

# Assessment of cumulative evidence on genetic associations: interim guidelines

John P A Ioannidis,<sup>1–3\*</sup> Paolo Boffetta,<sup>4</sup> Julian Little,<sup>5</sup> Thomas R O'Brien,<sup>6</sup> Andre G Uitterlinden,<sup>7</sup> Paolo Vineis,<sup>8</sup> David J Balding,<sup>8</sup> Anand Chokkalingam,<sup>9</sup> Siobhan M Dolan,<sup>10</sup> W Dana Flanders,<sup>11</sup> Julian P T Higgins,<sup>12</sup> Mark I McCarthy,<sup>13,14</sup> David H McDermott,<sup>15</sup> Grier P Page,<sup>16</sup> Timothy R Rebbeck,<sup>17</sup> Daniela Seminara<sup>18</sup> and Muin J Khoury<sup>19</sup>

**Accepted** 9 July 2007

Established guidelines for causal inference in epidemiological studies may be inappropriate for genetic associations. A consensus process was used to develop guidance criteria for assessing cumulative epidemiologic evidence in genetic associations. A proposed semi-quantitative index assigns three levels for the amount of evidence, extent of replication, and protection from bias, and also generates a composite assessment of 'strong', 'moderate' or 'weak' epidemiological credibility. In addition, we discuss how additional input and guidance can be derived from biological data. Future empirical research and consensus development are needed to develop an integrated model for combining epidemiological and biological evidence in the rapidly evolving field of investigation of genetic factors.

**Keywords** Epidemiologic methods, genetics, genomics, causality, evidence

Assessing the credibility of the proposed relationships between human genetic variation and various diseases and traits is a rapidly growing challenge. Here, we use the term 'credibility' to refer to the likelihood that an association exists after some evidence has been accumulated. However, evidence is continuously evolving. Currently over 6000 original articles reporting genetic epidemiology results are published annually.<sup>1</sup> This field started with assessments of small numbers of 'candidate'

genetic variants, but there are now increasing numbers of genome wide association studies (GWAs) that seek to discover novel genetic risk factors by testing several hundred thousand single nucleotide polymorphisms (SNPs) per participant. In the near future, whole genome sequencing data will provide information on millions of variants per individual. With the application of such massive genomic testing platforms to large case-control and cohort studies<sup>2</sup>, the amount of information per study and overall is increasing very rapidly. Reflecting this increase, approximately 500 Human Genome Epidemiology

<sup>1</sup> Clinical and Molecular Epidemiology Unit, University of Ioannina School of Medicine, Ioannina 45110, Greece.

<sup>2</sup> Biomedical Research Institute, Foundation for Research and Technology – Hellas, Ioannina 45110, Greece.

<sup>3</sup> Department of Medicine, Tufts University School of Medicine, Boston MA 02111, USA.

<sup>4</sup> International Agency for Research on Cancer, Lyon 69008, France.

<sup>5</sup> Department of Epidemiology and Community Medicine, Canada Research Chair in Human Genome Epidemiology, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada.

<sup>6</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville MD 02982, USA.

<sup>7</sup> Departments of Internal Medicine and Epidemiology & Biostatistics, Erasmus MC, Rotterdam 3000 CA, The Netherlands.

<sup>8</sup> Department of Epidemiology and Public Health, Imperial College, St Mary's Campus, London, W2 1PG London, UK.

<sup>9</sup> School of Public Health, University of California, Berkeley, CA 94707, USA.

<sup>10</sup> Department of Obstetrics and Gynecology and Women's Health, Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, NY 10461, USA.

<sup>11</sup> Emory University, Rollins School of Public Health, Department of Epidemiology, 1518 Clifton Rd, Atlanta, GA 30327, USA.

<sup>12</sup> MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 0SR, UK.

<sup>13</sup> Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Headington, Oxford, OX3 7LJ, UK.

<sup>14</sup> Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, Oxford, OX3 7BM, UK.

<sup>15</sup> Laboratory of Molecular Immunology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892, USA.

<sup>16</sup> Department of Biostatistics, University of Alabama at Birmingham School of Public Health, Birmingham, AL 35294, USA.

<sup>17</sup> Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, USA.

<sup>18</sup> Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Rockville, MD 20892, USA.

<sup>19</sup> National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA, USA.

\*Corresponding author. Professor J Ioannidis, Clinical and Molecular Epidemiology Unit, University of Ioannina School of Medicine, Ioannina 45110, Greece. E-mail: jioannid@cc.uoi.gr

(HuGE) reviews and meta-analyses have been published to date ([www.cdc.gov/genomics](http://www.cdc.gov/genomics)). These reviews typically integrate information on one or a few specific gene–disease associations at a time. The unknown extent of unpublished data and the potential biases that may influence the results of single studies threaten the credibility of the literature.<sup>3,4</sup> While some reported genetic associations have been confirmed to be credible, most have been refuted or remain ambiguous.<sup>5</sup> Many fear that the scientific literature has already become flooded with false or misleading information. It is important to develop mechanisms that can summarize and evaluate the current status of evidence of whole fields in genetic epidemiology. This undertaking requires regularly updated synopses of all adequate association studies on a particular disease or phenotype<sup>6</sup> based on widely accepted criteria for assessment of the cumulative evidence. Synopses would also be useful for planning future studies and eventually understanding the translational potential of genetic information for clinical and public health purposes.

There is special enthusiasm about the potential power of genomics to define the etiology of disease and phenotypes, because associations that arise from genetic epidemiology studies may be less likely to be confounded or biased than other types of epidemiologic studies. Guidelines for inferring causation from observational studies of associations between exposures and disease were proposed<sup>7,8</sup> in the 1960s and subsequently modified for various fields of epidemiology,<sup>9–11</sup> but these guidelines are not appropriate for the scale or specific challenges now being encountered in genetic epidemiology. Bradford Hill himself did not wish his nine items to be interpreted as strict criteria and for genetic epidemiology, many of these items are either irrelevant or problematic. *Temporality* is irrelevant for genetic factors fixed at birth and *experimental support* through randomization is impossible. *Analogy* and *coherence* ‘with generally known facts of the natural history and biology of the disease’ are impossible to use meaningfully yet, as we are still scratching the surface of complex trait biology. Genetic variants may exhibit *specificity* for highly circumscribed phenotypes but may also have pluripotent effects on multiple phenotypes. *Biological plausibility* carries considerable uncertainty, as we discuss subsequently. An association that fits to an additive dose–response model (*biological gradient*) is not clearly more reliable than one that follows a recessive model. *Strength* may still be relevant, but effects observed for most emerging associations are small. Finally, *consistency* needs to be redefined in the context of the genomic era.

