The HuGENet™ HuGE Review Handbook, version 1.0

Editors:

Julian Little[6]
Julian Higgins[2,3]

Writing group:

Molly Bray[1]
Julian Higgins[2,3]
John Ioannidis[4]
Muin Khoury[5]
Julian Little[6]
Teri Manolio[7]
Liam Smeeth[8]
Jonathan Sterne[9]

on behalf of the HuGE Net Systematic Review and Meta-Analysis Working Group (Betsy Anagnostelis, Adam Butterworth, John Danesh, Carol Dezateux, Doug Easton, John Gallacher, Marta Gwinn, Julian Higgins, John Ioannidis, Muin Khoury, Sarah Lewis, Julian Little, Teri Manolio, David Melzer, Cosetta Minelli, Paul Pharoah, Georgia Salanti, Simon Sanderson, Liam Smeeth, Lesley Smith, Jonathan Sterne, Donna Stroup, Emanuela Taioli, John Thompson, Simon Thompson, Neil Walker, Ron Zimmern)

and HuGE Net Executive Group (Molly Bray, Marta Gwinn, Julian Higgins, John Ioannidis, Muin Khoury, Julian Little, Teri Manolio, Ron Zimmern)

[1]USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX 77030, USA; [2]MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 2SR, UK; [3]Public Health Genetics Unit, Strangeways Research Laboratory, Cambridge CB1 8RN, UK; [4]Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina 45110, Greece; [5]Office of Genomics and Disease Prevention, CDC, Atlanta, GA 30333, USA; [6]Canada Research Chair in Human Genome Epidemiology, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada; [7]National Human Genome Research Institute, National Institutes of Health, Bethesda, USA; [8]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; [9]Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK.

CONTENTS

CITATION & FEEDBACK

This handbook should be cited as

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

We intend to revise periodically taking account of accumulating evidence about methodological issues regarding the investigation of genetic associations and related interactions, and feedback from users. Therefore, we would be very grateful for comments on this first version of the handbook, which may be posted on the above website.

1. INTRODUCTION

The synthesis of knowledge is crucial to the evidence-based integration of human genomics into the practice of medicine and public health in the 21st century (Khoury et al. 2004). There has been a rapid increase in the number of papers on gene-disease associations and related interactions. This trend is expected to accelerate as a result of the increasing availability of mapped single nucleotide polymorphism markers (Little et al. 2003) and improvements in genotyping technology, allowing massive testing of genetic variants in minimal time (Lawrence et al. 2005, Marchini et al. 2005, Thomas et al. 2005). There has been considerable concern about non-replication of gene-disease association studies (Cardon & Bell 2001, Gambaro et al. 2000, Hirschhorn et al. 2002, Ioannidis et al. 2001b, Lancet 2003, Lohmueller et al. 2003, Nature Genetics 1999, Tabor et al. 2002). In addition, the combination of high-throughput genotyping, exploratory statistical analyses in studies with limited sample sizes, and selective reporting potentially could generate a scientific literature replete with spurious findings, leading to wasteful investment of resources, unless mechanisms are put in place to evaluate accruing evidence in a timely manner (Ioannidis et al. 2005).

Systematic reviews and meta-analyses are one mechanism for evaluating the overall effect of a polymorphism and/or gene. The strengths and limitations of such overviews are well established for clinical trials, largely through the efforts of the Cochrane Collaboration. Systematic reviews and meta-analyses are increasingly being applied to observational studies (Stroup et al. 2000). Currently there are as many meta-analyses of observational data conducted as there are of clinical trials. The citation impact of both types of meta-analyses is equally high, the highest among all study designs in the health sciences (Patsopoulos et al. 2005). Meta-analyses of gene-disease association studies have been accepted as a key method for establishing the genetic components of complex diseases (Ioannidis et al. 2001b, Lohmueller et al. 2003). HuGE reviews were introduced as a means of integrating evidence from human genome epidemiologic studies, that is, population-based studies of the impact of human genetic variation on health and disease (Khoury & Dorman 1998). HuGE reviews are intended to be systematic, peer-reviewed synopses of epidemiologic aspects of human genes. Khoury & Dorman (Khoury & Dorman 1998) originally proposed a format for HuGE reviews, and more specific guidance for their content was provided in the *American Journal of Epidemiology* in 2000. In January 2001 an expert panel workshop, convened by the Centers for Disease Control and Prevention and the National Institutes of Health, led to recommendations regarding considerations that should be addressed in reporting studies of genotype prevalence in gene disease associations, both for individual investigators and systematic reviews (Little et al. 2002). Experience with the HuGE review process (Little et al. 2003) led to provision of revised guidance and formats for reviews (HuGE Net 2005). Four types of review were suggested: full reviews, association reviews, prevalence reviews and mini-reviews.

A limitation of these recommendations and guidance for HuGE reviews was lack of empirical evidence on which to base them. The widespread adoption of the CONSORT statement (Altman et al. 2001, Moher et al. 2001) on the reporting of randomized

controlled trials in part is due to recognition that the recommendations are evidence-based, and statements about the reporting of other types of investigation such as MOOSE (Stroup et al. 2000), QUOROM (Moher et al. 1999), STARD (Bossuyt et al. 2003) and TREND (Des Jarlais et al. 2004) have adopted similar principles. Poor reporting of observational studies (Pocock et al. 2004) is a general challenge in synthesizing evidence, and a **ST**rengthening the **R**eporting of **OB**servational studies in **E**pidemiology (STROBE) statement is under development (STROBE Group 2005, von Elm & Egger 2004). Therefore, it is timely to reconsider guidelines for systematic review and meta-analysis of gene disease association studies.

We now revise and extend the guidance for HuGE reviews on the basis of a systematic review methodology workshop, organised by HuGENet™ and the Cambridge Genetics Knowledge Park in November 2004. We focus on overviews of evidence from association studies. Appendix A presents a suggested format for overviews of data on the prevalence of gene variants (genotype prevalence and allele frequencies), and Appendix B presents a suggested revised format for full HuGE reviews.

HuGE reviews should use methods that are systematic and explicit. We also strongly urge investigators authoring HuGE reviews to use quantitative methodology (meta-analytic methods) for the analysis and synthesis of the evidence, as deemed appropriate in each occasion. This handbook aims to help authors make sensible decisions about the methods they use, rather than to dictate arbitrary standards.

2. FORMAT FOR HuGE ASSOCIATION REVIEWS AND META-ANALYSES

In this section, we present the format for what is likely to be the most common type of HuGE review and meta-analysis in the immediate future, i.e. of gene-disease associations. Formats for other types of HuGE reviews are presented in Appendices A and B.

2.1 General format

Cover sheet

Title (stating whether a meta-analysis was perfomed), whether new or update, contact details.

Abstract

Provide a one-page structured synopsis of the issues discussed in the items below with a brief statement on each of these items. If possible, supply keywords, including the name(s) of the gene(s), the name(s) of the disease(s) or disorder(s), the word 'epidemiology' and the term 'systematic review' or 'meta-analysis'.

Background

Gene(s)

Identify the gene(s) being reviewed and provide a brief review of chromosome location, gene product, and function, if known.

Gene variants and frequency

List known allelic variants with effects on gene product if known.
Summarize known information on the frequency of homozygosity and heterozygosity of these variants in different populations and ethnic groups. If a prevalence review exists, summarise its findings. If a prevalence review does not exist, briefly overview the available data with some key references. (Note that the statement on the specific gene variants on which the review is focused is made under 'selection criteria' in Methods.)

Disease(s) or other outcomes

Identify the disease(s) or other outcome(s) with which the gene(s) is/are believed to be associated. Briefly summarize the descriptive epidemiology and confirmed and suspected risk factors (including other genes). Refer to previous similar reviews, if available, and be succinct. Outline the rationale (if any) for the postulated association with the disease(s)/outcomes in the current review. (Note that the statement on the specific disease or outcome on which the review is focused is made under 'selection criteria' in Methods.)

Objectives

Provide a succinct summary of the objectives of the current review.

Methods

Selection criteria

State the gene(s), gene variant(s), disease and types of participants eligible for inclusion in the review.
State the types of study (e.g., design and conduct) eligible for investigation of association.
State the types of study (e.g., design and conduct), other gene(s) and environmental exposures eligible for investigation of interactions (if included in the review).

Identifying studies

Describe the methods used to identify relevant studies and/or other sources of information. List all electronic databases searched, with details of the search strategies and the periods for which they were searched, and describe any communication with investigators.

Data collection and analysis

Describe the methods for selection of studies, data collection (including data extraction from published reports and any attempt to retrieve unpublished or partially/selectively published data), assessment of risk of bias, methods for analysis of individual studies, methods for meta-analysis, and methods for dealing with heterogeneity and potential biases.

Results

Included studies

Include a table providing basic details of the included studies (location, date, design, types of participants (cases and controls)).

Quality and methodology of studies

Comment on the quality and methodology of studies.

Associations

Summarize the magnitude of the association between the allelic variants and the disease(s) and outcomes of interest in terms of relative, absolute, and/or attributable risks in different populations.

Interactions

Discuss whether the allelic variants interact with other risk factors for the disease, including other genes and environmental factors. Summarize the magnitude of such interactions, whenever possible. State variables adjusted for in any adjusted analyses.

Discussion

Main findings

Summarise the main findings of the review and the meta-analysis.

Limitations

Comment succinctly on the quality of the evidence. Discuss concerns over amount of relevant information, validity of individual studies, and other biases (e.g. publication bias and other reporting biases).

Biology

Comment on available mechanistic evidence relevant to the association.

Potential public health impact and other implications of results

(a)Potential public health impact

Summarize potential public health applications of human genome epidemiological information on the variants of the gene(s), e.g. available interventions, setting permissible exposure thresholds for individuals with specific genotypes.

(b) Implications for our understanding of disease

(c) Implications for research

Strengths and gaps in the evidence base should be identified. Recommendations should be made to stimulate research to fill any gaps, e.g. what research might be needed to give public health consequence to the summarized genetic knowledge.

Potential conflicts of interest

Any potential conflict of interest that might influence the judgments of reviewers should be noted (see 2.2). If none, this should be stated explicitly.

References

Internet sites

Include relevant links to various genetics databases, online resources, educational materials, consensus statements, policy statements, and support groups.

2.2 Potential conflicts of interest

HuGE reviews should be free of any bias introduced by the receipt of any benefit in cash or kind, any hospitality, or any subsidy derived from any source that may have or be perceived to have an interest in the outcome of the review.

Under the heading 'Conflict of Interest' reviewers should report any conflict of interest capable of influencing their judgments, including personal, political, academic and other possible conflicts. If specific funding has been obtained to support the review, the source of funding should be specified. Any other interest (such as personal conflicts) that might unduly influence judgments made in a review (concerning, for example, the inclusion or exclusion of studies, assessments of the validity of included studies or the interpretation of results) should be reported.

Disclosing a competing interest does not necessarily reduce the worth of a review and it does not imply dishonesty. However, competing interests can influence judgments in subtle ways.

2.3 Publication of HuGE reviews in peer-reviewed journals

HuGE reviews may be submitted to the American Journal of Epidemiology, Epidemiologic Reviews, or one of the collaborating journals listed on the HuGENet™ website (http://www.cdc.gov/genomics/hugenet/reviews/guidelines.htm). These submissions will be peer-reviewed and, if accepted, will be published in one of these journals as well as in the HuGENet™ knowledge base on the World Wide Web.

When submitting completed HuGE reviews to the editor of the appropriate journal, authors should explain in the covering letter that the manuscript is for consideration as a HuGE review and suggesting the names of up to three possible reviewers. The technical requirements of each journal for submission of manuscripts should be followed. Because of space constraints, it may not be possible for a journal to publish all of the tables and references prepared for a HuGE review. However, any supporting tables that cannot be published in a journal due to space limitations can be presented on the HuGENet™ website. It is expected that the text of a HuGE review in a journal and that available on the HuGENet™ website will be similar and that reference would be made to the additional tables and references held on the website. Accepted papers will be published both in the appropriate journal and on the HuGENet™ website.

3. THE PROTOCOL

The rationale for requiring a protocol is that HuGE reviews, like other reviews, are retrospective and so knowledge of study results potentially could influence the review process, for example, the criteria for study selection and the comparisons made in the analyses. The aim is that HuGE systematic reviews are in themselves valid pieces of research, and thus need to be subject to the same rigorous scientific process as other types of research.

Investigators who propose to carry out HuGE reviews should provide a brief protocol. The protocol should be sent as part of an electronic message of intent to sign up for a HuGE review to the US HuGENet™ coordinator at HUGE@cdc.gov. An updated list of HuGE reviews under development is kept on the HuGENet™ website at http://www.cdc.gov/genomics/hugenet/reviews.htm. The US HuGENet™ coordinator will advise about which journal may be the most appropriate for the proposed topic.

3.1 Brief protocol for an association review or/and meta-analysis

Cover sheet

Title (stating whether a meta-analysis is planned), type of review (association review, clarify if it is also a meta-analysis), contact details

Background

Indicate potential importance of the gene/gene variant to public health (or understanding of aetiology/pathogenesis). Briefly explain rationale for the type of review to be undertaken. State objectives of the review.

