Methods for meta-analysis in genetic association studies: A review of their potential and pitfalls

Article *in* Human Genetics · March 2008

DOI: 10.1007/s00439-007-0445-9 · Source: PubMed

2 authors:

Fotini Kavvoura
Royal Berkshire NHS Foundation Trust
44 PUBLICATIONS **3,591** CITATIONS

SEE PROFILE

John Ioannidis
Stanford University
1,014 PUBLICATIONS **143,394** CITATIONS

SEE PROFILE

Hum Genet (2008) 123:1–14
DOI 10.1007/s00439-007-0445-9

REVIEW

Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls

Fotini K. Kavvoura · John P. A. Ioannidis

Received: 27 May 2007 / Accepted: 29 October 2007 / Published online: 17 November 2007

Abstract Meta-analysis offers the opportunity to combine evidence from retrospectively accumulated or prospectively generated data. Meta-analyses may provide summary estimates and can help in detecting and addressing potential inconsistency between the combined datasets. Application of meta-analysis in genetic associations presents considerable potential and several pitfalls. In this review, we present basic principles of meta-analytic methods, adapted for human genome epidemiology. We describe issues that arise in the retrospective or the prospective collection of relevant data through various sources, common traps to consider in the appraisal of evidence and potential biases that may interfere. We describe the relative merits and caveats for common methods used to trace inconsistency across studies along with possible reasons for non-replication of proposed associations. Different statistical models may be employed to combine data and some common misconceptions may arise in the process. Several meta-analysis diagnostics are often applied or misapplied in the literature, and we comment on their use and limitations. An alternative to overcome limitations arising from retrospective combination of data from published studies is to create networks of research teams working in the same field and perform collaborative meta-analyses of individual participant data, ideally on a prospective basis. We discuss the advantages and the challenges inherent in such collaborative approaches. Meta-analysis can be a useful tool in dissecting the genetics of complex diseases and traits, provided its methods are properly applied and interpreted.

F. K. Kavvoura · J. P. A. Ioannidis (✉)
Clinical and Molecular Epidemiology Unit,
Department of Hygiene and Epidemiology,
University of Ioannina School of Medicine,
University Campus, Ioannina 45110, Greece
e-mail: jioannid@cc.uoi.gr

J. P. A. Ioannidis
Biomedical Research Institute,
Foundation for Research and Technology-Hellas,
Ioannina, Greece

J. P. A. Ioannidis
Institute for Clinical Research and Health Policy Studies,
Department of Medicine, Tufts-New England Medical Center,
Tufts University School of Medicine, Boston, MA, USA

Introduction

The development of high throughput techniques has resulted in an explosion of available genetic and genomic information. This creates challenges in analyzing, synthesizing and finally translating this rapidly accumulating evidence in useful clinical and public health applications (Burke et al. 2006; Guttmacher and Collins 2003; Higgins et al. 2007; Smith et al. 2006). Human genome epidemiology addresses associations between genetic variation and risk for complex common diseases. Currently, more than 6,000 original articles on human genome epidemiology findings are published annually, and the number is constantly increasing (Lin et al. 2006). Moreover, the average amount of data per article has increased steeply as we have moved from testing single polymorphisms to genome-wide association (GWA) studies (Ioannidis 2007b). We use the term meta-analysis here to define the quantitative methods for combining results of different studies on the same research question and for measuring and potentially explaining the extent of inconsistency among different studies.

In biomedical research, meta-analyses have been well established in the synthesis of clinical trials (Lau et al.

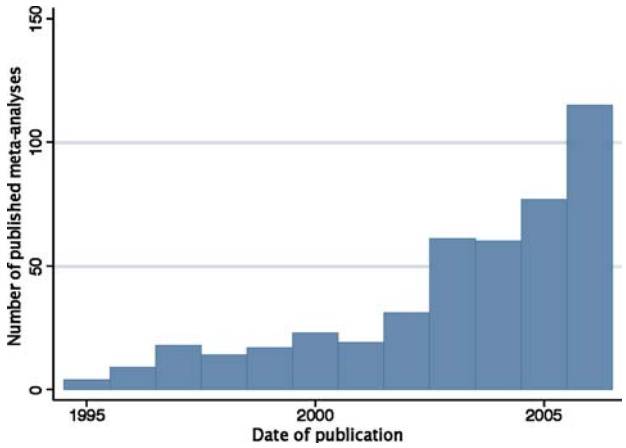

Fig. 1 Trend of published meta-analyses of genetic association studies overtime, starting from 1995 to 2006. The search was performed in Pubmed for 1995–2006

1998) and have increasingly been applied to observational studies (Stroup et al. 2000) with major impact in the literature (Patsopoulos et al. 2005). Meta-analyses of gene–disease association studies have also become popular (Ioannidis et al. 2001a; Lohmueller et al. 2003) (Fig. 1). Special initiatives such as the Human Genome Epidemiology Network (HuGENet) also support the conduct of high-quality meta-analyses on genetic associations (Khoury and Dorman 1998) (http://www.cdc.gov/genomics/hugenet/default.htm). Besides retrospective meta-analyses, there is mounting interest in prospective synthesis of evidence, currently pursued by several international consortia of investigators (Ioannidis et al. 2005, 2006). The advent of GWA investigations has also given a new twist to meta-analysis. Given that most genetic effects are small and require the coalition of many teams to generate large-scale evidence even at the discovery phase, meta-analyses of data from GWA studies and their replicating collaborating teams appear in the very first publication of genetic associations (Frayling et al. 2007; Scott et al. 2007; Zeggini et al. 2007).

As any research design, meta-analysis has strengths and weaknesses. This review discusses some critical areas where one can make a difference in maximizing the strengths and minimizing the weaknesses of meta-analyses on genetic association data. These issues should be ideally anticipated upfront when the explicit meta-analysis protocol is crystallized.

Sources of evidence

Ideally, meta-analyses should be prospective enterprises: all data that pass pre-set quality criteria are prospectively included, and prospective meta-analyses can be cumulatively updated, if and when new data appears. However, until now the vast majority of meta-analyses have been retrospective exercises. This creates a major challenge for retrieving the pertinent data in an unbiased way.

Published data

For the published literature, meta-analyses almost always use PubMed. EMBASE typically adds few or no relevant studies not indexed in PubMed, but this varies per topic (Royle et al. 2005). Insufficient attention in the search strategy or screening process may lead to missing useful data. Databases are also available that specialize on genetic association studies. The HuGE Published Literature database (http://www.cdc.gov/genomics/search/aboutHPLD.htm) (Lin et al. 2006) indexes genetic association studies published since October 1, 2000 (hosting almost 27,000 studies as of 15/4/2007). Disease-specific databases also exist for selected fields. For example, Alzgene (http://www.alzforum.org/res/com/gen/alzgene/) (Bertram et al. 2007) includes more than 1,000 published genetic association studies on Alzheimer's disease. Additional cross-linking may also be performed (e.g. screening of references or search in the Web of Science of the citations to the first studies that proposed an association).

Unpublished data

Many studies are not published in peer-reviewed journals, whereas others are not published at all, often due to publication bias (Dickersin et al. 1992; Hirschhorn et al. 2002; Munafo et al. 2004). Retrieving unpublished data retrospectively is difficult, if not impossible. Some may also question the validity of unpublished data that have not passed peer-review. Conversely, major deficiencies in the reporting of published genetic association studies (Bogardus et al. 1999; A.J. Yesupriya et al., unpublished data) suggest that the peer-review filter is imperfect. Empirical evidence from other fields suggests that inclusion of unpublished data may sometimes lead associations to lose their statistical significance, suggesting that the published literature is shaped by selective reporting biases (Kyzas et al. 2005). Moreover, there are numerous studies published in journals that are not indexed in any "mainstream" databases listed above.

