



Reporting, Appraising, and Integrating Data on Genotype Prevalence and Gene-Disease Associations

Julian Little¹, Linda Bradley², Molly S. Bray³, Mindy Clyne⁴, Janice Dorman⁵, Darrell L. Ellsworth⁶, James Hanson⁷, Muin Khoury⁴, Joseph Lau⁸, Thomas R. O'Brien⁷, Nat Rothman⁷, Donna Stroup⁹, Emanuela Taioli¹⁰, Duncan Thomas¹¹, Harri Vainio¹², Sholom Wacholder⁷, and Clarice Weinberg¹³

¹ Epidemiology Group, Department of Medicine and Therapeutics, University of Aberdeen, Aberdeen, Scotland.

² Sunrise Medical Laboratory, Hauppauge, NY.

³ University of Texas Health Science Center at Houston, Houston, TX.

⁴ Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA.

⁵ Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

⁶ National Heart, Lung, and Blood Institute, Bethesda, MD.

⁷ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

⁸ Division of Clinical Care Research, New England Medical Center, Boston, MA.

⁹ National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA.

¹⁰ Ospedale Policlinico IRCCS, Milano, Italy.

¹¹ Biostatistics Division, Department of Preventative Medicine, University of Southern California, Los Angeles, CA.

¹² International Agency for Research on Cancer, Lyon, France.

¹³ National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Received for publication September 5, 2001; accepted for publication June 13, 2002.

The recent completion of the first draft of the human genome sequence and advances in technologies for genomic analysis are generating tremendous opportunities for epidemiologic studies to evaluate the role of genetic variants in human disease. Many methodological issues apply to the investigation of variation in the frequency of allelic variants of human genes, of the possibility that these influence disease risk, and of assessment of the magnitude of the associated risk. Based on a Human Genome Epidemiology workshop, a checklist for reporting and appraising studies of genotype prevalence and studies of gene-disease associations was developed. This focuses on selection of study subjects, analytic validity of genotyping, population stratification, and statistical issues. Use of the checklist should facilitate the integration of evidence from these studies. The relation between the checklist and grading schemes that have been proposed for the evaluation of observational studies is discussed. Although the limitations of grading schemes are recognized, a robust approach is proposed. Other issues in the synthesis of evidence that are particularly relevant to studies of genotype prevalence and gene-disease association are discussed, notably identification of studies, publication bias, criteria for causal inference, and the appropriateness of quantitative synthesis. *Am J Epidemiol* 2002;156:300–10.

association; case-control studies; causality; cohort studies; epidemiologic methods; gene frequency; genetic techniques; meta-analysis

The recent completion of the first draft of the human genome sequence (1, 2) and advances in technologies for genomic analysis are generating tremendous opportunities for epidemiologic studies to evaluate the role of genetic vari-

ants in the etiology of human disease (3). The basis of this evaluation will be identification of the allelic variants of human genes, description of the frequency of these variants in different populations, identification of diseases influenced

Correspondence to Dr. Muin J. Khoury, Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30341-3724 (e-mail: muk1@cdc.gov).

by these variants, and assessment of the magnitude of the associated risk. The process of identifying DNA variation that may be associated with disease is now underway through the cataloging and mapping of single nucleotide polymorphisms throughout the genome. The analysis of genotype data on single nucleotide polymorphisms may aid in the identification of DNA alterations that result in or contribute to disease states.

Although the spectrum of the relation between genes and disease is very broad, ranging from single gene disorders to multifactorial conditions, many common methodological issues apply throughout this spectrum (4). These issues relate to the planning and analysis of original studies, to the critical appraisal of individual studies, and to the integration of evidence from diverse studies.

Many papers deal with critical appraisal. A number of national organizations have specified criteria for assessing evidence on which policies and guidelines are based, for example, the Preventive Services Task Force's *Guide to Community Preventive Services* (5), the Scottish Intercollegiate Guidelines Network (6), and the Australian National Health and Medical Research Council (7). Issues that are particularly important in the appraisal of studies of genotype prevalence and gene-disease associations include the analytic validity of genotyping, selection of subjects, confounding (especially as a result of population stratification), gene-environment and gene-gene interactions, statistical power, and multiple statistical comparisons. There appears to be a need to develop guidelines to assess evidence from these studies. For example, a checklist for studies of associations between asthma and candidate genes has been proposed (8). Issues regarding the integration of evidence include identification of studies, the adequacy of reporting methods and results from individual studies, publication bias, quality scoring schemes, and the appropriateness of quantitative synthesis of the evidence.

Critical appraisal and integration of evidence require that the evidence be adequately reported. This paper presents recommendations regarding considerations that should be addressed when reporting studies of genotype prevalence and gene-disease associations, both for individual investigations and for systematic reviews. Although the focus of this paper is on recommendations for reporting, these recommendations necessarily have implications for study conduct and analysis. The recommendations resulted from a meeting of an expert panel workshop convened by the Centers for Disease Control and Prevention and the National Institutes of Health in January 2001. The methods used to develop the recommendations are described in the accompanying commentary (9).

The checklist presented in table 1 is intended to guide investigators in the preparation of manuscripts, to guide those who need to appraise manuscripts and published papers, and to be useful to journal editors and readers. It should not be regarded as an exhaustive list of points that have to be presented in all journal articles. We recognize that it may not always be feasible to address all of the considerations, for example, in studies of rare conditions in clinical settings. However, it would be useful if a record were kept of the coverage of these points for each study to accommodate

syntheses from different studies. We suggest establishing a Web-based methods register to record such information.

Because many methodological issues relate to both studies of genotype prevalence and studies of gene-disease association, we discuss these in parallel. A number of issues about which there is ongoing methodological debate are discussed in the text but are not included in the checklist because resolution has not been achieved.