In all, as the vast majority of associations currently have small effect sizes their credibility may largely depend on the success of control for errors and biases. Several journals have policies or instructions regarding how they wish genetic associations to be supported for publication, but these refer to single studies and largely aim at screening the mass of data for publication in these specific journals.<sup>12–18</sup> The complete credibility picture should include evidence on data regardless of their priority for publication. Moreover, some generic themes on credibility in molecular epidemiology have been discussed,<sup>19</sup> but specifically operationalized criteria for genetic associations require further consensus across methodologists.<sup>19</sup>

Following the initial meeting of the Network of Networks,<sup>20</sup> a Human Genome Epidemiology Network (HuGENet) Working

Group on the Assessment of Cumulative Evidence was established. This group and a panel of experts in various relevant disciplines met in Venice, Italy, on November 9–10, 2006 to discuss these issues and draft guidelines. The panel discussed and assessed existing assessment schemes from other fields, experiences of developing synopses of cumulative evidence on diverse diseases, experiences of linking genetic epidemiology with biological plausibility in acute, infectious and chronic diseases, the new framework for causal inference in the genome era, and methods for the efficient assessment of quantity and quality of evidence.

In this report, we summarize the issues concerning the epidemiological assessment of cumulative evidence in genetic epidemiology for which consensus was attained as well as others for which substantial lack of agreement still remains. We propose interim guidelines for assessing the credibility of genetic epidemiological evidence; and additional points to consider on biological plausibility and the clinical (and public health) importance and relevance. We view this as a preliminary proposal that is likely to benefit from empirical research and scrutiny from other scientists.

## Epidemiological credibility

We consider the assessment of reported associations between a specific phenotype and a specific genetic variant. The variant could be a SNP, so that the data for each individual corresponds to which of three genotypes they carry, or it could be a haplotype or another type of polymorphism, such as a copy-number polymorphism. A more complex evidence assessment task arises when different studies use different data, for example distinct but tightly-linked SNPs, or (subtly) different phenotype definitions (Table 1).

The proposed assessment may be performed by any interested researcher, as well as networks or consortia, and we encourage its incorporation in systematic reviews and meta-analyses (including synthesized replication data from genome-wide association studies), where there is an opportunity for a systematic view of the available evidence. Such a systematic view may be a prerequisite for informed appraisal of the amount, replication and protection from bias. It is also possible that for a specific association, there may exist scattered studies, meta-analyses of published data and prospective standardized analyses by a consortium. In each case, one has to focus on the highest level of available evidence.

### Amount of evidence

Credibility is enhanced by a large amount of consistent evidence; cumulatively, evidence may be large by virtue of many studies, or by a more modest number of large studies. A large amount of evidence is required to ensure adequate power for detecting an association (if one is present) and reaching more stringent levels of statistical significance, or lower false-discovery rate.<sup>21</sup> Large sample sizes tend to also decrease the uncertainty in the magnitude of the observed genetic effect.<sup>22</sup> Larger studies may sometimes be performed by more experienced groups and may also be less likely to be affected by selective reporting biases than smaller studies, but this is not guaranteed.<sup>23</sup>

**Table 1** Considerations for epidemiologic credibility in the assessment of cumulative evidence on genetic associations

Criteria	Categories	Proposed operationalization
Amount of evidence	A: Large-scale evidence B: Moderate amount of evidence C: Little evidence	Thresholds may be defined based on sample size, power or false-discovery rate considerations. The frequency of the genetic variant of interest should be accounted for. As a simple rule, we suggest that category A requires a sample size over 1000 (total number in cases and controls assuming 1:1 ratio) evaluated in the least common genetic group of interest; B corresponds to a sample size of 100–1000 evaluated in this group, and C corresponds to a sample size <100 evaluated in this group (see ‘Discussion’ section in the text and Table 2 for further elaboration). <sup>a</sup>
Replication	A: Extensive replication including at least one well-conducted meta-analysis with little between-study inconsistency B: Well-conducted meta-analysis with some methodological limitations or moderate between-study inconsistency C: No association; no independent replication; failed replication; scattered studies; flawed meta-analysis or large inconsistency	Between-study inconsistency entails statistical considerations (e.g. defined by metrics such as $I^2$ , where values of 50% and above are considered large and values of 25–50% are considered moderate inconsistency) and also epidemiological considerations for the similarity/standardization or at least harmonization of phenotyping, genotyping and analytical models across studies. See ‘Discussion’ section in the text for the threshold (statistical or others) required for claiming replication under different circumstances (e.g. with or without including the discovery data in situations with massive testing of polymorphisms).
Protection from bias	A: Bias, if at all present, could affect the magnitude but probably not the presence of the association B: No obvious bias that may affect the presence of the association but there is considerable missing information on the generation of evidence C: Considerable potential for or demonstrable bias that can affect even the presence or absence of the association	A prerequisite for A is that the bias due to phenotype measurement, genotype measurement, confounding (population stratification) and selective reporting (for meta-analyses) can be appraised as not being high (as shown in detail in Table 3) plus there is no other demonstrable bias in any other aspect of the design, analysis or accumulation of the evidence that could invalidate the presence of the proposed association. In category B, although no strong biases are visible, there is no such assurance that major sources of bias have been minimized or accounted for because information is missing on how phenotyping, genotyping and confounding have been handled. Given that occult bias can never be ruled out completely, note that even in category A, we use the qualifier ‘probably’.

<sup>a</sup>For example, if the association pertains to the presence of homozygosity for a common variant and if the frequency of homozygosity is 3%, then category A amount of evidence requires over 30 000 subjects and category B between 3000 and 30 000. The sample size refers to subjects when genotype contrasts are used, and to alleles when alleles are contrasted.

The effective amount of evidence depends on factors that influence the power of the study to detect a true association, namely the total sample size, the underlying genetic model of the association, the frequency of the genetic variant of interest and the magnitude of the association. Even though magnitudes are likely to differ for different associations, most effect sizes identified to-date cluster in a narrow range of odds ratios (ORs) and are not reliably distinguishable from each other. Therefore, considerations of the amount of evidence that take also into account the exact magnitude of the effect are not very reliable (although the observed effect may be important in relationship to protection from bias, as discussed subsequently). The power of a study is driven largely by the sample size of the smallest genetic group of those contrasted. Thus, the size of the smallest genetic group of those contrasted ( $n_{\text{minor}}$ ) is a simple, but convenient, approach for assessing the amount of evidence (irrespective of whether it shows an association).

With large effect sizes, smaller sample sizes may theoretically suffice to reach extreme levels of statistical significance or Bayes factors, but some of these large effects may be over-estimates. Also, statistical significance may vary several log depending on the genetic contrast, set of data analysed (all, subgroups, different phenotypes), statistical model used, assumptions of the model (e.g. incorporating or not modest deviations in Hardy–Weinberg equilibrium), and—for

meta-analyses—the broader assumptions about the synthesis of the evidence (fixed effects, random effects, fully Bayesian models). The threshold of statistical significance or Bayes factor required to claim replication for an association is discussed in the Replication section subsequently.