Methods

Selection criteria

State the gene(s), gene variant(s), disease and types of participants to be included in the review.
State the types of study (e.g., design and conduct) for investigation of association.
State the types of study (e.g., design and conduct), other gene(s) and environmental exposures for investigation of interactions (if included in the review).

Identifying studies

Describe the proposed methods for identifying relevant studies - papers and/or other sources of information.

Data collection and analysis

Outline the proposed methods for: selection of studies; data collection (including data extraction from published reports and any attempt to retrieve unpublished or

partially/selectively published data); and assessment of risk of bias. Indicate whether meta-analysis is planned, for both associations and interactions. Specify details of primary hypothesis or hypotheses, and planned method of testing these.

Potential conflicts of interest

Any potential conflict of interest that might influence the judgments of authors should be noted (see 2.2). If none, this should be stated explicitly.

3.2 Planning the review

Clearly framed questions determine the focus of a systematic review and meta-analysis, guiding much of the review process, including inclusion criteria, strategies for locating and selecting studies or data, for critically appraising their relevance and validity, and for analysing variation among their results.

Inclusion criteria for association reviews in human genome epidemiology will need to specify the gene(s), allelic variant(s), specific disease(s) or other outcome(s) of interest and, ideally, which other exposures may potentially modify the association(s). Inclusion criteria may be different for different components of a review. In particular, different study designs may be suitable for assessing association and interaction.

3.2.1 The gene and allelic variants

A HuGE review will typically focus on a single gene, although groups of related genes might also be considered. It is usual for several loci on a gene to be addressed in a single review. The definition of the genotype(s) investigated should be clearly presented. The validity of grouping genotypes on the basis of putative functional effects depends on the availability and quality of functional studies of gene variants, and information on functional effects is likely to change over time. For multi-allelic systems, genotypes have been grouped according to functional effects in some investigations. For example, grouping according to inferred rapidity of acetylation has been done for the *NAT2* polymorphisms (Brockton et al. 2000).

Combining studies of different SNPs is not widely accepted, but it may be appropriate if the SNPs are in strong ($r^2 > 0.9$) linkage disequilibrium. An increasing number of studies which produce haplotype data will have to be considered in reviews and meta-analyses will have to target haplotypes (International HapMap Consortium 2003, Johnson et al. 2001, de Bakker et al. 2005). For haplotypes, it is important to clarify the method that has been used for their generation or imputation, the selected genetic contrasts and the methods for combing data across studies with due attention to avoid selective reporting and selective analyses. In addition, differences in linkage disequilibrium between ethnic groups may need to be considered.

3.2.2 Diseases and outcomes

A HuGE review will typically focus on a disease or a collection of closely related clinical outcomes. We encourage HuGE reviews to include intermediate outcomes (such as biomarkers or phenotypic traits) that may have a role in the pathway between gene and disease.

3.2.3 Interactions

Where possible, interactions should be considered, as these may enhance the interpretation of gene-disease associations. However, the available data may not be sufficient. The primary hypothesis or hypotheses should be clearly specified. Interactions can often only be appropriately addressed using individual participant data.

3.2.4 Study designs

Different study designs (e.g. family studies and studies of unrelated subjects) may be included if they are considered to be addressing the same research question (usually concerning associations in the population). However, different types of study may present results differently, making it difficult to combine across them. In the absence of a substantial body of empirical evidence, we suggest that inclusion criteria should generally veer towards being inclusive regarding study designs, sample size and susceptibility to bias. Possible effects of differences in these methodological features may be later addressed either by undertaking sensitivity analyses to assess robustness of conclusions, or carrying out meta-regression/subgroup analysis.

Study designs for addressing interactions may be different from study designs for examining associations. For example, case-only studies may be used for the former, but typically not for the latter.

Occasionally, it may be reasonable to exclude small studies (ideally with a pre-defined cut-point) in order to make the review more manageable when there is a large number of relevant studies. Also, smaller studies may be more likely to be biased, since sample size is sometimes correlated with efforts to minimise bias. In most cases, it should be possible to consider all studies regardless of sample size and have the benefit of examining whether small studies give different results than larger ones.

3.3 Published literature or individual participant data

Systematic reviews and meta-analyses of the published literature often run into obstacles. Reports frequently fail to present data in formats suitable for meta-analysis, and they may lack sufficient information on methods and key characteristics of the studies. Attempts should always be made to retrieve information from investigators (see 5.2), for example, by requesting results as grouped (aggregated) data. An alternative approach is a meta-analysis of individual participant data (MIPD), which involves the collection of detailed data on individual subjects.

The MIPD approach can offer many advantages over the meta-analysis of published data, including standardization of definitions of cases and variables, testing the assumptions of time-to-event models, better control of confounding, standardization of analyses of genetic loci that are in linkage disequilibrium, increased ability to evaluate alternative genetic models and multiple genes, consistent treatment of subpopulations, and assessment of sampling bias. Nevertheless, individual-level approaches require a much greater commitment of time and resources to collect primary data and to coordinate a large collaborative project (Steinberg et al. 1998). For questions that justify the required intensive effort, the MIPD approach is a useful tool to help to clarify the role of candidate genes in complex human diseases (Ioannidis et al. 2002a), and we recommend that this type of quantitative synthesis be done whenever possible. However, a meta-analysis based on published data can be useful to help decide whether resources should be invested in a MIPD.

In principle, a MIPD should reach the same conclusion as a meta-analysis of published data. This was the case for one HuGE review that required data on individual subjects in addition to performing a meta-analysis of grouped data (Engel et al. 2002). However, MIPD can reach different conclusions from meta-analyses of published data, and this is typically due to the inclusion of different studies or different data in the analysis. For example, an MIPD on the association of estrogen receptor alpha polymorphisms and osteoporosis outcomes found that none of three tested polymorphisms had an effect on bone mineral density, while the *XbaI* polymorphism had an effect on fracture risk (Ioannidis et al. 2004), while a meta-analysis of published data had suggested a modest effect on both bone mineral density and fracture risk for *XbaI* (Ioannidis et al. 2002b).

4. SEARCH METHODS

A comprehensive search is one of the key differences between a systematic review and a traditional review (Oxman 1992). It is especially important in a discipline that is seriously affected by reporting biases, because the most exciting results are likely to be published in places that are most accessible, and important complementary results may be available only through thorough searching.

4.1 Electronic databases

Searches generally start with electronic databases. A key source is PubMed (an online MEDLINE database that also includes up-to-date citations that have not yet been indexed), and this forms the basis of the HuGE Published Literature database (see Box 4.1). However, only a proportion of journals are indexed in PubMed. Other databases likely to be relevant include the ISI Science Citation Index, EMBASE and BIOSIS. There is a low degree of overlap in the journals covered by these databases. Experience in the field of clinical trials suggests that the degree of overlap varies considerably by topic (Higgins et al., 2005) (Higgins, Green 2005). Further evidence is required on the

distribution of gene-disease association references across different databases

Box 4.1 *The HuGE Published Literature database*
The HuGE Published Literature database (http://www.cdc.gov/genomics/hugenet) is built from references extracted from PubMed. This extraction process was started in October 2000 and involves review of abstracts for relevance to human genome epidemiology (specifically studies with information on one or more of genotype prevalence, gene-disease associations, gene-environment or gene-gene interactions, or evaluations of genetic tests). If relevance is unclear from the abstract, the full paper is checked. By January 17, 2006, this database contained over 19,000 articles concerning over 2000 genes and a similar number of diseases or outcomes. This is a helpful starting point for reviews but is not intended to be a comprehensive search tool.

4.1.1 The search strategy

Database searches for gene-disease association studies should generally take the form

(gene OR synonyms) AND (disease OR synonyms).

Search strategies should include both thesaurus terms (e.g. MeSH in MEDLINE) and free text. A methodological search filter (for example, adding 'AND (case-control study OR cohort study OR synonyms)') may be inappropriate, due to the inconsistent use of study design terms in papers and in indexing. However, such a filter might be considered in some instances when searches based on gene and disease yield overwhelming numbers of hits. This would need to be documented. On the other hand, when the number of hits is very limited, authors should consider eliminating completely the (disease OR synonyms) clause.

4.1.2 Geographical and language considerations

Journals published in languages other than English should be included in the search. There is compelling evidence that a large number of gene-disease association studies are published in non-English language journals not indexed in these common databases. In particular, there is evidence of a large Chinese genetic epidemiology literature (Pan et al. 2005) that appears in Chinese journals only indexed in the Chinese database of biomedical journals. The results of these Chinese studies suggest stronger genetic effects than the results of studies published in English language papers. This suggests the presence of an inverse language bias combined with possible publication bias. The situation is contrary to the language bias (tower of Babel bias) that has been described for randomized trials (Gregoire et al. 1995). Authors of reviews should balance all-inclusiveness against. the risk of introducing large bias from selectively published or reported analyses.

4.2 Other sources

Publication bias is potentially a serious problem for the integration of evidence. Numerous gene-disease association analyses are performed that are not written up in peer-reviewed papers. One method of minimizing the potential impact of publication bias is to identify such analyses through 'grey literature,' which includes conference proceedings, books, abstracts, technical reports, and journals that may not be identified by electronic searches. We recommend caution in using some types of 'grey literature' because the material may not be peer reviewed and may be subject to modification and revision and because the information on study methods may be insufficient to assess risk of bias. Evidence from randomised trials suggests that inclusion of grey literature tends to move the treatment effect towards a null result, but the direction of the effect is not always predictable (Burdett et al. 2003). However, given the strong risk of publication bias in the area of gene-disease association, we recommend inclusion of 'grey literature' providing that study quality can be assessed adequately.

Efforts are being made to foster the development of networks of investigators who are working on the same topic, who could maintain a register of ongoing work in the topic area (both by investigators within the network and by other investigators) (Ioannidis et al. 2005). This should in the future help control the problem of publication bias, and provide a resource for investigators undertaking systematic reviews.

In addition to these searches, it is valuable to check the reference lists of

1. any existing reviews on the topic (the HuGE Published Literature database (Box 4.1) provides a source of systematic reviews and meta-analyses of human genome epidemiologic studies published since October 2000;
2. relevant articles obtained for the review or meta-analysis.

4.3 Documenting the search strategy

In a review of 37 meta-analyses of gene-disease association studies published in the period 1966-2000, 65% stated the database used and 24% specified the search terms explicitly (Attia et al. 2003). We recommend that the details of the strategy used to identify relevant papers should be documented (Stroup et al. 2000).The databases and registries searched, the interface (e.g. PubMed, SilverPlatter or Ovid for MEDLINE), search strategy (including thesaurus and free-text terms, and time period) and date on which the search was performed should be specified. Further efforts made to identify studies including contact with authors and use of unpublished material should be described.

5. REVIEWING METHODS

5.1 Selecting studies

Studies should be selected to comply with stated criteria for inclusion, using the best information available from reports and through contact with investigators. In a literature-based systematic review, there a several steps in the process of reducing items identified by the search to those falling within the scope of the review. A typical process is as follows:

1. Merge the records from electronic reference databases using reference management software and remove duplications;
2. Scan the titles and abstracts to remove obviously irrelevant reports;
3. Retrieve full text of the potentially relevant reports that remain;
4. Assess each full text report for compliance with the inclusion criteria;
5. Examine other sources of studies (such as 'grey literature' and references of relevant publications) for compliance with the inclusion criteria;
6. Correspond with investigators, where appropriate, to clarify eligibility;
7. Make final decisions on inclusion.

Steps 2, 4, 5 and 7 should be undertaken by at least two people, independently, to reduce the risk of overlooking relevant studies.

Selection of studies for a meta-analysis of individual participant data may be different. For example, the investigators' willingness to be involved in the collaboration would be necessary. We recommend that such meta-analyses also seek and report on studies not included in the collaboration, and consider undertaking subsidiary analyses that additionally include any available data from such studies.

It must be remembered that the units of fundamental interest in a systematic reviews and meta-analyses are *studies* and not *reports*, even in reviews only of the literature. There have been numerous instances of sequential or multiple publications of analyses of the same or overlapping data sets. This should be carefully appraised and detected by comparison of geographic locations, author names, and period of study if specified. If it is clear that the reports relate to the same or overlapping data sets, then we recommend including results from the largest or most recent publication. However, it is not unlikely that the methodology or the study population is described in greater detail in an earlier publication. If so, we recommend including the reference to the earlier publication with the reference all publications from which information was abstracted.

5.2 Collecting Data

There are three common sources of data (including information on methodology and other characteristics of the studies) in HuGE reviews:

1. Data extracted from published reports;
2. Data obtained through correspondence with investigators;
3. Individual participant data in the context of a collaborative venture with the investigators.

Literature-based reviews typically use the first two sources, although individual participant data may be offered in response to correspondence.

Issues in data extraction are discussed in some detail in Section 7 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins, Green 2005).

6. ASSESSING AND ADDRESSING POTENTIAL BIASES

Each study should be assessed for susceptibility to bias. The relevant aspects of a study that need to be addressed may depend on the topic under review.

Quality assessment of individual studies that are summarised in systematic reviews is necessary to limit bias in conducting the systematic review, gain insight into potential comparisons, and guide interpretation of findings (Higgins, Green 2005). In relation to gene-disease associations and related interactions, the focus of attention at present is on internal validity as distinct from generalisability of results.