REPORTING AND APPRAISAL OF SINGLE STUDIES

Analytic validity of genotyping

Issues in the appraisal and reporting of the analytic validity of genotyping include the definition of the genotype, the type of samples, the timing of sample collection, the genotyping method used, and the quality control measures.

Definition and grouping of genotypes

We recommend that a clear definition of the genotype(s) investigated should be presented. The validity of grouping on the basis of putative functional effects depends on the availability and quality of functional studies of gene variants, and this is likely to change over time. For multiallelic systems, genotypes have been grouped according to functional effects in some investigations. For example, this has been done for the *N*-acetyltransferase 2 gene (*NAT2*) polymorphisms (10). Therefore, we recommend that, when there are multiple alleles, those tested for should be specified.

It is important to distinguish true functional variants from markers that are associated with a disease only because they are in linkage disequilibrium with a functional variant. It may be useful to type several polymorphisms throughout the region of a candidate gene in order to construct haplotypes, which could be tested for association with the phenotype of interest. The increasing availability of mapped single nucleotide polymorphism markers (11–15) offers the opportunity for such an approach and presents methodological challenges (see the section on statistical issues).

Type of samples and timing of collection

A variety of DNA sources can be used. Polymerase chain reaction methods are widely used to genotype DNA extracted from blood or a number of other sources (16–22). Theoretically, genotyping results using genomic DNA from different tissues should be identical, because the DNA should contain an identical sequence in all tissues. However, participation rates may be higher for studies based on mouthwash or hair samples than those based on blood samples (19), and DNA is often more difficult to obtain and purify from certain tissues than from whole blood.

Sometimes the source of DNA is peripheral blood for controls, while it is a different tissue (e.g., tumor specimen) for cases. Although this should not in theory affect the genotypic assay when germline mutations are being assessed, the technique may be easier to perform in DNA extracted from blood, and therefore the results of the genotype may be more accurate in controls than in cases. In studies on cancer in

TABLE 1. Proposed checklist for reporting and appraising studies of genotype prevalence and gene-disease associations

Item to be specified	Details by type of study	
	Genotype prevalence	Gene-disease associations
Purpose of study	Yes	Detect associations or estimate magnitude of association
Analytic validity of genotyping		
Types of samples used	Yes	For cases and for controls
Timing of sample collection and analysis, by study group*	Ethnic group†	Cases vs. controls†
Success rate in extracting DNA, by study group*	Ethnic group†	Cases vs. controls†
Definition of the genotype(s) investigated; when there are multiple alleles, those tested for should be specified	Yes	Yes
Genotyping method used (reference; for polymerase chain reaction methods—primer sequences,* thermocycle profile,* no. of cycles*)	Yes	Yes
Percentage of potentially eligible subjects for whom valid genotypic data were obtained, by study group	Ethnic group†	Cases vs. controls†
If pooling was used, strategy for pooling of specimens from cases and controls		Yes
Quality control measures*	Yes	Including blinding of laboratory staff
Degree of reproducibility between quality control replicates	Yes	Yes
Samples from each group of subjects compared (e.g., cases and controls) included in each batch analyzed*		Yes
Selection of study subjects		
Geographic area from which subjects were recruited	Yes	Yes
The recruitment period	Yes	Yes
Recruitment methods for subjects whose genotypes were determined, such as random population-based sampling, blood donors, and hospitalized subjects with reasons for hospitalization	Yes	Controls†
Definition of cases and method of ascertainment		Yes
No. of cases recruited from families and methods used to account for related subjects		Yes
Exclusion criteria for cases and controls		Yes
Recruitment rates	Where possible by sex, age, and ethnic group	For cases and controls
Mean age and standard deviation or age range of study subjects, and the distribution by sex	Yes	For cases and controls
If the subjects were controls from a case-control study, information on the disease under investigation and any matching criteria such as age, gender, and/or risk factor levels	Yes	
Ethnic group of study subjects	Yes	
Similarity of sociodemographic (or other) characteristics of subjects for whom valid genotypic data were obtained with characteristics of subjects for whom such data were not obtained*		Yes
Steps taken to ensure that controls are noncases*		Yes
Confounding, including population stratification		
Design		Yes
If other than a case-family control design, matching for ethnicity or adjustment for ethnicity in analysis		Yes
Potential correlates of the genotype identified and taken into consideration in design or analysis		Yes
Statistical issues		
Distinguish clearly a priori hypotheses and hypotheses generated		Yes
If haplotypes used, specify how these were constructed	Yes	Yes
No. of subjects included in the analysis	Yes	Yes
Method of analysis, with reference, and software used to do this		Yes
Confidence intervals	Of genotype frequency	Of measures of association with the genotype
Assessment of goodness-of-fit of the model used*		Yes

* Additional information recorded (ideally in Web-based methods register) but not necessarily presented in journal article.

† For example.

particular, the ability to extract DNA from tumor tissue depends on several factors, including the amount of necrosis and the quantity of tissue, and therefore can result in the inclusion of selected cases. Because the timing of sample collection and the storage period before DNA extraction and/or analysis may differ between cases and controls, we recommend that the number of specimens collected, the success rate and timing of DNA extraction, and the success rate and timing of genotyping should be reported by the study group. Differences in timing of the sample collection and the storage period may be of especial importance when the genotype is assessed on the basis of phenotype.

Genotyping by polymerase chain reaction methods

The results of genotyping by polymerase chain reaction methods may be affected by the technology used, operational conditions in the laboratory, and inadequate safeguards for scoring the results. A recent appraisal of 40 studies in which molecular genetic techniques were used demonstrates the need for universal standards for quality control. Molecular genetic analyses were repeated in a pertinent sample of specimens or the test was confirmed with another procedure in only 15 (37.5 percent) studies, and assays were conducted in a manner blind to pertinent characteristics of subjects or hypotheses in only 13 (32.5 percent) studies (23). With the application of high-throughput methods, a number of which are under development, quality control procedures are particularly important. Numerous polymerase chain reaction-based methods are available for single nucleotide polymorphism genotyping, including restriction fragment length polymorphism analysis (24), oligonucleotide ligation assays (25), the "TaqMan" assay (26), single-base-pair extension assays, and others. We recommend that the description of the genotyping assay should include the primer sequences, thermocycle profile, number of cycles, and reference.