Non-English literature

A particular issue is non-English language papers indexed in language-specific national databases. Local language

literature may sometimes account for a sizable quantum of evidence. An empirical assessment has highlighted this in the Chinese literature. Genetic effects of the Chinese language studies were consistently higher compared with those of, respective, PubMed-indexed studies from Europe or the USA (Pan et al. 2005). This problem may apply also to other national literatures; only the tip of the iceberg may be published in PubMed-indexed journals (Ariyaratnam et al. 2007). The ubiquitously large genetic effects in the Chinese literature probably reflect a stronger prevalence of publication and selective reporting biases in some research environments. Although in other clinical research domains, local language literature tends to accumulate non-significant results that are not attractive for the more competitive English language journals (Egger et al. 1997b), in genetics investigators may highlight significant results even more prominently in local journals. Inclusion of such data would yield spuriously strong effects in meta-analyses.

Implications

Given these sources of uncertainty, meta-analyses should report their exact sources of evidence, search strategies, and the results of the screening process with exclusions listed per type; flow diagrams may occasionally be helpful. Many genetic meta-analyses (80% in an empirical evaluation published in 2003), do not mention their search strategy (Attia et al. 2003). This makes the process impossible to reproduce by independent scientists. Depending on the selection criteria, databases screened, search algorithms, and publication periods covered, meta-analyses on the same association may end up including different studies and reaching even different conclusions. An example is shown in Table 1 for meta-analyses published in 2004–2007 on the association between three common polymorphisms and stroke.

Appraisal of evidence

The reliability of meta-analyses is affected by the reliability of the included data. The conduct of meta-analysis offers a unique opportunity to scrutinize the pieces of the data for errors and biases that may affect the validity of the results. A common misuse of meta-analysis is to ignore this step. At the other extreme, another common misuse is to stretch quality assessment beyond its true capabilities. For published data, reported quality is only a surrogate of the true quality of a study. If something is not reported, it cannot be always assumed that it was not done; if something is stated that it was done, it is not certain that it was done correctly. Sometimes composite quality scores are generated for the included studies that sum all the different aspects of the design, conduct, analysis and protection from bias or lack thereof, in each study. Such impressively accurate scores are misleading (Juni et al. 1999; Rothman and Greenland 1998). A single error may invalidate completely the results of a study, whereas a series of errors and biases may eventually cancel out among themselves.

Instead of spurious quality scores, meta-analyses should try to carefully record each potential error and bias. Biases can be broadly divided into study-specific and field-wide. The first category hosts selection bias, information bias, and confounding, whereas the second includes publication and selective outcome and analysis reporting biases (significance-chasing biases). Retrospective meta-analyses may suffer from both types of biases, while, in theory at least, prospective meta-analyses can avoid field-wide biases.

Selection bias refers to the differential selection of cases and controls, thus creating unequally representative groups within study populations (Hill and Kleinbaum 2000). Some empirical data support the notion that typically genotype differences do not influence response characteristics and participation rates (Bhatti et al. 2005), but specific forms of selection bias may still occur (Davey Smith and Ebrahim 2003; Little and Khoury 2003). If the genotype affects the severity of the disease, the strength of the association is affected when cases are selected based on severity. When the genotype affects also survival, prevalent cases yield different results compared with incident cases (Botto and Yang 2000). Similarly, assessment of gene–environment interactions should not be affected, unless genotype influences selection conditional on exposure and disease status (Morimoto et al. 2003; Wacholder et al. 2002a).

Information bias can affect genotypes, phenotypes, and/or other variables. Lack of quality control measures in the genotyping process, such as internal validation (by repeat testing and/or different genotyping techniques), quality control, and blinding (of laboratory personnel, clinical contributors or researchers) could cause bias (Little et al. 2002; Pompanon et al. 2005). Phenotype misclassification is possible when diagnostic criteria for the phenotype of interest have suboptimal sensitivity and specificity. Inclusion of control samples without rigorous exclusion of disease (e.g. blood donors) may cause spurious findings, when common but "silent" diseases (e.g. coronary artery disease) are studied (Aagaard-Tillery et al. 2006; Brockton et al. 2000; Cotton et al. 2000). Misclassification in the assessment of environmental exposures can be due to inaccurate measurements or lack of straightforward definitions (e.g. smoking habit can be assessed according to ever/never smokers or the number of cigarettes smoked per day) (Wong et al. 2004). Non-differential misclassification tends to shift effect sizes to the null (Garcia-Closas et al.

Table 1 Meta-analyses for risk of stroke and three common polymorphisms published between 2004 and 2007 showing differences in populations, number of studies, sample sizes and effect sizes

Author	Last search	Topic	Population	Polymorphism	Studies (*N*)	Total sample size	Effect size OR (95% CI)
Ariyaratnam et al. (2007)	1/2005	Stroke (any type)	As	ACE I/D	15	5,173	1.82 (1.28–2.60)
				MTHFR C677T	10	5,258	1.22 (0.98–1.52)
				APOE ε2/ε3/ε4	7	2,673	1.77 (1.30–2.39)
Banerjee et al. (2007)	8/2006	Stroke (any type)	As	ACE I/D	6	2,429	1.19 (0.91–1.56)
				MTHFR C677T	10	6,389	1.54 (1.13–2.10)
				APOE ε2/ε3/ε4	6	6,107	1.47 (1.00–2.15)
Sudlow et al. (2006)	10/2004	Stroke (per type)	As-Cauc	APOE ε2/ε3/ε4	24	20,113	1.11 (1.15–1.64)
Cronin et al. (2005)	3/2004	Stroke/TIA	As-Cauc	MTHFR C677T	32	14,780	1.37 (1.15–1.64)
Casas et al. (2005)	6/2003	Stroke (any type)	As-Cauc	MTHFR C677T	30	13,928	1.26 (1.14–1.40)
Casas et al. (2004)	1/2003	Stroke (any type)	Cauc	ACE I/D	11	14,295	1.21 (1.08–1.35)
				MTHFR C677T	22	7,984	1.24 (1.08–1.42)
				APOE ε2/ε3/ε4	10	12,726	0.96 (0.84–1.11)

ACE I/D Angiotensin converting enzyme insertion/deletion polymorphism, *APOE* Apolipoprotein E, *As* Asian, *Cauc* Caucasian, *CI* confidence interval, *MTHFR* 5,10-methylenetetrahydrofolate reductase, *OR* odds ratio, *TIA* Transient Ischemic Stroke

2004; Rothman et al. 1993). However, with massive testing of associations (e.g. GWA studies), differential misclassification may ensue for some pairs of polymorphisms and phenotypes and may generate select spurious genetic effects.

Confounding is a major threat. The classic example is *population stratification* (i.e. when the total population has been formed by admixture between subpopulations and when admixture proportions vary between individuals in cases and controls) (Hoggart et al. 2004). Family-based designs may overcome stratification. Moreover, there are several methods to deal with stratification (Devlin et al. 2001; Hoggart et al. 2004; Kohler and Bickeboller 2006; Price et al. 2006; Pritchard et al. 2000; Satten et al. 2001). Large biases are probably uncommon and, in theory, when many diverse studies are available, biases in opposite direction may tend to cancel out on average (Cardon and Palmer 2003; Wacholder et al. 2000, 2002b). However, small distortions cannot be excluded (Evangelou et al. 2006) without employing genomic control, principal components analyses, or other appropriate methods. With massive testing and large-scale evidence, small amounts of stratification may cause some spurious associations or dilute to the null some small effects.