Many papers deal with the critical appraisal of observational studies in general, and there has been concern both about validity (Benson & Hartz 2000, Concato et al. 2000, Concato & Horwitz 2004, Lawlor et al. 2004, Pocock & Elbourne 2000, Vandenbroucke 2004) and adequacy of reporting (Pocock et al. 2004, von Elm & Egger 2004). Checklists or guidelines for reporting gene-disease associations have been proposed (Cooper et al. 2002, Huizinga et al. 2004, Little et al. 2002, Nature Genetics 1999, Rebbeck et al. 2004, Weiss 2001, Ehm et al. 2005, Freimer & Sabatti 2005). Such checklists are not always used by study authors. An attempt should be made to quantify the potential direction and magnitude of any biases. In studies of gene-disease associations, and gene-environment and gene-gene interactions, there is the potential for selection bias, bias or error in outcome ascertainment, information bias (analytical validity of genotyping and, in studies of gene-environment interaction, the validity of exposure assessment) and confounding.

6.1 Selection bias

Selection bias occurs when the subjects included in a study are not representative of the source population, and/or when there are selective losses from the study population prior to data analysis (Hill & Kleinbaum 2000). If this were differential by genotype and/or exposure, it would distort the observed genetic and interaction effect sizes. If this were non-differential by these factors, the internal validity of the estimates would not be compromised, but it may not be possible to generalize to other populations and groups of patients.

Evaluation of potential selection bias requires consideration of the study design and setting of the study. Selection biases are similar to those described also for epidemiological studies of non-genetic risk factors (Breslow & Day 1980, Elwood 1998, Kelsey et al. 1996, dos Santos Silva 1999). However, some issues peculiar to gene association studies that may need to be taken into account are outlined below.

6.1.1 Extreme vs. unselected cases (spectrum of disease bias)

Some studies may select only ‘extreme’ cases for the disease or outcome of interest. If disease severity or heritability correlates with the strength of the association, then these

studies will obtain different estimates of effects compared with studies that include a broader spectrum of cases or a cohort unselected for disease severity or heritability. For example, cases may be over-selected from multiplex families with strong heritability; or only very severe cases may be eligible. Presenting the measure of association as an estimate of population association may be inappropriate in this situation. We should acknowledge, however, that we lack empirical evidence on the extent of spectrum of disease bias in genetic epidemiology of complex diseases.

6.1.2 Other biased selection of cases

In a number of studies the selection of cases is not well described (Botto & Yang 2000). In a review of type1 diabetes and HLA-DQ polymorphisms, the authors noted that many studies were based on convenience samples of cases in which persons with type 2 diabetes who used insulin in their treatment regimen had been included (Dorman & Bunker 2000).

Inclusion of prevalent cases would bias the estimate of effect if the genotype affected survival. For example, a positive association between the butyrylcholinesterase K variant and Alzheimer's disease among persons with an *APOE*4 allele has been observed in studies in which the age of cases was defined as age of onset, but not in studies where the age was defined as current age or age at death (Lehmann et al. 2001). In addition, bias could occur if genotypes were assayed by a phenotypic test, the outcome of which was influenced by disease progression or treatment as well as the underlying genotype. This may account for differences in the association between colorectal cancer and *NAT2* based on acetylator phenotype and genotype (Brockton 2002).

6.1.3 Biased selection of controls

Control groups are often described in even less detail than cases. In a number of case-control studies of gene-disease associations with unrelated controls, controls have not been selected from the same source population as the case-subjects (Botto & Yang 2000, Brockton et al. 2000, Cotton et al. 2000, Dorman & Bunker 2000). The potential problem of selecting controls who do not represent the population from which case-subjects arise is illustrated by the divergence in odds ratios for the association between colorectal cancer and the *GSTT1* null genotype (Chenevix-Trench et al. 1995), when the different control groups were analyzed (Cotton et al. 2000). This example also illustrates the potential effects of the method of recruitment of subjects in studies of genotype prevalence. Many early studies were based on convenience samples and not infrequently, little information was given about how the samples were selected (Wang et al. 1999) (Brockton et al. 2000, Dorman & Bunker 2000).

6.1.4 Differential participation in cases and controls

Selection bias also may occur as a result of non-participation that is differential between cases and controls. Incomplete participation can occur as a result of (1) refusal or inability to provide data; (2) refusal or inability to provide biological specimens; (3)

insufficient amount or quality of data, limiting the analysis; and (4) insufficient amount or quality of DNA (for example, in samples pulled from a repository or other archive), limiting numbers of assays. This would bias the odds ratio for gene-disease association only if the difference were related to genotype, and there would be bias in the odds ratio for interaction if the difference were related to genotype, exposure or both. There has been concern about a decline in participation rates (Morton et al. 2005, Olson 2001), especially in population-based studies. However, information on the potential effects of low participation rates is limited (Madigan et al. 2000, Morton et al. 2005).

6.1.5 Loss to follow-up

In cohort studies in general, bias would result if there were loss to follow-up that was differential by exposure or genotype, for example, if both loss to follow-up and exposure varied by socio-economic status, or if loss to follow-up and genotype prevalence varied by ethnic group. A related issue concerns the return of incomplete information during follow-up, i.e. item non-response. This has been shown to be associated with subsequent loss to follow-up (Deeg et al. 2002).

6.1.6 Selection bias considerations for gene-environment interactions

With regard to the possible effects of selection bias on estimates of gene-environment interaction, Morimoto et al. (Morimoto et al. 2003) showed that estimates of gene-environment interaction, defined as departure from a multiplicative joint effect, will not be subject to selection bias when genotype itself does not influence participation. This applies even when selection is influenced by exposure and disease status and genotype is associated with either or both of these. Similarly, in a hospital based case-control study, Wacholder et al. (2002) noted that even if exposure (therapy), or genotype, or both are associated with the condition that led to the hospitalization of control subjects, a departure from multiplicative effects can be estimated without bias. However, this is not the case for departure from additive effects. Although including controls with more than one type of disease might reduce a bias resulting with one disease being associated with exposure, genotype or both, pooling of controls with different diseases can lead to bias in assessing departure from multiplicative interaction, even if there is no such interaction in each individual disease-specific control set (Wacholder et al. 2002). These issues apply more generally to estimation of both interaction effects and gene-disease associations when controls are not selected from the same source population as the case-subjects (Little 2004, Wacholder et al. 2002)

6.2 Information bias

Information bias occurs as a consequence of errors in assessing factors of interest, here genotype, environmental exposures, or both, or in assessment of outcome. It is differential when the assessment of the factors of interest is influenced by the outcome under investigation, or vice versa.

6.2.1 Assessment of genotypes

Misclassification of genotype can bias genotype-disease associations, usually towards the null (Garcia-Closas et al. 2004, Rothman et al. 1993) – when it is non-differential. The most marked bias occurs when genotyping sensitivity is poor and genotype prevalence is high (>85%), or genotyping specificity is poor and genotype prevalence genotype is low (<15%) (Rothman et al. 1993). When exposure measurement error is substantial, genotyping errors of the order of 3% can substantially under-estimate an interaction effect (Wong et al. 2004). When there are systematic differences in genotyping according to outcome status (differential error), serious bias in any direction may occur.

Factors affecting the potential extent of misclassification (information bias) of genotype include the types and quality of samples, timing of collection, and the method used for genotyping (Little et al. 2002, Steinberg & Gallagher 2004). Quality control procedures are important in assessing the extent of possible misclassification and the extent to which this might be differential. Quality control procedures include, but are not limited to, re-analysis of random samples, analysis of samples with a different genotyping method, analysis of replicate samples, or sequencing, and analysis of samples at a central quality control facility.

For studies focused on specific genotypes, it is also very useful to record whether laboratory staff are blinded to outcome and to environmental exposures of interest, and whether genotyping of cases and controls was performed together or separately. Similarly, it should be recorded whether people appraising outcomes and exposures are blinded to the genotyping results. Unblinded assessment may lead to differential misclassification.

For genome-wide association studies, shifts between cases and controls in the point clusters corresponding to each genotype, thought to be due to differences in DNA processing, have been reported (Thomas et al. 2005). In this situation, using blinded samples to determine the parameters for allele calling could lead to differential misclassification. To minimize such misclassification, it would be necessary to calibrate the software separately for each group.

The extent of genotyping error has been reported to vary between about 1% and 30% (Akey et al. 2001). In Europe, in annual external quality assurance (EQA) schemes for cystic fibrosis run since 1996, the error rate for the detection of common *CFTR* mutations was 3.8% in 1996 and 1.3 % in 2000 (Dequeker et al. 2001). In the USA CAP/ACMG proficiency programme over the period 1995-2000, the error rate was estimated as 1.5%. For other disorders (Huntington's disease, Duchenne muscular dystrophy, Fanconi's anaemia, Charcot-Marie Tooth disease, BRCA and azoospermia factor) included in the European Molecular Genetics Quality Network, the genotyping error rate was 2-10% (Dequeker et al. 2001).

For microsatellites, global mean error rates in the range 0.55-5.9% per genotype have been reported (Mitchell et al. 2003). Because of high variance between markers, error rates for specific markers may be substantially higher. In high throughput centres, an

error rate of 0.5% per genotype for blind duplicates run on same gel was observed (Mitchell et al. 2003). However, the markers included in high-throughput panels have been specifically selected for inclusion, because they produce highly repeatable genotyping rates and their individual PCR reactions have been optimized. Issues therefore to consider are the type of marker and whether the genotyping was outsourced.

6.2.2 Assessment of exposures

Not surprisingly, exposure assessment is important in studies of gene-environment interaction. Points that need to be considered are the method of exposure assessment, and its validity and reproducibility. Exposure misclassification can bias the estimation of an interaction effect, the magnitude of which depends on the prevalence of the misclassified exposure and on the interaction model (Garcia-Closas et al. 1999, Garcia-Closas et al. 2004). If interaction is defined as lack of fit to a multiplicative model, a test for interaction will be conservative (Clayton & McKeigue 2001). In theory, case-control studies are more vulnerable to differential misclassification than are cohort studies (and the related case-cohort and nested case-control designs). However, provided that the extent of misclassification of exposure does not vary by genotype, differential misclassification between cases and controls is not a serious problem for the detection of departures from a multiplicative gene-environment joint effect (Clayton & McKeigue 2001).

Both differential and non-differential biases can result in over- or underestimation of an interaction effect (Greenland 1980). When genotype and exposure are independent in the source population, and the errors in the assessment of each are independent, both differential and non-differential misclassification of a dichotomous factor tends to underestimate departure from a multiplicative gene-exposure joint effect (Clayton & McKeigue 2001). The impact of misclassification on departures from additive effects is difficult to predict (Garcia-Closas et al. 1999).

6.3 Confounding

6.3.1 Population stratification

Concern has been raised about the possible effects of population stratification on the results of population-based association studies. Population stratification is due to differences between groups of variable ethnic origins that translate into different allele frequencies. The problem can arise even in populations of seemingly similar ethnic descent, especially when there has been limited admixture within these populations, such as in isolated populations and/or those with strong founder effects and insufficient time for mixture between founder groups.

The most discussed examples include the association between type 2 diabetes mellitus and the *Gm3;5,13,14* haplotype among residents of the Gila River Indian Community that was used to present the potential problem (Knowler et al. 1988), the relation between the dopamine receptor locus *(DRD2) A1* allele and alcoholism (Gelernter et al. 1993), and that between *CYP3A4* and prostate cancer in African Americans (Kittles et al. 2002). These results have generated controversy as to whether population stratification represents a fundamental problem for association studies, or whether it is part of more general issues about rigorous application of epidemiologic study design principles (Cardon & Palmer 2003, Thomas & Witte 2002, Wacholder et al. 2002). In an exploration of possible population stratification in US studies of cancer among non-Hispanic Americans of European descent, the effect was considered unlikely to be substantial when epidemiologic principles of study design, conduct, and analysis were rigorously applied (Wacholder et al. 2000). Similar conclusions were reached using data from case-unrelated control studies of non-Hispanic US whites with hypertension or type 2 diabetes, and Polish subjects with type 2 diabetes (Ardlie et al. 2002).

Considerations may vary across ethnic groups (Edland et al. 2004). For example, variations in the frequency of certain genotypes in African Americans appear to be much wider than those observed in persons of European origin and therefore the possibility of stratification may be higher (Garte 1998). Evidence was weak for an effect of population stratification in data from a case-unrelated control study of hypertension in African Americans, but this was no longer apparent when the study was restricted to persons with U.S.-born parents and grandparents (Ardlie et al. 2002). Millikan (2001) reported that bias was minimal in gene-disease associations in studies of African-Americans in which differences in ethnic composition were not taken into account, and Wang et al. (2004) did not identify substantial bias in hypothetical case-control studies of a candidate gene for prostate cancer in admixed populations such as African Americans. Moreover, meta-analyses of 43 gene-disease associations comprising 697 individual studies show consistent associations across groups of different ethnic origin (Ioannidis et al. 2004), and so provides evidence against a large effect of population stratification, hidden or otherwise.

