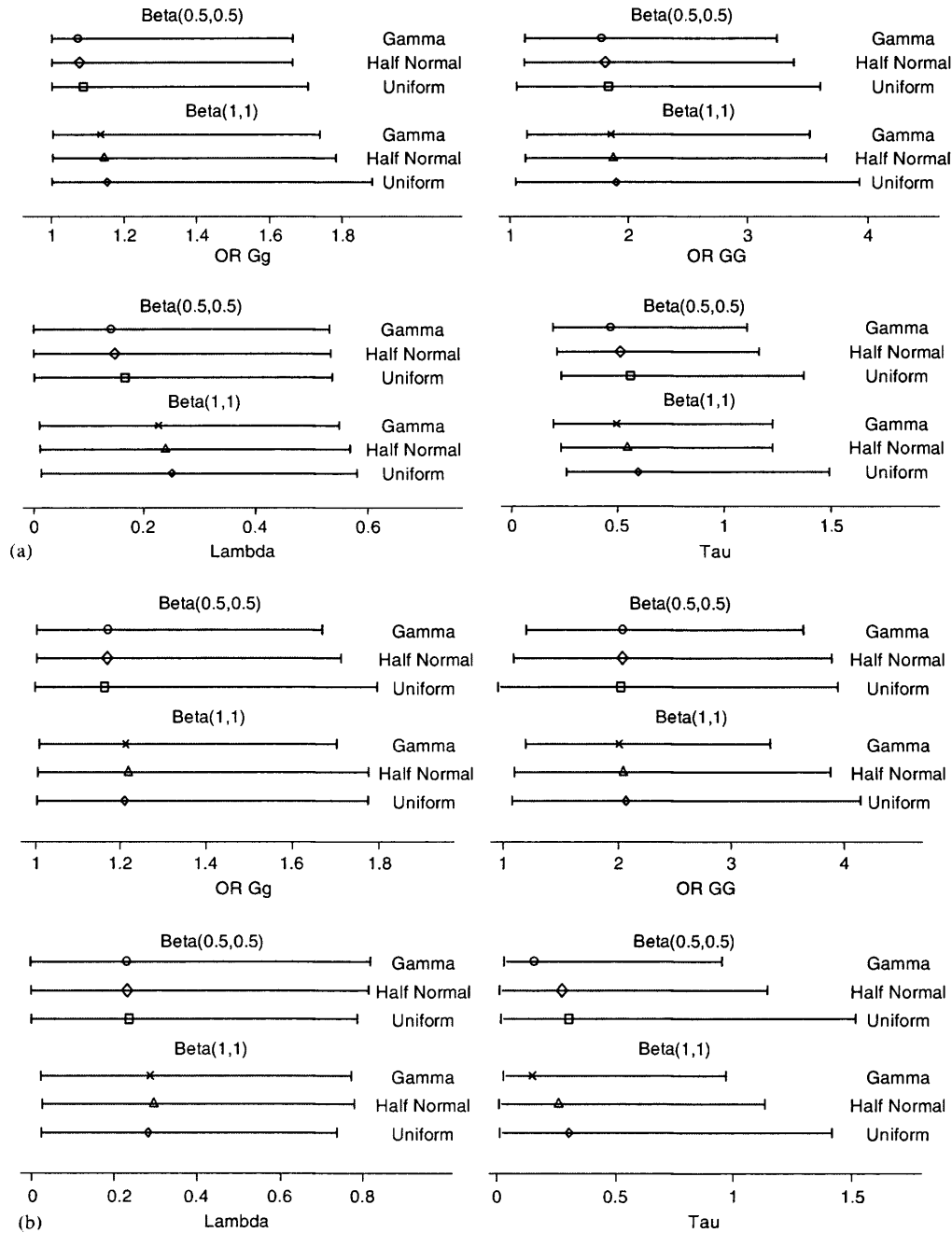
Figure 2. Plots of the results for the four parameters (OR_{GG}, OR_{Gg}, λ and τ) obtained by applying models with different prior distributions to the original meta-analyses by: (a) Kato; (b) Hani; and (c) Wheeler. The models are based on retrospective likelihood.
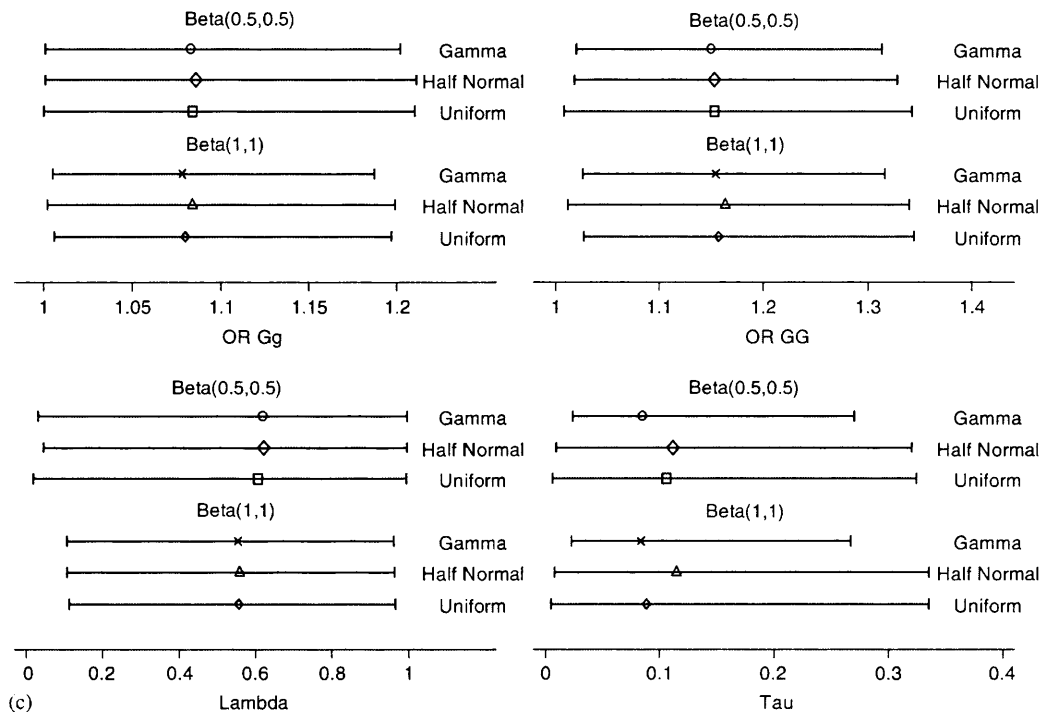
Figure 2. *Continued.*

estimates (medians) and 95 per cent credible intervals (CrI) of the four parameters of interest, OR_{Gg}, OR_{GG}, λ and τ, are plotted for each model.

The gamma distribution, with its spike close to zero, tends to produce lower estimates of τ with narrower credible intervals, which in turn tends to be reflected in the widths of the credible intervals for the odds ratios. This is particularly pronounced in Hani's meta-analysis, where the data are sparse because there are only 4 studies. Here the estimate of τ is 53 and 51 per cent lower with the gamma prior distribution compared with the uniform and the credible interval is 33 and 38 per cent narrower, for beta(1,1) and beta(0.5,0.5), respectively.

The impact of the two different beta prior distributions for λ varies according to the characteristics of the meta-analysis. As expected the beta(0.5,0.5) tends to pull the point estimates for λ towards the extremes, i.e. 0 and 1, and the beta(1,1) tends to provide more precise estimates of λ, when λ is near 0.5, as in Wheeler.

3. SIMULATIONS

Simulated data sets were created based on the three meta-analyses described in Section 2, in order to investigate the posterior parameter estimates in situations where the true values were known. The total number of studies and study sizes were kept the same as in the original meta-analyses, while values for the model parameters were taken from a profile maximum

likelihood approach previously used to analyse these data [19]. For each of the three meta-analyses, 1000 new data sets were randomly generated and each was analysed in WinBUGS using the different prior distributions described in Section 2.2. It was not possible to check the convergence of all 18 000 analyses so we selected data sets that gave a large discrepancy in results when analysed with different prior distributions and checked convergence for those by running longer chains ('burn-in' 50 000, chain length 100 000) with different starting values. In all cases the results confirmed the original analyses.

The median of the corresponding MCMC simulations was taken as the point estimate for each of the four parameters OR_{Gg}, OR_{GG}, λ and τ. The medians from the analyses with the different prior distributions were compared in terms of their mean, their root mean square error (RMSE), and the coverage of the 95 per cent CrIs, that is the percentage of intervals that included the true value. These three measures describe the average properties of the estimators across the 1000 data sets.

3.1. Results

For the four parameters OR_{Gg}, OR_{GG}, λ and τ, the mean, RMSE, and coverage of the 95 per cent CrIs are summarized in Table I. For the scenarios based on Wheeler's and Kato's meta-analyses, the number of data sets effectively analysed was in fact 995 and 998, respectively, since a few simulated meta-analyses contained studies with 0 cells for *both* cases and controls in a genotype group, and the MCMC algorithm did not converge.

In all cases the half-normal and uniform prior distributions caused the heterogeneity, τ, to be overestimated on average, although only in the case of the Hani-based simulations was the RMSE also appreciably larger. The beta prior distributions for λ caused the average estimate of λ to move towards 0.5, the more so in the presence of sparse data. This behaviour is caused by the constraint that λ must lie between 0 and 1, and the choice of symmetrical prior distributions such as the beta(0.5,0.5) or the beta(1,1). Such a situation is illustrated by the Kato-based simulations where the true value of λ is 0.13, so that underestimates had to lie between 0 and 0.13 while overestimates could lie between 0.13 and 1, and the average consequently tends to be too high. Under these circumstances the mean, or corresponding bias, is not an appropriate indicator of the quality of the estimator. A better indicator is the RMSE which favours the beta(0.5,0.5) prior distribution when λ is small and the beta(1,1) when λ is close to 0.5. On average the odds ratios are relatively insensitive to the choice of prior distributions.

Good average performance is reassuring but may not be a reliable guide to the sensitivity to the choice of prior distributions for any particular single data set. For this reason we used Bland–Altman style plots, originally described as a way to assess agreement between two methods of clinical measurement [28], in order to graphically evaluate the difference in results when comparing different prior distributions in all 1000 data sets (Figure 3). The difference in estimates based on any two prior distributions is plotted against the average of the two estimates. Plots for the Kato-based simulations are shown in Figure 3 and include a line drawn at the mean difference. Two dotted lines are drawn at the mean difference plus and minus 1.96 times the standard deviation of the difference, in order to both quantify the difference that can be observed when using different prior distributions on the same data set, and detect patterns in the difference which are related to the size of the parameter estimate. Plots for the other two scenarios showed similar results (data not shown).

Table I. Results of the sensitivity analyses to different prior distributions for λ and τ, for the simulated meta-analyses based on each of the 3 scenarios: (a) Kato; (b) Hani; and (c) Wheeler. RMSE = Root Mean Square Error.

(a) Parameter		OR$_{Gg}$			OR$_{GG}$			Lambda			Tau		
Assumed value		1.077			1.770			0.130			0.460		
Statistics													
Prior for λ	Prior for τ	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Beta(0.5,0.5)	Gamma	1.134	0.092	99.50	1.914	0.348	95.39	0.210	0.119	99.20	0.433	0.199	94.59
	Half-normal	1.144	0.098	99.30	1.931	0.357	97.39	0.222	0.126	99.00	0.511	0.174	96.59
	Uniform	1.152	0.104	99.60	1.945	0.366	98.10	0.234	0.134	98.50	0.564	0.210	94.89
Beta(1,1)	Gamma	1.186	0.118	98.09	1.979	0.386	93.88	0.273	0.148	97.39	0.454	0.209	94.48
	Half-normal	1.195	0.126	97.89	1.995	0.397	96.58	0.283	0.157	96.48	0.536	0.186	95.48
	Uniform	1.203	0.133	97.69	2.007	0.405	97.39	0.292	0.166	95.68	0.593	0.227	93.78

(b) Parameter		OR$_{Gg}$			OR$_{GG}$			Lambda			Tau		
Assumed value		1.177			2.030			0.230			0.134		
Statistics													
Prior for λ	Prior for τ	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Beta(0.5,0.5)	Gamma	1.170	0.111	99.60	2.057	0.373	99.30	0.248	0.135	98.09	0.206	0.076	99.60
	Half-normal	1.164	0.110	99.90	2.058	0.378	99.80	0.245	0.133	98.30	0.361	0.227	97.60
	Uniform	1.161	0.110	99.90	2.062	0.382	99.90	0.244	0.133	98.20	0.436	0.302	96.19
Beta(1,1)	Gamma	1.208	0.097	99.60	2.064	0.373	99.20	0.293	0.117	98.90	0.206	0.076	99.50
	Half-normal	1.203	0.096	99.90	2.067	0.380	99.80	0.291	0.116	98.70	0.363	0.229	97.10
	Uniform	1.200	0.095	99.90	2.072	0.381	99.90	0.289	0.115	98.60	0.435	0.301	95.60

(c) Parameter		OR$_{Gg}$			OR$_{GG}$			Lambda			Tau		
Assumed value		1.091			1.150			0.620			0.088		
Statistics													
Prior for λ	Prior for τ	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage	Mean	RMSE	Coverage
Beta(0.5,0.5)	Gamma	1.078	0.040	94.55	1.154	0.052	97.58	0.553	0.192	97.27	0.095	0.025	99.19
	Half-normal	1.076	0.041	94.74	1.155	0.053	97.88	0.54	0.195	96.26	0.113	0.040	98.69
	Uniform	1.076	0.041	94.43	1.154	0.053	97.87	0.539	0.196	96.46	0.114	0.040	98.79
Beta(1,1)	Gamma	1.078	0.037	94.75	1.16	0.054	97.17	0.532	0.153	97.98	0.097	0.026	99.19
	Half-normal	1.077	0.037	95.86	1.16	0.054	97.98	0.521	0.158	97.47	0.116	0.041	98.59
	Uniform	1.077	0.037	95.35	1.16	0.054	98.08	0.521	0.159	97.58	0.117	0.042	98.59

Figure 3. Plots of the difference in the estimates for Kato meta-analysis obtained by models with different prior distribution against their average value (Bland–Altman plot) for: (a) OR_{Gg}; (b) OR_{GG}; (c) λ; and (d) τ. Horizontal lines are drawn at the mean difference, and at the mean difference plus and minus 1.96 times the standard deviation of the differences. The models are based on retrospective likelihood.