Publication bias exists when there is a preference to publish studies with "positive" findings (i.e. statistical significant results or large association effects) (Calnan et al. 2006; Easterbrook et al. 1991; Hirschhorn et al. 2002). "Negative" (non-statistically significant) findings may never be published or may be published with delay (time-lag bias) (Ioannidis 1998). As discussed above, efforts to retrieve and include all the unpublished studies may or may not solve the problem—bias may occasionally be even aggravated by inclusion of some unpublished data or local literatures.

Selective outcome and analysis reporting bias (Chan and Altman 2005; Chan et al. 2004a, b; Contopoulos-Ioannidis et al. 2006) occurs when investigators publish only a subset of the analyses they have conducted, with preference to most impressive results. Almost 90% of observational studies report at least one statistically significant result in their abstract, and presented "positive" results typically outnumber non-statistically significant results (Kavvoura et al. 2007). Exploratory analyses are common in epidemiology (Michels and Rosner 1996) and may be even more prominent in modern discovery-oriented research (Ioannidis 2007a). The exploratory character may not be admitted in the published results. Overall, selective reporting is difficult to overcome, except for collaborative analyses with clearly a priori stated objectives and analyses (Ioannidis et al. 2005).

The term *significance-chasing bias* has recently been proposed (Ioannidis and Trikalinos 2007b) to collectively describe all biases stemming from the pursuit of nominal statistical significance. A meta-analysis that fails to take into account the possibility of significance-chasing biases may reach spuriously impressive summary estimates and put unjustified weight to their credibility.

Replication and inconsistency

An important issue in genetic associations is whether proposed effects are replicated or not by subsequent research (Ioannidis 2007b). Halfway between replication

and non-replication, one may have inconsistency, i.e. between-study heterogeneity in the observed effects. Proposed effects in the candidate gene era were often dissipated in following studies (Hirschhorn et al. 2002; Ioannidis et al. 2001a; Lohmueller et al. 2003; Trikalinos et al. 2004). Early studies may suggest stronger effects due to the "winner's curse" phenomenon. This may continue in the GWA era, where the discovery phase entails massive testing: the most significant polymorphisms are likely to exhibit some regression-to-the-mean upon further testing. In the "Proteus phenomenon", the first study gives the strongest effect ever observed, soon followed by a study showing the least strong (most opposite) effect ever observed (Ioannidis and Trikalinos 2005); subsequent studies have results that fall between these two extremes. The Proteus phenomenon probably reflects the strong attraction to publish impressive results and the equally strong attraction to refute them. Recently, Yu et al. (2007) proposed a method to correct for the effect of the "winner's curse phenomenon" in follow-up studies that try to replicate initially claimed associations.

Results of genetic association studies included in meta-analyses may vary due to additional reasons. The genetic variant may not be the true culprit one, but simply linked to the culprit with variable linkage across populations. For polymorphisms derived from GWA studies this might be the rule, since tag polymorphisms are usually selected for testing based on frequency, redundancy or coverage considerations—not candidate relevance (Hirschhorn and Daly 2005; Thomas et al. 2005). Similarly, the phenotype being studied may be correlated with the truly associated phenotype, and correlation of phenotypes may vary across studies. For example, an *FTO* polymorphism emerged in the search for type 2 diabetes genetic variation, but had inconsistent associations across populations for diabetes, while it had consistent associations with body mass index and obesity risk (Frayling et al. 2007). Common variants may regulate the risk of macroscopic clinical syndromes through molecular effects that are distant relatives of clinical risks. Therefore, diversity in the magnitude of clinical risks may be common in different populations, provided that large enough samples are available to document the diversity. Genuine diversity in effects across populations may also reflect population-specific gene–gene or gene–environment interactions (Hunter 2005) or complex pathways with exchangeable genetic variants.

Besides genuine diversity in the genetic effects, all the errors and biases discussed above may generate between-study heterogeneity. Therefore, heterogeneity should be carefully examined against all potential biases that may have affected each study or sets of studies. This scrutiny is still exploratory, but may affect interpretation of the evidence (Ioannidis 2006a; Ioannidis 2007b).

Table 2 lists commonly used statistical heterogeneity metrics. *Cochran's Q statistic* (Cochran 1954) cannot detect heterogeneity when there are few studies (a common problem) and may over-interpret unimportant heterogeneity when there are many studies (an uncommon situation) (Hardy and Thompson 1998; Higgins et al. 2002). A common misconception in GWA studies is to apply the Q statistic to compare the effect sizes in the GWA dataset results and in a few replication datasets, and to conclude that there is no heterogeneity, while the power to detect heterogeneity with such few studies is minimal.

Another common metric is the *between-study variance* τ^2 (Table 2). τ^2 depends on the respective effect size metric (e.g. standardized mean difference, odds ratio, hazard ratio) used; thus, it is not comparable among meta-analyses using different effect metrics (Huedo-Medina et al. 2006). The ratio of τ over the effect size conveys the extent of variability (between-study standard deviation) as compared with the effect size.

Finally, the I^2 metric is independent of the number of studies and can be compared across meta-analyses with different number of studies and metrics (Higgins and Thompson 2002) (Table 2). I^2 describes the percentage of total variation across studies due to heterogeneity rather than chance (Higgins et al. 2003). I^2 lies between 0 and 100%; values over 50% indicate large heterogeneity. However, I^2 also becomes uncertain when the meta-analysis includes few studies (Huedo-Medina et al. 2006). Confidence intervals for I^2 are easy to calculate (Higgins and Thompson 2002) and typically they are very large, unless many studies are available (Ioannidis et al. 2007b). Therefore, one should be aware that there can be large uncertainty in a meta-analysis about the presence or not and the extent of between-study heterogeneity. Strong inferences about heterogeneity or lack thereof are a common misconception when limited data are available.

Summary effect

Table 3 summarizes examples of commonly used effect size estimates in meta-analyses. With *fixed effects* models, it is assumed that there is a sole common effect estimate for all studies; observed between-study variability is attributed to chance only. Commonly used fixed effects' models include inverse-variance weighting, Mantel–Haenszel (1959) and Peto's methods (Yusuf et al. 1985). Peto's method may give inappropriate results when effect estimates are very large and when the numbers of cases and controls are not fairly equal (Greenland 1994). As we discussed in the previous section, with the range of data available in most meta-analyses, failure to reject the null hypothesis of homogeneity does not prove homogeneity. In

Table 2 Commonly used heterogeneity metrics

Heterogeneity metric	Calculation formula	Comments
Cochran's $\boldsymbol{Q}$ statistic	$Q = \sum w_i^F \left(d_i - d_+^F\right)^2$ d_+^F: summary effect size d_i : study-specific effect sizes w_i^F: weight of each study χ^2 distribution with k-1 degrees of freedom, typically considered significant at the $\alpha = 0.10$ level. k = number of studies	1. Influenced by the number of studies included in the meta-analysis thus 2. Underpowered when there are few studies (e.g. less than 20, the common situation), while it may overinterpret relatively unimportant heterogeneity when there are too many studies (e.g. over 40, an uncommon situation) 3. Non-statistically significant result is not proof of homogeneity
Between-study variance $\boldsymbol{\tau^2}$	$\tau^2 = \frac{Q-(k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$ Q: Cochran's Q statistic k: number of studies w_i: weight of each study	1. Reflects how much the true effect sizes estimated in the constituent studies of the meta-analysis differ 2. Depends on the respective effect size metric (e.g. standardized mean difference, odds ratio, hazard ratio) used 3. Not comparable among meta-analyses using different effect metrics
$\boldsymbol{I^2}$ statistic	$I^2 = \frac{Q-(k-1)}{Q} \times 100\%$ Q : Cochran's Q statistic k: number of studies	1. Values between 0 and 100% 2. >50% → large heterogeneity 3. >75% → very large heterogeneity 4. Confidence intervals are typically very large, unless many studies are available