As studies of association and interaction typically address moderate or small effects and hence require large sample sizes, small biases, such as due to population stratification, may be important (Marchini et al. 2004). Freedman et al. (2004) found no empirical evidence of stratification in data on 24-48 SNPs from eleven case-control and cohort studies in the USA, Poland and Portugal. However, when the number of SNPs was quadrupled, and the sample size increased by a factor of 5-6, statistically significant evidence of stratification was found in one of the studies in which a case-cohort design had been used. The effect of the degree of stratification in this study, which was in African Americans, would be to inflate the chi-square statistic for association by a factor of two in a study with 1000 cases and 1000 controls, and by a factor of 2-5 in a study with 2000 cases and 2000 controls. Khlat et al. (2004) found that bias from hidden population stratification and associated type I error were small whenever a high risk subpopulation is small (<10%) in proportion to the total population. In their hypothetical

scenarios, type I error rates were greater for rare alleles than for common alleles. Thus, there is controversy about the potential importance of population stratification for population-based studies of gene-disease association and for studies of gene-environment interaction.

Methods used to address the potential problem of population stratification in primary studies include (1) use of case-parental control and other family-based designs and (2) 'genomic control' using unlinked genetic markers that are assumed to have no effect on the disease being studied (Devlin & Roeder 1999, Hoggart et al. 2003, Thomas 2004). In the latter approach, the genomic control markers serve as control loci for non-independence due to population substructure for association tests of the genes under study. Adjustment for such markers would be expected to control for ethnic variation in disease risk attributable to genetic factors. However, residual confounding from other sources of ethnic variation in disease risk would be a potential issue (Lee 2004, Wacholder et al. 2002) and ethnic origin may also be a marker of bias from differential selective reporting and selective publication in different countries (Pan et al. 2005). Thus, it is useful to present routinely ethnic group-specific estimates. It is also useful to describe any strategies used to address possible population stratification in the design or analysis of included studies.

6.3.2 Confounding from other sources

In studies of gene-environment interaction, confounding of exposures is a potential problem. The principles regarding the control of confounding are the same as those for studying the relation between exposure and disease. In practice, the use of biomarkers of exposure may need care in interpretation, because the genotype may influence the presence or level of the biomarker. Rothman (1986) noted that an extraneous risk factor is a confounder only if its effect becomes mixed with the effect under study. For example, an exposure may cause altered physiology, which in turn causes disease. A biomarker of the altered physiology is a risk factor for the disease and is unrelated to exposure because it results from exposure. It is not confounding, because the effect of the exposure is mediated through the effect of the altered physiology, and therefore no effects are mixed. However, decisions about whether a biomarker represents an intermediate factor in aetiology or is a potential confounder are difficult when uncertainties exist about the mechanism of effect of the exposure. This would also apply to genes.

In case-cohort studies, controls are a random sample of the cohort, and the effect of age, which is the key time variable, is controlled for in the analysis only. In more traditional nested case-control designs, controls are selected to match the cases on a temporal factor, such as age, and the main comparisons are within the time-matched sets (Wacholder 1991). In appraising case-cohort studies, the method of age adjustment and, in appraising nested case-control studies, details of the matching on age or other temporal factors are important to consider.

6.4 Multiple testing and pre-study odds of true finding

It would be useful to interpret the results in the context of how many polymorphisms have been studied. However, adjustments for multiple corrections are not routinely adopted or accepted. Moreover, the problem is that it is often difficult or impossible to get a handle on the possible risk of selective reporting. Investigators may test a very large number of polymorphisms, but may mention only the most promising ones. Massive-testing as in whole-genome association studies (Thomas et al. 2005) may become increasingly common. It has been suggested that it would be useful to provide an estimate of the pre-study odds for an association, since this would be key for interpreting the credibility of a 'positive' finding (Ioannidis 2005, Wacholder et al. 2004, Whittemore 2005). It makes a tremendous difference if an association has been found as part of massive unselected screening of 100,000 polymorphisms (in which case, the pre-study odds is less than 1:1,000) or with a highly-targeted hypothesis supported by other data and specific line of reasoning (in which case, the pre-study odds will be much higher).

6.5 Approaches to summarising the validity of studies

There are a number of publications concerning the rating of the quality of analytic observational studies (Little et al. 2002). A number of authors have proposed quantitative quality scoring systems for critical appraisal (Dixon et al. 1997). Other schemes have been developed for the purposes of meta-analyses in which an attempt has been made to assess the importance of study quality in accounting for heterogeneity of results between studies (Berlin & Colditz 1990, Longnecker et al. 1988, Longnecker et al. 1990). This type of assessment has also been considered for meta- analysis of individual patient data (Friedenreich 1993, Friedenreich et al. 1994). Certain features of the assessment schemes are specific to the disease and/or the exposure under consideration. Some schemes assign equal weight to each aspect of the study, so that the summation of points might result in worse quality scores for a study with several minor flaws than for a study with one major flaw. Jüni et al. (1999) observed that the use of scores to identify clinical trials of high quality is problematic, and they recommended that relevant methodological aspects should be assessed individually and their influence on the magnitude of the effect of the intervention explored. We recommend a similar approach in consideration of studies of gene-disease associations. As in clinical trials, it is more appropriate to consider multidimensional domains than a single grade in the integration of evidence from observational studies. Reviewing investigators should also be aware that in many genetic epidemiology studies, there is no mention at all of the information that would be considered essential for understanding potential sources of bias. Quality of reporting and risk of bias ('study quality') are not necessarily the same and they may be very different. Lack of reporting should not be assumed to imply poor quality of the study. It is extremely important that efforts should be made to enhance the quality of reporting and scientific transparency of genetic epidemiologic studies.

6.6. Hardy-Weinberg equilibrium

Besides qualitative statements, there are limited tools that one can use to quantitatively

assess the potential for selection and information biases, and confounding. The one test that potentially could be applied is the evaluation of concordance with Hardy-Weinberg equilibrium (HWE). HWE has become widely accepted as an underlying model in population genetics after Hardy (Hardy 1908) proposed the concept that allele frequencies at a genetic locus are stable within one generation of random mating. Deviations from HWE may be due to a large variety of reasons, reflecting not only bias but other parameters (Salanti et al. 2005a). The assumptions underlying HWE, including random mating, lack of selection according to genotype, and absence of mutation or gene flow, are rarely met in human populations (Ayres & Balding 1998, Shoemaker et al. 1998). However, as a null hypothesis, HWE might be expected to be satisfied in populations of healthy controls. It might also be satisfied in the composite population of cases and controls in studies where both cases and controls are diseased (but differ in some other outcome, e.g. treatment response), but there is no evidence that the genetic variant is associated with the disease per se. For studies, where populations are split according to disease and no disease status and the disease phenotype is very common, a test can be used to detect whether deviations from HWE are due to some specific genetic model (Wittke-Thompson et al. 2005).

Deviations from HWE in healthy populations may be a sign of selection bias (Attia et al. 2003) or population stratification (Attia et al. 2003). It has also been suggested that deviation from HWE may be a sign of genotyping error (Hosking et al. 2004, Salanti et al. 2005a, Xu et al. 2002) but it has been noted that presence of HWE is not altered by the introduction of genotyping error (Gordon et al. 2002). Hypothesis testing for departure from HWE may be misleading, because without large numbers of participants, there is little chance of detecting even big departures from HWE, even though such departures may, for example, be indicative of major genotyping errors. On the other hand, in very large studies, quite small departures from HWE may be detected as “statistically significant” while being of no material importance.

We recommend that HuGE reviewers themselves evaluate departure from HWE in control subjects, and that, due to the low power of statistical tests, these evaluations focus on estimating the magnitude of departure rather than performing significance tests (Salanti et al. 2005a). When calculated by the original investigators, assessments of HWE are not always reliable (Salanti et al. 2005a, Xu et al. 2002).

7. STATISTICAL METHODS

7.1 Meta-analysis

We strongly encourage quantitative synthesis within HuGE reviews, with a specific focus on examining heterogeneity, rather than automatic production of a combined estimate of effect, which has been criticized as giving 'spurious precision' (Egger et al. 1998). In previous guidance, the use of meta-analysis or pooled analysis as a tool to synthesize evidence has been left to the discretion of the authors of HuGE reviews, in part because of concern about the lack of comparability of study methods and in part because of concern about the validity of meta-analysis of observational studies (Little et al. 2002, Little et al. 2003). However, some have argued that genetic association studies are more closely related to randomized trials than other types of epidemiological study because of 'Mendelian randomization', although there are important caveats (Davey Smith, Ebrahim 2003, Little, Khoury 2003). Meta-analytic methods have highlighted heterogeneity in gene-disease association studies, and investigation of possible causes of this has highlighted the problem of potential biases (Ioannidis 2003, Trikalinos et al. 2004). Given the ever increasing number of studies of gene-disease associations and related interactions (Khoury et al. 2004), there will be enhanced power to explore potential methodological reasons for heterogeneity through meta-regression methods and sensitivity analyses.

Meta-analysis may not be appropriate when there are few and clearly disparate studies or if data on methodology are too limited to allow investigation of heterogeneity. However, compilation of data and transparent presentation are always useful, whether or not a meta-analysis is undertaken.

In view of the controversy about the possible effects of population stratification, ethnic origin group-specific estimates should be presented whenever possible and should be analyzed separately in data syntheses. . Subsequent pooling of results from different groups may be warranted if the estimates of effect appear sufficiently similar.

Results should be extracted for each genotype whenever possible, and analyses based on comparing genotype groups or alleles without any further adjustment for confounding should generally be used. Adjusted estimates of association might be extracted for comparison with analyses based on unadjusted data. Methods for meta-analysis of clinical trials (differences in means; standardised differences in means) are applicable for two-level exposures (e.g. carriers versus non-carriers).

We recommend that, where possible, separate effect estimates for heterozygotes and for homozygotes be presented to allow readers to judge for themselves issues around the likely mode of inheritance and the choice of model. Primary studies would ideally be explicit about what genetic model has been used and why. Sometimes, there may be

strong biological or other rationale for preferring one specific model over others (e.g. when enzyme activity is known to be influenced by specific genotypes or alleles). When no such evidence exists, the options are: (i) a 'genetic model-free' approach (Minelli et al. 2005); (ii) two pair-wise comparisons; (iii) a specific choice of model (dominant, co-dominant or recessive); (iv) an allele-based analysis (corresponding to co-dominant modelling). Statistical tests based on the first two methods involve two degrees of freedom and make no assumptions about the genetic model. Tests based on the second two involve one degree of freedom (so typically have greater power) but make assumptions about the mode of inheritance. The model proposed in the first study in the literature should be tested in meta-analysis whenever this is clear and not likely to have been selectively reported based on its results, both including and excluding the initial study. Meta-analysts need to make a decision about whether to examine other models. This should not be based on the most favourable result. All comparisons undertaken should be reported.

Assessment of variation in associations across studies (heterogeneity) should be routine (Attia et al. 2003, Lohmueller et al. 2003, Munafo & Flint 2004, Salanti et al. 2005b, Thakkinstian et al. 2004). Statistical tests may be performed, but they are typically underpowered and have been criticised (Higgins & Thompson 2002). Tests might therefore be accompanied by an estimate of among-study variance or an I-squared statistic (Higgins & Thompson 2002), which measures the proportion of variation across studies that is due to genuine differences rather than random error. Sensitivity analyses should be considered to assess the impact of studies susceptible to bias, e.g. small studies, early studies, those with high risk of genotyping error.

It is common to see that the meta-analyses results change in a consistent direction over time, as more data accumulate, suggesting that early studies provide exaggerated estimates of effect, or that initial studies stimulate studies that may be substantially different in design or quality from the initial studies. A plot of studies ordered chronologically may reveal this phenomenon. Meta-regression, cumulative meta-analysis or recursive cumulative meta-analysis (Ioannidis & Lau 2001) might be considered as methods of investigating variations across studies further. Examination should routinely be made of whether the results of the first study(ies) differ from the results of subsequent research (Ioannidis 2003, Trikalinos et al. 2004). The 'Proteus phenomenon', where the first study gives the strongest effect ever observed, and this is immediately followed by a study showing the least strong (most opposite) effect) ever observed has been frequently encountered (Ioannidis & Trikalinos 2005).

As a sensitivity analysis, meta-analyses should also evaluate routinely the impact of deviations from HWE in the constituent studies. There is empirical evidence that some meta-analyses may lose their statistical significance when HWE-violating studies are excluded or when adjustments are made for deviations from HWE (Trikalinos et al. in press). Appropriate methods for adjustments in the presence of HWE deviations are available (Schaid & Jacobsen 1999, Trikalinos et al. in press).

7.2 Meta-analysis of individual participant data

Meta-analyses of individual participant data (MIPD) enable considerable flexibility in statistical analysis (see section 3.3). Stratification by original study is always important, otherwise an MIPD may be susceptible to Simpson's paradox and may provide less reliable results than a meta-analysis of group data that respects that data come from different studies. Both fixed and random effects models may still be applied for MIPD, allowing for study effects (Trikalinos & Ioannidis 2001). MIPD may also offer better opportunities to elucidate the causes of heterogeneity among the data sets being combined. It is hoped that the HuGENet™ 'network of networks' initiative (Ioannidis et al. 2005) will lead to further guidance on meta-analysis of individual participant data.

8. PRESENTING RESULTS

The magnitude of the association between the allelic variants and clinical outcomes in terms of relative and attributable risks in different populations should be summarized, and comments provided on the quality and methodology of studies.