Figure 3(b) shows the effects of the different prior distributions on the estimate of OR_{GG} in individual data sets generated under the conditions of the Kato meta-analysis. The posterior estimates of OR_{GG} with different prior distributions for τ are usually very close, mostly within
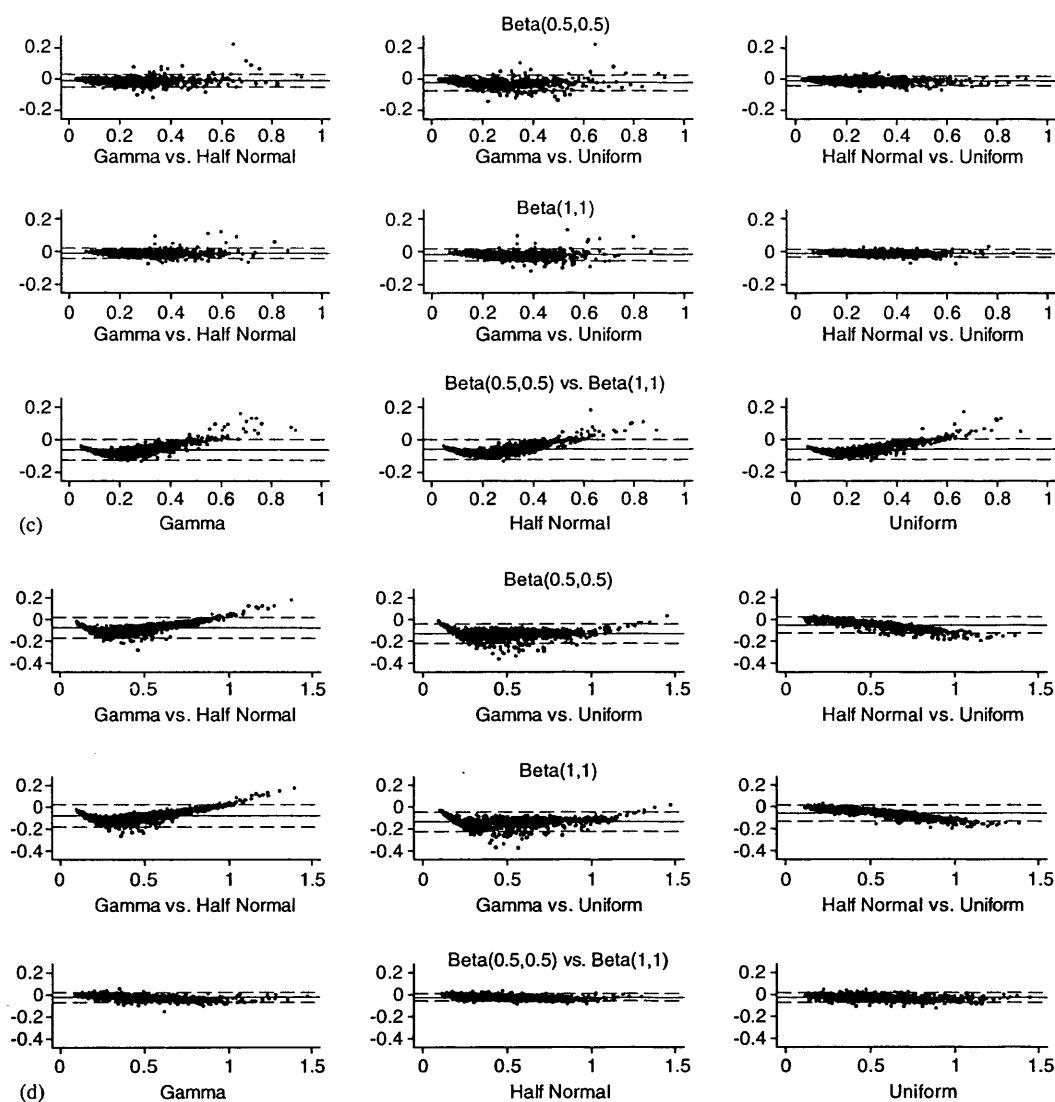
Figure 3. *Continued.*

±0.1 for estimates that are rarely over 3, and the agreement tends to be better in data sets where the posterior estimates of the odds ratio is close to 1. However, on rare occasions the difference can be as large as 0.3 when the average estimate is 3, a 10 per cent difference. Unfortunately there seems to be no way of distinguishing in advance if the prior distributions will have a large impact on a particular simulated data set. If a 10 per cent difference might be of importance, then sensitivity to the choice of prior distributions needs to be checked in any meta-analysis.

The impact on λ of the choice of the prior distribution is shown in Figure 3(c) and once again the differences tend to be small, but can, for particular data sets, be very large. For instance, data sets which produce estimates that average 0.6 can produce estimates that differ with the choice of prior distribution by up to 0.2. Figure 3(c) also shows some of the systematic effects noted in the average results. As might be anticipated, the impact of the choice of prior distribution is most marked in the estimate of τ, shown in Figure 3(d). Not only are there strong systematic patterns, but the differences can also be large. For instance, when comparing a gamma prior distribution and a uniform prior distribution the estimates can vary by as much as 0.4 when the average estimate is 0.5, that is one estimate is 0.3 while the other is 0.7.

4. PROSPECTIVE *VERSUS* RETROSPECTIVE LIKELIHOOD

Although the retrospective likelihood reflects the method of sampling in case–control studies, a prospective likelihood based on the probability of disease given exposure, gives the same odds ratio, in both maximum likelihood [21] and some Bayesian analyses [22]. The advantage of using the prospective likelihood is that the outcome variable, disease, is binary, whereas in the retrospective analysis the outcome, exposure, can have many levels. In the case of genetic association studies, the exposure, i.e. genotype, has three categories even in the simplest case of a bi-allelic polymorphism.

Although the equivalence of the two likelihoods for fixed effects meta-analyses follows from the analogy with a stratified case-control study, there is no reason to suppose that exactly equivalent results will be obtained with more complex hierarchical models. Nonetheless, where the heterogeneity is small or the data are not sparse, we might expect the results to be similar.

Denoting by y_{0j} and y_{1j} the number of controls and cases, respectively, in genotype group j, with $j = 1, 2, 3$ (i.e. gg, Gg and GG), the prospective likelihood (L_P) for each study included in the meta-analysis is derived from three binomial distributions, which leads to the following likelihood:

$$L_P(\alpha, \delta, y) = \prod_{j=1}^{3} \prod_{d=0}^{1} \left\{ \frac{\alpha^d \exp(d\delta_j)}{\sum_{k=0}^{1} \alpha^k \exp(d\delta_j)} \right\}^{y_{dj}} \tag{2}$$

where the parameter α is the baseline odds of disease (no exposure), i.e. the odds of disease when $j = 1$ (genotype gg), and δ is the log odds ratio of interest (δ_2 for $\log OR_{Gg}$ and δ_3 for $\log OR_{GG}$).

In the meta-analysis the full likelihood is obtained as the product of the likelihoods (2) over the i studies, assuming that the studies are independent. As in the retrospective meta-analysis (Section 2.2), the study-specific log odds ratios δ_{3i} are modelled as normally distributed random effects parameters, with an overall mean θ and between-study variance τ^2. The underlying study-specific log odds ratios, δ_{2i}, are again derived as the product of δ_{3i} and λ. A diffuse normal distribution, $\theta \sim \text{Normal}(0, 10\,000)$, is used in all models, while the different prior distributions for τ and λ are as discussed in Sections 2.2.1 and 2.2.2. Corresponding posterior distributions are obtained using MCMC methods implemented using WinBUGS 1.4.1 [26], and details on the WinBUGS code for fitting this model can be found on our website, www.hs.le.ac.uk/research/HCG/AppendixSiM2005.doc. The number of simulations was varied

Table II. Results of the genetic model-free approach applied to the three original meta-analyses, using both a retrospective and a prospective likelihood. The beta(0.5,0.5) and the uniform(0,2) are used as vague prior distributions for λ and τ, respectively.

Meta-analysis	Likelihood	OR_{Gg} (95% CrI)	OR_{GG} (95% CrI)	λ (95% CrI)	τ (95% CrI)
Kato	Retrospective	1.087 (1.000,1.705)	1.828 (1.057,3.598)	0.166 (0.001,0.538)	0.564 (0.236,1.373)
	Prospective	1.083 (1.000,1.736)	1.814 (1.053,3.660)	0.159 (0.000,0.542)	0.560 (0.239,1.425)
Hani	Retrospective	1.163 (0.997,1.797)	2.029 (0.959,3.956)	0.238 (0.002,0.785)	0.294 (0.017,1.513)
	Prospective	1.162 (0.998,1.773)	2.013 (0.968,4.095)	0.240 (0.001,0.815)	0.306 (0.015,1.492)
Wheeler	Retrospective	1.083 (1.001,1.207)	1.148 (1.019,1.337)	0.618 (0.028,0.997)	0.108 (0.009,0.332)
	Prospective	1.083 (1.000,1.210)	1.147 (1.009,1.330)	0.627 (0.032,0.998)	0.106 (0.005,0.331)

and the traces were inspected for evidence of non-convergence before deciding on a 'burn-in' of 5000 iterations followed by chains of length 10 000 for the retrospective model and 50 000 for the prospective model.

4.1. Results

We compared the retrospective and the prospective likelihoods by applying them to the three meta-analyses described in Section 2.1 and then to the simulated data sets described in Section 3.

Table II shows the results for the models with beta(0.5,0.5) and uniform(0,2) as vague prior distributions for λ and τ, respectively. The results are nearly identical for all meta-analyses and parameters of interest, both in terms of the point estimates (medians) and the width of the credible intervals. Different prior distributions for λ and τ gave similar results (data not shown). These findings were confirmed by the results of the simulations (data not shown). The only difference in the two approaches was a tendency to a slower convergence for the prospective models, even after the use of hierarchical centring in the specification of the prospective models [29]. This is the reason why the results for the prospective models (Table II) were based on longer chains, as described in the previous section.

5. DISCUSSION

The genetic model-free approach to the meta-analysis of genetic association studies is an integrated way of synthesizing the evidence on the genetic association, which captures both the magnitude of the genetic effect and information about the genetic mode of inheritance. Although the method can be implemented using maximum likelihood [19], the Bayesian framework is an attractive alternative with both philosophical and practical advantages. From

a philosophical point of view, Bayesian analysis allows explicit inclusion of prior distribution information on the genetic effect and on the genetic mode of inheritance. Although this possibility has not been explored in this paper, the models presented could incorporate prior knowledge when it is available. Such knowledge might be based on evidence from studies not included in the meta-analysis or on expert opinion. While in the first instance the inclusion of prior information would often be straightforward, in the second case it can be difficult to use expert opinion to derive appropriate probability distributions [8, 9]. From a practical point of view, the flexibility offered by Bayesian models estimated by MCMC algorithms makes it relatively straightforward to implement complex hierarchical models. The combination of increased computing power together with the availability of free software, particularly Win-BUGS [26], to implement MCMC models represents one of the main driving forces behind the increasing use of Bayesian methods in medical research. However, the increased flexibility leads to a greater requirement to consider the issue of model choice [4, 30].

In situations where external information is not available, prior distributions still have to be specified for all parameters. Although such prior distributions may be intended to be non-informative, this is in fact an impossible aim as Figure 1 illustrates. Rather, we must hope that the prior distributions will not be influential, in the sense that the use of alternative vague prior distributions will not change the conclusions. This may be an impossible aim if the data are sparse, especially when specifying prior distributions for scale parameters, such as the between-study heterogeneity in a random effects meta-analysis. If the results are sensitive to the choice of supposedly vague prior distributions, then we have no option but to consider that any prior distribution is informative and so must be chosen with care and subject to a sensitivity analysis [4]. The problem of statistical inference in the presence of sparse data is not limited to the Bayesian approach, and an analogous non-Bayesian analysis would find a very flat likelihood and would produce wide confidence intervals. In such circumstances, the ideal solution might be to incorporate subjective prior information or other external evidence in the Bayesian analysis [11].

The hierarchical model for the meta-analysis of genetic association studies involving the ratio, λ, requires many large studies if the choice of prior distribution for λ is not to have an undue influence on the posterior estimates. In much the same way, the prior distribution for the between-study variation, τ, can also influence the results. Recent research has suggested that the gamma prior distribution is not a good choice for the between-study precision in a hierarchical model and that the half-normal or uniform may be better [15]. Our findings do not support such a conclusion in this situation, where in general the gamma performs well. Of course, part of the problem is to do with the scaling of the prior distributions. Had we taken a half-normal with a smaller standard deviation, or a uniform distribution over a shorter range, then the corresponding estimates would have been more similar to those obtained using a gamma prior distribution. This very fact emphasizes the impossibility of defining a generic vague prior distribution when data are sparse, and the importance of careful specification even when using vague prior distributions.

Our analysis illustrates the importance of an investigation of sensitivity to the choice of prior distributions in any Bayesian analysis in which the prior distributions are not based on external knowledge. The sensitivity analysis will depend on the range of vague prior distributions that are considered reasonable in any given situation and the size of the change in the final estimate or its credible interval that is of practical importance. Thus, it will not be possible to find a *single* vague prior distribution that is always the least informative, so in

complex models the desire to use vague prior distributions does not free the researchers from the need to tailor their prior distributions to their particular problem.

In our analyses the prospective and retrospective likelihoods gave very similar results for all examples considered and for all the different vague prior distributions considered for τ and λ. It may well be that in practice the prospective likelihood could be used when synthesizing evidence from case–control studies. However, the approximate equivalence of the two likelihoods for a particular combination of data set and model can only be established by using both, which rather removes the benefits of the simpler, but theoretically inappropriate, prospective model. The retrospective likelihood has the further advantage that it can easily incorporate the assumption of Hardy–Weinberg equilibrium in the controls [31, 32]. Given these considerations it will often be more appropriate to use the retrospective likelihood unless there is considerable evidence of approximate equivalence from similar analyses.

Although not explored in this paper, informative prior distributions, based on expert opinion or external evidence, could be used for the different model parameters. In a more general meta-analysis context, empirical data-based prior distributions have been advocated for the heterogeneity term, τ, and might be an attractive option, especially when the number of studies included in the meta-analysis is small [33]. For the parameter λ, there might well be data from studies evaluating the effect of the same polymorphism on similar disease pathways. The increase in the precision of the estimated λ due to the use of an informative prior distribution would in turn increase the precision in the estimates of the odds ratios of interest, OR_{GG} and OR_{Gg}, and so might be very beneficial.

ACKNOWLEDGEMENTS

REFERENCES

1. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research.* Wiley: Chichester, U.K., 2000.
2. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**(6965):1351–1355.
3. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**(4):277–303.
4. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley: Chichester, U.K., 2004.
5. Brooks SP. Markov chain Monte Carlo method and its applications. *Journal of the Royal Statistical Society, Series D* 1998; **47**(1):69–100.
6. Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG. Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, Series A* 1996; **159**(2):323–342.
7. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**(6):619–629.
8. Garthwaite PH, Dickey JM. Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B* 1988; **50**(3):462–474.
9. Gokhale DV, Press SJ. Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Journal of the Royal Statistical Society, Series A* 1982; **145**(2):237–249.
10. Irony TZ, Singpurwalla ND. Noninformative priors do not exist: a discussion with Jose M. Bernardo. *Journal of Statistical Inference and Planning* 1997; **65**(1):159–189.

11. Berger JO. *Statistical Decision Theory and Bayesian Analysis* (2nd edn). Springer: New York, U.S.A., 1985.
12. Bernardo JM, Smith AFM. *Bayesian Theory*. Wiley: Chichester, U.K., 1994.
13. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 2005; 24:2401–2428.
14. Kass RE, Wasserman L. The selection of prior distributions by formal rules. *Journal of the American Statistical Association* 1996; 91(431):1343–1370.
15. Gelman A. Prior distributions for variance parameters in hierarchical models. *Technical Report*, Department of Statistics, Columbia University, 2004. Available at http://www.stat.columbia.edu/~gelman/research/unpublished/tau7.pdf
16. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; 20(3):453–472.
17. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; 20(3):435–452.
18. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *Journal of Clinical Epidemiology* 2003; 56(4):297–303.
19. Minelli C*, Thompson JR*, Abrams KR, Thakkinstian A, Attia J (* Joint first authors). The choice of a genetic model in the meta-analysis of molecular association studies. *International Journal of Epidemiology*, advance access published 22 August 2005.
20. Williams J, Spurlock G, Holmans P, Mant R, Murphy K, Jones L et al. Molecular Psychiatry 1998; 3(2): 141–149 (Erratum appears in *Molecular Psychiatry* 1998; 3(5):458).
21. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika* 1979; 66(3): 403–411.
22. Seaman SR, Richardson S. Equivalence of prospective and retrospective models in the Bayesian analysis of case-control studies. *Biometrika* 2004; 91(1):15–25.
23. Kato N, Sugiyama T, Morita H, Kurihara H, Yamori Y, Yazaki Y. Angiotensinogen gene and essential hypertension in the Japanese: extensive association study and meta-analysis on six reported studies. *Journal of Hypertension* 1999; 17(6):757–763.
24. Hani EH, Boutin P, Durand E, Inoue H, Permutt MA, Velho G et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* 1998; 41(12):1511–1515.
25. Wheeler JG, Keavney BD, Watkins H, Collins R, Danesh J. Four paraoxonase gene polymorphisms in 11212 cases of coronary heart disease and 12786 controls: meta-analysis of 43 studies. *Lancet* 2004; 363(9410): 689–695.
26. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS User Manual. Version 1.4.1*. MRC Biostatistics Unit, Cambridge, U.K., 2004.
27. Box GEP, Tiao GC. *Bayesian Inference in Statistical Analysis*. Addison-Wesley: Reading, MA, 1973.
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476):307–310.
29. Gelfand AE, Sahu SK, Carlin BP. Efficient parameterisations for normal linear mixed models. *Biometrika* 1995; 82(3):479–488.
30. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, U.K., 1996.
31. Cheng KF, Chen JH. Bayesian models for population-based case-control studies when the population is in Hardy–Weinberg equilibrium. *Genetic Epidemiology* 2005; 28(2):183–192.
32. Thompson D, Witte JS, Slattery M, Goldgar D. Increased power for case-control studies of single nucleotide polymorphisms through incorporation of family history and genetic constraints. *Genetic Epidemiology* 2004; 27(3):215–224.
33. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; 15(24):2733–2749.