Table 3 Examples of commonly used effect size estimates in meta-analysis, corresponding variances and weights

Type of effect size	Effect, d	Variance, var(d)	Weight, w
Dichotomous outcomes			
Log odds ratio (logOR), population-based study[a]	$\log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_2}{1-p_2}\right)$	$\frac{1}{p_1(1-p_1)n_1} + \frac{1}{p_2(1-p_2)n_2}$	$\frac{1}{\mathrm{var}(d)+\tau^2}$
Continuous outcomes			
Mean difference (Begg and Mazumdar)	$m_1 - m_2$	$\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}$	$\frac{1}{\mathrm{var}(d)+\tau^2}$
Standardized mean difference (Hedges's g)	$\frac{(m_1-m_2)\left(1-\frac{3}{4(n_1+n_2)-9}\right)}{\sqrt{\frac{(n_1-1)sd_1^2+(n_2-1)sd_2^2}{n_1+n_2-2}}}$	$\frac{n_1+n_2}{n_1n_2} + \frac{g^2}{2(n_1+n_2-3.94)}$	$\frac{1}{\mathrm{var}(d)+\tau^2}$

The indices 1 and 2 refer to the compared genotypes (for genotype-based comparisons) or the compared alleles (for allele-based comparisons)
p proportion with the genetic risk factor, n total number of people (or alleles), m mean of the quantitative trait, sd standard deviation of the quantitative trait, w weight of the study

[a] The variance estimator of the log odds ratio is one of several formulas that can be used. For details see Sutton et al. (2000)

genetic epidemiology, fixed-effects models often are also counterintuitive. It is implausible that effect modifications and biases are exactly the same across studies. In the presence of any between-study heterogeneity, fixed effects give tighter confidence intervals compared with random effects. This leads to spuriously lower levels of statistical significance for the summary effects; uncommon exceptions exist (Ntzani et al. 2007; Poole and Greenland 1999).

With *random effects*, it is assumed that there is a different underlying effect size for each study. The most popular estimator of the between-study variance is the DerSimonian–Laird estimator (1986). Random effects accommodate diversity between studies and thus are definitely preferable in the presence or anticipation of any between-study heterogeneity. This is the case in the majority of meta-analyses of genetic association studies, so that the use of random-effects models is generally preferable compared to fixed effects; when no heterogeneity exists, both models show similar effects anyhow. An important caveat exists when small studies in retrospective meta-analysis report very strong effects, whereas larger studies show no or minimal effects. Random effects give relatively more weight to smaller studies, and the summary effect may become stronger than that with fixed effects, while smaller studies may be less trustworthy. Extra caution is needed when effects are driven by mostly small studies which tend to show overestimated effect sizes (see also section on bias diagnostics). For prospective meta-

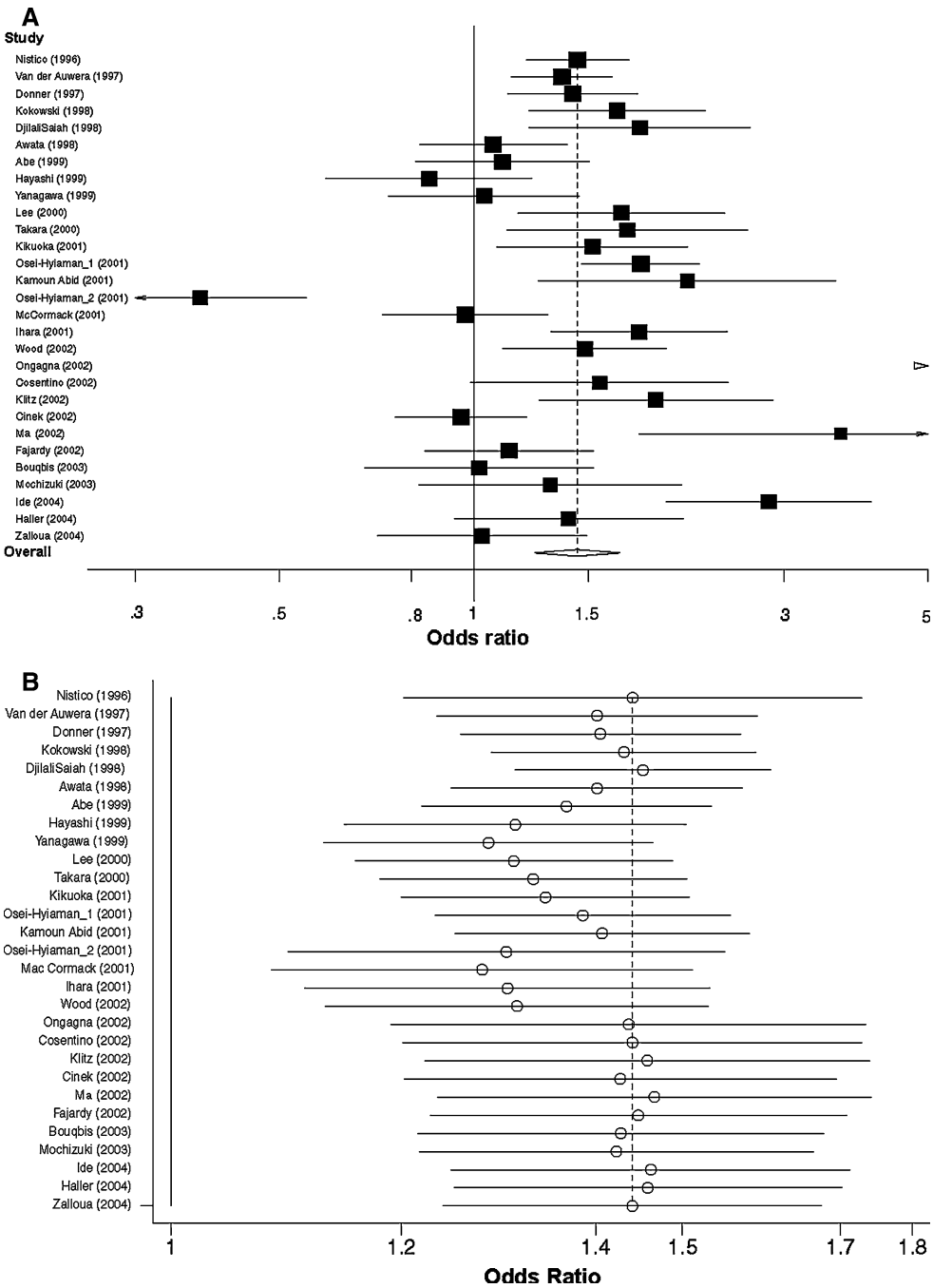

Fig. 2 **a** Forest plot showing the results of a meta-analysis for the association between the *G allele of the A49G polymorphism in the cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4) gene and susceptibility to type 1 diabetes mellitus. Individual studies (*circles*) are listed by year of publication from *top* to *bottom*. The *diamond* shows the summary random-effects odds ratio estimate from a meta-analysis. *Horizontal lines* 95% confidence interval. Data are adapted from Kavvoura and Ioannidis (2005). **b** Cumulative meta-analysis showing the change on the effect size as new studies accumulate, and their results are added to those previously published

analyses such bias is not an issue. Unfortunately, 70% of genetic meta-analyses have used fixed effects even in the presence of between-study heterogeneity (Attia et al. 2003). The practice of using fixed effects inappropriately continues even currently in meta-analyses of genome-wide association and their replication studies published even in the best journals (Ioannidis et al. 2007a, b).