The choice of an  $n_{\text{minor}}$  threshold is unavoidably arbitrary. For purely operational purposes, we propose a threshold of  $n_{\text{minor}}=1000$  to separate large-scale from moderate evidence and 10-times less ( $n_{\text{minor}}=100$ ) to separate moderate from little evidence. Table 2 shows what  $n_{\text{minor}}=1000$  means in diverse circumstances of different effects (ORs) and frequencies of the minor genetic group  $f_{\text{minor}}$ . For  $\alpha=0.05$ , as might be required in the independent replication of a proposed association (excluding the discovery data that may be based on massive testing of thousands of polymorphisms), the power remains high for a wide range of effect sizes (OR 1.2–5) and frequencies  $f_{\text{minor}}$  (0.01–0.50). For an OR as low as 1.10 (probably close to the limit of discriminating ability for observational epidemiology), power is about 18–32%. Conversely, with  $n_{\text{minor}}=500$ , there is major loss of power for odds ratios of 1.1–1.3, while with  $n_{\text{minor}}=1500$ , the gain in power for odds ratios in the whole range of 1.1–5.0 is relatively negligible compared with  $n_{\text{minor}}=1000$  (data not shown). We should note however that for  $n_{\text{minor}}=1000$ , for  $\alpha=10^{-7}$  (the genome-wide significance level for many current study designs,) power is steeply eroded

**Table 2** Power calculations for associations with  $n_{\text{minor}} = 1000$  for various ORs and various frequencies of the minor genetic group ( $f_{\text{minor}}$ )<sup>a</sup>

OR	$f_{\text{minor}}$	Power for $\alpha = 0.05$	Power for $\alpha = 10^{-7}$
1.10	0.01	0.32	<0.001
1.20	0.01	0.82	0.007
1.30	0.01	0.98	0.12
1.50	0.01	1.00	0.83
2.00	0.01	1.00	1.00
5.00	0.01	1.00	1.00
1.10	0.05	0.31	<0.001
1.20	0.05	0.80	0.006
1.30	0.05	0.98	0.09
1.50	0.05	1.00	0.78
2.00	0.05	1.00	1.00
5.00	0.05	1.00	1.00
1.10	0.10	0.30	<0.001
1.20	0.10	0.78	0.005
1.30	0.10	0.97	0.74
1.50	0.10	1.00	1.00
2.00	0.10	1.00	1.00
5.00	0.10	1.00	1.00
1.10	0.25	0.25	<0.001
1.20	0.25	0.69	0.002
1.30	0.25	0.94	0.04
1.50	0.25	1.00	0.52
2.00	0.25	1.00	1.00
5.00	0.25	1.00	1.00
1.10	0.50	0.18	<0.001
1.20	0.50	0.51	<0.001
1.30	0.50	0.81	0.006
1.50	0.50	0.99	0.15
2.00	0.50	1.00	0.96
5.00	0.50	1.00	1.00

<sup>a</sup>All calculations assume the same number of cases and controls; results are relatively robust to modest deviations in the allocation ratio. The minor genetic group is the smallest of the two groups contrasted and may have been selected based on genotype or allele considerations.

for ORs below 1.5 (Table 2). Thus, if the discovery data (testing many thousands of polymorphisms) are included in the amount of evidence, much higher sample sizes should be required to claim large-scale evidence.

### Replication

Credibility is enhanced when an association is found in different studies (replication) and when the magnitude is consistent across different study populations (homogeneity). Consistency can be assessed with statistical tests or measures of heterogeneity,<sup>24</sup> but qualitative aspects should also be considered.

The threshold of 'replication' is a matter of considerable debate.<sup>4,25,26</sup> The reader is referred to a recent excellent workshop for more detailed coverage.<sup>25</sup> In brief, with genome-wide testing of hundreds of thousands of polymorphisms, many would argue that  $P$ -values at least  $< 10^{-7}$  are

needed for main effects to be considered to even set strong candidacy for further replication (provided biases can be excluded, as discussed subsequently). Nevertheless, more conventional thresholds of statistical significance (e.g.  $P < 0.05$ ) may still be used appropriately for the replication phase of a proposed association, if the discovery data are excluded, the replication is limited to a specific polymorphism and a specific model of association is analysed. However, massive genome-wide testing may be performed increasingly by several teams on the same phenotype and the combination of these data may require again very stringent levels of statistical significance for associations to be considered credible in joint analyses. A Bayesian approach may be also used, with similar considerations and specification of prior probabilities (in the most simple approach Bayes factor =  $\alpha/\text{power}^4$ )

Lack of replication or between-study heterogeneity may signal underlying errors and biases, including genotyping error, phenotype misclassification, population stratification and selective reporting biases.<sup>22,27–29</sup> Lack of replication in a different study population does not necessarily refute the original reported association. It may reflect different linkage disequilibrium patterns across different populations, when the studied genetic variant is not causal, or it may be due to population-specific gene–gene epistasis, or gene–environment interactions.<sup>30,31</sup> Lack of precise comparability of the phenotype of interest may also lead to inconsistent results across studies. Thus, heterogeneity may point to genuine diversity in the genetic effect. Conversely, lack of heterogeneity does not exclude bias and may even reflect lack of independence in the replication process.<sup>32</sup>

Independence among studies in the replication process can occur to different degrees. Independence is enhanced when different teams of investigators test a proposed association separately using different samples drawn from distinct populations. In this respect, simply splitting a single population sample or the investigation of samples from different studies by one team of investigators increases the risk that the same latent biases (including, but not limited to population stratification and systematic genotyping errors) may operate across seemingly replicate assessments.<sup>33</sup> A split sample approach also reduces power, unless the two parts are then reassembled in a joint analysis.<sup>34</sup> Extensive replication by totally independent, even competing, teams of investigators may provide optimal evidence of the credibility of a putative genetic association.