The choice of a genetic model in the meta-analysis of molecular association studies

Cosetta Minelli,[1][†][*] John R Thompson,[1][†] Keith R Abrams,[1] Ammarin Thakkinstian[2,3] and John Attia[3]

Accepted	25 July 2005
Background	To evaluate gene–disease associations, genetic epidemiologists collect information on the disease risk in subjects with different genotypes (for a bi-allelic polymorphism: gg, Gg, GG). Meta-analyses of such studies usually reduce the problem to a single comparison, either by performing two separate pairwise comparisons or by assuming a specific underlying genetic model (recessive, co-dominant, dominant). A biological justification for the choice of the genetic model is seldom available.
Methods	We present a genetic model-free approach, which does not assume that the underlying genetic model is known in advance but still makes use of the information available on all genotypes. The approach uses OR_{GG}, the odds ratio between the homozygous genotypes, to capture the magnitude of the genetic effect, and λ, the heterozygote log odds ratio as a proportion of the homozygote log odds ratio, to capture the genetic mode of inheritance. The analysis assumes that the same unknown genetic model, i.e. the same λ, applies in all studies, and this is investigated graphically. The approach is illustrated using five examples of published meta-analyses.
Results	Analyses based on specific genetic models can produce misleading estimates of the odds ratios when an inappropriate model is assumed. The genetic model-free approach gives appropriately wider confidence intervals than genetic model-based analyses because it allows for uncertainty about the genetic model. In terms of assessment of model fit, it performs at least as well as a bivariate pairwise analysis in our examples.
Conclusions	The genetic model-free approach offers a unified approach that efficiently estimates the genetic effect and the underlying genetic model. A bivariate pairwise analysis should be used if the assumption of a common genetic model across studies is in doubt.
Keywords	Meta-analysis, population genetics, polymorphism, genetic models, association studies

Population-based genetic epidemiology, which evaluates the risk of a disease associated with a specific genetic polymorphism, often seeks to identify relatively small effects against a noisy background of biological and social complexity. Because of this,

most genetic association studies tend to be statistically under-powered.[1,2] While the need for large-scale population-based association studies has recently been recognized,[3,4] data from such studies will not be available in the near future. In the meantime, evidence synthesis from multiple small studies has the potential to play an important role in advancing biomedical knowledge by increasing the statistical power.[5] However, the appropriate use of meta-analysis within genetic epidemiology has been researched less than might be anticipated, and the general methodological quality of published meta-analyses of genetic association studies is poor.[6]

A recent review by Attia *et al.*[6] showed how meta-analyses of genetic association studies often fail to address general meta-analytical concerns and ignore important issues specific

[1] Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, University of Leicester, Leicester, UK.

[2] Clinical Epidemiology Unit, Faculty of Medicine, Mahidol University, Bangkok, Thailand.

[3] Centre for Clinical Epidemiology and Biostatistics, Faculty of Health, University of Newcastle, Newcastle, Australia.

[†] The first two authors contributed equally to this work.

[*] Corresponding author. Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, University of Leicester, 22–28 Princess Road West, Leicester LE1 6TP, UK. E-mail: cm109@le.ac.uk

to gene–disease associations. General concerns include the lack of explicit reporting of inclusion and exclusion criteria, a failure to explore possible sources of heterogeneity, and the absence of an investigation of publication bias. An important aspect of the inclusion criteria for a meta-analysis is outcome definition, since differences in the way outcome is defined and measured may well explain heterogeneity of study results.[3,7] Another important source of heterogeneity is diversity in the populations studied, in particular ethnic diversity.[3] Publication bias arises because studies showing either statistically significant results or large effect sizes are often more likely to be published than negative studies,[8,9] and thus the result of a meta-analysis based on published studies may be positively biased. Publication bias is particularly important in genetic epidemiology because it is possible to study many polymorphisms on the same subjects and then to select those that are submitted for publication.[3,10–13] Although simple graphical methods such as funnel plots can be used to detect publication bias,[8,9] in the review by Attia et al.[6] only 20% of the meta-analyses (7 out of 37) addressed this issue.

Methodological issues that are specific to genetic epidemiology include the checking of Hardy–Weinberg equilibrium and the choice of a genetic model.[6,7] In the meta-analysis of genetic association studies there are always at least three possible genotypes to compare. This contrasts with the two treatment groups characteristic of most biomedical meta-analyses. In practice, the number of possible comparisons between genotypes is often reduced by assuming a specific genetic model, such as dominant or recessive, but the conclusions might be sensitive to this assumption.[6]

In the simplest case of a polymorphism with two alleles (G and g), one of which is thought to be associated with a disease (G), association studies will usually collect information on the numbers of diseased and disease-free subjects with each of the three genotypes (gg, Gg, and GG). To date almost all meta-analyses of genetic association studies have reduced the three groups to two by (i) ignoring the heterozygotes and comparing gg with GG, (ii) performing separate pairwise comparisons, (iii) assuming a recessive model to justify combining the gg and Gg genotypes and comparing gg + Gg with GG, (iv) assuming a dominant model and comparing gg with Gg + GG, and (v) assuming a per-allele effect that places Gg mid-way between gg and GG, also called the co-dominant model. When unsure about the genetic model, some investigators fit multiple models and/or perform pairwise comparisons. However, adjustment for multiple testing is seldom made, and the pairwise estimates of the odds ratio of GG vs gg (subsequently referred to as OR_{GG}) and the odds ratio of Gg vs gg (subsequently referred to as OR_{Gg}) are usually obtained by carrying out two separate meta-analyses, thus ignoring the correlation between the two odds ratios induced by the common baseline group.

The review by Attia et al.[6] showed that 24 of 37 meta-analyses based their analysis on the assumption of an underlying genetic model, with half of these testing multiple modes of inheritance or multiple pairwise comparisons. A biological justification for the choice of the genetic model was provided in only eight meta-analyses. In nine of the meta-analyses the genetic effect was tested by comparing the allele frequency in cases and controls.

All of the methods of analysis in common use, with the exception of the pairwise comparisons, make the implicit assumptions that a particular genetic model applies in all studies,

and, more importantly, that the model is known in advance; for instance, the gene might be assumed to be recessive in all populations. Here we suggest a genetic model-free approach to the meta-analysis of genetic association studies that also assumes a common genetic model across studies but which does not specify the mode of inheritance in advance. The underlying genetic model is instead estimated from the data. Although no specific genetic model is assumed, the analyses are, of course, still based on an assumed statistical model. The model is based on a simple reparameterization and uses the odds ratio between the homozygous genotypes (OR_{GG}) to captures the magnitude of the genetic effect, and λ, the ratio of log OR_{Gg} and log OR_{GG}, to capture the genetic mode of inheritance. λ is assumed to be common across studies, but if this assumption is in doubt then pairwise comparisons obtained using bivariate random-effect meta-analysis methods, which take into account the correlation between OR_{GG} and OR_{Gg}, should be used.[14,15] We describe graphical and statistical ways of investigating whether the assumption of a common λ is reasonable.

Allowing λ to take any value (unbounded analysis), is equivalent to allowing the possibility of heterosis, i.e. the risk of the Gg group can be higher or lower than either of the homozygous groups. Although rare, heterosis has been described.[16,17] If this possibility can be excluded on biological grounds then it is better to constrain λ between 0 and 1 (bounded analysis); this restricts the mode of effect to the spectrum between dominant, through co-dominant, to recessive.

Methods

Genetic model-free approach: a common but unrestricted genetic model

Consider the meta-analysis of a bi-allelic polymorphism, in which G is the risk allele, and a dichotomous disease outcome is ascertained for each genotype. We define two parameters: the odds ratio between the two homozygous genotypes, OR_{GG}; and λ, the ratio of log OR_{Gg} and log OR_{GG}. The value of λ is not restricted, but values equal to 0, 0.5, and 1 correspond to the recessive, co-dominant, and dominant genetic model, respectively, and values >1 or <0 would suggest positive or negative heterosis.

Log OR_{GG} could be modelled as a fixed-effect or as a random effect that allows for heterogeneity across studies.[8] In the analyses presented, the log OR_{GG} has been modelled as a random effect except in those situations where the heterogeneity of log OR_{GG} was very close to 0. λ is modelled as a fixed-effect, that is, the genetic model is assumed to be the same in all studies. It is usually not possible to model both log OR_{GG} and λ as random effects because, without extra information, it is very difficult to disentangle the heterogeneity of λ from that of log OR_{GG}.

The two log odds ratios from each study are modelled as being bivariate normally distributed. The within study variances and covariances are obtained from the reports of the individual studies and are treated as known. Any heterogeneity is assumed to be normally distributed. Full details of the model are reported in the Appendix. In the examples presented the parameters were estimated by maximum likelihood using the ml command in Stata.[18] Interval estimates can be obtained either from the approximate standard errors obtained as part of the

maximization, or from the appropriate profile likelihood. The profile likelihoods were used for the bounded analysis and were obtained by considering selected values of one of the parameters and maximizing the likelihood over the others. The corresponding intervals are the range of estimates that had a profile likelihood within $1.92 = 1/2[\chi_1^2(95\%)]$ of the maximum. In the bounded analysis λ was restricted to the range 0–1, that is heterosis was excluded. To obtain intervals under these

conditions the maximization required for the profile likelihoods was performed over the restricted range. Values of Akaike's Information Criterion (AIC) are reported for model comparison,[19] with the best models showing the smallest AIC.

Prior to model fitting, it may be useful to plot, for each study, the log OR_{Gg} vs log OR_{GG}, as shown in Figure 1, in which the slope of the association between $\log OR_{Gg}$ and $\log OR_{GG}$ represents λ. Such a plot may help check the consistency of λ
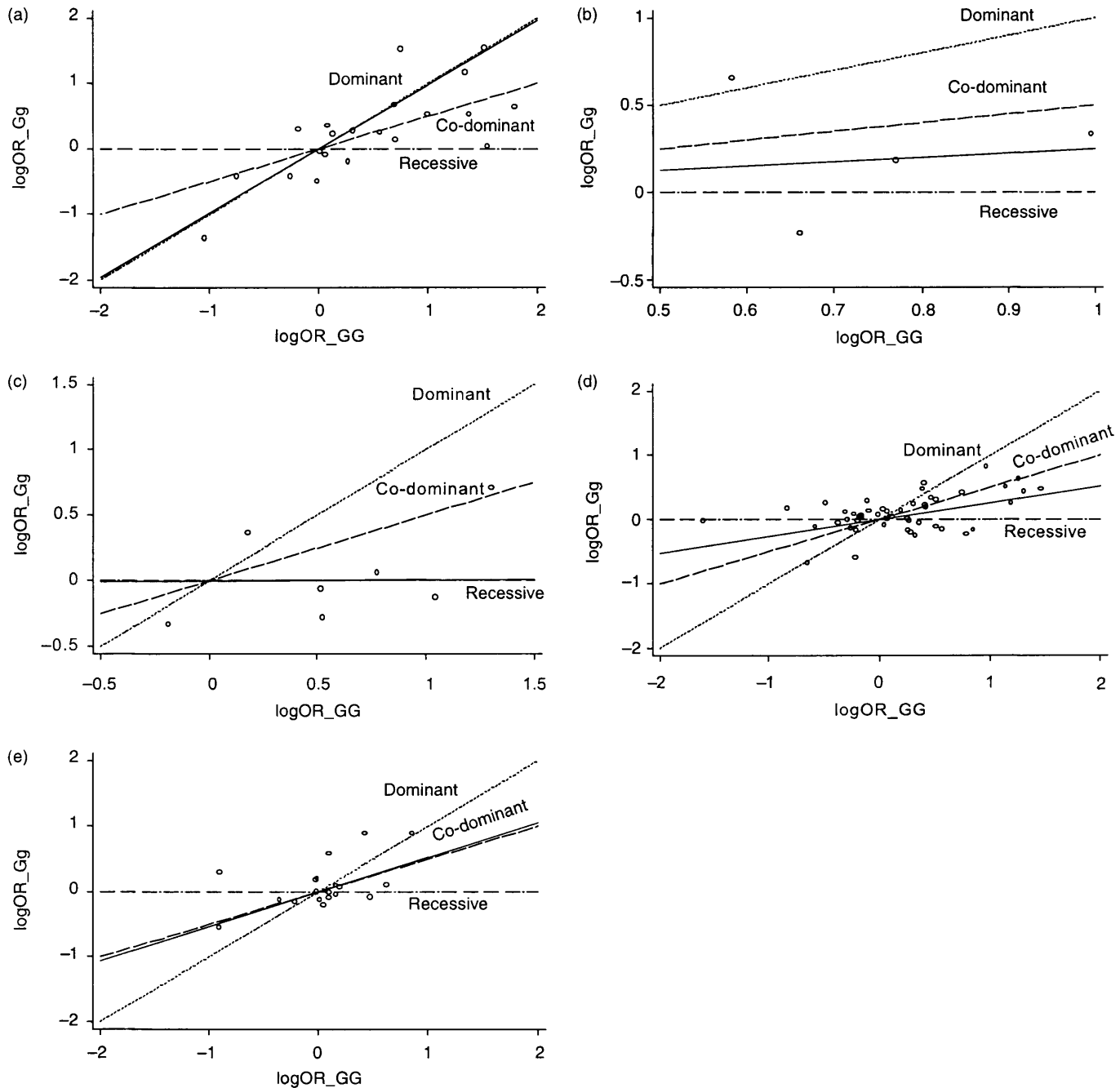
Figure 1 Plot of the log OR_{Gg} against the log OR_{GG} for: (a) *ACE* gene and diabetic nephropathy, (b) *KIR6.2* gene and Type II diabetes, (c) *AGT* gene and essential hypertension, (d) *MTHFR* gene and coronary heart disease, and (e) *PON1* Q192R polymorphism and myocardial infarction. The solid line represents the slope λ estimated by the genetic model-free approach; the three dotted lines correspond to the dominant, co-dominant, and recessive genetic models, respectively

across studies and identify outlying studies. Study-specific estimates of λ and bootstrapped 95% confidence intervals (CIs), as shown in Figure 2, help assess whether the variation in λ across studies might be explained by sampling error. Figure 2 is based on 1000 bootstrap samples from each study. If the genetic model does not seem to be consistent across studies then it may be better to perform joint pairwise comparisons using a general bivariate meta-analysis model,[14] which does not assume that λ is common but still takes into account the correlation between OR_{GG} and OR_{Gg}. Details of this model are also given in the Appendix.

Examples

The genetic model-free approach is illustrated using five published examples of the meta-analysis of genetic association studies. For each meta-analysis, the number of studies included, frequency of the risk allele, methods used by their authors, and main reported results, are given in Table 1.