The commonest way to present a meta-analysis graphically is a forest plot; each study is represented by its effect estimate along with its 95% confidence interval; the summary effect with its 95% confidence interval is also shown. In cumulative meta-analysis, studies are placed in order (e.g. chronological), and the summary effect and 95% confidence interval are plotted, as more studies are sequentially added to the calculations (Fig. 2). Visualization of cumulative meta-analysis may be particularly useful in genetic epidemiology, because many proposed genetic effects get dissipated over time (Ioannidis 2006b). Other graphical presentations are less popular (Sutton et al. 2000).

Cumulative meta-analysis is intuitively Bayesian (Lau et al. 1995): previous studies form the prior belief, and

estimates are updated with each new study to generate a posterior belief. More formal Bayesian methods (Spiegelhalter et al. 2003) are also applicable to meta-analysis (Brooks 1998; Spiegelhalter et al. 2004; WinBUGS 2003). A range of priors may be applied as a sensitivity analysis particularly for the uncertainty parameters (Lambert et al. 2005). Several prior elicitation methods have been described (Gartwaite et al. 2005). More complex models may also allow incorporation of the effect of deviations from Hardy–Weinberg equilibrium in the data synthesis (Salanti et al. 2007). Bayesian models may typically increase the uncertainty of the summary estimates, especially if more uncertainty is accommodated in the prior assumptions. Sometimes, however, borrowing strength from external prior evidence may lead to diminished uncertainty. Caution is warranted when inferences largely depend on the model used.

Another word of caution pertains to the interpretation of the summary effects. As in several other fields of molecular research, traditional levels of statistical significance ($P < 0.05$) may not suffice to reliably claim that an association is present (Ioannidis 2005; Sterne and Davey Smith 2001; Wacholder et al. 2004). Meta-analyses in the past have typically compiled evidence on a single genetic variant or a few ones. Currently it is possible to perform meta-analyses of GWA datasets with many thousands of polymorphisms (Evangelou et al. 2007) that are available in all datasets or that can be imputed through linked polymorphisms across datasets that have used different testing platforms. One might argue that a strict genome-wide significance level should be required ($P < 10^{-7}$) for polymorphisms that emerge from massive testing and undergo meta-analysis with subsequent studies (Skol et al. 2006). However, this may be an over-conservative requirement. False discovery rate and Bayesian approaches may be also considered (Benjamini and Hochberg 1995; Wacholder et al. 2004; Wakefield 2007). There is still no consensus on this issue, and we suggest that meta-analyses should report point estimates with 95% confidence (or credibility) intervals, avoiding strong statements about an association when such intervals reach close to the null.

Meta-analysis diagnostics

A number of diagnostics have been proposed in meta-analysis, that try to assess the robustness of the summary effects and explore between-study heterogeneity. *Sensitivity analyses* (excluding specific studies) are the simplest way to examine if the summary results and estimated heterogeneity depend on one or a few studies with perceived errors, biases, or special features. *Meta-regressions* relate the effect size to one or more characteristics of the studies (Thompson and Higgins 2002) that may explain the observed between-study heterogeneity. The covariates of interest should be pre-specified and few, to avoid largely inflated Type I errors. With few studies available in most occasions, meta-regressions are almost routinely abused in the literature. For group-level characteristics, meta-regressions also suffer from the potential of ecological fallacy (Chan et al. 2004b; Thompson and Higgins 2002).

Several meta-analysis diagnostics spuriously propose that they can test for publication bias (Rothstein et al. 2005). The typical premise for *funnel plot asymmetry* diagnostics is that small studies with non-significant or unfavorable results remain unpublished, whereas larger studies get published regardless. Unfortunately, this is an over-simplification. The funnel plot is a scatter plot of the meta-analyzed studies, with the treatment effect on the horizontal axis and a measure of precision (e.g. standard error, inverse variance, sample size) on the vertical axis. Despite the massive use of funnel plots in meta-analyses, visual inspection of funnel-plot asymmetry is entirely unreliable and should be abandoned (Lau et al. 2006; Tang and Liu 2000; Terrin et al. 2005). There are statistical test equivalents for asymmetry testing, including a rank-correlation test (Begg and Mazumdar 1994), a regression method (Egger et al. 1997a) and modified versions of the regression test with better calibration of type I and II errors (Harbord et al. 2006; Peters et al. 2006). Unfortunately, in most cases these tests are used either inappropriately or meaninglessly (Ioannidis and Trikalinos 2007a). These tests are appropriate and most meaningful when all the following are fulfilled: many studies (ideally 30 or more), not large between-study heterogeneity, presence of studies with formally significant results, and considerable variability in the range of study variances. These criteria are rarely fulfilled. Significant results in asymmetry tests do not mean that publication bias is certain, because there are many other reasons why small studies may yield different results from larger ones (Lau et al. 2006). The term "small-study effect" is more appropriate, when these diagnostics show significant asymmetry. Demonstration of small-study effects should lead to caution in the interpretation of the summary results. The most common, serious misunderstanding is to use these diagnostics and to conclude from their non-statistically significant results that publication bias is excluded. In the typical situation and for retrospectively collected data, it should be simply acknowledged that meta-analysts cannot really do anything to protect the meta-analysis from publication bias after the fact.

Similar caveats apply to other diagnostics such as the trim and fill (Duval and Tweedie 2000), but their detailed presentation is beyond the scope of this review (Rothstein et al. 2005). Some tests such as the failsafe N persist in the

literature, while it is clear that they should have been abandoned based on their poor mathematical and conceptual properties (Rothstein et al. 2005). A "significance chasing bias" test is also available (Ioannidis and Trikalinos 2007b) to test directly whether there is an excess of studies with nominally statistically significant results in a meta-analysis or many meta-analyses in the same field (Pan et al. 2005). Obviously, none of the "publication bias" diagnostics should be relevant for prospective meta-analyses.

Networks, consortia and prospective meta-analyses

Collaborative meta-analyses undertaken by consortia of investigators may have both retrospective and prospective features. HuGENet has created a Network of Investigators Networks (Ioannidis et al. 2005, 2006; Seminara et al. 2007). The represented networks comprise between 5 and 521 teams each, and accumulated sample sizes range from 3,000 to over half a million participants. Additional consortia are created continuously and play a key role in the validation of findings from GWA studies (Saxena et al. 2007; Zeggini et al. 2007). Networks form an open forum for communication and collaboration among separate research teams working in the same field. They can accumulate clinical, statistical, and laboratory expertise among the participating members, and this may improve both retrospective analyses and the planning of future studies.

Prerequisites

Elements deemed essential for the launch of a new consortium are a strong scientific rationale, the agreement of all teams to work together and contribute their data to address the research question, and the ability to support initial communication, coordination, identification, and recruitment of partners (Hunter et al. 2005). Networks may be open to inclusion of new partners. To avoid inclusion of flawed data, consortia may introduce inclusion criteria based on the appropriateness of study design, and phenotypic and genotypic accuracy (Seminara et al. 2007).

Data and quality control

Data are generated and/or gathered in coordinating centers where various data quality assurance practices and checks for logical errors and inconsistencies are used to guarantee adequate quality and transparency. Data standardization within the network aims to achieve agreement on common definitions to which all participating teams must conform (John et al. 2004). It is best implemented at the beginning of a "de novo" study, when definitions of data are formed. When standardization is not possible (different questions or criteria have already been used by established teams), some harmonization may still maximize the comparability of data from different teams and increase the credibility of derived findings compared with individual studies.