### Protection from bias

Bias may be caused by factors that lead to systematic deviations from the true effect of a genetic association. Biases may operate at the level of a single study, a collection of studies (e.g. meta-analysis), or a research field at large. They may arise in the study design (including participant recruitment, retrospective or prospective collection of DNA samples, and method of gathering information on phenotypes, exposures and covariates), DNA extraction method, production of genotype data, raw data management, data processing, data analysis, reporting of analyses, integration of studies through meta-analyses or integration of meta-analyses into field synopses.<sup>28,29,35–39</sup>

Two potential sources of bias are particularly widely recognized in genetic association analyses: population stratification and genotyping error. The magnitude of population

**Table 3** Typical biases and their typical impact on associations depending on the status of the evidence

Biases	Status of the evidence	Likelihood of bias to invalidate an observed association		
		Small OR <1.15	Typical OR 1.15–1.8	Large OR >1.8
Bias in phenotype definition	Not reported what was done	Unknown	Unknown	Unknown
	Unclear phenotype definitions	Possible/High	Possible/High	Possible/High
	Clear widely agreed definitions of phenotypes	Low/None	Low/None	Low/None
	Efforts for retrospective harmonization	Possible/High	Low	Low/None
	Prospective standardization of phenotypes	Low/None	Low/None	Low/None
Bias in genotyping	Not reported what was done	Unknown	Unknown	Unknown
	No quality control checks	Possible/High	Low	Low
	Appropriate quality control checks	Low	Low	Low/None
Population stratification	Not reported what was done	Unknown	Unknown	Unknown
	Nothing done <sup>a</sup>	Possible/High	Possible/High	Possible/High
	Same descent group <sup>b</sup>	Possible/High	Low	Low/None
	Adjustment for reported descent	Possible/High	Low	Low/None
	Family-based design	Low/None	Low/None	Low/None
	Genomic control, PCA or similar method	Low/None	Low/None	Low/None
Selective reporting biases	Meta-analysis of published data	Possible/High	Possible	Possible
	Retrospective efforts to include unpublished data	Possible/High	Possible	Possible
	Meta-analysis within consortium	Low/None	Low/None	Low/None

Category decreases from A to B, if the ‘Unknown’ are considered to be a major issue for the appraisal of the evidence. Any ‘Possible/high’ item confers category C status. ‘Possible’ (selective reporting biases for non-consortium/prospective meta-analysis) does not necessarily decrease the category grade (from A to C); this may need to be appraised separately in each field and may be facilitated by using tests for selective reporting biases (tests for small-study effects and excess of significant studies), although probably no test has high sensitivity and specificity for such biases. Clear demonstrable biases in other aspects of the design, conduct and analysis of the evidence (besides the four aspects considered in this table) also result in shift to category C for protection from bias.

<sup>a</sup>Including groups of clearly different descent without consideration to this diversity.

<sup>b</sup>The ethnic population structure may need to be considered also on a case-by-case basis.

OR, odds ratio; PCA, principal component analysis.

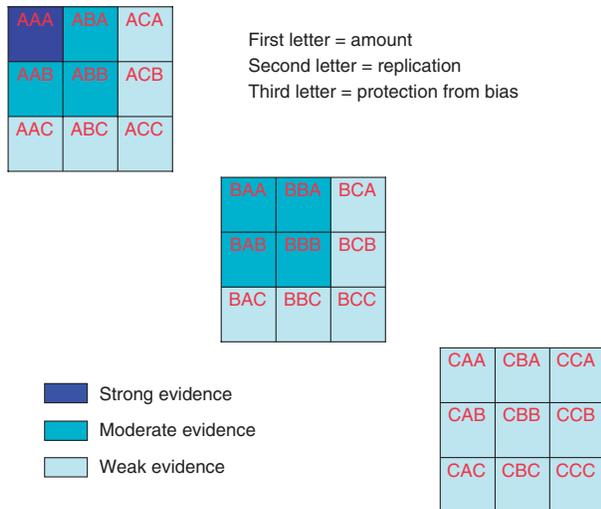
stratification effects remains a debated concern: they are expected to be small in well-designed studies, but subtle effects are always possible and can become relatively important when large sample sizes permit the investigation of small true effect sizes. Several statistical procedures are available to adjust for population structure effects, such as genomic control<sup>40</sup> and methods based on principal components analysis.<sup>41</sup>

Because cases and controls are typically ascertained separately, systematic genotyping error can have differential impact on cases and controls even when genotyping is performed blind to case–control status.<sup>28</sup> Methods to assess genotyping quality include blind replicate genotyping of some individuals, replicate studies using different genotyping platforms and testing for Hardy–Weinberg equilibrium, although this last method is not specific. A high rate of missing genotypes is suggestive of poor data quality, but a very low rate of missing genotypes can reflect overly-permissive genotype calling and is not a guarantee of high-quality genotype calls. Genotyping quality control methods may include analysis of missing data, e.g. tests of association between missing status and case–control phenotype or excess homozygosity.

Different biases may sometimes mitigate or magnify each other, according to whether or not they act in opposing directions. Adequate protection from bias can be evaluated when each level of evidence contributing to the putative association can be scrutinized, and the likelihood, direction and magnitude of bias that may affect the major conclusions about the proposed genetic association can be assessed. A prerequisite

for this kind of assessment is the availability of information concerning what was done at each step in the generation and accumulation of the evidence. Increased transparency can be achieved if detailed databases and protocols are publicly available.<sup>42,43</sup> and guidance on the reporting of genetic association studies in the literature is established and adopted in the field.<sup>44,45</sup> However, even if single studies are conducted and reported without bias and with full transparency, the cumulative evidence may still be biased if availability of information is driven by selective reporting or other publication biases.<sup>46,47</sup> Such selection biases can be reduced by the establishment of consortia of multiple teams that have explicit policies of analysing all eligible data from all participating teams.<sup>48</sup> Another approach is to encourage journals and investigators to publish high-quality null results.<sup>49</sup>

For practical purposes, a major decision in the proposed categorization is to distinguish biases that can affect only the magnitude of an association, from those that can invalidate the association. Table 3 lists some common biases (affecting single studies or meta-analyses of many studies) and whether they are likely to have such a major impact under different circumstances, where efforts are made or not to control for them. Whether bias can invalidate an association depends implicitly also on the magnitude of the association, e.g. the OR. Bias is more likely to create spurious small effects, although totally uncontrolled, major bias can also generate large effects. We, thus, also categorize the protection from bias for associations based on the observed effects. We consider that



**Figure 1** Categories for the credibility of cumulative epidemiological evidence. The three letters correspond (in order) to amount of evidence, replication and protection from bias. Evidence is categorized as strong, when there is A for all three items, and is categorized as weak when there is a C for any of the three items. All other combinations are categorized as moderate

investigators should assess biases in the four major aspects of a genetic association: phenotype, genotype, population stratification and (for meta-analyses) selective reporting. These cover the two variables involved in the association, study-specific confounding and field-wide bias. However, as we discussed earlier, bias can lurk at any other step in the process, but we suggest that unless bias in these steps is demonstrable, the uncertainty about our ability to probe in detail all other biases should not affect the practical categorization. Given that unknown bias can never be ruled out completely, note that even in category A we use the qualifier ‘probably’.

### Combination of criteria—suggested guidance and examples

Merging all considerations into a common credibility grading scheme is not straightforward. Figure 1 is a preliminary proposal for such a scheme for epidemiological credibility using three categories: weak, moderate and strong cumulative evidence for an association. However, it should be recognized that overall grading of epidemiological evidence has been difficult even in relatively straightforward questions, such as the literature on the effectiveness of medical interventions.<sup>50,51</sup> For example, there may not be consensus on whether there could be some further sub-categorization of evidence, e.g., splitting the ‘strong’ category into ‘very strong’ (e.g. HLA in type 1 diabetes) and merely ‘strong’; or splitting the ‘moderate’ category into two sub-categories.