ACE gene and diabetic nephropathy

This meta-analysis was carried out to evaluate the controversial association of the I/D polymorphism of the *ACE* gene with
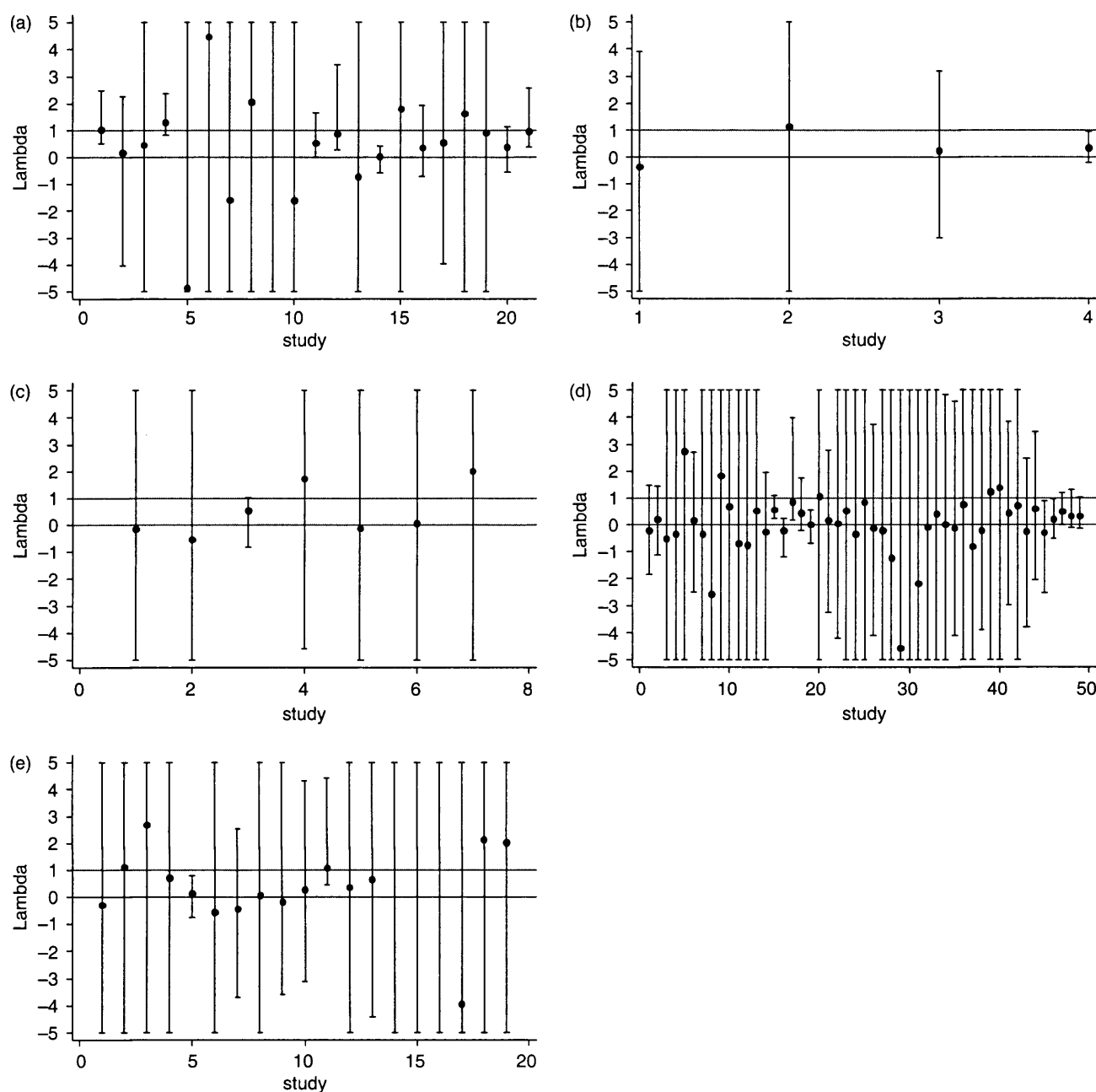
Figure 2 Plot of the study-specific estimates of λ (with 95% CI) for: (a) *ACE* gene and diabetic nephropathy, (b) *KIR6.2* gene and Type II diabetes, (c) *AGT* gene and essential hypertension, (d) *MTHFR* gene and coronary heart disease, and (e) *PON1* Q192R polymorphism and myocardial infarction. To better investigate the region in the middle, where the two lines correspond to the recessive and dominant models, the 95% CIs have been truncated at ±5

Table 1 Five published meta-analyses used for illustration, with methods and results reported in the original articles

Author, year	Association evaluated	Number of studies	Risk allele frequency	Reported analysis	
				Method	Results
Fujisawa, 1998[20]	*ACE* gene and diabetic nephropathy	21	0.46	Assumed dominant genetic model	1.32 (1.15–1.51)
Hani, 1998[21]	*KIR6.2* gene and Type II diabetes	4	0.34	Only *P*-value, under dominant and recessive genetic models	Dominant: $P < 0.05$ Recessive: $P < 0.01$
Kato, 1999[22]	*AGT* gene and essential hypertension	7	0.75	Allele frequencies cases vs controls	1.22 (1.05–1.42)
Wald, 2002[23]	*MTHFR* gene and coronary heart disease	49	0.32	Heterozygotes ignored, pairwise comparison for OR_{GG}	1.21 (1.06–1.39)
Wheeler, 2004[25]	*PON1* Q192R polymorphism and myocardial infarction	19	0.33	Per-allele relative risk	1.12 (1.15–1.51)

diabetic microangiopathy (nephropathy and retinopathy).[20] Here we consider only the meta-analysis assessing the effect on nephropathy. A dominant model was assumed and 21 studies were pooled to give an odds ratio of 1.32 (95% CI 1.15–1.51). The average allele frequency for the genetic variant was 0.46.

KIR6.2 gene and Type II diabetes

The K+ inwardly rectifier (KIR) channel is a protein that plays a major role in glucose-stimulated insulin secretion. Its encoding gene, *KIR6.2*, has been suggested as a candidate for inherited defects in Type II diabetes. This meta-analysis was carried out assuming dominant, recessive, and co-dominant models with *P*-values corrected for multiple testing.[21] The result of the meta-analysis, based on four studies, was a significant association between *KIR6.2* and Type II diabetes. The average frequency for the risk allele was 0.34.

AGT gene and essential hypertension

The genetic variant Thr235 of the angiotensinogen (*AGT*) gene has been found to be associated with hypertension in some linkage and association studies. This meta-analysis of seven Japanese case–control studies reported an odds ratio for the Thr235 allele of 1.22 (95% CI 1.05–1.42), with an average allele frequency of 0.75.[22]

MTHFR gene and coronary heart disease

The 677C→T is a polymorphism of the *MethyleneTetraHydroFolate Reductase* (*MTHFR*) gene involved in folate metabolism, which causes elevated homocysteine levels and has been associated with an increased risk of coronary heart disease. This meta-analysis of 49 studies reported an odds ratio of 1.21 (95% CI 1.06–1.39) for the TT vs CC comparison,[23] in close agreement with another meta-analysis published around the same time.[24] The average frequency for the T allele was 0.32.

PON1 Q192R polymorphism and myocardial infarction

PON1 is one of the genes encoding for paraoxonase, a serum enzyme that has been implicated in the prevention of atherogenesis and coronary heart disease through its association with high-density-lipoprotein particles. This recent meta-analysis of 19 studies investigated the effect of the Q192R polymorphism in the *PON1* gene on the risk of myocardial infarction.[25] The reported per-allele relative risk was 1.08 (95% CI 1.02–1.14), and the average allele frequency was 0.33.

Results

Figure 1 shows, for each meta-analysis, a plot of log OR_{Gg} against log OR_{GG}. All meta-analyses show variation in the genetic effect as represented by the two log odds ratios. This might be explained by a number of factors, including sampling error, differences in the study methods and differences in the true genetic risk across study populations. In the absence of heterogeneity in the genetic model and sampling error, all studies would be expected to lie along a straight line with slope λ. The solid line in Figure 1 represents the slope, λ, estimated by the genetic model-free approach, while the three dotted lines corresponding to the dominant, co-dominant, and recessive genetic models are plotted for comparison. The figure allows visual identification of any outliers or influential studies. Figure 2 plots the study-specific estimates of λ and their 95% bootstrap CIs, and is used to investigate whether any departures from linearity in Figure 1 are consistent with sampling error. Within individual studies λ is often poorly estimated, but there is little indication in any of the meta-analyses that the genetic models are not common across studies.

Table 2 summarizes the results for the different meta-analytical methods in common use; namely, separate pairwise comparisons, where log OR_{Gg} is pooled independently of log OR_{GG}, and methods based on assumed genetic models. In these analyses the log OR_{GG} has been modelled as a random effect except in two cases, marked in Table 2, where the heterogeneity of log OR_{GG} was very close to zero. The result for the ACE example when assuming a dominant model (Table 2) differs from the published result, which also assumed a dominant model (Table 1), because the main result in the original paper was based on a fixed-effect meta-analysis rather than our random effect meta-analysis.[8] The choice of the genetic model in model-based methods can have a marked impact on the estimates of OR_{GG} and OR_{Gg}. For instance, in the *KIR6.2* example, the estimates of OR_{GG} vary between 1.38 (95% CI 1.04–1.82) and 1.94 (95% CI 1.30–2.90). Separate pairwise comparisons give a consistent estimate of OR_{GG} of 2.21, but with an unnecessarily wide CI (95% CI 1.43–3.40) because they do not incorporate any of the information on OR_{Gg} when estimating OR_{GG}. Values of the AIC can be used to identify genetic models that are not consistent with the data. For instance, in the ACE example the possibility of a recessive model can be eliminated.

Table 3 presents the results of the genetic model-free approach, with λ unbounded and bounded between 0 and 1, and of the joint pairwise comparisons. The pooled estimates of λ obtained from the genetic model-free approach tend not to be very precise,

Table 2 Results of currently used meta-analytical methods for the five meta-analyses

Meta-analysis	Method	OR$_{GG}$ (95% CI)	OR$_{Gg}$ (95% CI)	Implicit λ	AIC
ACE gene and diabetic nephropathy	Separate pairwise comparisons	1.44 (1.07–1.93)	1.23 (0.94–1.60)	–	94.4
	Recessive model	1.16 (1.01–1.32)	–	0	89.8
	Co-dominant model	1.42 (1.09–1.87)	1.19 (1.04–1.37)	0.5	76.2
	Dominant model	1.29 (1.00–1.66)	–	1	68.2
KIR6.2 gene and Type II diabetes[a]	Separate pairwise comparisons	2.21 (1.43–3.40)	1.22 (0.91–1.64)	–	8.4
	Recessive model	1.93 (1.29–2.88)	–	0	8.0
	Co-dominant model	1.94 (1.30–2.90)	1.39 (1.14–1.70)	0.5	7.9
	Dominant model	1.38 (1.04–1.82)	–	1	13.3
AGT gene and essential hypertension	Separate pairwise comparisons	1.58 (1.06–2.35)	1.16 (0.77–1.76)	–	26.0
	Recessive model	1.64 (1.17–2.29)	–	0	20.7
	Co-dominant model	2.15 (1.26–3.65)	1.47 (1.12–1.91)	0.5	24.2
	Dominant model	1.41 (0.95–2.09)	–	1	40.6
MTHFR gene and coronary heart disease	Separate pairwise comparisons	1.19 (1.04–1.36)	1.05 (0.99–1.12)	–	88.6
	Recessive model	1.16 (1.02–1.31)	–	0	76.2
	Co-dominant model	1.18 (1.05–1.32)	1.08 (1.02–1.15)	0.5	75.8
	Dominant model	1.08 (1.01–1.16)	–	1	88.3
PON1 Q192R polymorphism and myocardial infarction[a]	Separate pairwise comparisons	1.16 (1.02–1.32)	1.08 (1.00–1.17)	–	22.4
	Recessive model	1.13 (1.00–1.27)	–	0	20.0
	Co-dominant model	1.17 (1.05–1.31)	1.08 (1.03–1.14)	0.5	15.8
	Dominant model	1.10 (1.02–1.18)	–	1	17.4

[a] Fixed-effect model.

Table 3 Results of the proposed genetic model-free approach, for both unbounded and bounded λ, and the joint pairwise comparisons obtained using bivariate meta-analysis

Meta-analysis	Method	OR$_{GG}$ (95% CI)	OR$_{Gg}$ (95% CI)	λ (95% CI)	AIC
ACE gene and diabetic nephropathy	Genetic model-free approach				
	Unbounded λ	1.30 (0.98–1.72)	1.29 (1.01–1.66)	0.98 (0.61–1.34)	70.2
	Bounded λ	1.30 (1.00–1.77)	1.29 (1.00–1.69)	0.98 (0.61–1.00)	70.2
	Joint pairwise comparisons	1.39 (1.07–1.81)	1.23 (0.96–1.58)	–	71.4
KIR6.2 gene and Type II diabetes[a]	Genetic model-free approach				
	Unbounded λ	2.14 (1.39–3.29)	1.21 (0.90–1.63)	0.25 (−0.11 to 0.61)	8.4
	Bounded λ	2.14 (1.43–3.29)	1.21 (1.08–1.63)	0.25 (0.00–0.69)	8.4
	Joint pairwise comparisons	2.14 (1.39–3.29)	1.21 (0.90–1.63)	–	8.4
AGT gene and essential hypertension	Genetic model-free approach				
	Unbounded λ	1.64 (0.99–2.72)	1.00 (0.66–1.53)	0.01 (−0.83 to 0.85)	22.7
	Bounded λ	1.64 (1.15–3.05)	1.00 (1.00–1.62)	0.01 (0.00–0.52)	22.7
	Joint pairwise comparisons	1.86 (1.14–3.05)	1.16 (0.77–1.76)	–	24.2
MTHFR gene and coronary heart disease	Genetic model-free approach				
	Unbounded λ	1.20 (1.05–1.37)	1.05 (0.99–1.11)	0.26 (0.04–0.47)	73.6
	Bounded λ	1.20 (1.05–1.38)	1.05 (1.01–1.12)	0.26 (0.04–0.49)	73.6
	Joint pairwise comparisons	1.20 (1.05–1.37)	1.06 (0.99–1.13)	–	75.3
PON1 Q192R polymorphism and myocardial infarction[a]	Genetic model-free approach				
	Unbounded λ	1.17 (1.04–1.33)	1.08 (1.00–1.17)	0.53 (−0.03 to 1.13)	17.8
	Bounded λ	1.17 (1.04–1.33)	1.08 (1.01–1.17)	0.53 (0.09–1.00)	17.8
	Joint pairwise comparisons	1.17 (1.04–1.33)	1.08 (1.00–1.17)	–	17.8

[a] Fixed-effect model.

but like the AIC, they can usually rule out some of the commonly assumed genetic models. For example, the *KIR6.2* gene and the *ACE* gene examples rule out the dominant and recessive models, respectively, while the *MTHFR* gene example suggests that λ is different from any of the values corresponding to the standard genetic models. In the example of the *ACE* gene, the estimate of λ is very close to 1, that is, close to dominant. Compared with an assumed dominant model, the model-free approach gives very similar estimates of OR_{GG}, but the CI is wider reflecting uncertainty about the true mode of inheritance.

In all of the examples, the AIC shows that the genetic model-free approach fits at least as well as the joint pairwise comparisons. Since the two approaches only differ for the assumption of common λ, these findings support those in Figure 2, and suggest that there is no evidence against the assumption of common λ in any of the five examples.

Under a fixed-effect assumption there is no between-study heterogeneity and so the model-free approach is exactly equivalent to the joint pairwise comparison as both models adjust for within-study correlation. For a random-effects model they give different answers because the model-free approach implies a structured covariance pattern as well as assuming a common mode of inheritance (see Appendix). The bounded analysis, in which λ must lie between 0 and 1, did not alter the point estimates of any of the parameters in our examples, because the maximum likelihood estimates of λ were all within the required range. The intervals for λ in the bounded analysis are truncated at 0 and 1 and are based on profile likelihoods rather than approximate standard errors, which accounts for some small differences from the unbounded analysis. The bounded analysis can have an effect on the interval estimates. For instance, in the *AGT* example, where the fitted model is very close to recessive, the restriction on λ implies that OR_{Gg} cannot fall <1.00 as this would either require a negative λ or a protective effect of the GG genotype; the bound rules out the former and the data contradict the latter.

The *AGT* example appears to be close to recessive, $\lambda = 0.01$, but with the largest study pointing to a co-dominant effect, as shown in Figure 2c. If the constancy of λ is doubted then joint pairwise comparisons could be used; such an analysis does not down weight the OR_{GG} and OR_{Gg} estimates from the largest study to the same extent and so produces larger pooled estimates. The AIC prefers the genetic model-free approach because it requires three parameters instead of four.

Discussion

When synthesizing the evidence on the association between a genetic polymorphism and a disease the main issue is the size of any association, but an important additional question is the mode of action of the gene. In practice, the estimate of the size of the association is influenced by our assumptions about the underlying genetic model. A review of the literature on meta-analysis of genetic association studies reveals how currently used approaches fail to address this issue.[6] Investigators often base their meta-analyses on the assumption of a specific genetic model and ignore their uncertainty about the mode of inheritance. Moreover, since it may be that no a priori biological evidence is available to justify the choice, different common genetic models are sometimes tested and the different results reported. Apart from the problem of

multiple testing, this leaves the reader with a set of estimates and significance tests to interpret, all based on different assumptions. A number of investigators compare allele frequencies between cases and controls; however, this method yields a per-allele effect that is equivalent to assuming a co-dominant model with Hardy–Weinberg equilibrium. Additionally, the issue of whether the genetic model is actually common across populations does not seem to have been addressed.

The results for the five meta-analysis examples show that adopting the wrong genetic model can lead to erroneous pooled estimates with deceptively high precision. The only meta-analytical approach currently in use that does not assume a common known underlying genetic model is analysis by separate pairwise comparisons, i.e. independent meta-analyses comparing genotype groups two at a time. This method ignores the correlation between the two estimated odds ratios induced by the common baseline group and thus is inefficient, as the estimates cannot 'borrow strength' from one another as they would in a multivariate meta-analysis.[14,15] The genetic model-free approach is likely to be particularly beneficial compared with pairwise comparisons when either of the alleles is rare. Moreover, separate pairwise comparisons run into the problem of multiple testing, which becomes especially important when a polymorphism with more than two alleles is considered.

As Table 1 illustrates, published meta-analysis of genetic association studies have used a variety of methods for presenting their results. The genetic model-free approach offers a single method that could have been used in all of these examples giving a consistent presentation and avoiding the pitfall of overly strong assumptions about the genetic model or of inefficient estimates.