The accuracy of polymerase chain reaction methodology is generally quite high, although different types of laboratory errors can occur. Poor or nonspecific amplification of polymerase chain reaction products or lack of complete enzymatic function of restriction enzymes, leading to incomplete digestion in restriction fragment length polymorphism assays, can occur with measurable frequency. These errors can be controlled for by optimization of assays prior to genotyping and the use of internal controls and repetitive experiments. A data set containing less than 95 percent reproducibility between replicates indicates a potential problem and should not be considered as accurate. Rothman et al. (27) noted that misclassification of the genotype can bias measures of association between the genotype and disease, especially when the prevalence of the genotype is either very high or very low.

Some genotypic tests require visual inspection and interpretation of electrophoresis gels, and therefore, observer variability may be important (23). Observer variability can be minimized by double blind scoring and data entry, followed by electronic comparison of the blind entries.

Discrepancies are then flagged automatically and adjudicated by a third (experienced) person.

We recommend that authors specify the quality measures used for the genotyping analysis and provide information on the degree of reproducibility between quality control replicates. Quality control measures include 1) internal validation for analytic validity; 2) blinding of laboratory personnel to pertinent characteristics of the samples, donor subjects, and hypotheses being investigated; 3) procedures for establishing duplicates and quality control numbers from blind duplicates; 4) test failure rate, by study group; 5) inspection of whether genotype frequencies conform to Hardy-Weinberg equilibrium (in controls in case-control studies) and are consistent with other reports for the same population (this criterion should not be binding); and 6) blind or automated data entry and third party adjudication. If a large number of samples do not produce acceptable genotyping results, comparability data should be provided with samples that yield acceptable results.

Genotyping on the basis of phenotype

In some studies, the genotype has been inferred on the basis of a phenotypic test. A potential advantage of this approach is that the phenotypic assay reflects the enzyme activity level and therefore may provide a direct measure of the functional significance of the underlying genetic polymorphism. The genotype is one determinant of the long-term enzyme activity level. Nevertheless, an enzyme activity assay provides a measure only at a single point in time and potentially may be distorted by systematic influences (e.g., effects of disease stress on metabolism and inducing factors), as well as by random measurement error. Contrasting results between studies based on phenotypic and polymerase chain reaction methods have been observed, for example, for the acetylator polymorphism and colorectal cancer (10) and for the glutathione *S*-transferase μ polymorphism and lung cancer (28). These differences underline the need to specify the genotyping method.

Phenotypic methods have also been used because enormous sequence variability within some genes presents challenges to genotyping by polymerase chain reaction methods. For example, more than 240 different constitutional neurofibromatosis type 1 mutations have been documented (29). As the majority of these lead to a truncated protein product, a protein truncation test was developed. This test is commercially available, but its sensitivity, specificity, and predictive value have not been established. A "significant" proportion of cases identified by the protein truncation test are not confirmed by sequencing, suggesting a problem of false positives for the protein-based assay. False negative results may also be a problem as the protein truncation test had a sensitivity of around 70 percent in small series of clinically diagnosed neurofibromatosis type 1 (29).

As for genotyping by polymerase chain reaction methods, we recommend that quality control measures and the degree of reproducibility between quality control replicates be specified.

Selection of subjects

Evaluation of potential selection bias requires consideration of the aim and design of the study and fieldwork. To date, most studies of gene-disease associations for late-onset diseases have used the case-control design. It is important to distinguish between studies whose aim is the *detection* of an association and those whose aim is the *estimation* of an association. In the former situation, cases may be "overselected" from multiplex families to increase the power to detect an association; it would be inappropriate to present the measure of association as an estimate of population association. In the latter situation, the principles underlying study design are essentially the same as for the investigation of the magnitude of association with environmental risk factors, including the minimization of the potential for selection bias emphasized in many epidemiologic textbooks (30–33). In a number of studies, the selection of cases has not been well described (34). In a review of type 1 diabetes and human leukocyte-DQ antigen locus (*HLA-DQ*) polymorphisms, it was noted that many studies were based on convenience samples of cases in which type 2 diabetic persons who use insulin in their treatment regimen have been included (35). In several studies of cancer, prevalent cases have been included to varying extents (28). In these studies, bias in both detecting and estimating association would occur if the genotype affected survival and thereby ascertainment or if genotypes were assayed by a phenotypic test that was influenced by disease progression and/or treatment.

A recurrent problem in case-control studies of gene-disease associations with unrelated controls has been that the controls were not selected from the same source population as the cases (10, 34–36). The potential problem of selecting controls who do not represent the population from which cases arise is illustrated by the divergence in odds ratios for the association between colorectal cancer and the glutathione *S*-transferase T1 gene (*GSTT1*) null genotype (37), when the different control groups were analyzed (36). Thus, sufficient information should be presented to assess whether controls would have become cases if they had developed the disease. In regard to genotype prevalence, many early studies were based on convenience samples, and not infrequently, little information was given on how the samples were selected (10, 34, 36, 38).

There are a number of ongoing cohort studies in which DNA samples have been collected. Compared with case-control studies, cohort studies have a number of advantages, including the capacity to examine age at onset distributions and multiple disease outcomes. The use of case-cohort and nested case-control analysis of archived samples that are suitable for genotypic analysis potentially can minimize the disadvantages of the cost of genotyping an entire cohort (39–41). A major advantage of the case-cohort design for studies where expensive biologic markers are collected is that the same comparison group can be used for several different disease outcomes. Therefore, this design is likely to be used increasingly.