8.1 Presenting individual studies

The following table provides the suggested information for reporting each gene-disease association study. We recommend that results of individual studies be presented in forest plots, where possible, irrespective of whether a meta-analysis is performed.

Geographic area in which the study was carried out and, if available, period of recruitment of study subjects.

Case type and number. The type should include the condition under investigation (so that differences between studies can be identified), the diagnostic procedures and case definition applied, the method of ascertainment (e.g., register, one hospital, several hospitals, family practitioners), age range, and gender distribution. If a nested case-control or case cohort design was used, this should be specified.

Controls (or cohort) type and number. Sufficient information should be given for the reader to understand whether controls were derived from the population from which cases arose. If matching was used, the matching variables should be specified.

For loci for which there are multiple alleles, the alleles investigated and the baseline (reference) allele for the relative risk estimate should be specified.

It may be helpful to include a column specifying the proportion of controls with the genotype of interest. This is particularly the case when the studies of genotype-disease associations have been carried out in a limited number of geographic areas, or if selection bias is thought to be an issue in the interpretation of the studies

Data

Summary data for each genotype should be provided. For case/control outcomes these typically comprise the number of cases and the number of controls for each of the genotypes. For quantitative traits such as biomarkers, a sample size, mean and standard deviation should typically be presented for each genotype.

Relative risk and 95 percent confidence interval for the stated comparison. Data with numbers per genotype should be presented as this enables verification and further analysis. When adjustment has been carried out, these should be specified. When no estimate is presented and insufficient data are presented for calculation, but the authors have commented on an association, this should be stated.

For some genotype-disease associations, subgroup analyses have been reported by disease or socio-demographic characteristics. So far, the statistical power of many of

these analyses has been low. It may be helpful to include a column indicating what sub-groups have been analyzed (e.g., subsite of tumor, tumor histology, age, ethnic group), but usually one should discuss the results in the text rather than trying to summarize all subgroup analyses in a table, unless subgroups form one of the main pre-specified analyses of the systematic review.

The results of any analysis of possible interaction (effect modification) should be summarized. When this concerns possible joint effects of genotype and exposure, brief details of the exposure assessment should also be included. Assertions of interaction should be based on P values or confidence intervals for interaction terms and not on comparisons of statistical significance within different subgroups.

Reference The in-text reference should be specified in the format already described.

8.2 Presenting meta-analyses

Meta-analyses should usually be presented as forest plots. Multiple subgroup analyses and sensitivity analyses may be conveniently presented as forest plots without including the individual studies.

It will be helpful to summarize the magnitude of association between the allelic variants and various diseases in terms of both relative and attributable risks in different populations. Expressing results from retrospective studies as attributable risks may require external information on disease risks. Comments on the overall quality and methodology of studies should be associated with principal findings from meta-analyses.

9. INTERPRETING RESULTS

For gene-disease associations and related interactions, interpretative issues include the overall quantity and quality of the evidence base (9.1), consideration of possible publication and selective reporting biases (9.2) and application of guidelines for causal inference (9.3). In addition, there is the question of the implications of the results of the review, including any potential public health impact (9.4).

9.1 Overall quantity and quality of evidence base

Results from small or few studies should be interpreted with great caution, as it is common to see dissipation of early claimed effects (Ioannidis et al. 2001a) and early studies have no predictive power for the subsequent picture of the evidence (Trikalinos et al. 2004). It is unclear what would be a sufficient amount of evidence that would lead to high credibility for a gene-disease association. One should note that for relatively rare alleles and genotypes, even large cumulative sample sizes may not represent a lot of evidence. For example, if the genotype of interest occurs in only 1% of the general population, then even in a meta-analysis of 10,000 subjects only about 100 would carry the allele of interest, i.e. evidence would be quite thin despite the deceptively large total sample size, depending on the magnitude of the effect.

Investigators conducting HuGE reviews should also comment on the overall quality and methodology of studies, and the possible effects this would have on the study findings. One should address whether methodological flaws or other potential flaws are strong enough that they may invalidate the magnitude or even the presence of a postulated observed effect.

9.2 Publication and selective reporting biases

Publication bias is the selective publication of studies on the basis of the magnitude and direction of their findings (Stroup & Thacker 2000). Research with statistically significant results is more likely to be submitted and published than work with null or non-significant results (Easterbrook et al. 1991), and this has led to a preponderance of potentially spurious results in the literature (Begg & Berlin 1989). Therefore, publication bias is a potentially serious problem for the integration of evidence on gene-disease associations (Hirschhorn et al. 2002, Ioannidis et al. 2001b), and the problem may be even more prominent for more complex analyses, such as those encountered in relation to gene-environment and gene-gene interactions. In addition to the larger number of potential comparisons implicit in the concept of multiple interacting variables, authors face the problem that large tables of gene-environment interaction estimates are very cumbersome and difficult to assemble in publishable format. This inevitably increases the potential for publication bias.

There is no way to be certain about the presence of publication bias. A funnel plot may

be drawn and its asymmetry may be formally assessed using regression or correlation tests (Begg & Mazumdar 1994, Egger et al. 1997). Funnel plots in the absence of formal testing can be very misleading (Terrin et al. 2005) and should be avoided. Even with formal testing, these tests simply evaluate whether small studies give different results from larger studies. This may be due to reporting bias, other biases or genuine heterogeneity, and it may be difficult to determine which is the case. Moreover, these tests are largely underpowered with the typical number of studies included in most meta-analysis. Their performance (sensitivity and specificity) for detection of publication bias is unknown and is likely to be very poor. One of the most common mistakes is to interpret these tests as tests for publication bias (Egger et al. 1997).

In theory, another potential method of identifying publication bias is to search research registries. However, managing research registries on studies of genotype prevalence and gene-disease associations is challenging, because data for each additional allele genotyped would need to be added to the database. It is even more difficult for studies of gene-environment and gene-gene interactions, because of the diversity of joint effects that can be investigated. We propose that registries of studies or investigators through a network of networks, each with a defined structure, might be a feasible method of partially controlling the problem of publication bias in this field (Ioannidis et al. 2005). Analyses based on pre-defined consortia of investigators/teams would bypass the problem of publication bias, even if these consortia do not include all teams working on a specific topic; they would at least be inclusive within their network.

A further issue is the selective reporting of the results of studies. This is likely to be an increasingly important problem with the application of high throughput genotyping methods in gene-disease association studies, and would be expected to increase the prevalence of spurious results. There is empirical evidence of selective reporting of randomised controlled trials. Some evidence is also available in pharmacogenetics (Contopoulos-Ioannidis DG, Alexiou G, Gouvias T, Ioannidis JPA. personal communication). Comparisons of primary outcomes defined in trial protocols and those defined in published articles has shown that the reporting of trial outcomes is frequently incomplete, biased, with statistically significant efficacy outcomes being more likely to be reported that non-significant efficacy outcomes, and inconsistent with protocols (Chan et al. 2004a, Chan et al. 2004b, Chan & Altman 2005). This problem could be addressed at two levels:
(i) explicit documentation of *a priori* study hypotheses in reporting of primary studies (Little et al. 2002, STROBE Group 2005, von Elm & Egger 2004);
(ii) through maintaining registries of data and sample collections through networks of investigators on specific diseases or pathways (Ioannidis et al. 2005).

9.3 Causal inference

Observed associations may be real or spurious. Real associations may be due to a direct causal variant, to a variant in linkage disequilibrium with a direct causal variant, or to confounding. Spurious associations may be due to chance, bias within studies or bias across studies (reporting bias).

9.3.1 Linkage disequilibrium

Linkage disequilibrium is the tendency for the alleles of two separate but already linked loci on the same chromosome to be found together more than would be expected by chance in the general population. In consequence, when an allele at a specific locus appears to be associated with a disease, an issue is whether the allele is causal, or whether the association exists only because the allele is associated with a truly causal allele at another locus. Linkage disequilibrium depends on population history and on the genetic make-up of the founders of that population (Ardlie et al. 2002, Hirschhorn et al. 2002, Salanti et al. 2005b). Linkage disequilibrium varies between populations (Ardlie et al. 2002) and therefore potentially could be a source of heterogeneity between studies of gene-disease associations and related interactions. The use of informative haplotypes considering many gene variants may bypass this problem, although population differences in haplotype structure may persist (Terwilliger et al. 2002).

9.3.2 Establishing causality

The most frequently quoted guidelines for inferring causation from observational studies of associations between exposures and disease were proposed in the 1960s (Hill 1965, Surgeon General (Advisory Committee) 1964). In a review of the application of these guidelines in cancer epidemiology, consistency, strength of association, dose-response and biologic plausibility were the most frequently used, in descending order (Weed & Gorelic 1996). In the literature on epidemiologic methods, the most often mentioned guidelines are, in descending order, strength of association, temporality, consistency, biologic plausibility, dose-response and specificity (Potischman & Weed 1999).

In the literature on gene-disease associations, consistency (replication) has received the greatest emphasis (Hirschhorn et al. 2002, Ioannidis et al. 2001b, Lohmueller et al. 2003). In regard to strength of association, the relative risks for common diseases associated with most of the genetic variants at single loci so far observed have been low. Apart from methodological factors, the strength of association also is not a biologically consistent feature but rather a characteristic that depends on the relative prevalence of other causes (Rothman 1986) In regard to temporality, germline variants obviously must be present before a disease occurs. However the timing of gene expression may not be known. For example, this may be important in the interpretation of associations between infant genotype and congenital anomalies, as some enzyme systems do not appear to be expressed in the developing fetus.

Data on gene-function and on the tissue(s) in which a gene is expressed is relevant to consideration of dose-response and specificity. Considerable importance is being accorded by many commentators to biologic data, including data on gene-function and mechanisms, in making causal inference about gene-disease associations (Cloninger 2002, Glazier et al. 2002, Harrison & Owen 2003, Nature Genetics 2001, Page et al. 2003, Rebbeck et al. 2004, Weiss & Terwilliger 2000). Weed (1997) noted that consideration of biologic plausibility bridges the gap between epidemiologic evidence

and diverse forms of biologic evidence, and that it was likely to become increasingly important in causal assessments as molecular epidemiology permits more precise measurements of intracellular causal effects (further comment in section on use of biologic measures as data in epidemiologic studies). On the other hand, there has been concern that some form of mechanistic evidence might be identified and used selectively to reinforce an assertion of causality. Cardon & Bell (2001) note that a biologic argument can be constructed for virtually any associated allele because of the 'relative paucity of current understanding of the mechanisms of action of complex trait loci'. While candidate gene studies are often based on some biological knowledge of the candidate gene, genome-wide linkage and association studies initially identify variants without regard to biological function. However, the absence of mechanistic evidence, or evidence of high quality, would not exclude concluding that an association was causal, if other guidelines for causation were satisfied. As knowledge of the genome is not complete, biological plausibility may not be apparent (Begg 2005, Page et al. 2003, Pharoah et al. 2005; Ioannidis JPA & Kavvoura FK, personal communication).

9.4 Potential public health impact and other implications of results

9.4.1 Laboratory tests

If data are available, the sensitivity, specificity, and positive and negative predictive values (including 95 percent confidence intervals) of different tests available for the gene(s), including biochemical, molecular, and other tests in different populations should be summarized, as well as the type of study subjects in which the analytic or clinical validity of the tests were investigated.

9.4.2 Population testing

Population-specific data on the magnitude and determinants of testing for allelic variants of the gene(s) and the impact of testing on public health (morbidity, mortality, disability), including policy statements, recommendations, and legislation (including mention of available interventions) should be summarized. This section should summarize the quality of evidence regarding population testing and any associated intervention that might affect the relation between the gene and the disease.

9.4.3 Other potential public health applications

Potential public health applications of human genome epidemiologic information on the variants of the gene(s), e.g. setting permissible exposure thresholds, including quality of information should be summarized.

9.4.4 Implications for our understanding of disease

9.4.5 Implications for research

Strengths and gaps in the evidence base regarding the public health implications of human gene variants should be identified. Recommendations should be made to stimulate research to fill the gaps.

Appendix A. HuGE PREVALENCE REVIEWS

A.1 General format

Cover sheet

Title, whether new or update, contact details

Abstract

Provide a one-page structured synopsis of the issues discussed in the items below with a brief statement on each of these items. Supply keywords, including the name(s) of the gene(s), the name(s) of the disease(s) or disorder(s), and the word ‘epidemiology’.

Background

Gene(s)

Identify the gene(s) being reviewed and provide a brief review of chromosome location, gene product, and function, if known.

Gene variants

List known allelic variants with effects on gene product if known.

Disease(s) or other outcomes

This section is optional for HuGE prevalence reviews; systematic review on this aspect is not required.

Identify the disease(s) or other outcome(s) with which the Gene(s) is/are believed to be associated. Briefly summarize the descriptive epidemiology and confirmed and suspected risk factors (including other genes).

Objectives

Succinct summary of objectives of the current review.

Methods

Selection criteria

State the gene(s), gene variant(s), types of populations and types of study eligible for inclusion in the review.

Identifying studies

Describe the methods used to identify relevant papers and/or other sources of information, including all electronic databases searched and the periods for which they were searched. Brief details of any hand searches should be given.

Data collection and analysis

Describe the methods for: selection of studies; data extraction; any tools used for

assessing risk of bias; any methods to combine findings across studies (meta-analysis).