## Examples of application of the epidemiological credibility criteria to specific genetic associations

We show examples of the application of this scheme to recently proposed associations in age-related macular degeneration and

obesity. The examples also show how credibility can change over time for the same association and how credibility may vary, as different phenotypes and study populations are studied for the same genetic variant.

### Example 1A

The association between the *CFH* Y420H variant and age-related macular degeneration in populations of European descent was identified by a genome-wide association study.<sup>52</sup> A well-conducted meta-analysis<sup>53</sup> of 11 studies ( $n=8991$ ) shows a summary OR of 2.49 and 6.15 for heterozygotes, and homozygotes, respectively, without any clear between-study inconsistency in effect sizes or other heterogeneity among populations of European descent. The association was also replicated in a subsequent publication from the large Rotterdam cohort ( $n=5681$ ).<sup>54</sup> The H allele is very common in Caucasians (e.g. 36.2%; i.e.  $n=4116$  in the Rotterdam cohort alone) and therefore, the evidence easily passes the  $n=1000$  threshold for category A of amount of evidence. There is no demonstrable inconsistency across studies, therefore the replication category is also A. Finally, the accumulated large-scale evidence is transparent enough and meticulous to give reasonable assurance that there is adequate protection from bias (category A). The overall scheme is thus AAA, which results in a characterization as ‘strong’ evidence.

### Example 1B

On the same Y420H variant, several studies on Asian populations find no significant association with age-related macular degeneration.<sup>55,56</sup> Asian populations have a different predominant form for age-related macular degeneration compared with European populations (wet vs dry phenotype). The Y420H variant is uncommon (~3%) in Asians, and all studies are underpowered to find the OR seen in European populations. In all six studies combined, the total frequency of the minor allele is less than 1000, thus amount of evidence category is B. Replication category is C (scattered studies without meta-analysis). Protection from bias is B, since several aspects in the reporting of these scattered studies are not fully transparent and thus considerable bias cannot be excluded. The overall schema is thus BCB, which results in a characterization as ‘weak’ evidence.

### Example 2A

The association between rs7566605 (10 kb upstream of the transcription start site of *INSIG2*) and obesity was found in a genome-wide association study and it was replicated in a recessive genetic model in another three of four populations in the same publication in *Science*.<sup>57</sup> Excluding the discovery (genome-wide association screen) data, at the time of the *Science* publication, the evidence from case-control designs pertained to 9881 genotyped people, and the frequency of CC homozygote (the smallest genetic group) was  $n=1040$ , with some additional consistent evidence from a family-based study ( $n=368$ , of which  $n=52$  had the CC genotype). Therefore, the amount of evidence category is A (more than 1000 subjects genotyped in the smallest genetic group of those compared). The replication category is B, because one of the populations

found no evidence of association, actually with a trend in the opposite direction, and thus there is moderate between-study inconsistency. The protection from bias category is A: this was a well-conducted investigation with transparent reporting of the designs of the constituent studies and prospective meta-analysis (apparently no selective reporting). The overall schema is thus ABA, which results in characterization as ‘moderate evidence’.

### Example 2B

Several months later, a series of Technical Reports in *Science*<sup>58–60</sup> presented evidence from three different teams of investigators (total of seven study populations) that found absolutely no association between the rs7566605 variant and obesity. The new evidence pertained to over 21 000 genotyped individuals. Based on the newer update, the amount of evidence is A (more than 1000), replication is C (failed replication), and protection from bias is still A. The overall schema is thus ACA, which results in characterization as ‘weak evidence’.

These examples also illustrate that the bar that we set for ‘strong’ evidence is quite high, and some further calibration work would be useful. Moreover, even for an association that is set at ‘strong evidence’, further work may lead to a change in grading. Illustratively, for example 1A above, it is increasingly recognized that there is extensive disequilibrium in the implicated *CFH* region and it is not clear that Y420H is truly causative or the only causative allele in the region. Conversely, pursuing further associations that are likely to be false (‘weak’ credibility) may be a low priority when there are many associations with higher credibility to pursue. However, we acknowledge that the threshold of interest may vary between researchers who try to find associations and explain what they mean and those who try to make use of this knowledge for practical purposes.

## Stages of accumulation of evidence across diverse fields

Different disease content areas of genetic epidemiology may have attained different stages in the accumulation of evidence. For most diseases, the currently available published evidence consists of fewer than 100 studies of mostly single-gene, single-disease assessments, one or a few meta-analyses, if any, and no strong established consortia of investigators. Other fields may already have many thousands of published (and unpublished) studies, many meta-analyses of group-level data and even several comprehensive rigorous consortia of investigators utilizing the latest genome-wide technologies and combining data (see selected examples in Table 4).

A detailed synopsis of the cumulative evidence can readily be performed at the level of each single study in ‘early evidence’ fields such as pre-term birth and childhood leukaemia (Table 4). For such fields, however, the cumulative evidence is likely to be rated as insufficient until substantially more data become available. Such a synopsis is mostly helpful not so much to tell us exactly how insufficient the evidence is, but to create a comprehensive basis upon which the field may expand towards more credible evidence. This work can facilitate the

conduct of meta-analyses and HuGE reviews, the creation of consortia, and improved organization of research in the field.<sup>48</sup>

Conversely, for fields where more evidence already exists (e.g. type 2 diabetes, osteoporosis and cancer; Table 5), it may be appropriate to ignore scattered studies of small sample size and doubtful quality. Focusing on the stronger parts of the evidence may suffice, e.g. data for associations where several studies exist,<sup>61</sup> or even better evidence exists from large-scale studies, well-conducted meta-analyses and consortia with documented adequate protection from bias.<sup>62–64</sup> Synopses of the literature may thus be a way to continuously raise the standards of research in specific fields of human genome epidemiology.

An additional important issue is to quantify how credible associations are likely to be. Even when there is strong supporting evidence, it may still be difficult to assess whether these associations are 60%, 80% or >99% likely to be true. One empirical possibility is to continue testing in ever larger and less-biased studies. Stopping rules in this field are an intriguing consideration and need more discussion. In theory, replication can be continued even for associations that have reached the level of being assigned to strong evidence. Such open-ended replication<sup>65</sup> is not ethically prohibited, and it is unlikely to be very expensive from the perspective of laboratory analysis, even if very large sample sizes are contemplated, since relatively few variants will likely reach the point of such testing. An additional reason for continuing replication is the expected heterogeneity of phenotype for most complex diseases and the need to consider gene–gene and gene–environment interactions. The main obstacle to this approach is the availability of samples. By the time an association has reached the category of strong evidence, then most if not all well-conducted studies and consortia may have tested it already. One option is to anticipate that such associations may be prospectively tested in the very large biobanks<sup>66</sup> and are expected to accumulate large enough sample sizes for common disease events in the next few decades.