Standardization or harmonization of phenotypes and other non-genetic variables may be a major challenge in some fields; e.g. in 21 pharmacogenetic studies in asthma 483 different outcomes were analyzed (Contopoulos-Ioannidis et al. 2006). Nevertheless, a network has the best chance of achieving some consensus across teams (Seminara et al. 2007). Standardization of genotypes may be achieved through central genotyping of all samples (Andrulis et al. 2002). When this is not feasible, quality control of genotyping facilities of each team is usually performed to assure that systematic errors do not occur. In the absence of central genotyping control, one may use post hoc analyses, such as deviation from HWE in the controls, to identify possible genotyping (or other) errors, but these tests are generally underpowered (Salanti et al. 2007; Zou and Donner 2006). Unmeasured errors may still exist and are more likely in retrospective designs and less stringent quality control.

Data availability

Regarding the availability of data, networks can develop policies to make their resources and findings accessible to the larger scientific community. It is important that "positive" and "negative" results are both reported (Ioannidis 2006c). Consortia constitute one of the last lines of defense against publication and selective reporting bias; they should strive to include all high-quality data (Seminara et al. 2007).

Group-level versus individual-level data

Networks typically allow analyses to be performed in more detail compared to what is traditionally available in published data. Meta-analysis of individual level data (Ioannidis et al. 2002; Stewart and Tierney 2002) has definitive advantages in time-to-event phenotype analyses, multivariate modeling, examination of effect modification (subgroup analyses) and interaction effects, besides the possible advantages in standardization and harmonization of information. The drawback is the need for more extensive resources and the difficulties inherent in large-scale collaborative projects where many investigators need to agree on common plans. Timing can also be an issue, when the rate-limiting step is dictated by the slowest team.

Comparisons of meta-analyses of published group-level data versus meta-analyses of individual-level data in genetic epidemiology show examples where they largely agree with the results (Ioannidis et al. 1998, 2001b): partial agreement (Efstathiadou et al. 2001; Ralston et al. 2006), or disagreement (Cooper and Umbach 1996; Uitterlinden et al. 2006). Typically, in disagreements, individual-level data and particularly prospective designs may find no effect when a meta-analysis of published data had suggested strong effects, potentially based on biased evidence. This has been previously documented also in other fields of clinical research (Stewart and Parmar 1993).

GWA meta-analyses and multiple consortia

Currently for some research fields and diseases, there are already several different operating consortia. Each new GWA investigation along with its replicating teams (Chanock et al. 2007) may be seen as a consortium. Well-established, existing consortia may be used to replicate GWA-proposed genetic variants (Easton et al. 2007). Alternatively, a GWA investigation may give the opportunity to recruit a number of teams for examining replication, thus creating a new consortium. Some teams of investigators may participate in more than one such coalition, as the need arises to replicate newly proposed associations. As many GWA investigations may be performed on the same disease/phenotype, a major challenge is to combine all data from all these networks as well as additional independent teams that do not yet participate in any of these coalitions. Meta-analyses combining several GWA-related investigations have already been performed for some diseases, such as type 2 diabetes (Zeggini et al. 2007, Ioannidis et al. 2007b) and Parkinson's disease (Evangelou et al. 2007) and may become the norm in the near future. A major threat of bias in this situation is if results become available only for the most statistically significant of each GWA investigation. Public availability of all data is very essential in this regard (Manolio et al. 2007). Even though each GWA investigation and its replicating datasets form a prospective study, meta-analyses of all GWA investigations, collaborating replicating teams and other independent datasets still has the characteristics of a retrospective effort and may be affected by selective reporting biases, as described above.

Conclusion

Meta-analysis is a methodological tool that offers the opportunity to reach stronger conclusions by combining evidence from published studies or unpublished and prospectively generated data. Meta-analysis requires considerable expertise and careful adherence to a number of methodological steps. Otherwise, unreliable results may lead to misconceptions. We have highlighted some of the key pitfalls in the application and interpretation of meta-analysis methods in human genome epidemiology. The range of applications of meta-analysis is continuously expanding in this field, and the consolidation and expansion of networks of investigators create new opportunities and challenges.

Acknowledgment Dr. Kavvoura is supported by a PENED grant co-financed by the European Union-European Social Fund (75%) and the Greek Ministry of Development-General Secretariat of Research and Technology (25%).

References

Aagaard-Tillery K, Sibai B, Spong CY, Momirova V, Wendel G Jr, Wenstrom K, Samuels P, Cotroneo M, Moawad A, Sorokin Y, Miodovnik M, Meis P, O'Sullivan MJ, Conway D, Wapner RJ (2006) Sample bias among women with retained DNA samples for future genetic studies. Obstet Gynecol 108:1115–1120

Andrulis IL, Anton-Culver H, Beck J, Bove B, Boyd J, Buys S, Godwin AK, Hopper JL, Li F, Neuhausen SL, Ozcelik H, Peel D, Santella RM, Southey MC, van Orsouw NJ, Venter DJ, Vijg J, Whittemore AS (2002) Comparison of DNA- and RNA-based methods for detection of truncating BRCA1 mutations. Hum Mutat 20:65–73

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

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

Banerjee I, Gupta V, Ganesh S (2007) Association of gene polymorphisms with genetic susceptibility to stroke in Asian populations: a meta-analysis. J Hum Genet 52:205–219

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50:1088–1101

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23

Bhatti P, Sigurdson AJ, Wang SS, Chen J, Rothman N, Hartge P, Bergen AW, Landi MT (2005) Genetic variation and willingness to participate in epidemiologic research: data from three studies. Cancer Epidemiol Biomarkers Prev 14:2449–2453

Bogardus ST Jr, Concato J, Feinstein AR (1999) Clinical epidemiological quality in molecular genetic research: the need for methodological standards. JAMA 281:1919–1926

Botto LD, Yang Q (2000) 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am J Epidemiol 151:862–877

Brockton N, Little J, Sharp L, Cotton SC (2000) N-acetyltransferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 151:846–861

Brooks S (1998) Marcov chain Monte-Carlo method and its application. Statistician 47:69–100

Burke W, Khoury MJ, Stewart A, Zimmern RL (2006) The path from genome-based research to population health: development of an international public health genomics network. Genet Med 8:451–458

Calnan M, Smith GD, Sterne JA (2006) The publication process itself was the major cause of publication bias in genetic epidemiology. J Clin Epidemiol 59:1312–1318

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–604

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

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

Chan AW, Altman DG (2005) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ 330:753

Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG (2004a) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 291:2457–2465

Chan AW, Krleza-Jeric K, Schmid I, Altman DG (2004b) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 171:735–740

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS (2007) Replicating genotype–phenotype associations. Nature 447:655–660

Cochran WG (1954) The combination of estimates from different experiments. Biometrics 10:101–129

Contopoulos-Ioannidis DG, Alexiou GA, Gouvias TC, Ioannidis JP (2006) An empirical evaluation of multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. Pharmacogenet Genomics 16:705–711

Cooper GS, Umbach DM (1996) Are vitamin D receptor polymorphisms associated with bone mineral density? A meta-analysis. J Bone Miner Res 11:1841–1849

Cotton SC, Sharp L, Little J, Brockton N (2000) Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 151:7–32

Cronin S, Furie KL, Kelly PJ (2005) Dose-related association of MTHFR 677T allele with risk of ischemic stroke: evidence from a cumulative meta-analysis. Stroke 36:1581–1587