The genetic model-free approach provides an integrated way of synthesizing the evidence on genetic associations, which yields not only the magnitude of the genetic effect (OR), but also an indication of the operating genetic model based on the available data. The underlying genetic model is not constrained to correspond to one of the classical modes of inheritance (recessive, co-dominant, dominant), in recognition of the fact that the gene's mode of action in complex diseases might differ from that found in Mendelian traits, where the association between genotype and disease tend to be of a deterministic nature and, hence, the mode of inheritance is relatively clearly apparent. For example, a value of 0.26 for λ, as in the *MTHFR* meta-analysis, might be interpreted in two ways:

(i) The polymorphism is recessive in some studies and co-dominant in others, so that the average result is between the two.

(ii) In complex diseases, the genotype is only one of many factors acting in a complex causal cascade leading to the disease. Although, at the molecular level, the polymorphism of interest might act in a clearly Mendelian manner on some intermediate phenotype, that Mendelian 'signal' may be 'diluted' or 'distorted' when measured at the level of the final step in the cascade. Hence, λ may be a more flexible and appropriate way to discuss genetic models in complex disease.

In the meta-analysis of genetic association studies there are two important types of heterogeneity that need to be addressed: heterogeneity in the genetic effect and heterogeneity in the genetic model. There are a number of reasons why we might see heterogeneity in the genetic effect, including differences in study

methods and differences in the underlying genetic risk associated with gene–gene or gene–environment interactions. Heterogeneity of the genetic effect might also arise if the polymorphism under study does not act directly on the disease risk, i.e. it is not a 'functional' or 'causal' polymorphism but is simply a marker, which tends to be inherited together with the causal polymorphism (linkage disequilibrium). Populations may have different patterns of linkage disequilibrium, which lead to differences in the marker association with disease. It is important to note that causes of heterogeneity in the genetic effect will not necessarily cause heterogeneity in the genetic model. In fact, in order to act on the genetic model, interactions need to influence the disease risk in heterozygotes to a different extent to the risk in homozygotes.

The absence of heterogeneity in the genetic model is an important assumption of the genetic model-free analysis and, although this assumption is likely to hold in most cases, it still needs to be assessed. For example, the effect of genotype on allergy to pollens appears to follow different modes of inheritance for different ethnic groups and different forms of allergy.[26,27] Although these studies are based on segregation analyses, and are relatively weak, they do raise the possibility that the mode of action may vary from study to study, perhaps owing to complex gene–environment interactions that have different impact on the disease risk in heterozygotes compared with homozygotes for the polymorphism. Thus, the assumption of a common genetic model should be checked before applying the genetic model-free approach, for instance by using the graphs presented in Figures 1 and 2. Should this assumption be in doubt, then the best approach would be to carry out joint pairwise comparisons using a multivariate meta-analysis, where the correlations between the odds ratios for the different genotype groups are taken into account. In addition to the graphical investigation, the difference in fit, as measured by AIC, between the model with common λ and the corresponding pairwise analysis offers a guide to the appropriateness of the assumption of a common genetic model. In general the random-effects model-free approach is easier to fit than the corresponding pairwise bivariate model because it contains two fewer parameters. Only in very large meta-analyses will it be possible to estimate the correlation in the heterogeneities required for the pairwise model. So, even when the assumptions of the model-free analysis are not met exactly, the model-free analysis may still be the best way of summarizing the data and obtaining CIs that are not falsely optimistic.

All of the models considered in this paper have been based on the normal approximation to the distribution of the log odds ratio. In examples where some of the studies have very few subjects within one of the genotypes, as might happen with a rarer allele, it would be better to use a multinomial likelihood. In the case of a random-effects model this adds to the complexity because of the need to numerically integrate over the random effect before maximization. Within this multinomial framework we can still use the λ parameterization basic to the genetic model-free approach and interpret the results in the same way as with the normal approximation.

The results presented in this paper have been obtained using maximum likelihood methods, but a Bayesian approach with non-informative prior distributions gave very similar results to those in Tables 2 and 3 (data not shown). The choice of a Bayesian approach to implement the method might be more desirable when there is external information regarding the magnitude of the genetic effect and/or mode of inheritance, which might come from studies not included in the meta-analysis or from expert opinion.[28] When Markov chain Monte Carlo methods are used, it also makes the generalization to multinomial likelihoods with random effects more straight-forward.[29]

In conclusion, we propose a new meta-analytical method based on a re-parameterization of the classical representation of genetic association studies, where the new parameters are biologically meaningful and informative. The approach makes maximum use of the information available by quantifying the magnitude of the genetic effect and estimating the genetic mode of action at the same time. The genetic model is estimated on the basis of the data rather than assumed, and this is important in all cases where no a priori knowledge about the underlying genetic model is available.

Acknowledgements

KEY MESSAGES

- Meta-analysis of molecular association studies is often based on the assumption of a specific genetic model (recessive, co-dominant, or dominant).

- Biological justification for the choice of the genetic model is seldom available, and results can be misleading when an inappropriate model is assumed.

- Specification of the genetic model is sometimes avoided by comparing genotype groups two at a time, but this is inefficient.

- We propose a genetic model-free approach where the information available on all genotypes is used and the genetic model is estimated rather than assumed.

- The approach assumes that all studies share the same unknown genetic model, and we suggest ways of investigating whether this assumption might hold.

References

[1] Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;**4**:45–61.

[2] Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001;**358**:1356–60.

[3] Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;**361**:865–72.

[4] Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;**2**:91–99.

[5] Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177–82.

[6] Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: Methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003;**56**:297–303.

[7] Salanti G, Sanderson S, Higgins JP. Obstacles and opportunities in meta-analysis of genetic association studies. *Genet Med* 2005;**7**:13–20.

[8] Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. Chichester: Wiley, 2000.

[9] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;**323**:101–5.

[10] The Lancet. In search of genetic precision. *Lancet* 2003;**361**:357.

[11] Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001;**29**:306–9.

[12] Morgan TM, Coffey CS, Krumholz HM. Overestimation of genetic risks owing to small sample sizes in cardiovascular studies. *Clin Genet* 2003;**64**:7–17.

[13] Agema WR, Jukema JW, Zwinderman AH, van der Wall EE. A meta-analysis of the angiotensin-converting enzyme gene polymorphism and restenosis after percutaneous transluminal coronary revascularization: evidence for publication bias. *Am Heart J* 2002;**144**:760–68.

[14] Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Stat Med* 2003;**22**:2309–33.

[15] van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;**21**:589–624.

[16] Williams J, Spurlock G, Holmans P et al. A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia [erratum in Mol Psychiatr 1998;3:458] *Mol Psychiatr* 1998;**3**:141–49.

[17] Juengel JL, Quirke LD, Tisdall DJ, Smith P, Hudson NL, McNatty KP. Gene expression in abnormal ovarian structures of ewes homozygous for the inverdale prolificacy gene. *Biol Reprod* 2000;**62**:1467–78.

[18] Gould W, Sribney W. *Maximum Likelihood Estimation with Stata*. College Station, Texas: Stata Press, 1999.

[19] Akaike H. Fitting autoregressive models for prediction. *Ann Inst Stat Math* 1969;**21**:243–47.

[20] Fujisawa T, Ikegami H, Kawaguchi Y et al. Meta-analysis of association of insertion/deletion polymorphism of angiotensin I-converting enzyme gene with diabetic nephropathy and retinopathy. *Diabetologia* 1998;**1**:47–53.

[21] Hani EH, Boutin P, Durand E et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* 1998;**41**:1511–15.

[22] Kato N, Sugiyama T, Morita H, Kurihara H, Yamori Y, Yazaki Y. Angiotensinogen gene and essential hypertension in the Japanese: extensive association study and meta-analysis on six reported studies. *J Hypertens* 1999;**17**:757–63.

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

[24] Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, Schouten EG. MTHFR 677C→T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA* 2002;**288**:2023–31.

[25] Wheeler JG, Keavney BD, Watkins H, Collins R, Danesh J. Four paraoxonase gene polymorphisms in 11212 cases of coronary heart disease and 12786 controls: meta-analysis of 43 studies. *Lancet* 2004;**363**:689–95.

[26] Marsh DG, Huang SK. Molecular genetics of human immune responsiveness to pollen allergens. *Clin Exp Allergy* 1991;**21**(suppl.1):168–72.

[27] Sasazuki T, Nishimura Y, Muto M, Ohta N. HLA-linked genes controlling immune response and disease susceptibility. *Immunol Rev* 1983;**70**:51–75.

[28] Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley, 2004.

[29] Minelli C, Thompson JR, Abrams KR, Lambert PC. Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Stat Med* (in press).

Appendix

Bivariate meta-analysis

Consider the meta-analysis of a set of case–control association studies of a bi-allelic polymorphism. Let z_{1i} represent the value of log OR_{Gg} estimated from the ith study and z_{2i} the log OR_{GG}. Assuming approximate bivariate normality

$$\begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \end{bmatrix}, \begin{bmatrix} v_{1i} & v_{12i} \\ v_{12i} & v_{2i} \end{bmatrix} \right\},$$

where the μ_is are the true log odds ratios for that study. The values of the variances and covariances are treated as known and can be derived from the number of cases and controls in each genotype in that study. If we assume that the studies come from a population in which the log odds ratios are also normally distributed, then

$$\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \tau_1 & \tau_{12} \\ \tau_{12} & \tau_2 \end{bmatrix} \right\},$$

where the τ's represent the heterogeneities between studies. The distribution of the observed data in the meta-analysis is thus

$$\begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} v_{1i} + \tau_1 & v_{12i} + \tau_{12} \\ v_{12i} + \tau_{12} & v_{2i} + \tau_2 \end{bmatrix} \right\}.$$

From which a likelihood can be formed and the parameters estimated. Unless the meta-analysis includes a large number of studies, the covariance between the heterogeneities is difficult to estimate, but the results for the other parameters are not very sensitive to τ_{12} so using an assumed value will not be misleading. In our analyses we used $\tau_{12} = 0.9\sqrt{\tau_1 \tau_2}$ and checked the results in a sensitivity analysis. A fixed-effects model assumes that $\tau_1 = \tau_2 = \tau_{12} = 0$.

Genetic model-free analysis

The genetic model-free analysis is similar to the general bivariate meta-analysis. First we assume that

$$\begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \lambda\mu_{2i} \\ \mu_{2i} \end{bmatrix}, \begin{bmatrix} v_{1i} & v_{12i} \\ v_{12i} & v_{2i} \end{bmatrix} \right\},$$

where the parameter, λ, which describes the genetic model is common across studies. The heterogeneity between studies will be

$$\begin{bmatrix} \lambda\mu_{2i} \\ \mu_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \lambda\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \lambda^2\tau & \lambda\tau \\ \lambda\tau & \tau \end{bmatrix} \right\}.$$

The distribution of the observed data in the meta-analysis is thus

$$\begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \lambda\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} v_{1i} + \lambda^2\tau & v_{12i} + \lambda\tau \\ v_{12i} + \lambda\tau & v_{2i} + \tau \end{bmatrix} \right\},$$

and once again the likelihood can be formed and maximized to estimate the parameters. In this model the covariance between the heterogeneities is controlled by λ and can thus be estimated. It is advisable to inspect the profile likelihood of each parameter as in small meta-analyses the log-likelihood can be far from quadratic. A fixed-effects model assumes that $\tau = 0$.

In some meta-analyses it may be appropriate to restrict λ to lie in the range (0,1), that is, to exclude heterosis. In this case, the overall maximization and the profile likelihood maximizations are over the restricted range.

HUMAN GENOME EPIDEMIOLOGY (HuGE) REVIEW

Systematic Review and Meta-Analysis of the Association between β2-Adrenoceptor Polymorphisms and Asthma: A HuGE Review

Ammarin Thakkinstian[1], Mark McEvoy[2], Cosetta Minelli[3], Peter Gibson[4], Bob Hancox[5], David Duffy[6], John Thompson[3], Ian Hall[7], Joel Kaufman[8], Ting-fan Leung[9], Peter Joseph Helms[10], Hakon Hakonarson[11], Eva Halpi[11], Ruth Navon[12], and John Attia[2]

[1] Clinical Epidemiology Unit, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.
[2] Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, Newcastle, New South Wales, Australia.
[3] Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, United Kingdom.
[4] Department of Respiratory and Sleep Medicine, Hunter Medical Research Institute, John Hunter Hospital, Newcastle, New South Wales, Australia.
[5] Department of Respiratory Medicine, Waikato Hospital, Hamilton, New Zealand.
[6] Queensland Institute of Medical Research, Brisbane, Queensland, Australia.

[7] Division of Therapeutics and Molecular Medicine, University Hospital, Nottingham, United Kingdom.
[8] Department of Environmental and Occupational Health Sciences, School of Public Health and Community Medicine, University of Washington, Seattle, WA.
[9] Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong, Special Administrative Region, China.
[10] Department of Child Health, Medical School, University of Aberdeen, Aberdeen, United Kingdom.
[11] deCODE Genetics, Inc., Reykjavik, Iceland.
[12] Department of Genetics, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

A number of studies have investigated two common polymorphisms in the β2-adrenoceptor gene, *Arg/Gly16* and *Gln/Glu27*, in relation to asthma susceptibility. The authors performed a meta-analysis of each polymorphism, as well as haplotype analysis, for adult and pediatric populations separately, using published data, supplemented by additional data requested from the original authors. Individual analysis detected no effect of *Arg/Gly16* in adults but did suggest a recessive protective effect of *Gly16* for children, with an odds ratio of 0.71 (95% confidence interval (CI): 0.53, 0.96) compared with the other genotypes. Results for *Gln/Glu27* in adults seem to indicate that heterozygotes are at decreased risk of asthma than either homozygote (odds ratio = 0.73, 95% CI: 0.62, 0.87), although the studies are heterogeneous; in children, the *Glu/Glu* genotype has a decreased risk of asthma (odds ratio = 0.60, 95% CI: 0.35, 0.99) compared with the other genotypes. Despite the proximity of these two polymorphic sites, the linkage disequilibrium coefficient of 0.41 was not high ($p < 0.001$). Haplotype analysis suggests that there may be an interaction between the two sites, with a lower risk of asthma associated with the *Glu27* allele (compared with *Gln27*), and that this risk is modified by the allele at position 16.

asthma; epidemiology; genetics; haplotypes; linkage disequilibrium; meta-analysis; polymorphism, genetic; receptors, adrenergic

Abbreviations: CI, confidence interval; LR, likelihood ratio; OR, odds ratio; SNP, single-nucleotide polymorphism.

Editor's note: This paper is also available on the website of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/).

One of the main thrusts of genetic epidemiology is to understand the genetic contribution to complex diseases such as cardiac disease, diabetes, and asthma. One of the

most popular study designs in this area is a molecular association study in which a polymorphism is linked to the disease outcome, either in cases and controls or in a cohort. These studies are often limited by small sample sizes (1), so there is a role for meta-analysis in pooling these studies, particularly to detect the small effect sizes that may be associated with these polymorphisms.

The β_2-adrenoceptor gene is a key gene to study in asthma. β_2-Adrenoceptors are present on many airway cells, including smooth muscle cells which are hyperreactive in asthma, and β_2-adrenoceptor agonists form a major treatment class in asthma. Functional polymorphisms of this gene may influence both disease susceptibility and treatment response in asthma.

A number of studies have investigated polymorphisms in the β_2-adrenoceptor gene in relation to asthma. Two common polymorphisms are *Arg/Gly16* and *Gln/Glu27*; in the former polymorphism, glycine is substituted for arginine at codon 16 (Arg16→Gly) and, in the latter, glutamic acid is substituted for glutamine at codon 27 (Gln27→Glu) (2, 3). In vitro studies indicate that the *Gly16* allele enhances agonist-induced down regulation of the receptor, whereas the *Glu27* allele enhances resistance to down regulation (4, 5). It is plausible that these differences in receptor regulation influence the reactivity of airway smooth muscle in response to airway inflammation and thereby alter the risk of asthma. However, epidemiologic studies have yielded conflicting results, with the direction of the effects not always congruent with the in vitro results. Several narrative reviews of these two polymorphisms and asthma (4–6) have been conducted; however, neither a magnitude nor a mode of gene effect was provided in these reviews. Furthermore, new studies that examine this association have been reported since those reviews, and there have been new developments in the methodology of meta-analysis of genetic studies (1, 7, 8). We therefore performed a systematic review of the association between *Arg/Gly16* and *Gln/Glu27* and asthma with the following objectives: first, to estimate allele frequencies; second, to ascertain if there is an effect of these polymorphisms on asthma susceptibility, and if so to estimate the magnitude of that effect and the possible mode of inheritance (1, 7, 8); third, to determine linkage disequilibrium between these two polymorphisms; and fourth, to infer haplotypes of these polymorphisms and link them with asthma susceptibility.