Results

Included studies

Include a table providing basic details of the included studies (location, date, types of participants. Ideally should include links to the proposed HuGENet™ 'registry of platforms'.

Prevalences

Summarize the prevalences in a table, providing numbers are percentages for each genotype where available, along with allele frequencies. Comment on the quality and methodology of studies.

Discussion

Main findings

Summarise the main findings of the review.

Limitations of the review

Discuss concerns over completeness of information, validity of individual studies, reporting bias.

Conclusions and recommendations for research

Strengths and gaps in the evidence base regarding the prevalence of human gene variants in different populations should be identified. Recommendations should be made to stimulate research to fill the gaps.

Potential conflicts of interest

Any potential conflict of interest that might influence the judgments of reviewers should be noted (see 2.2). If none, this should be stated explicitly.

References

Internet sites

Include relevant links to various genetics databases, online resources, educational materials, consensus statements, policy statements, and support groups.

A.2 Brief protocol for a prevalence review

Cover sheet

Title, type of review (prevalence review), contact details

Background

Indicate potential importance of the gene/gene variant to public health (or understanding of aetiology/pathogenesis). Briefly explain rationale for the type of review to be undertaken. State objectives of the review.

Methods

Selection criteria

State the gene(s), gene variant(s), types of populations and types of study to be included in the review.

Identifying studies

Describe the proposed methods for identifying relevant papers and/or other sources of information.

Data collection and analysis

Outline the proposed methods for: selection of studies; data extraction; any tools used for assessing risk of bias. Indicate whether pooled analysis or meta-analysis is planned.

Potential conflicts of interest

Any potential conflict of interest that might influence the judgments of authors should be noted (see 2.2). If none, this should be stated explicitly.

A.3 Presenting results

Summarize known information on the frequency of homozygosity and heterozygosity of these variants in different populations and ethnic groups. The summary of variation in genotype frequency should include a table and a commentary on this in the text. The date of preparation of the table should be indicated in a footnote on each page. It is recommended that the following information be included:

Geographic area in which the study was carried out. Provide a brief description of how the subjects whose genotypes were determined were sampled, e.g., subjects selected randomly from a population-based sampling frame, blood donors, hospitalized subjects (give reasons). When possible, the description should include the mean age (standard deviation) or age range and the distribution by gender. If the subjects were controls from a case-control study, the disease under investigation and any matching criteria should be

specified (e.g., subjects matched to lung cancer patients on age, sex, and smoking history). When relevant, the ethnic group should be specified.

If genotyping in a substantial proportion of studies was inferred on the basis of phenotypic tests, these studies should be grouped in a specific section of the table. If more than one type of test was used, a column should be used to indicate the type of test used in the study.

Number of subjects whose genotypes were determined. When there are multiple alleles, those tested for should be specified.

Allele and/or genotype frequency, with 95 percent confidence interval.
Presenting allele frequencies cuts all the amount of genotypic information that should be presented, especially for multi allelic systems.

Presenting genotype frequencies enables allele frequencies to be calculated, whereas genotype frequencies cannot be calculated if only allele frequencies are presented. Information on genotype may be useful when there is information on the functional effects of genotypes.

Reference The in-text reference should be specified in the following format: single author—last name, year of publication (reference number); two authors—last name 1 and last name 2, year of publication (reference number); more than two authors—last name et al., year of publication (reference number).

It is suggested that order of the articles reviewed be first by continent or major geographic area (e.g., Americas, Africa, Asia, Europe, Oceania) and then alphabetically by location within country (multicentre studies and studies in which the location within the country is not specified first). This order makes it relatively easy to see sequential publications relating to the same population.

A.4 Interpretation

A.4.1 Publication and selective reporting biases

With regard to the population prevalence of genetic variants, a potential source of data is control series from case-control studies. However, if there were publication or reporting bias in favor of positive associations between specific genetic variants and disease, it is possible that genotype frequencies in the general population would be under-estimated.

Appendix B. FULL HuGE REVIEWS

Cover sheet

Title, whether new or update, contact details

Abstract

Provide a structured synopsis of the issues discussed in the items below with a brief statement on each of these items. Supply keywords, including the name(s) of the gene(s), the name(s) of the disease(s) or disorder(s), and the word ‘epidemiology’.

Gene(s)

Identify the gene(s) being reviewed and provide a brief review of chromosome location, gene product, and function (with a description of the source of knowledge, e.g. the type of studies generated the data), if known.

Gene variants

List known allelic variants with effects on function (with a description of the source of knowledge, e.g. the type of studies generated the data), if known.

Gene variant frequency

Investigators conducting HuGE reviews should decide on the extent of detail that is appropriate for the evaluation of gene variant frequency data.

Summarize known information on the frequency of homozygosity and heterozygosity of these variants in different populations and ethnic groups. If a prevalence review exists, summarise its findings. If a prevalence review does not exist, briefly overview the available data with some key references.

When information on genotype frequencies in control subjects is combined, e.g. in order to calculate attributable fractions, control frequencies should be estimated within ethnic groups.

Disease(s) or other outcomes

Identify the disease(s) or other outcome(s) with which the Gene(s) is/are associated. Briefly summarize the descriptive epidemiology and confirmed and suspected risk factors (including other genes).

Associations

Summarize the magnitude of the association between the allelic variants and various diseases in terms of absolute, relative, and attributable risks in different populations. Comment on the quality and methodology of studies (further details, section 8.2).

Interactions

Discuss whether the allelic variants interact with any of the known risk factors for the disease, including other genes and environmental factors. Summarize the magnitude of such interactions.

Laboratory tests

If data available, summarize the sensitivity, specificity, and positive and negative predictive values (including 95 percent confidence intervals) of different tests available for the gene(s), including biochemical, molecular, and other tests in different populations.

Summarize the type of study subjects in which the analytic or clinical validity of the tests were investigated.

Population testing

Summarize population-specific data on the magnitude and determinants of testing for allelic variants of the gene(s) and the impact of testing on public health (morbidity, mortality, disability), including policy statements, recommendations, and legislation (including mention of available interventions). This section should summarize the quality of evidence regarding population testing and any associated intervention that might affect the relation between the gene and the disease.

Other potential public health applications

Summarize potential public health applications of human genome epidemiologic information on the variants of the gene(s), e.g. setting permissible exposure thresholds, including quality of information.

Conclusions and recommendations for research

Strengths and gaps in the evidence base regarding the public health implications of human gene variants should be identified. Recommendations should be made to stimulate research to fill the gaps.

Potential conflicts of interest

Any potential conflict of interest that might influence the judgments of reviewers should be noted (see 2.2). If none, this should be stated explicitly.

References (according to journal format)

Internet sites

Include relevant links to various genetics databases, online resources, educational materials, consensus statements, policy statements, and support groups.

Appendix: Methods

The methods used to identify, select, appraise and synthesize evidence should be summarized in an Appendix. This should refer to any HuGE association reviews and

prevalence reviews on which the full review is based.

References

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,* vol. 68, no. 6, pp. 1447-1456.

Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P.C., Lang, T. & CONSORT GROUP (Consolidated Standards of Reporting Trials). 2001, "The revised CONSORT statement for reporting randomized trials: explanation and elaboration", *Annals of Internal Medicine,* vol. 134, no. 8, pp. 663-694.

Ardlie, K.G., Kruglyak, L. & Seielstad, M. 2002, "Patterns of linkage disequilibrium in the human genome", *Nat Rev Genet,* vol. 3, no. 4, pp. 299-309.

Attia, J., Thakkinstian, A. & D'Este, C. 2003, "Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology", *Journal of clinical epidemiology,* vol. 56, no. 4, pp. 297-303.

Ayres, K.L., Balding, D.J. 1998, "Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient", *Heredity,* vol. 80 (Pt 6), no. Pt 6, pp. 769-777.

Begg, C.B. 2005, "Reflections on publication criteria for genetic association studies", *Cancer epidemiology, biomarkers & prevention,* vol. 14, no. 6, pp. 1364-1365.

Begg, C.B., Berlin, J.A. 1989, "Publication bias and dissemination of clinical research", *Journal of the National Cancer Institute,* vol. 81, no. 2, pp. 107-115.

Begg, C.B., Mazumdar, M. 1994, "Operating characteristics of a rank correlation test for publication bias", *Biometrics,* vol. 50, pp. 1088-1101.

Benson, K., Hartz, A.J. 2000, "A comparison of observational studies and randomized, controlled trials", *New England Journal of Medicine,* vol. 342, no. 25, pp. 1878-1886.

Berlin, J.A., Colditz, G.A. 1990, "A meta-analysis of physical activity in the prevention of coronary heart disease", *American Journal of Epidemiology,* vol. 132, pp. 612-628.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H.C. & Standards for Reporting of Diagnostic Accuracy Group. 2003, "Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative", *Annals of Internal Medicine,* vol. 138, no. 1, pp. 40-44.

Botto, L., Yang, Q. 2000, "5, 10-Methylenetetrahydrofolate Reductase Gene Variants and Congenital Anomalies: A HuGE Review", *American Journal of Epidemiology*, vol. 151, no. 9, pp. 862-877.

Breslow, N. & Day, N. 1980, *Statistical Methods in Cancer Research. Volume 1. The Analysis of Case-Control Studies*, IARC, Lyon.

Brockton, N., Little, J., Sharp, L. & Cotton, S.C. 2000, "N-acetyltransferase polymorphisms and colorectal cancer: a review", *American Journal of Epidemiology*, vol. 151, no. 9, pp. 846-861.

Brockton, N.T. 2002, *Genetic polymorphisms in folate and xenobiotic metabolism and susceptibility to colorectal cancer*, PhD thesis, University of Aberdeen.

Burdett, S., Stewart, L.A. & Tierney, J.F. 2003, "Publication bias and meta-analyses: a practical example", *International Journal of Technology Assessment in Health Care*, vol. 19, no. 1, pp. 129-134.

Cardon, L.R., Bell, J.I. 2001, "Association study designs for complex diseases", *Nat Rev Genet*, vol. 2, no. 2, pp. 91-99.

Cardon, L.R., Palmer, L.J. 2003, "Population stratification and spurious allelic association", *Lancet*, vol. 361, no. 9357, pp. 598-604.

Chan, A.W., Altman, D.G. 2005, "Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors", *BMJ*, vol. 330, no. 7494, pp. 753.

Chan, A.W., Hrobjartsson, A., Haahr, M.T., Gotzsche, P.C. & Altman, D.G. 2004a, "Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles", *JAMA*, vol. 291, no. 20, pp. 2457-2465.

Chan, A.W., Krleza-Jeric, K., Schmid, I. & Altman, D.G. 2004b, "Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research", *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, vol. 171, no. 7, pp. 735-740.

Chenevix-Trench, G., Young, J., Coggan, M. & Board, P. 1995, "Glutathione S-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset", *Carcinogenesis*, vol. 16, no. 7, pp. 1655-1657.

Clayton, D., McKeigue, P.M. 2001, "Epidemiological methods for studying genes and environmental factors in complex diseases", *Lancet*, vol. 358, no. 9290, pp. 1356-1360.

Cloninger, C.R. 2002, "The discovery of susceptibility genes for mental disorders", *Proceedings of the National Academy of Sciences of the USA*, vol. 99, pp. 13365-

13367.

Concato, J., Horwitz, R.I. 2004, "Beyond randomised versus observational studies.", *The Lancet,* vol. 363, pp. 1660-1661.

Concato, J., Shah, N. & Horwitz, R.I. 2000, "Randomized, controlled trials, observational studies, and the hierarchy of research designs", *New England Journal of Medicine,* vol. 342, no. 25, pp. 1887-1892.

Cooper, D.N., Nussbaum, R.L. & Krawczak, M. 2002, "Proposed guidelines for papers describing DNA polymorphism-disease associations", *Human Genetics,* vol. 110, no. 207, pp. 208.

Cotton, S.C., Sharp, L., Little, J. & Brockton, N. 2000, "Glutathione S-transferase polymorphisms and colorectal cancer", *American Journal of Epidemiology,* vol. 151, no. 1, pp. 7-32.

Davey Smith, G., Ebrahim, S. 2003, "'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?", *International Journal of Epidemiology,* vol. 32, pp. 1-22.

de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. & Altshuler, D. 2005, "Efficiency and power in genetic association studies", *Nature genetics,* vol. 37, no. 11, pp. 1217-1223.

Deeg, D.J., van Tilburg, T., Smith, J.H. & de Leeuw, E.D. 2002, "Attrition in the longitudinal aging study Amsterdam. The effect of differential inclusion in side studies", *Jounal of Clinical Epidemiology,* vol. 55, no. 4, pp. 319-328.

Dequeker, E., Ramsden, S., Grody, W.W., Stenzel, T.T. & Barton, D.E. 2001, "Quality control in molecular genetic testing", *Nature Reviews Genetics,* vol. 2, no. 9, pp. 717-723.

Des Jarlais, D.C., Lyles, C., Crepaz, N. & TREND Group. 2004, "Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement", *American Journal of Public Health,* vol. 94, no. 3, pp. 361-366.

Devlin, B., Roeder, K. 1999, "Genomic control for association studies", *Biometrics,* vol. 55, no. 4, pp. 997-1004.

Dixon, R.A., Munro, J.F. & Silcocks, P.B. 1997, *The Evidence Based Medicine Workbook. Critical appraisal for clinical problem solving,* Butterworth-Heinemann, Oxford.