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

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7:177–188

Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155–166

Dickersin K, Min YI, Meinert CL (1992) Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 267:374–378

Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56:455–463

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991) Publication bias in clinical research. Lancet 337:867–872

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R; SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X; kConFab; AOCS Management Group, Mannermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447:1087–1093

Efstathiadou Z, Tsatsoulis A, Ioannidis JP (2001) Association of collagen Ialpha 1 Sp1 polymorphism with the risk of prevalent fractures: a meta-analysis. J Bone Miner Res 16:1586–1592

Egger M, Davey Smith G, Schneider M, Minder C (1997a) Bias in meta-analysis detected by a simple, graphical test. BMJ 315:629–634

Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G (1997b) Language bias in randomised controlled trials published in English and German. Lancet 350:326–329

Evangelou E, Trikalinos TA, Salanti G, Ioannidis JP (2006) Family-based versus unrelated case–control designs for genetic associations. PLoS Genet 2:e123

Evangelou E, Maraganore DM, Ioannidis JP (2007) Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. PLoS ONE 2:e196

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316:889–894

Garcia-Closas M, Wacholder S, Caporaso N et al (2004) Inference issues in cohort and case–control studies of genetic effects and gene–environment interactions. Oxford University Press, New York

Gartwaite PH, Kadane JB, O'Hagan A (2005) Statistical methods for eliciting probability distributions. J Am Stat Assoc 100:680–701

Greenland S (1994) Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol 140:290–296

Guttmacher AE, Collins FS (2003) Welcome to the genomic era. N Engl J Med 349:996–998

Harbord RM, Egger M, Sterne JA (2006) A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 25:3443–3457

Hardy RJ, Thompson SG (1998) Detecting and describing heterogeneity in meta-analysis. Stat Med 17:841–856

Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539–1558

Higgins J, Thompson S, Deeks J, Altman D (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 7:51–61

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560

Higgins JP, Little J, Ioannidis JP, Bray MS, Manolio TA, Smeeth L, Sterne JA, Anagnostelis B, Butterworth AS, Danesh J, Dezateux C, Gallacher JE, Gwinn M, Lewis SJ, Minelli C, Pharoah PD, Salanti G, Sanderson S, Smith LA, Taioli E, Thompson JR, Thompson SG, Walker N, Zimmern RL, Khoury MJ (2007) Turning the pump handle: evolving methods for integrating the evidence on gene–disease association. Am J Epidemiol 166:863–866

Hill HA, Kleinbaum DG (2000) Bias in observational studies. Wiley, Chichester

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM (2004) Design and analysis of admixture mapping studies. Am J Hum Genet 74:965–978

Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J (2006) Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 11:193–206

Hunter DJ (2005) Gene–environment interactions in human diseases. Nat Rev Genet 6:287–298

Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, Haynes RB, Henderson BE, Kaaks R, Stram DO, Thomas G, Thun MJ, Blanche H, Buring JE, Burtt NP, Calle EE, Cann H, Canzian F, Chen YC, Colditz GA, Cox DG, Dunning AM, Feigelson HS, Freedman ML, Gaziano JM, Giovannucci E, Hankinson SE, Hirschhorn JN, Hoover RN, Key T, Kolonel LN, Kraft P, Le Marchand L, Liu S, Ma J, Melnick S, Pharaoh P, Pike MC, Rodriguez C, Setiawan VW, Stampfer MJ, Trapido E, Travis R, Virtamo J, Wacholder S, Willett WC (2005) A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer 5:977–985

Ioannidis JP (1998) Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 279:281–286

Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2:e124

Ioannidis JP (2006a) Commentary: grading the credibility of molecular evidence for complex diseases. Int J Epidemiol 35:572–578 (discussion 593–596)

Ioannidis JP (2006b) Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. J Natl Cancer Inst 98:1350–1353

Ioannidis JP (2006c) Journals should publish all "null" results and should sparingly publish "positive" results. Cancer Epidemiol Biomarkers Prev 15:186

Ioannidis JP (2007a) Molecular evidence-based medicine: evolution and integration of information in the genomic era. Eur J Clin Invest 37:340–349

Ioannidis JP (2007b) Non-replication and inconsistency in the genome-wide association setting. Hum Hered 64:203–13

Ioannidis JP, Trikalinos TA (2005) Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. J Clin Epidemiol 58:543–549

Ioannidis JP, Trikalinos TA (2007a) The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ 176:1091–1096

Ioannidis JP, Trikalinos TA (2007b) An exploratory test for an excess of significant findings. Clin Trials 4:245–253

Ioannidis JP, O'Brien TR, Rosenberg PS, Contopoulos-Ioannidis DG, Goedert JJ (1998) Genetic effects on HIV disease progression. Nat Med 4:536

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001a) Replication validity of genetic association studies. Nat Genet 29:306–309

Ioannidis JP, Rosenberg PS, Goedert JJ, Ashton LJ, Benfield TL, Buchbinder SP, Coutinho RA, Eugen-Olsen J, Gallart T, Katzenstein TL, Kostrikis LG, Kuipers H, Louie LG, Mallal SA, Margolick JB, Martinez OP, Meyer L, Michael NL, Operskalski E, Pantaleo G, Rizzardi GP, Schuitemaker H, Sheppard HW, Stewart GJ, Theodorou ID, Ullum H, Vicenzi E, Vlahov D, Wilkinson D, Workman C, Zagury JF, O'Brien TR (2001b) Effects of CCR5-Delta32, CCR2–64I, and SDF-1 3'A alleles on HIV-1 disease progression: an international meta-analysis of individual-patient data. Ann Intern Med 135:782–795

Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR (2002) Commentary: meta-analysis of individual participants' data in genetic epidemiology. Am J Epidemiol 156:204–210

Ioannidis JP, Bernstein J, Boffetta P, Danesh J, Dolan S, Hartge P, Hunter D, Inskip P, Jarvelin MR, Little J, Maraganore DM, Bishop JA, O'Brien TR, Petersen G, Riboli E, Seminara D, Taioli E, Uitterlinden AG, Vineis P, Winn DM, Salanti G, Higgins JP, Khoury MJ (2005) A network of investigator networks in human genome epidemiology. Am J Epidemiol 162:302–304

Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ (2006) A road map for efficient and reliable human genome epidemiology. Nat Genet 38:3–5

Ioannidis JP, Patsopoulos NA, Evangelou E (2007a) Uncertainty in heterogeneity estimates in meta-analyses. BMJ 335:914-916

Ioannidis JP, Patsopoulos NA, Evangelou E (2007b) Heterogeneity in meta-analyses of genome-wide association investigations. PLoS ONE 2:e841

John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara D (2004) The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res 6:R375–R389

Juni P, Witschi A, Bloch R, Egger M (1999) The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 282:1054–1060

Kavvoura FK, Ioannidis JP (2005) CTLA-4 gene polymorphisms and susceptibility to type 1 diabetes mellitus: a HuGE Review and meta-analysis. Am J Epidemiol 162:3–16

Kavvoura FK, Liberopoulos G, Ioannidis JP (2007) Selection in reported epidemiological risks: an empirical assessment. PLoS Med 4:e79

Khoury MJ, Dorman JS (1998) The Human Genome Epidemiology Network. Am J Epidemiol 148:1–3

Kohler K, Bickeboller H (2006) Case–control association tests correcting for population stratification. Ann Hum Genet 70:98–115

Kyzas PA, Loizou KT, Ioannidis JP (2005) Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 97:1043–1055

Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR (2005) How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med 24:2401–2428

Lau J, Schmid CH, Chalmers TC (1995) Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. J Clin Epidemiol 48:45–57(discussion 59–60)