In case-cohort studies, comparison subjects are a random sample of the cohort, and the effect of age, which is the key time variable, is controlled for in the analysis. In more tradi-

tional 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 (42). We recommend that the method of age adjustment be specified in case-cohort studies and that in nested case-control studies details of the matching on age or other temporal factors be presented.

Population stratification

There has been concern about the possible effects of population stratification on the results of population-based case-control studies (43–46). Population stratification includes differences between groups in ethnic origin, and it can also arise because of differences between groups of similar ethnic origin but between which there has been limited admixture, such as in isolated populations. For example, a population might comprise the descendants of waves of immigrants from the same source but differ generally because of founder effects. The differences may then be apparent because insufficient time has elapsed for mixture between the groups. In an exploration of the possible degree of bias from population stratification in US studies of cancer among non-Hispanics of European descent, it was concluded that this bias is unlikely to be substantial when the epidemiologic principles of study design, conduct, and analysis are rigorously applied (47). Variations in the frequency of certain genotypes in African Americans appear to be much wider than those observed in subjects of European origin, and therefore the possibility of stratification may be higher (48).

Concern about the possible effects of population stratification has stimulated the development of family-based case-control designs, which essentially eliminate potential confounding from this source (49, 50). The most commonly used examples of such designs involve the use of siblings or parents as controls. Sibling controls are derived from the same gene pool as cases. However, selection bias could arise from the fact that a sibling may not be available for every case. Bias would arise if determinants of availability, for example, sibship size, were associated with genotype, particularly if a substantial proportion of cases had to be excluded because no sibling control was available. In addition, because of overmatching on genotype, there is a loss of statistical power compared with the use of unrelated controls (51). This loss of power generally does not occur for case-parental control studies (50), which have been advocated for the identification of modest gene-disease associations (52). However, the need to obtain samples from parents is a practical problem that limits the applicability of the design for diseases of late onset. Therefore, it is important that the study design be reported so that the possible impact of population stratification can be assessed.

Another approach proposed to minimize the potential problem of population stratification when using unrelated controls is to measure and adjust for genetic markers of ethnicity that are not linked to the disease under investigation (53–56). This 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. It is

unlikely that a single measure will capture the important sources of ethnic variation (57). In case-control studies with unrelated controls or in cohort studies, we recommend that the details of matching for ethnicity or adjustment for ethnicity in analysis be reported.

Statistical issues

We recommend that genotype frequencies be presented when available or inexpensive to obtain, rather than allele frequencies alone, both for studies of genotype prevalence and for studies of gene-disease associations. We make this recommendation because it is the genotype that determines risk and because allele frequencies can be calculated from genotype data (e.g., to determine Hardy-Weinberg equilibrium), whereas if only allele frequencies are presented, genotype frequencies cannot be calculated. Clearly, this recommendation would not be appropriate for studies based on DNA pooling, which may be a valuable approach in estimating allele frequency distributions in many loci in multiple populations (58), for initial investigation of disease loci, or for follow-up to confirm regions identified in linkage studies (59). In studies in which pooling is used, the strategy for pooling specimens from cases and controls should be specified.

Methodological issues relating to haplotype analysis are still under development. In particular, in studies based on unrelated individuals, haplotypes can only be estimated probabilistically based on allele frequencies. If external estimates of haplotype frequencies in the population are used instead of estimating them within the study, inference may be affected by the quality and availability of the data on haplotype frequencies in the relevant population. As more single nucleotide polymorphism loci are identified, the number of possible haplotypes will become very large, in turn raising the issues of multiple comparisons and sparse data for many haplotypes (60–62). We recommend that, when haplotypes are used, the method of construction should be specified.

Calculation of risk difference (i.e., the risk of the disease in those with the genotype under investigation minus the risk of the disease in those without the genotype) in the context of gene-disease associations may be useful in that it measures the potential impact of the association in public health terms (33). However, the magnitude of the risk difference is less likely to be generalizable to other populations than is the relative risk (30), because it depends on the baseline risk in those without the genotype, which is likely to vary between populations.

A small study size is a limitation of many studies that test a priori hypotheses about gene-disease associations (e.g., 36, 63). A possible solution is pooled analysis (see below). One research strategy proposed for the future is large-scale testing by genome-wide association mapping (52, 60, 64, 65). It is important to note that this strategy is hypothesis generating rather than hypothesis testing, and thus it may require additional safeguards against type 1 error. For example, Risch and Merikangas (52) suggested specifying a higher significance level. However, increasing the significance level will increase the number of subjects required to

have adequate statistical power (52), which may significantly increase recruitment costs and make some studies unfeasible. An alternative approach is to emphasize replication of findings and to obtain data on biologic plausibility, for example, from *in vitro* studies. We recommend that all tests performed should be reported, as long as the tests have adequate statistical power, not just the “significant” ones. This would require reviewers and editors to give importance (and journal space) to negative results.

Associations between several genes and a disease can be tested according to a priori hypotheses based, for example, on the biologic mechanism of these genes in determining the disease. It is recognized that it is becoming the usual practice in human genome epidemiology studies to initiate a study to test hypotheses that are current at that time and also to establish a resource to test additional hypotheses proposed later, on the basis of knowledge external to the resource. These are all a priori hypotheses. We reiterate the need to distinguish between hypothesis testing and hypothesis generation.

INTEGRATION OF EVIDENCE

There are established principles for the integration of observational evidence in relation to causal inference (66, 67). Since these principles were documented, there has been considerable work on the identification of admissible evidence to which the principles should be applied.