MATERIALS AND METHODS

Search strategy

Embase and Medline databases (from January 1966 to March 2004) were searched using the Embase, PubMed, and Ovid search engines. The search strategy for allele frequency was as follows: beta2* AND prevalence AND gene. The search strategy for association between gene polymorphisms and asthma was the following: asthma AND (beta receptor or beta-2 or adrenoceptor) AND (polymorph* or mutation* or variant* or genotype*). Searching was performed in duplicate by two independent reviewers (A. T. and M. M.).

Inclusion criteria

For allele frequency, any human studies that estimated the prevalence of β_2-adrenoceptor polymorphisms at codon 16 (*Arg/Gly16*) and/or codon 27 (*Gln/Glu27*) and reported on ethnically homogeneous populations were included, regardless of size. For assessing association, human studies, regardless of sample size, were included if they met the following criteria:

- β_2-Adrenoceptor polymorphisms at codon 16 (*Arg/Gly16*) and/or codon 27 (*Gln/Glu27*) were determined. The wild-type alleles for these two polymorphisms were *Arg* and *Gln*, respectively.
- The outcome was asthma (incident or prevalent), and there were at least two comparison groups, for example, asthma versus control (nonasthma) groups.
- Participants could be either children or adults, but results should be reported separately.
- There were sufficient results for extraction of data, that is, number of subjects for each genotype in asthma and control groups. Where eligible papers had insufficient information, we contacted authors by e-mail for additional information.

The reference lists of the articles retrieved were also reviewed to identify publications on the same topic. The most complete and recent results were used where there were multiple publications from the same study group.

Data extraction

Data were extracted independently and in duplicate by two reviewers (A. T. and M. M.) who used a standardized data extraction form. Any disagreement was adjudicated by a third author (J. A.). Covariables, such as mean age, gender, and ethnicity, were also extracted for each study.

Quality score assessment

The quality of studies was also independently assessed by the same two reviewers who used quality assessment scores that were modified from our previous meta-analysis of molecular association studies (7) (appendix table 1). These scores were based on both traditional epidemiologic considerations and genetic issues (1). Total scores ranged from 0 (worst) to 13 (best).

Statistical analysis

Data analyses were performed as follows. First, the frequency of *Arg16* and *Gln27* alleles in various ethnic groups was estimated by the inverse variance method, as described in the Appendix.

Second, estimation of the gene effect on asthma was performed by a logistic regression approach described previously (8). In brief, the steps were as follows. Hardy-Weinberg equilibrium was assessed for each study by use of the χ^2 test or Fisher's exact test, where appropriate, and only in control groups. A Q test for heterogeneity was performed separately for three odds ratios (ORs), that is, *Gly/Gly* versus

Arg/Arg (OR_1), *Arg/Gly* versus *Arg/Arg* (OR_2), and *Gly/Gly* versus *Arg/Gly* (OR_3) for the *Arg/Gly16* polymorphism and *Glu/Glu* versus *Gln/Gln* (OR_1), *Gln/Glu* versus *Gln/Gln* (OR_2), and *Glu/Glu* versus *Gln/Glu* (OR_3) for the *Gln/Glu27* polymorphism. If there was heterogeneity on at least one of these odds ratios, the cause of heterogeneity was explored by fitting a covariable (e.g., ethnicity, age, gender, or quality score) in a meta-regression model (9–11). If there was no heterogeneity, logistic regression analysis with the fixed-effect model was used to determine the gene effect; otherwise, the random-effect model was used to pool. A likelihood ratio test was used to gauge whether the overall gene effect was significant. If the main effect of the genotype was statistically significant, further comparisons of OR_1, OR_2, and OR_3 were explored. These pairwise differences were used to indicate the most appropriate genetic model as follows.

1. If $OR_1 = OR_3 \neq 1$ and $OR_2 = 1$, then a recessive model is suggested.
2. If $OR_1 = OR_2 \neq 1$ and $OR_3 = 1$, then a dominant model is suggested.
3. If $OR_2 = 1/OR_3 \neq 1$ and $OR_1 = 1$, then a complete overdominant model is suggested (also referred to as a "homozygous model" or "heterosis").
4. If $OR_1 > OR_2 > 1$ and $OR_1 > OR_3 > 1$ (or $OR_1 < OR_2 < 1$ and $OR_1 < OR_3 < 1$), then a codominant model is suggested.

Third, the gene effect was estimated by use of a newer, "parsimonious" approach detailed elsewhere (C. Minelli et al., University of Leicester, unpublished manuscript). In brief, this approach summarizes the genetic model in terms of a parameter lambda (λ), which is the ratio between $\log(OR_1)$ (*Glu/Glu* vs. *Gln/Gln*) and $\log(OR_2)$ (*Gln/Glu* vs. *Gln/Gln*). This parameter, which represents the heterozygote effect as a proportion of the homozygote variant effect, captures information about the genetic mode of action as follows: a recessive model if $\lambda = 0$, a dominant model if $\lambda = 1$, a codominant model if $\lambda = 0.5$, and homozygous or over-dominant if λ is greater than 1 or less than 0. The two log odds ratios are modeled as either fixed or random effects, as described in the second statistical analysis enumerated above.

Once the best genetic model is identified, this model is used to collapse the three genotypes into two groups (except in the case of a codominant model) and to pool the results again. Sensitivity analyses were performed by including or excluding studies not in Hardy-Weinberg equilibrium.

Fourth, with haplotype analysis, the haplotype frequencies of *Arg/Gly16* and *Gln/Glu27* polymorphisms were inferred using the expectation-maximization algorithm (12). The odds ratio was then estimated by use of the profile likelihood. The linkage disequilibrium coefficient was then estimated (13). The likelihood ratio test was used to test whether the linkage disequilibrium was significant.

All analyses were performed using Stata software, version 8.0 (14), apart from the parsimonious approach, for which WinBugs 1.4 (15) with vague prior distributions was used. A *p* value of less than 0.05 was considered statistically significant, except for tests of heterogeneity where a level of 0.10 was used.

RESULTS

For pooling allele frequency, 67 studies were identified, of which 16 (16–31) reported separate information for defined ethnic groups. Fourteen studies (2, 3, 32–43) retrieved from the search for gene effect were also included.

Allele frequencies

Arg allele. To estimate the pooled frequency, we used data only from control groups where a case-control design was used or from the entire group where a cohort design was used. Twenty-six studies (2, 3, 17, 18, 20–22, 24, 26–43) reported *Arg* allele frequencies (table 1), with 13 studies of Caucasian adults, three of Caucasian children, four of Black adults, six of Oriental adults, two of Oriental children, and one of Semite (Jews/Arabs) adults. Of these, six were not in Hardy-Weinberg equilibrium, leaving 12 studies of Caucasians, three of Blacks, and seven of Orientals for pooling.

There was heterogeneity among the 12 Caucasian studies ($\chi^2_{11} = 109.96, p < 0.001$). The pooled frequency using the random effects model was 42.0 percent (95 percent confidence interval (CI): 38.4, 45.7). The pooled frequency among Blacks was 49.2 percent (95 percent CI: 45.7, 52.7), and this estimate was homogeneous ($\chi^2_2 = 0.24$, $p = 0.89$). There was heterogeneity among Oriental studies ($\chi^2_6 = 18.79$, $p = 0.01$), and the pooled frequency was 56.2 percent (95 percent CI: 51.9, 60.6).

Gln allele. Twenty-six studies (3, 16, 17, 19, 20, 22–25, 27–43) reported the frequency of the *Gln/Glu27* polymorphism, 12 studies of Caucasian adults, three of Caucasian children, three of Black adults, seven of Oriental adults, two of Oriental children, one of Jewish adults, and one of Polynesian adults (table 2). Three studies, all of Caucasians, did not observe Hardy-Weinberg equilibrium and were not included in pooling.

There was heterogeneity among the 10 Caucasian studies ($\chi^2_9 = 437.77$, $p < 0.001$), and the pooled frequency was 59.6 percent (95 percent CI: 53.6, 65.6). All Black studies were homogeneous ($\chi^2_2 = 1.08, p = 0.58$), and the pooled frequency with the fixed model was 81.3 percent (95 percent CI: 79.7, 83.0). Seven Oriental studies were also homogeneous ($\chi^2_6 = 7.26, p = 0.30$), and the pooled frequency was 91.9 percent (95 percent CI: 90.9, 92.9).

Assessing association between gene polymorphisms and asthma

Across both Embase and Medline databases, 435 studies were identified in total, of which 113 were duplicates, leaving 322 study abstracts that were reviewed. From these, 30 studies seemed to be relevant, and therefore the full papers were retrieved. Sixteen studies were judged to have met the inclusion criteria, of which eight provided complete data in the paper. Requests for additional data on the other eight studies were made, of which four were granted. Two additional studies (36, 43) were identified by a known expert (D. D.), and the authors provided additional data. The characteristics of the adult and pediatric study populations, for

TABLE 1. Estimation of the pooled prevalence of the *Arg* allele

Subjects, first author (reference no.)	Hardy-Weinberg equilibrium (*p* value)	Total no.	*Arg* allele frequency (no.)	% with *Arg* allele
Caucasian adults*				
Santillan (3)	0.07	1,208	520	43
Barr (2)	0.50	274	164	60
Holloway (32)	0.30	182	73	40
Dewar (33)	0.32	1,268	489	39
Arnaiz (34)	0.02†	102	42	41
Reihsaus (35)	0.04†	112	30	27
Hakonarson (36)	0.75	362	127	35
Rosmond (17)	<0.001†	534	238	45
Dallongeville (20)	0.53	2,258	857	38
Tang (21)	0.06	248	95	38
Aynacioglu (22)	0.84	208	84	40
Weir (30)	1.00	168	102	61
Xie (31)	<0.05†	376	172	46
Caucasian children*				
Martinez (37)	0.90	538	206	38
Binaei (38)	0.13	310	135	44
Hopes (39)	0.12	838	279	33
Black adults‡				
Kotanko (18)	<0.05†	162	63	33
Tang (21)	0.51	286	144	50
Candy (24)	0.37	246	119	48
Xie (31)	1.00	246	120	49
Oriental adults§				
Wang (42)	0.50	272	140	51
Sugaya (26)	<0.05†	414	165	40
Chang (27)	0.11	260	137	53
Kim (28)	0.37	178	115	65
Iwamoto (29)	0.71	238	115	48
Xie (31)	0.69	208	122	59
Oriental children§				
Leung (41)	0.48	140	81	58
Lin (40)	0.06	298	182	61
Jewish/Arab adults				
Shachor (43)	0.45	222	101	45.5

* Pooled prevalence (%): 42 (95% confidence interval (CI): 38.4, 45.7).
† Not included in pooled prevalence.
‡ Pooled prevalence: 49.2 (95% CI: 45.7, 52.7).
§ Pooled prevalence: 56.2 (95% CI: 51.9, 60.6).

TABLE 2. Estimation of the pooled prevalence of the *Gln* allele

Subjects, first author (reference no.)	Hardy-Weinberg equilibrium (*p* value)	Total no.	*Gln* allele frequency (no.)	% with *Gln* allele
Caucasian adults*				
Arnaiz (34)	<0.05†	102	54	53
Santillan (3)	0.12	1,208	972	81
Holloway (32)	0.13	182	107	59
Dewar (33)	0.58	1,260	656	52
Reihsaus (35)	0.19	112	57	51
Hakonarson (36)	0.09	398	208	52
Rosmond (17)	0.45	532	314	45
Heckbert (19)	0.81	8,882	5,069	57
Dallongeville (20)	<0.001†	2,982	1,321	44
Aynacioglu (22)	0.81	208	142	68
Weir (30)	1.00	168	90	54
Xie (31)	0.06	376	245	65
Caucasian children*				
Martinez (37)	0.69	538	343	64
Hopes (39)	0.38	838	433	52
Binaei (38)	<0.05†	310	250	81
Black adults‡				
Heckbert (19)	0.64	1,616	1,315	81
Candy (24)	0.75	246	204	83
Xie (31)	0.78	246	195	79
Oriental adults§				
Wang (42)	1.00	272	248	91
Kawamura (16)	0.16	838	772	92
Kahara (23)	1.00	248	233	94
Chang (27)	0.55	260	240	92
Kim (28)	0.59	176	156	89
Iwamoto (29)	1.00	238	221	93
Xie (31)	1.00	208	193	93
Oriental children§				
Leung (41)	1.00	140	125	89
Lin (40)	1.00	298	267	90
Jewish/Arab adults				
Shachor (43)	0.66	218	150	69
Polynesian				
Duarte (25)	0.51	2,044	1,944	95

* Pooled prevalence (%): 59.6 (95% confidence interval (CI): 53.6, 65.6).
† Not included in pooled prevalence.
‡ Pooled prevalence: 81.3 (95% CI: 79.7, 83.0).
§ Pooled prevalence: 91.9 (95% CI: 90.9, 92.9).

example, mean age, gender, ethnicity, type of subjects, and allele frequency, are given in table 3.

Arg/Gly16 polymorphism. Adult asthma. Nine studies (2, 3, 32–36, 42, 43) determined the association between *Arg/Gly16* and asthma in adults (table 4). Total sample sizes

for asthma and control groups were 1,331 and 1,872, respectively. Within the asthma group, the mean age was 41 (standard deviation: 11) years, and 49 percent were females. Within the control group, the mean age was 39 (standard deviation: 11) years, and 35 percent were females.

TABLE 3.   General characteristics of studies included in pooling gene effects

Subjects, first author (reference no.)	Year	Study design	Race	Mean age (years)	% female	Quality score
Adults						
Shachor (43)	2003	Case-control	Jewish/Arab	38	53.0	5
Arnaiz (34)	2003	Cohort	Caucasian	28	1.9	9
Santillan (3)	2003	Case-control	Caucasian	37.3	15.0	13
Barr (2)	2001	Case-control	Caucasian	58.4	64.9	10
Wang (42)	2001	Case-control	Asian	33.0	61.7	13
Hakonarson (36)	2001	Case-control	Caucasian	47.5	56.3	6
Holloway (32)	2000	Case-control	Caucasian	31.4	54.9	6
Dewar (33)	1998	Cross-sectional	Caucasian	18–70*	54.0	6
Reihsaus (35)	1993	Case-control	Unknown	46		5
Children						
Martinez (37)	1997	Cross-sectional	Caucasian	10.8		9
Hopes (39)	1998	Cross-sectional	Caucasian	10.5		5
Leung (41)	2002	Case-control	Asian	10.8	55.0	5
Binaei (38)	2003	Case-control	Caucasian			1
Lin (40)	2003	Cross-sectional	Asian	13.9		9

* Range.

The seven studies (2, 3, 32, 33, 36, 42, 43) that observed Hardy-Weinberg equilibrium were pooled. Heterogeneity was checked for OR₁ (*Gly/Gly* vs. *Arg/Arg*), OR₂ (*Arg/Gly* vs. *Arg/Arg*), and OR₃ (*Gly/Gly* vs. *Arg/Gly*). Results indicated heterogeneity for OR₁ and OR₂ but not for OR₃ (for OR₁: $\chi^2_6 = 14.14, p = 0.03$; for OR₂: $\chi^2_6 = 13.98, p = 0.03$; for OR₃: $\chi^2_6 = 10.38, p = 0.11$). Race was explored as a potential cause; however, heterogeneity was still present in all odds ratios after excluding the one study of Asians (42) and the one study of Semites (for OR₁: $\chi^2_4 = 8.85, p = 0.07$; for OR₂: $\chi^2_4 = 9.76, p = 0.04$; for OR₃: $\chi^2_4 = 7.84, p = 0.10$). Hence, these seven studies were pooled by use of logistic regression with the random-effects model. The overall gene effect was not significant (likelihood ratio (LR) = $0.01, p = 0.99$), with the estimated OR₁, OR₂, and OR₃ being 1.00 (95 percent CI: 0.80, 1.24), 0.99 (95 percent CI: 0.81, 1.22), and 1.01 (95 percent CI: 0.85, 1.20), respectively (table 5). Analysis using the parsimonious approach yielded very similar results: OR₁ = 1.01 (95 percent CI: 0.79, 1.32), OR₂ = 1.00 (95 percent CI: 0.79, 1.30), and λ = 0.15 (95 percent CI: −4.15, 4.99).

Sensitivity analysis was performed by including the two studies (34, 35) that did not observe Hardy-Weinberg equilibrium; the results were similar in showing no genetic effect (LR₂ = $0.41, p = 0.96$).

Childhood asthma. Five studies (37–41) determined the association between the *Arg/Gly16* polymorphism and asthma in children (table 4), and all observed Hardy-Weinberg equilibrium. The total sample size was 334 with asthma and 842 controls.