## Evolution of evidence

Any assessment of cumulative evidence is only temporary and needs to be continuously refined, as new data are gathered.<sup>67</sup> The evolution of the evidence over time may be of particular interest. Genuine associations tested under protection from bias are expected to fluctuate over time due to random error, and may show some early diminution.<sup>68</sup> However, the association effects should eventually stabilize and remain virtually unaltered with further replication efforts. Associations due to bias are expected to be discredited, sooner or later, with further replication efforts. Finally, for some associations, genuine variability in the existence and magnitude of associations across populations may exist. Of particular note, in this respect, are the well-known changes over time of exposure patterns, such as diet and physical activity. These can influence gene–environment interactions and thereby, hide and/or enhance certain genetic associations. Thus, even with adequate protection from bias the strength of these associations may vary across successive replications. In this case, the population characteristics of each study sample and the biological support

**Table 4** Variation in the volume of human genome epidemiology evidence for selected diseases, 2001–6<sup>a</sup>

Disease	Papers	Genes studied	Meta-analyses (HuGE reviews)	Investigators	Consortia
Type 2 diabetes	2246	555	56 (7)	9320	Established
Breast cancer	1020	275	24 (3)	4499	Established
Osteoporosis <sup>b</sup>	503	128	11 (2)	2041	Established
Pre-term birth <sup>c</sup>	176	112	1 (1)	855	Emerging
Childhood leukaemia	102	76	1 (1)	668	Emerging

<sup>a</sup>Data from the HuGE published literature database run November 27, 2006; does not include data on genome-wide associations that started appearing for some of these phenotypes (e.g. type 2 diabetes or breast cancer) in early 2007.

<sup>b</sup>Includes studies on bone mineral density.

<sup>c</sup>Includes studies on gestational age.

**Table 5** Considerations for assessment of clinical and public health relevance and importance of genetic associations**Magnitude of effect**

Effect size  
Frequency of genetic variant in population

**Clinical and public health importance**

Type of phenotype: biological, endophenotype and hard clinical outcome  
Disease burden: incidence, severity and mortality  
Interaction with identified modifiable environmental exposures  
Potential to prevent disease through intervention (e.g. through Mendelian randomization insights)

for these associations should be scrutinized again and redefined in an effort to understand the sources of heterogeneity.

Given that the evolution of observed effects is likely to happen in a narrow range of effect sizes, it may sometimes be difficult to appreciate which of the three patterns is operating. Learning to live with some uncertainty is thus unavoidable, but at a minimum, we should be able to decipher the more reliable and consistent associations from the less believable ones.

**Biological evidence**

By 'biological' evidence, we mean evidence as to the specific function of a variant or associated gene, which may make it a plausible candidate for association with the phenotype under study. It includes whether a variant generates a synonymous, non-synonymous, or nonsense amino acid change, or is located in an exon, intron, splice site or regulatory region, as well as information about conservation across species. Biological evidence may also be gained from gene knock-out experiments in model species, or gene expression microarray experiments. Such experiments may be conducted specifically to examine a postulated association, but substantial information about known gene function is also recorded routinely in genome annotation databases. The assessment of biological plausibility of genetic associations is complex and a variety of sources of relevant evidence is available. Some are limited to the study of specific genes or diseases, while others may be more broadly applicable.

In appraising evidence for biological plausibility, the strength and consistency of biological effects, the amount of data, the number of different lines of corroborating biological evidence,

and the relevance of the biological system to the phenotype may be considered. For experimental data, additional points to consider are the extent of replication (i.e. using the same type of experiments vs. approximate corroboration with different experimental methods), and whether there was protection from bias.<sup>69</sup> Given the diversity of potential biological evidence, it is difficult to generalize on the relative importance of each piece of experimental data in each disease and situation. Empirical data on the concordance of biological and epidemiological evidence may be useful,<sup>70–72</sup> although it may be difficult to construct a consistent algorithm that applies across different complex diseases.

Unfortunately, given an imperfect knowledge base, the use of this experimental and non-experimental evidence to support or refute an association has been often misleading. For example, associations have been proposed with considerable support for synonymous SNPs, for SNPs in introns and for SNPs with no clear functional role despite many attempts to elicit some functional data. While part of this puzzle may be explained by linkage disequilibrium of the identified SNPs with the culprit ones, the biological relevance may be more difficult to ascertain than has been supposed. The panel members do not advocate that these lines of evidence be ignored. However, most members felt that for common variants, non-epidemiological evidence alone is unlikely to be sufficient to make an association highly credible, if it is not already highly credible on epidemiological grounds.

**Clinical and public health importance**

In addition to epidemiological credibility and biological plausibility, it is important to assess the potential public health impact and clinical relevance of genetic associations, but only after the credibility of an association reaches a high level (practical importance is unrelated to causality). The group did not formulate a specific assessment scheme but proposed items to consider in assessing clinical (and wider public health) relevance and importance (as shown in Table 5). No categories are provided for each item, because the empirical evidence-base for the assessment of clinical relevance and importance of specific associations is limited.

The attributable fraction due to a specific variant(s) depends on the effect size and the frequency of the variant(s) of interest; and is a direct measure of the population impact. For quantitative traits, the proportion of variance explained may be

considered. However, the cumulative impact of many variants on the same phenotype needs further empirical study with large-scale data.<sup>73</sup>

For many common genetic variants that are involved in regulation of transcription and protein levels,<sup>74</sup> the end-effects on clinically meaningful outcomes may be minimal or absent. Even for clinical phenotypes, the severity of disease may vary substantially.<sup>75</sup> The importance of an association may also be related to whether there are identified interactions with modifiable strong environmental risk factors or whether it can point to modifiable acquired exposures, e.g. through Mendelian randomization.<sup>76</sup> Deciphering and quantifying interaction effects may be difficult, however, given that misclassification error tends to be larger for environmental exposures than for genotypes.

Effect size requires a special note about whether it also influences the credibility of an association. As discussed earlier, biases more easily generate spurious small effects rather than large effects. From the perspective of gaining insights into disease pathogenesis, any effect, regardless how small, may provide useful information. Moreover, current evidence suggests that, with few notable exceptions, most genetic associations of common genetic variants with common complex diseases have small effect sizes, typically less than 1.3 in the OR scale.<sup>77–83</sup>

Small main effects may be associated with considerably larger effects when considered in gene–gene and gene–environment interactions. Moreover, traditional epidemiology has struggled with the discriminating ability of its analytical tools. Different views have been expressed, ranging from the claim that epidemiology should abandon all efforts to dissect OR below 2,<sup>84</sup> to claims that even ORs below 1.1 are measurable and potentially credible, if the evidence has been otherwise strong.<sup>85</sup> The debate becomes very pertinent for genetic associations, where ORs of 1.1 or smaller may be common for main effects. There is no consensus on whether there is a cap of maximal credibility that cannot be exceeded for small effects, even under the best circumstances. Credibility is important to assess even for small effects that have no clinical importance, since they may still be useful for understanding biology and etiology.