Lau J, Ioannidis JP, Schmid CH (1998) Summing up evidence: one answer is not always enough. Lancet 351:123–127

Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I (2006) The case of the misleading funnel plot. BMJ 333:597–600

Lin BK, Clyne M, Walsh M, Gomez O, Yu W, Gwinn M, Khoury MJ (2006) Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. Am J Epidemiol 164:1–4

Little J, Khoury MJ (2003) Mendelian randomisation: a new spin or real progress? Lancet 362:930–931

Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury M, Lau J, O'Brien TR, 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. Am J Epidemiol 156:300–310

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182

Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, Donnelly P, Faraone SV, Frazer K, Gabriel S, Gejman P, Guttmacher A, Harris EL, Insel T, Kelsoe JR, Lander E, McCowin N, Mailman MD, Nabel E, Ostell J, Pugh E, Sherry S, Sullivan PF, Thompson JF, Warram J, Wholley D, Milos PM, Collins FS (2007) New models of collaboration in genome-wide association studies: the genetic association information network. Nat Genet 39:1045–1051

Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719–748

Michels KB, Rosner BA (1996) Data trawling: to fish or not to fish. Lancet 348:1152–1153

Morimoto LM, White E, Newcomb PA (2003) Selection bias in the assessment of gene–environment interaction in case–control studies. Am J Epidemiol 158:259–263

Munafo MR, Clark TG, Flint J (2004) Assessing publication bias in genetic association studies: evidence from a recent meta-analysis. Psychiatry Res 129:39–44

Ntzani EE, Rizos EC, Ioannidis JP (2007) Genetic effects versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large-scale evidence. Am J Epidemiol 165:973–984

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

Patsopoulos NA, Analatos AA, Ioannidis JP (2005) Relative citation impact of various study designs in the health sciences. JAMA 293:2362–2366

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L (2006) Comparison of two methods to detect publication bias in meta-analysis. JAMA 295:676–680

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859

Poole C, Greenland S (1999) Random-effects meta-analyses are not always conservative. Am J Epidemiol 150:469–475

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Ralston SH, Uitterlinden AG, Brandi ML, Balcells S, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Scollen S, Bustamante M, Husted LB, Carey AH, Diez-Perez A, Dunning AM, Falchetti A, Karczmarewicz E, Kruk M, van Leeuwen JP, van Meurs JB, Mangion J, McGuigan FE, Mellibovsky L, del Monte F, Pols HA, Reeve J, Reid DM, Renner W, Rivadeneira F, van Schoor NM, Sherlock RE, Ioannidis JP (2006) Large-scale evidence for the effect of the COLIA1 Sp1 polymorphism on osteoporosis outcomes: the GENOMOS study. PLoS Med 3:e90

Rothman KJ, Greenland S (1998) Modern epidemiology, 2nd edn. Lippincott Williams & Wilkins, Philadelphia

Rothman N, Stewart WF, Caporaso NE, Hayes RB (1993) Misclassification of genetic susceptibility biomarkers: implications for case–control studies and cross-population comparisons. Cancer Epidemiol Biomarkers Prev 2:299–303

Rothstein HR, Sutton AJ, Borestein M (2005) Publication bias in meta-analysis—prevention, assessment and adjustments. Wiley, Chichester

Royle PL, Bain L, Waugh NR (2005) Sources of evidence for systematic reviews of interventions in diabetes. Diabet Med 22:1386–1393

Salanti G, Higgins JP, Trikalinos TA, Ioannidis JP (2007) Bayesian meta-analysis and meta-regression for gene–disease associations and deviations from Hardy–Weinberg equilibrium. Stat Med 26:553–567

Satten GA, Flanders WD, Yang Q (2001) Accounting for unmeasured population substructure in case–control studies of genetic association using a novel latent-class model. Am J Hum Genet 68:466–477

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S (2007) Genome-wide association analysis identifies loci for Type 2 diabetes and triglyceride levels. Science 316:1331–1336

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of Type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

Seminara D, Khoury MJ, O'Brien TR, Manolio T, Gwinn ML, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam AP, Danesh

J, Davey Smith G, Dolan S, Duncan R, Gruis NA, Hashibe M, Hunter D, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, Riboli E, Salanti G, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Ioannidis JP (2007) The emergence of networks in human genome epidemiology: challenges and opportunities. Epidemiology 18:1–8

Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38:209–213

Smith GD, Gwinn M, Ebrahim S, Palmer LJ, Khoury MJ (2006) Make it HuGE: human genome epidemiology reviews, population health, and the IJE. Int J Epidemiol 35:507–510

Spiegelhalter DJ, Thomas A, Best NG, Lunn D (2003) WinBUGS version 1.4 users manual

Spiegelhalter DJ, Abrams KR, Myles PJ (2004) Evidence synthesis. Wiley, Chichester

Sterne JA, Davey Smith G (2001) Sifting the evidence-what's wrong with significance tests? BMJ 322:226–231

Stewart LA, Parmar MK (1993) Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 341:418–422

Stewart LA, Tierney JF (2002) To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 25:76–97

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283:2008–2012

Sudlow C, Martinez Gonzalez NA, Kim J, Clark C (2006) Does apolipoprotein E genotype influence the risk of ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage? Systematic review and meta-analyses of 31 studies among 5961 cases and 17,965 controls. Stroke 37:364–370

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000) Methods for meta-analysis in medical research. Wiley, Chichester

Tang JL, Liu JL (2000) Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol 53:477–484

Terrin N, Schmid CH, Lau J (2005) In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol 58:894–901

Thomas DC, Haile RW, Duggan D (2005) Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet 77:337–345

Thompson SG, Higgins JP (2002) How should meta-regression analyses be undertaken and interpreted? Stat Med 21:1559–1573

Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP (2004) Establishment of genetic associations for complex diseases is independent of early study findings. Eur J Hum Genet 12:762–769

Uitterlinden AG, Ralston SH, Brandi ML, Carey AH, Grinberg D, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Reeve J, Reid DM, Amedei A, Bassiti A, Bustamante M, Husted LB, Diez-Perez A, Dobnig H, Dunning AM, Enjuanes A, Fahrleitner-Pammer A, Fang Y, Karczmarewicz E, Kruk M, van Leeuwen JP, Mavilia C, van Meurs JB, Mangion J, McGuigan FE, Pols HA, Renner W, Rivadeneira F, van Schoor NM, Scollen S, Sherlock RE, Ioannidis JP (2006) The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. Ann Intern Med 145:255–264

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

Wacholder S, Chatterjee N, Hartge P (2002a) Joint effect of genes and environment distorted by selection biases: implications for hospital-based case–control studies. Cancer Epidemiol Biomarkers Prev 11:885–889

Wacholder S, Rothman N, Caporaso N (2002b) Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 11:513–520

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. J Natl Cancer Inst 96:434–442

Wakefield J (2007) A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 81:208–227

WinBUGS (2003) 1.4 edn. MRC Biostatistic Unit, Cambridge

Wong MY, Day NE, Luan JA, Wareham NJ (2004) Estimation of magnitude in gene–environment interactions in the presence of measurement error. Stat Med 23:987–998

Yu K, Chatterjee N, Wheeler W, Li Q, Wang S, Rothman N, Wacholder S (2007) Flexible design for following up positive findings. Am J Hum Genet 81:540–551

Yusuf S, Peto R, Lewis J, Collins R, Sleight P (1985) Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 27:335–371

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT (2007) Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. Science 316:1336–1341

Zou GY, Donner A (2006) The merits of testing Hardy–Weinberg equilibrium in the analysis of unmatched case–control data: a cautionary note. Ann Hum Genet 70:923–933