Identification of studies

A comprehensive search is one of the key differences between a systematic review and a traditional review (68). We recommend that the details of the strategy used to identify relevant papers should be specified as described by Stroup et al. (69). There have been several instances of sequential or multiple publications of analyses of the same or overlapping data sets. An aid to identifying this problem is to organize evidence tables first by geographic area and then by study period within a specified area. If it is clear that the reports relate to the same or overlapping data sets, then we recommend including data only from the largest or most recent publication. It is possible that, under these circumstances, details of the methodology are described in greater detail in an earlier publication. If so, we recommend including the reference to the earlier publication with the reference to the publication from which the data were abstracted in the evidence tables.

Publication bias

Publication bias is potentially a serious problem for the integration of evidence. One method of minimizing the potential impact of publication bias is to identify and include “gray literature,” which includes abstracts, technical reports, and non-English journals (70) that may not be identified by electronic searches. We recommend caution in using various types of “gray 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 study quality. We suggest that consid-

eration should be given to including “gray literature” if the study quality can be assessed adequately.

A labor-intensive method to minimize publication bias is to establish a research register for studies of gene-disease association similar to the Cochrane Collaboration, which maintains a register of controlled trials (71), and the *Directory of On-going Research in Cancer Prevention* (72).

In other fields, quantitative and qualitative methods of detecting publication bias have been used, such as the fail-safe technique where the number of new studies averaging a null result needed in order to bring the overall effect to nonsignificance is calculated (73, 74). After this is calculated, a judgment can be made as to whether it is realistic to assume that so many unpublished studies exist in the field of investigation. If the assumption were realistic, then there would be doubt about the validity of conclusions based on potential evidence. Other quantitative and qualitative methods have been reviewed by Sutton et al. (75) and Thornton and Lee (76). In general, all the methods have limitations. Therefore, it seems appropriate to take into account the fact that the evidence base may be skewed toward positive results in drawing conclusions about causal relations.

Quality scoring

There are a number of publications concerning the rating of the quality of analytic observational studies. Several relate to case-control studies (30, 32, 77–81). Some (77, 78) are part of a series of articles documenting the deficiencies of epidemiologic research; they have been challenged on the grounds of technical errors, failure to distinguish important from unimportant biases, and ignoring the need to weight the totality of the evidence about a relation (82, 83). Other issues include possible overemphasis of the potential problems of case-control studies as compared with cohort studies (80) and difficulty in assessing differences between methods applied in the case and control groups or between different exposure (prognostic) groups (81, 84).

A number of authors have proposed quantitative quality scoring systems for critical appraisal (84). 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 (85–87). This type of assessment has also been considered for pooled analysis (88, 89). Certain features of the assessment schemes are specific to the disease and/or the exposure under consideration, and each aspect of the study is given equal weight. Thus, summation of points might result in worse quality scores for a study with several minor flaws than for a study with one major flaw. Although empirical studies on a large number of primary investigations might suggest an overall relation between a specific aspect of study design and the reported results, this relation is ecologic and may not be true for a specific investigation. Therefore, it is very difficult to isolate specific noncausal factors, which might affect the interpretation of a single investigation. Jüni et al. (90) 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 similar caution in consideration of studies of gene-disease associations. As in clinical trials, it may be more appropriate to consider multidimensional domains than a single grade in the integration of evidence from observational studies.

Currently, there is little or no empirical evaluation of the quality scoring of association studies. However, we recognize that many users of data on genotype prevalence and gene-disease associations need a robust means of grading evidence. We recommend following the approach of the Scottish Intercollegiate Guidelines Network (6). In this approach, studies of the gene-disease association in which all or most of the criteria specified in table 1 are satisfied would be graded as “++.” Criteria that have not been fulfilled would not affect the grade if it were thought that the conclusions of the study would be *very unlikely* to be affected by their omission. Studies in which some of the criteria have been fulfilled, and those that were not fulfilled would be thought to be *unlikely* to alter the conclusions, would be graded as “+.” Studies in which few or no criteria were fulfilled, and the conclusions of the study would be thought *likely* or *very likely* to be altered by multiple omissions in required criteria for an acceptable study would be graded as “-.” For studies of genotype prevalence, similar considerations would be applied.

QUALITATIVE SYNTHESIS

Hierarchy of evidence

In many schemes of qualitative synthesis of evidence, there is a hierarchy whereby certain study designs are considered inherently superior to others. In general, analytic epidemiologic designs are stronger than ecologic designs and studies of case series or reports. Although it has been argued that cohort studies may be less subject to bias than case-control studies, there are important issues about the quality of follow-up and case ascertainment. Therefore, it seems more rigorous to weight the evidence from specific studies of these types on the basis of a full critical appraisal rather than solely on the basis of general design. In addition, case reports may lead to novel hypotheses and be of value in considering biologic plausibility (see below).

Causal inference

There are well-established criteria for causal inference (66, 67). In relation to the consistency of gene-disease associations, heterogeneity between studies is frequent (91). Consideration has to be given to methodological factors that might account for inconsistency. For example, in a meta-analysis of 370 studies that assessed 36 gene-disease associations, a small sample size of the first publication was a predictor of inconsistent results (91). In addition, it is important to consider that differences between studies in distributions of subjects by age and gender will be sources of heterogeneity. For example, hormonal alterations can affect ligand binding, enzyme activity, gene expression, and the metabolic pathways influenced by gene expression. Some

inconsistency among the results of gene-disease association studies may be secondary to variation among studies in the prevalence of interacting environmental factors that have not been assessed (D. J. Hunter, Channing Laboratory, unpublished manuscript). It would be appropriate to test a priori hypotheses about differences in gene-disease associations and genotype frequencies between studies that may arise from these sources. We recommend that information on the age distribution of subjects be presented and that consideration be given to presenting data on gene-disease associations by gender. As noted in the section on analytic validity, contrasting results between studies in which genotyping was based on polymerase chain reaction methods and those in which it was based on phenotypic assays have been observed. However, there is a need to consider the possibility that these differences may have been due to reasons other than the genotyping method, such as selection or participation bias of cases and/or controls. In addition, a given gene may metabolize multiple environmental substrates, and thus phenotypic assays, using one chemical to induce the gene, may not truly reflect the metabolizing activity of that gene. It is also possible that other DNA variants may alter enzyme function or activity.