## Combining epidemiologic credibility, biological plausibility and clinical importance

While the panel managed to reach consensus on the grading scheme for epidemiologic evidence (Tables 1 and 3),

the panel did not agree on similarly detailed guidance for rating biological plausibility and clinical relevance. More work is clearly needed to examine if this is feasible.

Weakly credible epidemiological evidence does not merit an in-depth evaluation of biological plausibility or clinical relevance, although it is arguable that if very strong biological plausibility exists, the question would merit more study. Associations with moderate epidemiological credibility deserve more study and additional biological and clinical assessment. Those with strong epidemiological credibility may also require active pursuit of understanding the details of biological pathways and also whether this information can usefully be applied for clinical and public health benefits.

The HuGENet network of investigator networks is conducting pilot studies on a few selected diseases to assess and calibrate the proposed preliminary guidelines. These efforts will also be useful in developing a template for online field synopses on genetic associations that could become part of an updatable encyclopedia on genetic variation and human diseases and to refine the criteria outlined here and their combination in an overall assessment of the evidence. Given that human genome epidemiology is a rapidly moving field, we encourage investigators in different fields to use, experiment and adapt these guidelines for specific diseases. Such an endeavour is essential in making sense of the anticipated explosion of genetic information in the coming years.

## Acknowledgement

This work was partly supported by ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), a network of excellence operating within the European Union 6th Framework Program (Contract No 513943). This research was also supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics and National Institute of Allergy and Infectious Diseases, Laboratory of Molecular Immunology. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

**Conflict of interest:** None declared.

### KEY MESSAGES

- We used a consensus process to develop guidance criteria for assessing cumulative epidemiological evidence in genetic associations.
- The criteria assign three levels for the amount of evidence, extent of replication and protection from bias.
- A composite assessment results in categories of ‘strong’, ‘moderate’ or ‘weak’ epidemiological credibility.

## References

- <sup>1</sup> Lin BK, Clyne M, Walsh M *et al.* Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol* 2006;**164**:1–4.
- <sup>2</sup> Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet* 2006;**2**:e150.
- <sup>3</sup> Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;**2**:e124.
- <sup>4</sup> Wacholder S, Chanock S, Garcia-Closas M *et al.* Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;**96**:434–42.
- <sup>5</sup> Ioannidis JP. Genetic associations: false or true? *Trends Mol Med* 2003;**9**:135–38.
- <sup>6</sup> Anonymous. Embracing risk. *Nat Genet* 2006;**38**:1.
- <sup>7</sup> Surgeon General (Advisory Committee). *Smoking and Health*, 1964.
- <sup>8</sup> Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**58**:295–300.
- <sup>9</sup> Weed DL, Gorelic LS. The practice of causal inference in cancer epidemiology. *Cancer Epidemiol Biomarkers Prev* 1996;**5**:303–11.
- <sup>10</sup> Potischman N, Weed DL. Causal criteria in nutritional epidemiology. *Am J Clin Nutr* 1999;**69**:1309S–14S.
- <sup>11</sup> Coglianò VJ, Baan RA, Straif K *et al.* The science and practice of carcinogen identification and evaluation. *Environ Health Perspect* 2004;**112**:1269–74.
- <sup>12</sup> Weiss ST. Association studies in asthma genetics. *Am J Resp Crit Care Med* 2001;**164**:2014–15.
- <sup>13</sup> Cooper DN, Nussbaum RL, Krawczak M. Proposed guidelines for papers describing DNA polymorphism-disease associations. *Hum Genet* 2002;**110**:208.
- <sup>14</sup> Huizinga TW, Pisetsky DS, Kimberly RP. Associations, populations, and the truth: recommendations for genetic association studies in Arthritis & Rheumatism. *Arthritis Rheum* 2004;**50**:2066–71.
- <sup>15</sup> Rebbeck TR, Martinez ME, Sellers TA *et al.* Genetic variation and cancer: improving the environment for publication of association studies. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:1985–86.
- <sup>16</sup> Freimer NB, Sabatti C. Guidelines for association studies in human molecular genetics. *Hum Mol Genet* 2005;**14**:2481–83.
- <sup>17</sup> Wacholder S. Publication environment and broad investigation of the genome. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:1361.
- <sup>18</sup> Anonymous. Framework for a fully powered risk engine. *Nat Genet* 2005;**37**:1153.
- <sup>19</sup> Ioannidis JP. Commentary: grading the credibility of molecular evidence for complex diseases. *Int.J.Epidemiol* 2006;**35**:572–78.
- <sup>20</sup> Ioannidis JP, Bernstein J, Boffetta P *et al.* A network of investigator networks in human genome epidemiology. *Am J Epidemiol* 2005;**162**:302–04.
- <sup>21</sup> Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J Royal Stat Soc B* 1995;**57**:289–300.
- <sup>22</sup> Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;**7**:781–91.
- <sup>23</sup> Ioannidis JP, Trikalinos TA, Ntzani EE *et al.* Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003;**361**:567–71.
- <sup>24</sup> Higgins JP, Thompson SG, Deeks JJ *et al.* Measuring inconsistency in meta-analyses. *Br. Med J* 2003;**327**:557–60.
- <sup>25</sup> Chanock SJ, Manolio T, Boehnke M *et al.* Replicating genotype-phenotype associations. *Nature* 2007;**447**:655–60.
- <sup>26</sup> Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 2007;**64**:203–13.
- <sup>27</sup> Pompanon F, Bonin A, Bellemain E *et al.* Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 2005;**6**:847–59.
- <sup>28</sup> Clayton DG, Walker NM, Smyth DJ *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005;**37**:1243–46.
- <sup>29</sup> Page GP, George V, Go RC *et al.* 'Are we there yet?': Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet* 2003;**73**:711–19.
- <sup>30</sup> Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet* 2005;**6**:287–98.
- <sup>31</sup> Evans DM, Marchini J, Morris AP *et al.* Two-stage two-locus models in genome-wide association. *PLoS Genet* 2006;**2**:e157.
- <sup>32</sup> Ioannidis JP, Trikalinos TA, Zintzaras E. Extreme between-study homogeneity in meta-analyses could offer useful insights. *J Clin Epidemiol* 2006;**59**:1023–32.
- <sup>33</sup> Rosenbaum PR. Replicating effects and biases. *Am Statist* 2001;**55**:223–27.
- <sup>34</sup> Skol AD, Scott LJ, Abecasis GR *et al.* Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;**38**:209–13.
- <sup>35</sup> Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet* 2005;**366**:1315–23.
- <sup>36</sup> Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:124–32.
- <sup>37</sup> Newton-Cheh C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. *Mutat Res* 2005;**573**:54–69.
- <sup>38</sup> Cordell HJ, Clayton DG. Genetic association studies. *Lancet* 2005;**366**:1121–31.
- <sup>39</sup> Pan Z, Trikalinos TA, Kavvoura FK *et al.* Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005;**2**:e334.
- <sup>40</sup> Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;**55**:997–1004.
- <sup>41</sup> Price AL, Patterson NJ, Plenge RM *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet* 2006;**38**:904–9.
- <sup>42</sup> Genetic Association Information Network. Accessible at: [http://www.fnih.org/GAIN/GAIN\\_home.shtml](http://www.fnih.org/GAIN/GAIN_home.shtml)
- <sup>43</sup> Welcome Trust Case Control Consortium. Accessible at: <http://www.wtccc.org.uk/>
- <sup>44</sup> Ioannidis JP, Gwinn M, Little J *et al.* A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006;**38**:3–5.
- <sup>45</sup> Little J, Bradley L, Bray MS *et al.* Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002;**156**:300–10.
- <sup>46</sup> Chan AW, Hrobjartsson A, Haahr MT *et al.* Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;**291**:2457–65.
- <sup>47</sup> Contopoulos-Ioannidis DG, Alexiou GA, Gouvas TC *et al.* An empirical evaluation of multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenet Genomics* 2006;**16**:705–711.
- <sup>48</sup> Seminara D, Khoury MJ, O'Brien TR *et al.* The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* 2007;**18**:1–8.

