The number of genetic association studies is growing rapidly, and this growth is likely to accelerate in the future. The correct interpretation of such studies has important implications for our understanding of disease causation, variability in drug response and drug side effects and the biology of populations. There is increasing awareness that the literature of association studies requires the application of stricter standards, to prevent the promulgation of false positive results. Although there is no single agreed upon set of standards for such studies, we believe that journals must, through the manuscripts that they choose to publish, promote an improvement in the quality of association studies. Guidelines can be an instrument for both authors and reviewers, to make such improvements. The subsequent guidelines should be followed for all future submissions to Human Molecular Genetics. The references provide background which may be useful for authors.

As the popularity of association studies increases, two difficulties in interpreting their results emerge. On the one hand, the large number of markers typed and the variety of phenotype definitions considered lead to a multiplicity of statistical tests, whose significance is meaningless unless appropriately corrected for multiple comparisons. On the other hand, interpretation of association studies depends in part on the degree of justification for the scope of the study: is the number of individuals typed sufficient for the identified goal? What led to the choice of the studied candidate genes? Is the evidence strong enough that we can really exclude consideration of the rest of the genome or are the candidate genes studies more appropriately considered as a random subset of genes?

We have recently published a review article in which we discussed these issues extensively (1) and we reviewed the consequences for the interpretation of the study results of ‘searching too much’ as well as ‘searching too little’. The present guidelines aim at helping authors and reviewers to make sure that the published paper is self-contained in terms of providing all the details needed for readers to be able to interpret the results. In this spirit, we require the authors to justify their study design and to provide a context in which correction of multiple comparisons of their significance results can be carried out.

Acknowledging the need for authors to address these issues, does not, however, translate into specifying one particular threshold of significance that studies must meet for publication. Two factors, at least, argue against making such a specification. On the one hand, there is not a clear consensus on what statistical approach is most appropriate for the problem and imposing one would be immature and reductive; on the other hand, we believe that a study may be informative, even if it fails to map the gene for a trait, provided that some standards in its design are met. With regard to the first point, there are different statistical definitions of global errors and in different contexts one or the other may be appropriate. Traditionally, in linkage studies, the family wise error rate has been controlled and this strategy has proved adequate for mapping Mendelian diseases. In the context of complex diseases, and particularly for association studies, it may be appropriate to consider a less stringent definition of error rates, as for example the false discovery rate. It goes without saying that different definitions of global errors translate into different cutoffs on the nominal $P$-values. (2,3).

Bayesian procedures have inspired some of the most traditional significance cutoffs in statistical genetics (e.g. the lod score of 3) and may be particularly useful in the context of candidate gene studies. The identification of significance results in the Bayesian context and the treatment of multiple comparisons do not correspond to a universal $P$-value cutoff. Association tests involving different markers and different

---

*To whom correspondence should be addressed. Tel: +310 7949571; Fax: +310 7949613; Email: nfreimer@mednet.ucla.edu

© The Author 2005. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oupjournals.org
phenotypes may be dependent (due to correlation in the phenotypes or linkage disequilibrium between the markers) and that dependency can translate into an effective reduction of the number of multiple comparisons, with consequent variation of the appropriate significance cutoff. Unfortunately, we do not have an adequate model for dependence among association tests at different markers, for example, and increase in power due to the consideration of dependence can be achieved only on a case-by-case basis using resampling techniques. This lack of a general model also makes it also difficult to define a common significance threshold to be required in all studies.

The subsequent guidelines are written with the view that there is a ‘burden of proof’ that either the detected association is significant or the study was able to exclude association with the typed polymorphisms, and this burden of proof rests on the authors’ shoulders. Although there are a variety of ways of addressing this problem, simply presenting nominal $P$-values is not one of them. The following guidelines outline the expectations of the journal in this regard. We have tried to be as synthetic as possible to provide a practical instrument that may be used as a checklist by authors and reviewers. The guidelines reflect the multiple options open to authors and require the usual scholarly contribution of reviewers in assessing the quality of the paper. It goes without saying that more attention should be paid to the spirit of the guidelines than the letter.

Candidate gene and genome-wide association studies face different statistical issues, so we consider the guidelines for these categories separately.

**Candidate genes**

For these studies, authors must provide explicit justification for the choice of the candidate gene. Use of the literature for such justification must include a rigorous accounting of the level of statistical evidence from prior studies, including both studies supportive of the candidate hypothesis and negative studies.

(A) Authors must make a clear distinction between candidate genes for which there is previous statistical evidence of linkage or association with the specific disease under study and genes that are proposed solely on the basis of a specific biological hypothesis.

(B) For genes for which there is prior statistical evidence of linkage or association to the trait in question, authors should report the $P$-values obtained in previous studies. In addition, they should (a) specify any differences between the phenotypes employed in prior cited studies and those employed in the current manuscript; (b) specify the populations employed in prior, cited studies as well as the size of the samples employed and (c) provide details about the extent of the genomic region implicated in the prior, cited studies as well as the location and allele frequencies of variants.

(C) Over all, authors should provide a quantitative estimate (using the cited evidence) of the prior probability that the list of genes considered contains a relevant one. Procedures for such estimation are described in Wacholder et al. (4). This prior probability can be used in a Bayesian procedure to obtain a posterior probability of association (17,18). The Bayesian analysis that controls a measure of global error is, indeed, an option that authors may want to consider.

(D) Authors should correct the $P$-values in their results for multiple comparisons due to the multiple genes studied, obtaining an adjusted $P$-value. They should provide a justification for the specific statistical method used for this correction. Furthermore, we encourage authors to present, in addition to $P$-values, an indication of the magnitude of the genetic effect and its associated standard error; this facilitates the combination of results across different studies.

(E) In a conservative, least favorable scenario, it has been suggested that a significant $P$-value should be lower than $\sim 10^{-7}$ (1). Authors should use their arguments made in (B) and (C) to argue why a variant with a nominal association $P$-value higher than this level should be considered noteworthy.

**Genome screens**

For genome-wide association studies, authors must indicate why they believe that the study design is appropriate. They must do the following.

1. Justify the sample size employed, including estimation of the power to detect significant association results for given effect sizes; the basis for considering such effect sizes to be realistic must also be provided.
2. Justify the marker density employed for the particular study population.
3. Implement a procedure to correct for multiple comparisons. Either FWER or FDR are acceptable criteria, as well as global error thresholds less stringent than 0.05. In addition, Bayesian procedures can be adopted. What is crucial is that authors provide an evaluation of the global error rate achieved when they have deemed an association to be significant. It may also be useful for authors to discuss the rationale for their choice of either Bayesian or frequentist interpretations of their results.
4. Attempt to quantify the uncertainty associated with the estimated location of trait-associated variant(s). Although the methods for estimating such uncertainty are not as straightforward as the ones used to determine confidence intervals for lod scores in linkage analysis, several papers have suggested reasonable assumptions that authors can make in association studies (5–12).

Other factors that can contribute to incorrect interpretation of association results should also be accounted for explicitly. For example, authors should state how they have addressed the possible impact on their findings of population substructure (14–16). Colhoun et al. (13) suggested a checklist for referees to evaluate the study design of association studies. Although several of these points are incorporated in the previous guidelines, others of their points should be considered by authors in preparation of manuscripts, in particular those relating to sampling and sub-group analyses.
Acknowledgements

We thank Susan Service for helpful comments. The authors are supported by grants from NIH and NSF.

References