No heterogeneity was detected for OR₁ (*Gly/Gly* vs. *Arg/Arg*), OR₂ (*Arg/Gly* vs. *Arg/Arg*), or OR₃ (*Gly/Gly* vs. *Arg/Gly*) (for OR₁: $\chi^2_4 = 1.97, p = 0.74$; for OR₂: $\chi^2_4 = 1.38$,

$p = 0.85$; for OR₃: $\chi^2_4 = 4.92, p = 0.30$). Logistic regression with the fixed-effect model was used to assess the overall gene effect, and this was close to the formal significance level (LR₂ = $5.15, p = 0.08$). The estimated OR₁, OR₂, and OR₃ were 0.75 (95 percent CI: 0.50, 1.12), 1.08 (95 percent CI: 0.76, 1.55), and 0.70 (95 percent CI: 0.51, 0.96) (table 5). These estimates suggest a recessive protective effect of the *Gly* allele, and therefore *Arg/Arg* and *Arg/Gly* were combined and compared with *Gly/Gly*. The estimated odds ratio was 0.71 (95 percent CI: 0.53, 0.96); that is, children with the *Gly/Gly* genotype had about 29 percent lower risk of having asthma than did children with the *Arg/Arg* and *Arg/Gly* genotypes. Using the parsimonious approach gave similar results: OR₁ and OR₂ of 0.88 (95 percent CI: 0.52, 1.20) and 1.04 (95 percent CI: 0.76, 1.54), respectively. The estimated λ was −0.16 (95 percent CI: −3.85, 4.39), close to what would be expected for a recessive model, that is, 0, although the confidence interval was wide.

Gln/Glu27 polymorphism. Adult asthma. Eight studies (3, 32–36, 42, 43) assessed the association between the *Gln/Glu27* polymorphism and asthma in adult patients (table 6). The sample size was 1,162 for asthma and 1,745 for control groups. All studies except one (34) observed Hardy-Weinberg equilibrium, and seven studies were therefore pooled to assess gene effect.

Heterogeneity tests were negative for OR₁ (*Glu/Glu* vs. *Gln/Gln*) and OR₃ (*Glu/Glu* vs. *Gln/Glu*) but significant for OR₂ (*Gln/Glu* vs. *Gln/Gln*) (for OR₁: $\chi^2_6 = 2.33, p = 0.89$; for OR₃: $\chi^2_6 = 8.15, p = 0.23$; for OR₂: $\chi^2_6 = 18.47, p = 0.01$). A number of factors were explored, including race, but we could not identify the source of heterogeneity. We then pooled these studies by logistic regression with the random-effects model to assess the gene effect. The likelihood ratio

TABLE 4. Genotype frequencies of the *Arg/Gly16* polymorphism between asthma and control groups

Subjects, first author (reference no.)	Asthma group					Control group				
	No.	% with Arg allele	Genotype (no.)			No.	% with Arg allele	Genotype (no.)		
			Arg/Arg	Arg/Gly	Gly/Gly			Arg/Arg	Arg/Gly	Gly/Gly
Adults										
Arnaiz (34)*	12	54	4	5	3	39	37	9	11	19
Santillan (3)	303	45	56	163	84	604	43	101	318	185
Barr (2)	171	49	36	97	38	137	60	51	62	24
Wang (42)	128	62	52	54	22	136	51	38	64	34
Holloway (32)	154	34	29	47	78	91	40	17	39	35
Dewar (33)	117	33	14	50	53	517	40	74	263	180
Reihsaus (35)*	51	28	5	19	27	56	27	7	16	33
Hakonarson (36)	323	37	45	151	127	181	35	21	85	75
Shachor (43)	72	46	13	40	19	111	46	25	51	35
Total	1,331		254	626	451	1,872		343	909	510
Children										
Martinez (37)	38	37	5	18	15	231	35	35	108	88
Leung (41)	76	58	25	38	13	70	58	22	37	11
Binaei (38)	38	10	7	24	7	155	44	34	67	54
Lin (40)	80	58	34	35	11	69	57	27	25	17
Hopes (39)	102	37	11	54	37	317	32	28	147	142
Total	334		82	169	83	842		146	384	312

* Arnaiz and Reihsaus were not included in the pooled gene effect.

test indicated that the overall gene effect was significant (LR = 14.64, $p < 0.05$). The estimated OR_1, OR_2, and OR_3 were 0.88 (95 percent CI: 0.68, 1.14), 0.72 (95 percent CI: 0.60, 0.85), and 1.22 (95 percent CI: 0.94, 1.60) (table 5).

The estimated OR_1, OR_2, and λ by the parsimonious approach were 0.97 (95 percent CI: 0.75, 1.27), 0.88 (95 percent CI: 0.63, 1.18), and 0.61 (95 percent CI: -4.66, 5.54), respectively. Sensitivity analysis was performed by adding the one study (34) not observing Hardy-Weinberg equilibrium, and the gene effect was robust: The estimated OR_1, OR_2, and OR_3 were 0.88 (95 percent CI: 0.68, 1.13), 0.71 (95 percent CI: 0.60, 0.84), and 1.22 (95 percent CI: 0.95, 1.59), respectively This seems to indicate a homozygous or overdominant mode of effect, with heterozygotes being at lower risk of asthma than either homozygote. Pooling according to this model yielded an odds ratio of 0.73 (95 percent CI: 0.62, 0.87); that is, the chance of having asthma was about 27 percent less with *Gln/Glu* compared with *Gln/Gln* + *Glu/Glu*. Although this is a nonintuitive model, there is precedent for other genes acting in this manner (see Discussion); alternatively, this may be a spurious result due to the distribution of data and the possibility of interaction between the two polymorphic sites. We address this possibility further in the next section using haplotype analysis.

Childhood asthma. There were five studies (37–41) addressing the association between the *Gln/Glu27* polymorphism and asthma in children (table 6). All studies observed Hardy-Weinberg equilibrium except one (38).

The four studies observing Hardy-Weinberg equilibrium were pooled (37, 39–41). Since the studies by Lin et al. (40) and Leung et al. (41) had cells with no counts, we added 1 for each cell for these two studies. There was no evidence of heterogeneity for OR_1 (*Glu/Glu* vs. *Gln/Gln*), OR_2 (*Gln/Glu* vs. *Gln/Gln*), or OR_3 (*Glu/Glu* vs. *Gln/Glu*) (for OR_1: $\chi^2_3 = 0.47$, $p = 0.93$; for OR_2: $\chi^2_3 = 2.24$, $p = 0.53$; for OR_3: $\chi^2_3 = 1.51$, $p = 0.68$). Logistic regression with the fixed-effect model was then used to pool; the estimated OR_1 and OR_3 of 0.62 (95 percent CI: 0.36, 1.07) and 0.59 (95 percent CI: 0.35, 0.99), respectively, were similar, whereas the estimated OR_2 of 1.05 (95 percent CI: 0.75, 1.48) was close to one (table 5). Although the overall gene effect was not significant ($p = 0.12$), there is the suggestion of a recessive protective effect. The *Gln/Gln* and *Gln/Glu* genotypes were therefore combined and compared with *Glu/Glu*. We found that the estimated odds ratio was 0.60 (95 percent CI: 0.37, 1.00); that is, children who had the *Glu/Glu* genotype were about 40 percent less likely to have asthma than were children who had genotype *Gln/Glu* or *Gln/Gln*. Sensitivity analysis was performed by including the study not in Hardy-Weinberg equilibrium; this did not change the indication of a recessive protective effect (OR = 0.61, 95 percent CI: 0.38, 0.98). The parsimonious model was compatible with this effect, with an OR_1 of 0.90 (95 percent CI: 0.49, 1.22), an OR_2 of 1.02 (95 percent CI: 0.76, 1.40), and an estimated λ of -0.04 (95 percent CI: -3.63, 4.30). Hence, these results suggested a recessive protective effect of *Glu*, although neither model was statistically significant.

TABLE 5. Determination of the genetic effects of *Arg/Gly16* and *Gln/Glu27* polymorphisms on asthma

Genotype	Logistic regression		Model-free approach	
	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval
Arg/Gly16				
Adults				
Gly/Gly vs. *Arg/Arg*	1.00	0.80, 1.24	1.01	0.79, 1.32
Arg/Gly vs. *Arg/Arg*	0.99	0.81, 1.22	1.00	0.79, 1.30
Gly/Gly vs. *Arg/Gly*	1.01	0.85, 1.20	$\lambda = 0.15$	−4.15, 4.99
Children				
Gly/Gly vs. *Arg/Arg*	0.75	0.50, 1.12	0.88	0.52, 1.20
Arg/Gly vs. *Arg/Arg*	1.08	0.76, 1.55	1.04	0.76, 1.54
Gly/Gly vs.*Arg/Gly*	0.70	0.51, 0.96	$\lambda = -0.16$	−3.85, 4.39
Gly/Gly vs. *Arg/Arg* + *Arg/Gly* (recessive effect)	0.71	0.53, 0.96		
Gln/Glu27				
Adults				
Glu/Glu vs. *Gln/Gln* (OR₁*)	0.88	0.68, 1.14	0.97	0.75, 1.27
Gln/Glu vs. *Gln/Gln* (OR₂*)	0.72	0.60, 0.85	0.88	0.63, 1.18
Glu/Glu vs. *Gln/Glu* (OR₃*)	1.22	0.94, 1.60	$\lambda = 0.61$	−4.66, 5.54
Gln/Glu vs. *Gln/Gln* + *Glu/Glu* (overdominant effect)	0.73	0.62, 0.87		
Glu/Glu vs. *Gln/Gln* (OR₁)	0.62	0.36, 1.07	0.90	0.49, 1.22
Gln/Glu vs. *Gln/Gln* (OR₂)	1.05	0.75, 1.48	1.02	0.76, 1.40
Glu/Glu vs. *Gln/Glu* (OR₃)	0.59	0.35, 0.99	$\lambda = -0.04$	−3.63, 4.30
Glu/Glu vs. *Gln/Glu* + *Gln/Gln* (recessive effect)	0.60	0.37, 1.00		

* OR₁, odds ratio of asthma with the preceding comparison of genotypes (OR₂ and OR₃ defined similarly).

Haplotype analysis of *Arg/Gly16* and *Gln/Glu27* polymorphisms

Three studies of adults provided data for haplotype analysis (3, 33, 36). The study by Weir et al. (30) reported inferred haplotype data among subjects who had only homozygous wild or mutant genotypes at one locus, so this study was not included in the present analysis. The expectation-maximization algorithm was applied to infer haplotypes for the three studies, and linkage disequilibrium was assessed. The estimated linkage disequilibrium coefficient was 0.48 ($p < 0.001$).

The haplotype frequency in asthmatics and controls is described in table 7. The three most common haplotypes were *Arg/Gln* (37.5 percent), *Gly/Glu* (31.7 percent), and *Gly/Gln* (28.2 percent). The estimated odds ratios were 0.39 (95 percent CI: 0.29, 0.58), 0.99 (95 percent CI: 0.74, 1.49), and 0.83 (95 percent CI: 0.62, 1. 24) for haplotypes *Arg/Glu*, *Gly/Gln*, and *Gly/Glu* compared with *Arg/Gln*. These numbers seem to indicate that, when *Gln* is present at position 27, the risk of asthma is the same regardless of what allele is present at position 16. However, with *Glu* at position 27, the risk of asthma is lower, and this

decreased risk is modified by the allele at position 16, being lower with *Arg16* than with *Gly16*.

This effect modification is marked, and the confidence interval of the odds ratio for the *Arg/Glu* haplotype does not overlap with that of the *Gly/Glu* haplotype (table 7). Subjects who had haplotypes *Arg/Glu* and *Gly/Glu* were 61 percent and 17 percent less likely to have asthma than were subjects who had haplotype *Arg/Gln*. However, subjects with haplotype *Gly/Gln* had the same chance of asthma as did subjects with *Arg/Gln*.

DISCUSSION

The various results of the individual single-nucleotide polymorphism (SNP) analyses and haplotype analyses are complex, but synthesizing the data overall seems to indicate the following. First, the *Glu27* allele appears to be protective against asthma, reducing the risk of asthma by approximately 27 percent. This makes biologic sense because the *Glu* variant is resistant to down regulation in vitro, and it is possible that these individuals express higher β₂-receptor levels in the context of inflammation. This was suggested

TABLE 6.   Genotype frequencies of the *Gln/Glu27* polymorphism between asthma and control groups

Subjects, first author (reference no.)	Asthma group					Control group				
	No.	% with *Gln* allele	Genotype (no.)			No.	% with *Gln* allele	Genotype (no.)		
			Gln/Gln	*Gln/Glu*	*Glu/Glu*			*Gln/Gln*	*Gln/Glu*	*Glu/Glu*
Adults										
Arnaiz (34)*	12	58	6	2	4	39	51	14	12	13
Santillan (3)	303	88	241	53	9	604	80	385	202	17
Wang (42)	128	92	108	19	1	136	91	113	22	1
Holloway (32)	153	87	49	76	28	91	59	35	37	19
Dewar (33)	119	49	33	51	35	511	53	134	271	106
Reihsaus (35)	51	39	13	26	12	56	51	17	23	16
Hakonarson (36)	324	55	92	173	59	199	52	48	112	39
Shachor (43)	72	73	38	29	5	109	69	50	50	9
Total	1,162		580	429	153	1,745		796	729	220
Children										
Martinez (37)	38	64	16	17	5	231	64	95	104	32
Hopes (39)	102	54	24	63	15	317	51	83	156	78
Leung (41)	76	92	64	12	0	70	89	55	15	0
Binaei (38)*	37	78	23	12	2	155	81	107	36	12
Lin (40)	80	91	65	15	0	69	88	54	14	1
Total	333		192	119	22	842		394	325	123

* Not in Hardy-Weinberg equilibrium and not pooled.

in both adult and pediatric populations, although the genetic model in each was different.

Second, the protective effect of *Glu27* may be due to the haplotype. It is probable that this is not an effect of this SNP in isolation but, instead, reflects a common haplotype that includes this allele. Drysdale et al. (44) investigated 13 SNPs in the human β_2-adrenergic receptor gene promoter and coding regions in relation to responsiveness to β_2 agonists. They found that, although there was no association when SNPs were analyzed individually, there was a clear relation between one of the common haplotypes (haplotype 2 in their paper, which included *Glu27*) and good response to β_2 agonists in vivo, as well as increased messenger RNA levels and gene expression in vitro. Haplotypes that included *Gln27* (e.g., haplotype 4 in their paper) had overall poorer response to β_2 agonists and lower expression levels. Presumably,

good response to exogenous agonists also reflects good response to endogenous agonists and, hence, a protective effect against asthma.

Third, the genetic model suggested by the data appears to be an overdominant protective effect of *Glu27*. This model is also called heterozygote advantage or positive heterosis, and although it may appear counterintuitive, a recent review indicates that this mode of action is perhaps more common than previously thought and cites numerous examples (45). Indeed, the *IL12B* promoter polymorphism has been associated with severity of asthma in children, and this also seems to observe a pattern of heterozygote advantage (46). The mechanism of such a model is still speculative but may include 1) advantages in having variation in a multimeric protein, such as better V_{max} (47); 2) an allele with a selective advantage that is detrimental when homozygous (e.g., sickle

TABLE 7.   Distribution of haplotype frequency of *Arg/Gly16* and *Gln/Glu27* polymorphisms between asthma and control groups

Haplotype	Control group (n = 2,331)		Asthma group (n = 950)		Adjusted odds ratio*	95% confidence interval
	No.	%	No.	%		
Arg/Gln	978	37	573	39	1.00	
Arg/Glu	91	3	18	1	0.39	0.29, 0.58
Gly/Gln	741	28	428	29	0.99	0.74, 1.49
Gly/Glu	852	32	461	31	0.83	0.62, 1.24

* Adjusted for study effect.

cell and falciparum malaria); and 3) a greater range of expression of gene products and plasticity with heterozygotes than homozygotes (45). Alternatively, this may be a spurious result due to other untyped loci in the haplotypes analyzed.

Fourth, there may be interaction or synergism between different SNPs. The haplotype analysis raises the possibility that the position 16 polymorphism may be an effect modifier: The protective effect of *Glu27* was accentuated with *Arg16* compared with *Gly16*, although there was no independent effect of the position 16 polymorphism on its own. This would indicate that it may be difficult to predict a haplotype effect from its constituent SNPs.

Fifth, the linkage disequilibrium between position 16 and 27 polymorphisms is not high. This may be surprising given that they are only 30 nucleotides apart and there are no intervening introns. However, this is congruent with other studies indicating that recombination frequency is not strictly proportional to chromosomal distance, and it is sensitive to ancestral effects; for example, Drysdale et al. found that "some pairs of close sites have reduced levels of linkage disequilibrium relative to more spaced pairs of sites" (44, p. 10485).