Regarding the strength of association, many of the genetic variants so far identified as influencing susceptibility to common diseases are associated with a low relative and absolute risk (92). Therefore, exclusion of noncausal explanations for associations is crucial.

Biologic plausibility is a particularly important issue. It is linked with consideration of 1) whether a known function of the gene product can be linked to the observed phenotype; 2) whether the gene is expressed in the tissue of interest; and 3) temporal relations, including the time window of gene expression in relation to age-specific gene-disease relations. Thus, the gene should be in the disease pathway and/or involved in the mechanism that is responsible for the development of the disease. If not, then the effect of the gene may be indirect. It may also be relevant to consider maternally mediated effects of the maternal genotype and parental imprinting. Case reports may provide clues that could not be obtained from epidemiologic designs. For example, evidence from a heteropaternal twin pair provided a lead to genetic differences in the metabolism of phenytoin that accounted for a lack of concordance for teratogenic effect (93).

In regard to temporality, it is possible that the disease could influence the result of a phenotypic assay of the genotype under investigation. This should not be a problem with polymerase chain reaction methods. Methods to analyze longitudinal phenotypes, such as changes in blood pressure or obesity (94) over time, are being developed. If data were available on the time window of gene expression, it would be relevant to consider this in relation to the age specificity of gene-disease relations.

Experimental support for a gene-disease association is most likely to be derived from studies of gene expression in knockout or other experimental animals, from *in vitro* data on gene function, or from experimental interactions based on clinical protocols aimed at normalizing the levels of a product regulated by the gene (e.g., with gene therapy in cystic fibrosis).

Quantitative synthesis

There are two types of quantitative synthesis of evidence: 1) meta-analysis of the results of studies and 2) pooled analysis of data on individual subjects obtained in several studies. There has been debate about the validity of meta-analysis of observational studies (69, 95). On the one hand, meta-analysis may indicate a "spurious precision," and it has been suggested that either meta-analysis of observational studies should be abandoned altogether (96) or the focus of attention should be the consideration of possible sources of heterogeneity between studies (91, 97). On the other hand, meta-analysis can help to clarify whether or not an association exists and to provide an indication of the quantitative relation between the dependent and independent variables (98). The indication of the quantitative relation, although potentially biased, may be of value in consideration of the public health effects of interventions based on knowledge of the genetic factor and/or its interactions.

Pooled analysis requires *data* on individual subjects. This approach offers many advantages over the meta-analysis of the *results* of studies, 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, evaluation of alternative genetic models and multiple genes, consistent treatment of subpopulations, and assessment of sampling bias (88). For example, this approach has been used successfully to study the effect of chemokine and chemokine receptor alleles on human immunodeficiency virus type 1 disease progression (99). Nevertheless, pooling approaches require a much greater commitment of time and resources to collect primary data and to coordinate a large collaborative project (100). For questions that justify the required intensive effort, the pooling approach is a useful tool to help to clarify the role of candidate genes in complex human diseases (101).

We recommend that this type of quantitative synthesis be done whenever possible in preference to meta-analysis of the results of studies when a high degree of accuracy of the measures of effect is required. However, stratification by original study may still be important, to allow for and elucidate the causes of heterogeneity among the data sets being pooled.

CONCLUSION

Analytic epidemiologic methods remain a critical issue in studies of genotype prevalence and gene-disease associations. While recognizing the innovative aspects of early work, we note that many studies have had limited value because one or more of the issues of analytic validity of genotyping, possible selection bias, confounding, possible gene-environment and gene-gene interactions, and statistical power were inadequately addressed. We recognize the particular importance of interdisciplinary collaboration in this quickly expanding field and recommend that both individual studies and systematic reviews involve epidemiologic input.

ACKNOWLEDGMENTS

This paper represents consensus from the Centers for Disease Control and Prevention-National Institutes of Health Human Genome Epidemiology Workshop, January 2001.