- <sup>49</sup> Shields PG. Publication bias is a scientific problem with adverse ethical outcomes: the case for a section for null results. *Cancer Epidemiol Biomarkers Prev* 2000;**9**:771–72.
- <sup>50</sup> Atkins D, Best D, Briss PA *et al*. Grading quality of evidence and strength of recommendations. *Br. Med. J* 2004;**328**:1490.
- <sup>51</sup> Atkins D, Briss PA, Eccles M *et al*. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005;**5**:25.
- <sup>52</sup> Klein RJ, Zeiss C, Chew EY *et al*. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;**308**:385–89.
- <sup>53</sup> Conley YP, Jakobsondtir J, Mah T *et al*. CFH, ELOVL4, PLEKHA1 and LOC387715 genes and susceptibility to age-related maculopathy: AREDS and CHS cohorts and meta-analyses. *Hum Mol Genet* 2006;**15**:3206–18.
- <sup>54</sup> Despriet DD, Klaver CC, Witteman JC *et al*. Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration. *JAMA* 2006;**296**:301–9.
- <sup>55</sup> Uka J, Tamura H, Kobayashi T *et al*. No association of complement factor H gene polymorphism and age-related macular degeneration in the Japanese population. *Retina* 2006;**26**:985–87.
- <sup>56</sup> Gotoh N, Yamada R, Hiratani H *et al*. No association between complement factor H gene polymorphism and exudative age-related macular degeneration in Japanese. *Hum Genet* 2006;**120**:139–43.
- <sup>57</sup> Herbert A, Gerry NP, McQueen MB *et al*. A common genetic variant is associated with adult and childhood obesity. *Science* 2006;**312**:279–83.
- <sup>58</sup> Roszkopf D, Bornhorst A, Rimbach C *et al*. Comment on “A common genetic variant is associated with adult and childhood obesity”. *Science* 2007;**315**:187.
- <sup>59</sup> Loos RJ, Barroso I, O’rahilly S *et al*. Comment on A common genetic variant is associated with adult and childhood obesity. *Science* 2007;**315**:187.
- <sup>60</sup> Dina C, Meyre D, Samson C *et al*. Comment on “A common genetic variant is associated with adult and childhood obesity”. *Science* 2007;**315**:187.
- <sup>61</sup> Bertram L, McQueen MB, Mullin K *et al*. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 2007;**39**:17–23.
- <sup>62</sup> Breast Cancer Association Consortium. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. *J Natl Cancer Inst* 2006;**98**:1382–96.
- <sup>63</sup> Ioannidis JP. Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. *J Natl Cancer Inst* 2006;**98**:1350–53.
- <sup>64</sup> Ntzani E, Rizos V, Ioannidis JP. Genetic effects versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large scale evidence. *Am J Epidemiol* 2007;**165**:973–84.
- <sup>65</sup> Khoury MJ, Little J, Gwinn M, Ioannidis JP. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol* 2006;**36**:439–45.
- <sup>66</sup> Giles J. Huge Biobank project launches despite critics. *Nature* 2006;**440**:263.
- <sup>67</sup> Ioannidis JP, Ntzani EE, Trikalinos TA *et al*. Replication validity of genetic association studies. *Nat Genet* 2001;**29**:306–9.
- <sup>68</sup> Lohmueller KE, Pearce CL, Pike M *et al*. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177–82.
- <sup>69</sup> Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet* 2004;**5**:589–97.
- <sup>70</sup> Ioannidis JPA, Kavvoura FK. Concordance of functional in vitro biological data with epidemiological associations for complex diseases. *Genet Med* 2006;**8**:583–93.
- <sup>71</sup> Jais PH. How frequent is altered gene expression among susceptibility genes to human complex disorders? *Genet Med* 2005;**7**:83–96.
- <sup>72</sup> Zhu Y, Spitz MR, Amos CI *et al*. An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. *Cancer Res* 2004;**64**:2251–57.
- <sup>73</sup> Weedon MN, McCarthy MI, Hitman G *et al*. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med* 2006;**3**:e374.
- <sup>74</sup> ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004;**306**:636–40.
- <sup>75</sup> Lopez AD, Mathers CD, Ezzati M *et al*. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 2006;**367**:1747–57.
- <sup>76</sup> Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2003;**33**:30–42.
- <sup>77</sup> Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol* 2006;**164**:609–14.
- <sup>78</sup> Zeggini E, Weedon MN, Lindgren CM *et al*. Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336–41.
- <sup>79</sup> Scott LJ, Mohlke KL, Bonnycastle LL *et al*. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;**316**:1341–45.
- <sup>80</sup> Saxena R, Voight BF, Lyssenko V *et al*. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;**316**:1331–36.
- <sup>81</sup> McPherson R, Pertsemlidis A, Kavaslar N *et al*. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007;**316**:1488–91.
- <sup>82</sup> Easton DF, Pooley KA, Dunning AM *et al*. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;**447**:1087–93.
- <sup>83</sup> Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
- <sup>84</sup> Shapiro S. Looking to the 21st century: have we learned from our mistakes, or are we doomed to compound them? *Pharmacoepidemiol Drug Saf* 2004;**13**:257–65.
- <sup>85</sup> Willett W, Greenland S, MacMahon B *et al*. The discipline of epidemiology. *Science* 1995;**269**:1325–26.