Dorman, J.S., Bunker, C.H. 2000, "HLA-DQ and Type 1 Diabetes", *Epidemiologic Reviews,* vol. 22, no. 2, pp. 218-227.

dos Santos Silva, I. 1999, *Cancer epidemiology: Principles and methods*, IARC, Lyon.

Easterbrook, P.J., Berlin, J.A., Gopalan, R. & Matthews, D.R. 1991, "Publication bias in clinical research", *Lancet*, vol. 337, no. 8746, pp. 867-872.

Edland, S.D., Slager, S. & Farrer, M. 2004, "Genetic association studies in Alzheimer's disease research: challenges and opportunities", *Statistics in medicine*, vol. 23, no. 2, pp. 169-178.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. 1997, "Bias in meta-analysis detected by a simple, graphical test", *British medical journal*, vol. 315, pp. 629-634.

Egger, M., Schneider, M. & Davey Smith, G. 1998, "Meta-analysis spurious precision? Meta analysis of observational studies", *British medical journal*, vol. 316, pp. 140-144.

Ehm, M.G., Nelson, M.R. & Spurr, N.K. 2005, "Guidelines for conducting and reporting whole genome/large-scale association studies", *Human molecular genetics*, vol. 14, no. 17, pp. 2485-2488.

Elwood, M. 1998, *Critical appraisal of epidemiological studies and clinical trials*, Oxford University Press, Oxford.

Engel, L.S., Taioli, E., Pfeiffer, R., Garcia-Closas, M., Marcus, P.M., Lan, Q., Boffetta, P., Vineis, P., Autrup, H., Bell, D.A., Branch, R.A., Brockmöller, B., Daly, A.K., Heckbert, S.R., Kalina, I., Kang, D., Katoh, T., Lafuente, A., Lin, H.J., Romkes, M., Taylor, J.A. & Rothman, N. 2002, "Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review", *American Journal of Epidemiology*, vol. 156, no. 2, pp. 95-109.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N., Pato, M.T., Petryshen, T.L., Kolonel, L.N., Lander, E.S., Sklar, P., Henderson, B., Hirschhorn, J.N. & Altshuler, D. 2004, "Assessing the impact of population stratification on genetic association studies", *Nature Genetics*, vol. 36, no. 4, pp. 388-393.

Freimer, N.B., Sabatti, C. 2005, "Guidelines for association studies in Human Molecular Genetics", *Human molecular genetics*, vol. 14, no. 17, pp. 2481-2483.

Friedenreich, C.M. 1993, "Methods for pooled analyses of epidemiologic studies", *Epidemiology*, vol. 4, pp. 295-302.

Friedenreich, C.M., Brant, R.F. & Riboli, E. 1994, "Influence of methodologic factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber", *Epidemiology*, vol. 5, no. 1, pp. 66-79.

Gambaro, G., Anglani, F. & D'Angelo, A. 2000, "Association studies of genetic polymorphisms and complex disease", *Lancet*, vol. 355, no. 9200, pp. 308-311.

Garcia-Closas, M., Rothman, N. & Lubin, J. 1999, "Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size", *Cancer Epidemiology Biomarkers and Prevention*, vol. 8, no. 12, pp. 1043-1050.

Garcia-Closas, M., Wacholder, S., Caporaso, N. & Rothman, N. 2004, "Inference issues in cohort and case-control studies of genetic effects and gene-environment interactions" in *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease*, eds. M.J. Khoury, J. Little & W. Burke, Oxford University Press, New York.

Garte, S. 1998, "The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1", *Carcinogenesis*, vol. 19, no. 8, pp. 1329-1332.

Gelernter, J., Goldman, D. & Risch, N. 1993, "The A1 allele at the D2 dopamine receptor gene and alcoholism: a reappraisal", *Journal of the American Medical Association*, vol. 269, no. 13, pp. 1673-1677.

Glazier, A.M., Nadeau, J.H. & Aitman, T.J. 2002, "Finding genes that underlie complex traits", *Science*, vol. 298, pp. 2345-2349.

Gordon, D., Finch, S.J., Nothnagel, M. & 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*, vol. 54, no. 1, pp. 22-33.

Greenland, S. 1980, "The effect of misclassification in the presence of covariates", *American Journal of Epidemiology*, vol. 112, no. 4, pp. 564-569.

Gregoire, G., Derderian, F. & Le Lorier, J. 1995, "Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias?", *Journal of clinical epidemiology*, vol. 48, no. 1, pp. 159-163.

Hardy, G.H. 1908, "Mendelian proportions in a mixed population", *Science*, vol. 28, pp. 49-50.

Harrison, P.J., Owen, M.J. 2003, "Genes for schizophrenia? Recent findings and their pathophysiological implications", *Lancet*, vol. 361, pp. 417-419.

Higgins, J.P., Thompson, S.G. 2002, "Quantifying heterogeneity in a meta-analysis", *Statistics in medicine*, vol. 21, no. 11, pp. 1539-1558.

Higgins, J.P.T. & Green, S. (eds) 2005, *Cochrane Handbook for Systematic Reviews of Interventions 4.2.4*, The Cochrane Library, Issue 2, 2005. John Wiley & Sons, Ltd., Chichester, UK.

Hill, A.B. 1965, "The environment and disease: association or causation?", *Proceedings of the Royal Society of Medicine*, vol. 58, pp. 295-300.

Hill, H.A. & Kleinbaum, D.G. 2000, "Bias in observational studies" in *Encyclopedia of Epidemiologic Methods*, eds. M.H. Gail & J. Benichou, John Wiley & Sons, Ltd, Chichester.

Hirschhorn, J.N., Lohmueller, K., Byrne, E. & Hirschhorn, K. 2002, "A comprehensive review of genetic association studies", *Genetics in Medicine*, vol. 4, no. 2, pp. 45-61.

Hoggart, C.J., Parra, E.J., Shriver, M.D., Bonilla, C., Kittles, R.A., Clayton, D.G. & McKeigue, P.M. 2003, "Control of confounding of genetic associations in stratified populations", *American Journal of Human Genetics*, vol. 72, no. 6, pp. 1492-1504.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I. & Xu, C.F. 2004, "Detection of genotyping errors by Hardy-Weinberg equilibrium testing", *European Journal of Human Genetics*, vol. 12, no. 5, pp. 395-399.

HuGE Net 2005, *, Revised Guidelines and Format for HuGE Reviews, February 2005*. Available: http://www.cdc.gov/genomics/hugenet/reviews/guidelines2.htm [2005, september/20].

Huizinga, T.W., Pisetsky, D.S. & Kimberly, R.P. 2004, "Associations, populations, and the truth: recommendations for genetic association studies in Arthritis & Rheumatism", *Arthritis & Rheumatism*, vol. 50, no. 7, pp. 2066-2071.

International HapMap Consortium 2003, "The international HapMap project", *Nature*, vol. 426, pp. 789-796.

Ioannidis, J.P. 2003, "Genetic associations: false or true?", *Trends in molecular medicine*, vol. 9, no. 4, pp. 135-138.

Ioannidis, J.P. 2005, "Why most published research findings are false", *PLoS Med.*, vol. 2, no. 8, pp. e124.

Ioannidis, J., Lau, J. 2001, "Evolution of treatment effects over time: empirical insight from recursive cumulative meta-analyses", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 3, pp. 831-836.

Ioannidis, J.P., Trikalinos, T.A. 2005, "Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials", *Journal of clinical epidemiology*, vol. 58, no. 6, pp. 543-549.

Ioannidis, J.P.A., Ntzani, E.E., Trikalinos, T.A. & Contopoulos-Ioannidis, D.G. 2001a, "Replication validity of genetic association studies", *Nature Genetics*, vol. 29, no. 3, pp. 306-309.

Ioannidis, J.P.A., Ntzani, E.E., Trikalinos, T.A. & Contopoulos-Ionnadis, D.G. 2001b, "Replication validity of genetic association studies", *Nature Genetics*, vol. 29, no. 3, pp. 306-309.

Ioannidis, J.P., Rosenberg, P.S., Goedert, J.J. & O'Brien, T.R. 2002a, "Commentary: meta-analysis of individual participants' data in genetic epidemiology", *American Journal of Epidemiology*, vol. 156, no. 3, pp. 204-210.

Ioannidis, J.P., Stavrou, I., Trikalinos, T.A., Zois, C., Brandi, M.L., Gennari, L., Albagha, O., Ralston, S.H., Tsatsoulis, A. & ER-alpha Genetics Meta-Analysis. 2002b, "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*, vol. 17, no. 11, pp. 2048-2060.

Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. 2004, "'Racial' differences in genetic effects for complex diseases", *Nature genetics*, vol. 36, no. 12, pp. 1312-1318.

Ioannidis, J.P., Ralston, S.H., Bennett, S.T., Brandi, M.L., Grinberg, D., Karassa, F.B., Langdahl, B., van Meurs, J.B., Mosekilde, L., Scollen, S., Albagha, O.M., Bustamante, M., Carey, A.H., Dunning, A.M., Enjuanes, A., van Leeuwen, J.P., Mavilia, C., Masi, L., McGuigan, F.E., Nogues, X., Pols, H.A., Reid, D.M., Schuit, S.C., Sherlock, R.E., Uitterlinden, A.G. & GENOMOS Study. 2004, "Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes", *JAMA*, vol. 292, no. 17, pp. 2105-2114.

Ioannidis, J.P., Bernstein, J., Boffetta, P., Danesh, J., Dolan, S., Hartge, P., Hunter, D., Inskip, P., Jarvelin, M.R., Little, J., Maraganore, D.M., Bishop, J.A., O'Brien, T.R., Petersen, G., Riboli, E., Seminara, D., Taioli, E., Uitterlinden, A.G., Vineis, P., Winn, D.M., Salanti, G., Higgins, J.P. & Khoury, M.J. 2005, "A network of investigator networks in human genome epidemiology", *American Journal of Epidemiology*, vol. 162, no. 4, pp. 302-304.

Johnson, G.C., Esposito, L., Barratt, B.J., Smith, A.N., Heward, J., Di Genova, G., Ueda, H., Cordell, H.J., Eaves, I.A., Dudbridge, F., Twells, R.C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S.C., Clayton, D.G. & Todd, J.A. 2001, "Haplotype tagging for the identification of common disease genes", *Nature genetics*, vol. 29, no. 2, pp. 233-237.

Juni, P., Witschi, A., Bloch, R. & Egger, M. 1999, "The hazards of scoring the quality of clinical trials for meta-analysis", *JAMA*, vol. 282, no. 11, pp. 1054-1060.

Kelsey, J.L., Whittemore, A.S., Evans, A.S. & Thompson, W.D. 1996, *Methods in observational epidemiology*, Oxford University Press, Oxford.

Khlat, M., Cazes, M.H., Genin, E. & Guiguet, M. 2004, "Robustness of case-control studies of genetic factors to population stratification: magnitude of bias and type I error", *Cancer epidemiology, biomarkers & prevention*, vol. 13, no. 10, pp. 1660-1664.

Khoury, M.J., Dorman, J.S. 1998, "The Human Genome Epidemiology Network", *American Journal of Epidemiology*, vol. 148, no. 1, pp. 1-3.

Khoury, M.J., Millikan, R., Little, J. & Gwinn, M. 2004, "The emergence of epidemiology in the genomics age.", *International Journal of Epidemiology,* vol. 33, pp. 936-944.

Kittles, R.A., Chen, W., Panguluri, R.K., Ahaghotu, C., Jackson, A., Adebamowo, C.A., Griffin, R., Williams, T., Ukoli, F., Adams-Campbell, L., Kwagyan, J., Isaacs, W., Freeman, V. & Dunston, G.M. 2002, "CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification?", *Human Genetics,* vol. 110, no. 6, pp. 553-560.

Knowler, W.C., Williams, R.C., Pettitt, D.J. & Steinberg, A.G. 1988, "Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture", *American Journal of Human Genetics,* vol. 43, no. 4, pp. 520-526.

Lancet 2003, "In search of genetic precision", *Lancet,* vol. 361, pp. 357-357: 12094.

Lawlor, D.A., Smith, G.D., Bruckdorfer, K.R., Kundu, D. & Ebrahim, S. 2004, "Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?", *The Lancet,* vol. 363, pp. 1724-1727.

Lawrence, R.W., Evans, D.M. & Cardon, L.R. 2005, "Prospects and pitfalls in whole genome association studies", *Philosophical transactions of the Royal Society of London. Series B: Biological sciences,* vol. 360, no. 1460, pp. 1589-1595.

Lee, W.C. 2004, "Case-control association studies with matching and genomic controlling", *Genetic Epidemiology,* vol. 27, pp. 1-13.

Lehmann, D.J., Williams, J., McBroom, J. & Smith, A.D. 2001, "Using meta-analysis to explain the diversity of results in genetic studies of late-onset Alzheimer's disease and to identify high-risk subgroups.", *Neuroscience,* vol. 108, pp. 541-554.

Little, J. 2004, "Reporting and review of human genome epidemiology studies" in *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease,* eds. M.J. Khoury, J. Little & W. Burke, Oxford University Press, New York.