REFERENCES

- McPherson JD, Marra M, Hillier A, et al. A physical map of the human genome. *Nature* 2001;409:934–41.
- Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304–51.
- Shpilberg O, Dorman JS, Ferrell RE, et al. The next stage: molecular epidemiology. *J Clin Epidemiol* 1997;50:633–8.
- Khoury MJ, Little J. Human Genome Epidemiologic reviews: the beginning of something HuGE. (Editorial). *Am J Epidemiol* 2000;151:2–3.
- Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based guide to community preventive services—methods. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;18(suppl):35–43.
- Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50: a guideline developers' handbook. Edinburgh, United Kingdom: SIGN, 2001. (SIGN publication no. 50). (<http://www.sign.ac.uk/guidelines/fulltext/50/index.html>).
- National Health and Medical Research Council (NHMRC). A guide to the development, implementation and evaluation of clinical practice guidelines. Canberra, Australia: NHMRC, 1999.
- Weiss ST. Association studies in asthma genetics. *Am J Respir Crit Care Med* 2001;164:2014–15.
- Khoury MJ. Commentary: epidemiology and the continuum from genetic research to genetic testing. *Am J Epidemiol* 2002;156:297–9.
- Brockton N, Little J, Sharp L, et al. *N*-Acetyltransferase polymorphisms and colorectal cancer: a HuGE review. *Am J Epidemiol* 2000;151:846–61.
- Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;409:928–33.
- Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature* 2001;411:199–204.
- Altshuler D, Pollara VJ, Cowles CR, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000;407:513–16.
- Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 2000;9:2403–8.
- Porter CJ, Talbot CC, Cuticchia AJ. Central mutation databases—a review. *Hum Mutat* 2000;15:36–44.
- Yang M, Hendrie HC, Hall KS, et al. Improved procedure for eluting DNA from dried blood spots. *Clin Chem* 1996;42:1115–16.
- Gale KB, Ford AM, Repp R, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A* 1997;94:13950–4.
- Walker AH, Najarian D, White DL, et al. Collection of genomic DNA by buccal swabs for polymerase chain reaction-based biomarker assays. *Environ Health Perspect* 1999;107:517–20.
- Harty LC, Shields PG, Winn DM, et al. Self collection of oral epithelial cell DNA under instruction from epidemiologic interviewers. *Am J Epidemiol* 2000;151:199–205.
- Garcia-Closas M, Egan K, Abruzzo J, et al. Collection of genomic DNA from adults in epidemiologic studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001;10:687–96.
- Pfeiffer H, Huhne J, Ortmann C, et al. Mitochondrial DNA typing from human axillary, pubic and head hair shafts—success rates and sequence comparisons. *Int J Legal Med* 1999;112:287–90.
- Schneider A, Borgnat S, Lang H, et al. Evaluation of microsatellite analysis in urine sediment for diagnosis of bladder cancer. *Cancer Res* 2000;60:4617–22.
- Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research. The need for methodological standards. *JAMA* 1999;281:1919–26.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with *Hha*I. *J Lipid Res* 1990;31:545–8.
- Tobe VO, Taylor SL, Nickerson DA. Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. *Nucleic Acids Res* 1996;24:3728–32.
- Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation polymerase chain reaction with fluorogenic probes. *Nucleic Acids Res* 1993;21:3761–6.
- Rothman N, Stewart WF, Caporaso NE, et al. Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons. *Cancer Epidemiol Biomarkers Prev* 1993;2:299–303.
- d'Errico A, Malats N, Vineis P, et al. Chapter 23. Review of studies of selected metabolic polymorphisms and cancer. In: Vineis P, Malats N, Lang M, et al, eds. *Metabolic polymorphisms and susceptibility to cancer*. Lyon, France: International Agency for Research on Cancer, 1999:323–93.
- Rasmussen SA, Friedman JM. *NF1* gene and neurofibromatosis type 1. *Am J Epidemiol* 2000;151:33–40.
- Breslow NE, Day NE, eds. *Statistical methods in cancer research. Vol I. The analysis of case-control studies*. Lyon, France: International Agency for Research on Cancer, 1980. (IARC scientific publication no. 32).
- Kelsey JL, Whittemore AS, Evans AS, et al. *Methods in observational epidemiology*. Oxford, United Kingdom: Oxford University Press, 1996.
- Elwood M. *Critical appraisal of epidemiological studies and clinical trials*. Oxford, United Kingdom: Oxford University Press, 1998.
- dos Santos Silva I. *Cancer epidemiology: principles and methods*. Lyon, France: International Agency for Research on Cancer, 1999.
- Botto LD, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am J Epidemiol* 2000;151:862–77.
- Dorman JS, Bunker CH. *HLA-DQ* locus of the human leukocyte antigen complex and type 1 diabetes mellitus: a HuGE review. *Epidemiol Rev* 2000;22:218–27.
- Cotton SC, Sharp L, Little J, et al. Glutathione *S*-transferase polymorphisms and colorectal cancer. *Am J Epidemiol* 2000;151:7–32.
- Chenevix-Trench G, Young J, Coggan M, et al. Glutathione *S*-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset. *Carcinogenesis* 1995;16:1655–7.
- Wang SS, Fernhoff PM, Hannon WH, et al. Medium chain acyl-CoA dehydrogenase deficiency human genome epidemiology review. *Genet Med* 1999;1:332–9.
- Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City cohort, ALIVE