The pooled allele frequencies at both the *Arg16* and *Gln27* sites confirm the presence of significant variation between racial groups and are similar to values generally recognized, for example, in ALFRED (Allele Frequency Database) (48). Although crude, these results do support a role of these polymorphisms in asthma susceptibility, given the varying incidence of asthma in these racial groups. Interestingly, the variation was more marked at the *Gln27* locus than at *Arg16*, and it was the former that was more strongly implicated in asthma susceptibility in our results.

These findings must be taken with caution at the present time for a number of reasons. First, these estimates are obtained by pooling despite heterogeneity.

Second, the asthma phenotype was often not fully specified, and details of asthma diagnoses were often scanty. Future studies should clearly identify whether asthma cases were diagnosed from symptoms or on population screening, and they should include results of atopic testing, spirometry, or methacholine challenge. Without sufficient information in individual studies, the condition labeled as asthma in this meta-analysis is likely to be heterogeneous and may be contributing to the inconsistency of results.

Third, the haplotype results are very different from those found in the longitudinal Normative Aging Study cohort (49), where the *Gly16/Gln27* haplotype had a protective effect compared with *Arg16/Glu27* (a different reference genotype), whereas in our study there was an increased risk. This discrepancy, however, may be due to the fact that, in the latter, the outcome was airway hyperresponsiveness (which does not always correspond to asthma) and that the population was general, community-dwelling males screened with a methacholine challenge test, not diagnosed asthmatics.

Fourth, these findings do not take into account smoking status, since data were available from only two studies (3, 42). There are some indications that the genotype effects may be more apparent among nonsmokers (49).

Fifth, the findings in childhood and adult asthma are inconsistent. This may be due to chance, or, alternatively,

there may be a genuinely different mode of action in adults compared with children, in that asthma is a clinically different disease in these two populations. Asthma in late childhood, which was the age range studied in these papers, is predominantly atopic in nature, more likely to be eosinophilic, more likely to be symptom diagnosed and episodic, and less likely to be associated with persistent airway hyperresponsiveness (49–52). Since the *Glu27* polymorphism is associated with less airway hyperresponsiveness (53), this may explain differences between the associations in adults and children. Alternatively, given the incomplete understanding of asthma pathogenesis, there may be pleiotropic effects of the β2-receptor at different stages or etiologies of disease. Indeed, one of us has observed such an age-specific association for another gene candidate in a population of children followed from childhood into early adult life (54).

In summary, these results are suggestive of a protective effect of the *Glu27* allele, probably as part of a haplotype, and they raise the possibility of interactions with the position 16 alleles and possibly other SNPs. This warrants further investigation in larger studies. The clinical implications of these findings are not clear. These polymorphisms may be involved in both conferring the risk to develop asthma and influencing the response to β2-agonists; this has been the subject of a recent randomized crossover trial (55) and is the topic of an ongoing meta-analysis (56).

REFERENCES

1. Attia J, Thakkinstian A, D'Este C. Systematic review of meta-analyses of molecular association studies: methodologic lesson for genetic epidemiology. J Clin Epidemiol 2003;56: 297–303.
2. Barr RG, Cooper DM, Speizer FE, et al. Beta(2)-adrenoceptor polymorphism and body mass index are associated with adult-onset asthma in sedentary but not active women. Chest 2001;120:1474–9.
3. Santillan AA, Camargo CA Jr, Ramirez-Rivera A, et al. Association between beta2-adrenoceptor polymorphisms and asthma diagnosis among Mexican adults. J Allergy Clin Immunol 2003;112:1095–100.
4. Hall IP. Beta2-adrenoceptor polymorphisms and asthma. Clin Exp Allergy 1999;29:1151–4.
5. Hall IP. Genetics and pulmonary medicine: asthma. Thorax 1999;54:65–9.
6. Bracken MB, Belanger K, Cookson WO, et al. Genetic and perinatal risk factors for asthma onset and severity: a review and theoretical analysis. Epidemiol Rev 2002;24:176–89.
7. Thakkinstian A, D'Este C, Eisman J, et al. Meta-analysis of molecular association studies: vitamin D receptor gene polymorphisms and BMD as a case study. J Bone Miner Res 2004; 19:419–28.
8. Thakkinstian A, McElduff P, D'Este C, et al. A method for meta-analysis of molecular association studies. Stat Med 2005; 24:1291–306.

9. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351–5.

10. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18: 2693–708.

11. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 1997;16:2741–58.

12. Thakkinstian A, D'Este C, Attia J. Haplotype analysis of VDR gene polymorphisms: a meta-analysis. Osteoporos Int 2004; 15:729–34.

13. Chapman NH, Thompson EA. Linkage disequilibrium mapping: the role of population history, size, and structure. In: Rao DC, Province MA, eds. Genetic dissection of complex traits. New York, NY: Academic Press, 2001:417–28.

14. StataCorp. Stata statistical software: release 8.0. College Station, TX: Stata Corporation, 2003.

15. Spiegelhalter D, Thomas A, Best N. WinBUGS user manual. Cambridge, United Kingdom: MRC Biostatistics Unit, Institute of Public Health, 2001.

16. Kawamura T, Egusa G, Fujikawa R, et al. Gln27Glu variant of the beta2-adrenergic receptor gene is not associated with obesity and diabetes in Japanese-Americans. Metabolism 2001; 50:443–6.

17. Rosmond R, Ukkola O, Chagnon M, et al. Polymorphisms of the beta2-adrenergic receptor gene (ADRB2) in relation to cardiovascular risk factors in men. J Intern Med 2000;248:239–44.

18. Kotanko P, Binder A, Tasker J, et al. Essential hypertension in African Caribbeans associates with a variant of the beta2-adrenoceptor. Hypertension 1997;30:773–6.

19. Heckbert SR, Hindorff LA, Edwards KL, et al. Beta2-adrenergic receptor polymorphisms and risk of incident cardiovascular events in the elderly. Circulation 2003;107:2021–4.

20. Dallongeville J, Helbecque N, Cottel D, et al. The Gly^{16}→ Arg^{16} and Gln^{27}→Glu^{27} polymorphisms of β$_2$-adrenergic receptor are associated with metabolic syndrome in men. J Clin Endocrinol Metab 2003;88:4862–6.

21. Tang W, Devereux RB, Kitzman DW, et al. The Arg16Gly polymorphism of the β2-adrenergic receptor and left ventricular systolic function. Am J Hypertens 2003;16:945–51.

22. Aynacioglu AS, Cascorbi I, Gungor K, et al. Population frequency, mutation linkage and analytical methodology for the Arg16Gly, Gln27Glu and Thr164Ile polymorphisms in the beta2-adrenergic receptor among Turks. Br J Clin Pharmacol 1999;48:761–4.

23. Kahara T, Hayakawa T, Nagai Y, et al. Gln27Glu polymorphism of the beta2 adrenergic receptor gene in healthy Japanese men is associated with the change of fructosamine level caused by exercise. Diabetes Res Clin Pract 2004;64:207–12.

24. Candy G, Samani N, Norton G, et al. Association analysis of beta2 adrenoceptor polymorphisms with hypertension in a Black African population. J Hypertens 2000;18:167–72.

25. Duarte NL, Colagiuri S, Palu T, et al. Obesity, type II diabetes and the β2 adrenoceptor gene Gln27Glu polymorphism in the Tongan population. Clin Sci (Lond) 2003;104:211–15.

26. Sugaya K, Nishijima S, Yamada T, et al. Molecular analysis of adrenergic receptor genes and interleukin-4/interleukin-4 receptor genes in patients with interstitial cystitis. J Urol 2002; 168:2668–71.

27. Chang TJ, Tsai MH, Jiang YD, et al. The Arg16Gly polymorphism of human beta2-adrenoreceptor is associated with type 2 diabetes in Taiwanese people. Clin Endocrinol (Oxf) 2002;57:685–90.

28. Kim SH, Kim DJ, Seo IA, et al. Significance of beta2-adrenergic receptor gene polymorphism in obesity and type 2 diabetes mellitus in Korean subjects. Metabolism 2002;51: 833–7.

29. Iwamoto N, Ogawa Y, Kajihara S, et al. Gln27Glu β2-adrenergic receptor variant is associated with hypertriglyceridemia and the development of fatty liver. Clin Chim Acta 2001;314:85–91.

30. Weir TD, Mallek N, Sandford AJ, et al. β$_2$-Adrenergic receptor haplotypes in mild, moderate and fatal/near fatal asthma. Am J Respir Crit Care Med 1998;158:787–91.

31. Xie HG, Stein CM, Kim RB, et al. Frequency of functionally important beta-2 adrenoceptor polymorphisms varies markedly among African-American, Caucasian and Chinese individuals. Pharmacogenetics 1999;9:511–16.

32. Holloway JW, Dunbar PR, Riley GA, et al. Association of beta2-adrenergic receptor polymorphisms with severe asthma. Clin Exp Allergy 2000;30:1097–103.

33. Dewar JC, Wheatley AP, Venn A, et al. Beta2-adrenoceptor polymorphisms are in linkage disequilibrium, but are not associated with asthma in an adult population. Clin Exp Allergy 1998;28:442–8.

34. Arnaiz NO, Kaufman JD, Daroowalla FM, et al. Genetic factors and asthma in aluminum smelter workers. Arch Environ Health 2003;58:197–200.

35. Reihsaus E, Innis M, MacIntyre N, et al. Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects. Am J Respir Cell Mol Biol 1993;8:334–9.

36. Hakonarson H, Bjornsdottir US, Ostermann E, et al. Allelic frequencies and patterns of single-nucleotide polymorphisms in candidate genes for asthma and atopy in Iceland. Am J Respir Crit Care Med 2001;164:2036–44.

37. Martinez FD, Graves PE, Baldini M, et al. Association between genetic polymorphisms of the β$_2$-adrenoceptor and response to albuterol in children with and without a history of wheezing. J Clin Invest 1997;100:3184–8.

38. Binaei S, Christensen M, Murphy C, et al. β$_2$-Adrenergic receptor polymorphisms in children with status asthmaticus. Chest 2003;123:375S.

39. Hopes E, McDougall C, Christie G, et al. Association of glutamine 27 polymorphism of β$_2$ adrenoceptor with reported childhood asthma: population based study. BMJ 1998;316:664.

40. Lin YC, Lu CC, Shen CY, et al. Roles of genotypes of beta2-adrenergic receptor in the relationship between eosinophil counts and lung function in Taiwanese adolescents. J Asthma 2003;40:265–72.

41. Leung TF, Tang NL, Chan IH, et al. Distribution in allele frequencies of predisposition-to-atopy genotypes in Chinese children. Pediatr Pulmonol 2002;34:419–24.

42. Wang Z, Chen C, Niu T, et al. Association of asthma with β$_2$-adrenergic receptor gene polymorphism and cigarette smoking. Am J Respir Crit Care Med 2001;163:1404–9.

43. Shachor J, Chana Z, Varsano S, et al. Genetic polymorphisms of the beta-2 adrenergic receptor in Israelis with severe asthma compared to non-asthmatic Israelis. Isr Med Assoc J 2003; 5:821–4.

44. Drysdale CM, McGraw DW, Stack CB, et al. Complex promoter and coding region β$_2$-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci U S A 2000;97:10483–8.

45. Comings DE, MacMurray JP. Molecular heterosis: a review. Mol Genet Metab 2000;71:19–31.

46. Morahan G, Huang D, Wu M, et al. Association of IL12B promoter polymorphism with severity of atopic and non-atopic asthma in children. Lancet 2002;360:455–9.

47. Gobeil S, Rodrigue MA, Moisan S, et al. Intracellular sequestration of hetero-oligomers formed by wild-type and

glaucoma-causing myocilin mutants. Invest Ophthalmol Vis Sci 2004;45:3560-7.

48. ALFRED: the ALlele FREquency Database. New Haven, CT: Department of Genetics, Yale University School of Medicine, 2005. (http://alfred.med.yale.edu/alfred/index.asp).

49. Litonjua AA, Silverman EK, Tantisira KG, et al. β2-Adrenergic receptor polymorphisms and haplotypes are associated with airways hyperresponsiveness among nonsmoking men. Chest 2004;126:66–74.

50. Gibson PG, Simpson JL, Saltos N. Heterogeneity of airway inflammation in persistent asthma: evidence of neutrophilic inflammation and increased sputum interleukin-8. Chest 2001; 119:1329–36.

51. Douwes J, Gibson P, Pekkanen J, et al. Non-eosinophilic asthma: importance and possible mechanisms. Thorax 2002; 57:643–8.

52. Gibson PG, Simpson JL, Hankin R, et al. Relationship between induced sputum eosinophils and the clinical pattern of childhood asthma. Thorax 2003;58:116–21.

53. Ramsay CE, Hayden CM, Tiller KJ, et al. Polymorphisms in the beta2-adrenoreceptor gene are associated with decreased airway responsiveness. Clin Exp Allergy 1999;29:1195–203.

54. Srivastava P, Helms PJ, Stewart D, et al. Association of CCR5Δ32 with reduced risk of childhood but not adult asthma. Thorax 2003;58:222–6.

55. Israel E, Chinchilli VM, Ford JG, et al. Use of regularly scheduled albuterol treatment in asthma: genotype-stratified, randomised, placebo-controlled cross-over trial. Lancet 2004; 364:1505–12.

56. Hancox B, Gibson P. Beta2-receptor polymorphism for improving long-term beta-agonist efficacy in asthma (protocol for a Cochrane review). In: The Cochrane Library, issue 2. Oxford, United Kingdom: Cochrane Database of Systematic Reviews, 2005. (http://cochrane.bireme.br/cochrane/show. php?db=&mfn=&id=_ID_CD003276&lang=pt&dblang=).

APPENDIX

The pooled prevalence was calculated as

$$\bar{p} = \frac{\sum w_i p_i}{\sum w_i},$$

where $\bar{p}$ was the pooled prevalence of the allele, p_i was the prevalence of the allele in each study, and w_i was $1/\mathrm{var}(p_i)$, which was the weight of each study.

Heterogeneity of prevalences across studies was checked as follows:

$$Q = \sum w_i(p_i - \bar{p})^2.$$

The Q statistic follows a χ^2 distribution with number of studies $(k) - 1$ df. If heterogeneity was present, between-study variation was then estimated as follows:

$$\tau^2 = \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \text{ if } Q > k - 1 \text{ or } 0 \text{ otherwise.}$$

This was used to calculate a weight term that accounted for between-study variation:

$$w_i^* = \frac{1}{\mathrm{var}(p_i) + \tau^2},$$

and the pooled prevalence was estimated as follows:

$$\bar{p}^* = \frac{\sum w_i^* p_i}{\sum w_i^*}.$$

The 95 percent confidence interval was estimated as follows:

$$95 \text{ percent CI} = \bar{p}^* \pm \frac{1.96}{\sqrt{\sum w_i^*}}.$$

APPENDIX TABLE 1. Scale for quality assessment of molecular association studies of asthma

Criteria	Score
Representativeness of cases	
Consecutive/randomly selected from case population with clearly defined sampling frame	2
Consecutive/randomly selected from case population without clearly defined sampling frame or with extensive inclusion/exclusion criteria	1
No method of selection described	0
Representativeness of controls	
Controls were consecutive/randomly drawn from the same sampling frame (ward/community) as cases	2
Controls were consecutive/randomly drawn from a different sampling frame as cases	1
Not described	0
Ascertainment of asthma	
Clearly described objective criteria for diagnosis of asthma	2
Diagnosis of asthma by patient self-report or by patient history	1
Not described	0
Ascertainment of controls	
Controls were tested to screen out asthma, i.e., measured FEV₁* or PEFR*	2
Controls were subjects who did not report asthma; no objective testing	1
Not described	0
Genotyping examination	
Genotyping done under "blinded" condition	1
Unblinded or not mentioned	0
Hardy-Weinberg equilibrium	
Hardy-Weinberg equilibrium in control group	2
Hardy-Weinberg disequilibrium in control group	1
No checking for Hardy-Weinberg equilibrium	0
Association assessment	
Assess association between genotypes and asthma with appropriate statistics and adjustment for confounders	2
Assess association between genotypes and asthma with appropriate statistics without adjustment for confounders	1
Inappropriate statistics used	0
Response rate	
Response rates for both groups are the same, i.e., to within 5%	2
Response rates are different, between 5% and 10%	1
Response rates are more than 10% different, or no mention of response rates	0
Total	

* FEV₁, forced expiratory volume in 1 second; PEFR, peak expiratory flow rate.