Little, J., Bradley, L., Bray, M.S., Clyne, M., Dorman, J., Ellsworth, D.L., Hanson, J., Khoury, M., Lau, J., O'Brien, T.R., Rothman, N., Stroup, D., Taioli, E., Thomas, D., Vainio, H., Wacholder, S. & Weinberg, C. 2002, "Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations", *American Journal of Epidemiology,* vol. 156, no. 4, pp. 300-310.

Little, J., Khoury, M.J. 2003, "Mendelian randomisation: a new spin or real progress?", *Lancet,* vol. 362, pp. 930-931.

Little, J., Khoury, M.J., Bradley, L., Clyne, M., Gwinn, M., Lin, B., Lindegren, M.L. & Yoon, P. 2003, "The human genome project is complete. How do we develop a handle for the pump?", *American Journal of Epidemiology,* vol. 157, no. 8, pp. 667-

673.

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. 2003, "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease", *Nature genetics,* vol. 33, pp. 177-182.

Longnecker, M.P., Berlin, J.A., Orza, M.J. & Chalmers, T.C. 1988, "A meta-analysis of alcohol consumption in relation to risk of breast cancer", *Journal of the American Medical Association,* vol. 260, no. 5, pp. 652-656.

Longnecker, M.P., Orza, M.J., Adams, M.E., Vioque, J. & Chalmers, T.C. 1990, "A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer", *Cancer Causes and Control,* vol. 1, no. 1, pp. 59-68.

Madigan, M.P., Troisi, R., Potischman, N., Brogan, D., Gammon, M.D., Malone, K.E. & Brinton, L.A. 2000, "Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women", *International Journal of Epidemiology,* vol. 29, no. 5, pp. 793-798.

Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. 2004, "The effects of human population structure on large genetic association studies", *Nature Genetics,* vol. 36, no. 5, pp. 512-517.

Marchini, J., Donnelly, P. & Cardon, L.R. 2005, "Genome-wide strategies for detecting multiple loci that influence complex diseases", *Nature genetics,* vol. 37, no. 4, pp. 413-417.

Millikan, R.C. 2001, "Re: population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias", *Journal of the National Cancer Institute,* vol. 93, no. 2, pp. 156-157.

Minelli, C., Thompson, J.R., Abrams, K.R., Thakkinstian, A. & Attia, J. 2005, "The choice of a genetic model in the meta-analysis of molecular association studies", *International journal of epidemiology,* e-pub ahead of print .

Mitchell, A.A., Cutler, D.J. & Chakravarti, A. 2003, "Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test", *American Journal of Human Genetics,* vol. 72, no. 3, pp. 598-610.

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. Quality of Reporting of Meta-analyses", *Lancet,* vol. 354, no. 9193, pp. 1896-1900.

Moher, D., Schultz, K.F. & Altman, D. 2001, "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials", *Journal of the American Medical Association,* vol. 285, no. 15, pp. 1987-1991.

Morimoto, L.M., White, E. & Newcomb, P.A. 2003, "Selection bias in the assessment of gene-environment interaction in case-control studies", *American Journal of Epidemiology*, vol. 158, pp. 259-263.

Morton, L.M., Cahill, J. & Hartge, P. 2005, "Reporting Participation in Epidemiologic Studies: A Survey of Practice", *American Journal of Epidemiology*, e-pub ahead of print.

Munafo, M.R., Flint, J. 2004, "Meta-analysis of genetic association studies", *Trends in Genetics*, vol. 20, no. 9, pp. 439-444.

Nature Genetics 1999, "Freely associating (editorial)", *Nature Genetics*, vol. 22, pp. 1-2.

Nature Genetics 2001, "Challenges for the 21st century", *Nature genetics*, vol. 29, pp. 353-354.

Olson, S.H. 2001, "Reported participation in case-control studies: changes over time", *American Journal of Epidemiology*, vol. 154, no. 6, pp. 574-581.

Oxman, A.D. 1992, *The Cochrane Collaboration handbook: preparing and maintaining systematic reviews.*, 2nd edn, The Cochrane Collaboration, Oxford, UK.

Page, G.P., George, V., Go, R.C., Page, P.Z. & Allison, D.B. 2003, ""Are we there yet?": deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits", *American Journal of Human Genetics*, vol. 73, pp. 711-719.

Pan, Z., Trikalinos, T.A., Kavvoura, F.K., Lau, J. & Ioannidis, J.P.A. 2005, "Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature.", *PLoS Medicine*, vol. 2, pp. e334.

Patsopoulos, N.A., Analatos, A.A. & Ioannidis, J.P. 2005, "Relative citation impact of various study designs in the health sciences", *JAMA*, vol. 293, no. 19, pp. 2362-2366.

Pharoah, P.D., Dunning, A.M., Ponder, B.A. & Easton, D.F. 2005, "The reliable identification of disease-gene associations", *Cancer epidemiology, biomarkers & prevention*, vol. 14, no. 6, pp. 1362.

Pocock, S.J., Collier, T.J., Dandreo, K.J., de Stavola, B.L., Goldman, M.B., Kalish, L.A., Kasten, L.E. & McCormack, V.A. 2004, "Issues in the reporting of epidemiological studies: a survey of recent practice", *BMJ*, vol. 329, no. 7471, pp. 883.

Pocock, S.J., Elbourne, D.R. 2000, "Randomized trials or observational tribulations? ", *New England Journal of Medicine*, vol. 342, no. 25, pp. 1907-1909.

Potischman, N., Weed, D.L. 1999, "Causal criteria in nutritional epidemiology", *The American Journal of Clinical Nutrition*, vol. 69, no. 6, pp. 1309S-1314S.

Rebbeck, T.R., Martinez, M.E., Sellers, T.A., Shields, P.G., Wild, C.P. & Potter, J.D.

2004, "Genetic variation and cancer: improving the environment for publication of association studies", *Cancer Epidemiology, Biomarkers & Prevention,* vol. 13, no. 12, pp. 1985-1986.

Rothman, K.J. 1986, *Modern epidemiology*, Little, Brown and Company, Boston/Toronto.

Rothman, N., Stewart, W.F., Caporaso, N.E. & Hayes, R.B. 1993, "Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons", *Cancer Epidemiology Biomarkers and Prevention,* vol. 2, no. 4, pp. 299-303.

Salanti, G., Amountza, G., Ntzani, E.E. & Ioannidis, J.P. 2005a, "Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power", *European journal of human genetics,* vol. 13, pp. 840-848.

Salanti, G., Sanderson, S. & Higgins, J.P. 2005b, "Obstacles and opportunities in meta-analysis of genetic association studies", *Genetics in Medicine,* vol. 7, no. 1, pp. 13-20.

Schaid, D.J., Jacobsen, S.J. 1999, "Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions", *American Journal of Epidemiology,* vol. 149, no. 8, pp. 706-711.

Shoemaker, J., Painter, I. & Weir, B.S. 1998, "A Bayesian characterization of Hardy-Weinberg disequilibrium", *Genetics,* vol. 149, no. 4, pp. 2079-2088.

Steinberg, K. & Gallagher, M. 2004, "Assessing genotypes in human genome epidemiology studies" in *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease*, eds. M.J. Khoury, J. Little & W. Burke, Oxford University Press, New York.

Steinberg, K.K., Smith, S.J., Stroup, D.F., Olkin, I., Lee, N.C., Williamson, G.D. & Thacker, S.B. 1998, "Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies", *American Journal of Epidemiology,* vol. 145, no. 10, pp. 917-925.

STROBE Group 2005, 25 October 2005-last update, *STROBE statement, version 3*. Available: http://www.strobe-statement.org/.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B. for the Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. 2000, "Meta-analysis of observational studies in epidemiology: A proposal for reporting", *Journal of the American Medical Association,* vol. 283, no. 15, pp. 2008-2012.

Stroup, D.F. & Thacker, S.B. 2000, "Meta-analysis in epidemiology" in *Encyclopedia of*

Epidemiologic Methods, eds. M.H. Gail & J. Benichou, Wiley & Sons Publishers, New York.

Surgeon General (Advisory Committee) 1964, *Smoking and health*, US Department of Health, Education and Welfare, Washington, DC.

Tabor, H.K., Risch, N.J. & Myers, R.M. 2002, "Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations", *Nat Rev Genet*, vol. 3, pp. 391-397.

Terrin, N., Schmid, C.H. & Lau, J. 2005, "In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias", *Journal of clinical epidemiology*, vol. 58, no. 9, pp. 894-901.

Terwilliger, J.D., Haghighi, F., Hiekkalinna, T.S. & Goring, H.H. 2002, "A bias-ed assessment of the use of SNPs in human complex traits", *Current opinion in genetics & development*, vol. 12, no. 6, pp. 726-734.

Thakkinstian, A., D'Este, C., Eisman, J., Nguyen, T. & Attia, J. 2004, "Meta-analysis of molecular association studies: vitamin D receptor gene polymorphisms and BMD as a case study", *Journal of Bone & Mineral Research*, vol. 19, no. 3, pp. 419-428.

Thomas, D.C. 2004, *Statistical methods in genetic epidemiology*, Oxford University Press, New York.

Thomas, D.C., Haile, R.W. & Duggan, D. 2005, "Recent developments in genomewide association scans: a workshop summary and review", *American Journal of Human Genetics*, vol. 77, no. 3, pp. 337-345.

Thomas, D.C., Witte, J.S. 2002, "Point: population stratification: a problem for case control studies of candidate-gene associations?", *Cancer Epidemiology Biomarkers and Prevention*, vol. 11, no. 6, pp. 505-512.

Trikalinos, T.A., Ioannidis, J.P. 2001, "Predictive modeling and heterogeneity of baseline risk in meta-analysis of individual patient data", *Journal of clinical epidemiology*, vol. 54, no. 3, pp. 245-252.

Trikalinos, T.A., Ntzani, E.E., Contopoulos-Ioannidis, D.G. & Ioannidis, J.P. 2004, "Establishment of genetic associations for complex diseases is independent of early study findings", *European Journal of Human Genetics*, vol. 12, no. 9, pp. 762-769.

Trikalinos, T.A., Salanti, G., Khoury, M.J. & Ioannidis, J.P.A. in press, "Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations.", *American Journal of Epidemiology*.

Vandenbroucke, J.P. 2004, "When are observational studies as credible as randomised trials?", *The Lancet*, vol. 363, pp. 1728-1731.

von Elm, E., Egger, M. 2004, "The scandal of poor epidemiological research", *BMJ*, vol. 329, no. 7471, pp. 868-869.

Wacholder, S. 1991, "Practical considerations in choosing between the case-cohort and nested case-control designs", *Epidemiology*, vol. 2, no. 2, pp. 155-158.

Wacholder, S., Chanock, S., Garcia-Closas, M., El ghormli, L. & Rothman, N. 2004, "Assessing the probability that a positive report is false: an approach for molecular epidemiology studies", *Journal of the National Cancer Institute*, vol. 96, no. 6, pp. 434-442.

Wacholder, S., Chatterjee, N. & Hartge, P. 2002, "Joint effects of genes and environment distorted by selection biases: implications for hospital-based case-control studies", *Cancer Epidemiology Biomarkers and Prevention*, vol. 11, no. 9, pp. 885-889.

Wacholder, S., Rothman, N. & Caporaso, N. 2000, "Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias", *Journal of the National Cancer Institute*, vol. 92, no. 14, pp. 1151-1158.

Wang, S.S., Fernhoff, P.M., Hannon, W.H. & Khoury, M.J. 1999, "Medium chain acyl-CoA dehydrogenase deficiency human genome epidemiology review", *Genetics in Medicine*, vol. 1, no. 7, pp. 332-339.

Wang, Y., Localio, R. & Rebbeck, T.R. 2004, "Evaluating bias due to population stratification in case-control association studies of admixed populations", *Genetic Epidemiology*, vol. 27, pp. 14-20.

Weed, D.L. 1997, "On the use of causal criteria", *International Journal of Epidemiology*, vol. 26, no. 6, pp. 1137-1141.

Weed, D.L., Gorelic, L.S. 1996, "The practice of causal inference in cancer epidemiology", *Cancer Epidemiology Biomarkers and Prevention*, vol. 5, no. 4, pp. 303-311.

Weiss, K.M., Terwilliger, J.D. 2000, "How many diseases does it take to map a gene with SNPs?", *Nature Genetics*, vol. 26, pp. 151-157.

Weiss, S.T. 2001, "Association studies in asthma genetics", *American Journal of Respiratory and Critical Care Medicine*, vol. 164, pp. 2014-2015.

Whittemore, A.S. 2005, "Genetic association studies: time for a new paradigm?", *Cancer epidemiology, biomarkers & prevention*, vol. 14, no. 6, pp. 1359-1360.

Wittke-Thompson, J.K., Pluzhnikov, A. & Cox, N.J. 2005, "Rational inferences about departures from Hardy-Weinberg equilibrium", *American Journal of Human Genetics*, vol. 76, no. 6, pp. 967-986.

Wong, M.Y., Day, N.E., Luan, J.A. & Wareham, N.J. 2004, "Estimation of magnitude in

gene-environment interactions in the presence of measurement error", *Statistics in medicine,* vol. 23, no. 6, pp. 987-998.

Xu, J., Turner, A., Little, J., Bleecker, E.R. & Meyers, D.A. 2002, "Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error?", *Human genetics,* vol. 111, pp. 573-574.