- Study. *Science* 1996;273:1856–62.
40. Michael NL, Chang G, Louie LG, et al. The role of viral phenotype and *CCR-5* gene defects in HIV-1 transmission and disease progression. *Nat Med* 1997;3:338–40.
 41. Langholz B, Rothman N, Wacholder S, et al. Cohort studies for characterizing measured genes. *J Natl Cancer Inst Monogr* 1999;26:39–42.
 42. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991;2:155–8.
 43. Knowler WC, Williams RC, Pettitt DJ, et al. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 1988;43:520–6.
 44. Gelernter J, Goldman D, Risch N. The *A1* allele at the D2 dopamine receptor gene and alcoholism: a reappraisal. *JAMA* 1993;269:1673–7.
 45. Khoury M, Beaty TH, Cohen BL. *Fundamentals of genetic epidemiology*. New York, NY: Oxford University Press, 1993.
 46. Caporaso N, Rothman N, Wacholder S. Case-control studies of common alleles and environmental factors. *J Natl Cancer Inst Monogr* 1999;(26):25–30.
 47. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000;92:1151–8.
 48. Garte S. The role of ethnicity in cancer susceptibility gene polymorphisms: the example of *CYP1A1*. *Carcinogenesis* 1998;19:1329–32.
 49. Teng J, Risch N. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human disease. II. Individual genotyping. *Genome Res* 1999;9:234–41.
 50. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999;149:693–705.
 51. Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. *J Natl Cancer Inst Monogr* 1999;26:31–7.
 52. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516–17.
 53. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997–1004.
 54. Pritchard J, Stephens M, Rosenberg NA, et al. Association mapping in structured populations. *Am J Hum Genet* 2000;67:170–81.
 55. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;20:4–16.
 56. Satten GA, Flanders WD, Yang Q. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. *Am J Hum Genet* 2001;68:446–77.
 57. Lin SS, Kelsey JL. Use of race and ethnicity in epidemiologic research: concepts, methodological issues, and suggestions for research. *Epidemiol Rev* 2000;22:187–202.
 58. Shaw SH, Carrasquillo MM, Kashuk C, et al. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res* 1998;8:111–23.
 59. Barcellos LF, Klitz W, Field LL, et al. Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am J Hum Genet* 1997;61:734–47.
 60. Schork NJ, Fallin D, Lanchbury S. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* 2000;58:250–64.
 61. Fallin D, Cohen A, Essioux L, et al. Genetic analysis of case/control data using estimated haplotype frequencies: application to *APOE* locus variation and Alzheimer's disease. *Genome Res* 2001;11:143–51.
 62. Thomas DC, Morrison J, Clayton DG. Bayes estimates of haplotype effects. *Genet Epidemiol* 2001;21(suppl 1):S712–17.
 63. Boffetta P, Pearce N. Chapter 10. Epidemiological studies on genetic polymorphisms: study design issues and measures of occurrence and association. In: Vineis P, Malats N, Lang M, et al, eds. *Metabolic polymorphisms and susceptibility to cancer*. Lyon, France: International Agency for Research on Cancer, 1999:97–108.
 64. Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* 1998;95:11389–93.
 65. Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998;8:1273–88.
 66. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
 67. Office of the Surgeon General, Public Health Service. *Smoking and health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Department of Health, Education, and Welfare, 1964. (PHS publication no. 1103).
 68. Oxman AD. *The Cochrane Collaboration handbook: preparing and maintaining systematic reviews*. 2nd ed. Oxford, United Kingdom: The Cochrane Collaboration, 1992.
 69. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283:2008–12.
 70. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995;48:159–63.
 71. Chalmers I, Sackett D, Silag C. *The Cochrane Collaboration*. In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research: on the 25th anniversary of Archie Cochrane's effectiveness and efficiency*. London, United Kingdom: British Medical Journal Books, 1997:231–49. (also see <http://www.cochrane.org/cochrane/cc-broch.htm>CCTR).
 72. Sankaranarayanan R, Becker N, Démaret E, eds. *Directory of on-going research in cancer prevention*. Lyon, France: International Agency for Research on Cancer, 2000. (<http://www-dep.iarc.fr/direct/prevent.htm>).
 73. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979;86:638–41.
 74. Stroup DF, Thacker SB. Meta-analysis in epidemiology. In: *Encyclopedia of epidemiologic methods*. Chichester, United Kingdom: Wiley & Sons Publishers, 2000:557–70.
 75. Sutton AJ, Abrams KR, Jones DR, et al. Systematic reviews of trials and other studies. *Health Technol Assess* 1998;2:1–276.
 76. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol* 2000;53:207–16.
 77. Feinstein AR. Methodologic problems and standards in case-control research. *J Chronic Dis* 1979;32:35–41.
 78. Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 1979;66:556–64.
 79. Kopec JA, Esdaile JM. Bias in case-control studies. A review. *J Epidemiol Community Health* 1990;44:179–86.
 80. Crombie IK. *The pocket guide to critical appraisal*. London, United Kingdom: BMJ Publishing Group, 1996.
 81. Liddle J, Williamson M, Irwig L. *Method for evaluating research and guideline evidence (MERGE)*. Sydney, Australia: NSW Health Department, 1996.
 82. Savitz DA, Greenland S, Stolley PD, et al. Scientific standards of criticism: a reaction to "Scientific standards in epidemiologic studies of the menace of daily life," by A. R. Feinstein.

- Epidemiology 1990;1:78–83.
83. Weiss NS. Scientific standards in epidemiologic studies. *Epidemiology* 1990;1:85–6.
 84. Dixon RA, Munro JF, Silcocks PB. *The evidence based medicine workbook: critical appraisal for clinical problem solving*. Oxford, United Kingdom: Oxford University Press, 1997.
 85. Longnecker MP, Berlin JA, Orza MJ, et al. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988;260:652–6.
 86. Longnecker MP, Orza MJ, Adams ME, et al. A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. *Cancer Causes Control* 1990;1:59–68.
 87. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol* 1990;132:612–28.
 88. Friedenreich CM. Methods for pooled analyses of epidemiologic studies. *Epidemiology* 1993;4:295–302.
 89. Friedenreich CM, Brant RF, Riboli E. Influence of methodologic factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber. *Epidemiology* 1994;5:66–79.
 90. Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
 91. Ioannidis JP, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet* 2001;29:306–9.
 92. Caporaso N. Selection of candidate genes for population studies. In: Vineis P, Malats N, Lang M, et al, eds. *Metabolic polymorphisms and susceptibility to cancer*. Lyon, France: International Agency for Research on Cancer, 1999:23–36.
 93. Phelan MC, Pellock JM, Nance WE. Discordant expression of fetal hydantoin syndrome in heteroparental dizygotic twins. *N Engl J Med* 1982;307:99–101.
 94. Ellsworth DL, Coady SA, Chen W, et al. Influence of the β_2 -adrenergic receptor Arg16Gly polymorphism on longitudinal changes in obesity from childhood through young adulthood in a biracial cohort: the Bogalusa Heart Study. *Int J Obes Relat Metab Disord* 2002;26:928–37.
 95. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999;28:1–9.
 96. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;140:771–8.
 97. Egger M, Schneider M, Smith GD. Meta-analysis spurious precision? Meta analysis of observational studies. *BMJ* 1998;316:140–4.
 98. Doll R. The use of meta-analysis in epidemiology: diet and cancers of the breast and colon. *Nutr Rev* 1994;52:233–7.
 99. Ioannidis JPA, Rosenberg PS, Goedert JJ, et al. 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* 2001;135:782–95.
 100. Steinberg KK, Smith SJ, Stroup DF, et al. 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. *Am J Epidemiol* 1998;145:917–25.
 101. Ioannidis JPA, Rosenberg PS, Goedert JJ, et al. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002;156:204–10.