Commentary

Publication Policy or Publication Bias?

Graham Byrnes, Lyle Gurrin, James Dowty, and John L. Hopper

Centre for Genetic Epidemiology, The University of Melbourne, Carlton, Victoria, Australia

Although ''a large number of studies have appeared in the literature that report associations of low penetrance genetic variants with disease,'' as suggested in the recent *CEBP* editorial (1), hardly any of these reports have been translated into solid results by replication studies, at least for the common cancers. To some extent this is inevitable, as false-positive associations are an inherent part of science. Because the great majority of tested hypotheses will be unfounded, false positives will outnumber correct findings in the literature (2).

Large-scale genotyping is likely to compound these problems in the same way RNA microarrays have done in laboratory studies. The true type 1 error rate in most genetic association studies is far higher than 5% due to data dredging. This is in spite of the development of new techniques developed to limit false discovery (3, 4) that were not referenced in the editorial (1). This situation will only be compounded by giving license to the rampant multiple testing of interactions: why scan only 100,000 single nucleotide polymorphisms for association when 5 billion pairwise interactions can be tested for the same cost?

This throws greater responsibility onto scientific journals to select which reports should be published; see Altman's editorial about ''The scandal of poor medical research'' written more than 10 years ago (5). The editors have bravely tried to build a foundation for publication choice on three bases: (*a*) rigorous review of statistical and laboratory methods; (*b*) encouraging replication studies by prioritizing their publication; and (*c*) encouraging the investigation of hypotheses with greater prior probability of verification.

The first two of these deserve applause: we do not need any more low-powered studies reporting results that are entirely consistent with the null hypothesis once properly adjusted for multiple comparisons. The recognition of the fallibility of genotyping and the need for detailed reporting of quality control is also timely.

The third is a mixed bag. Selecting hypotheses for testing according to a variety of current information is simple common sense. However, the editors also state ''it is likely that…genes do not act alone but interact with other genes or biomarkers,'' without any supporting references. This may be true, but is it relevant? Are we such clever scientists that we can posit interactions from our understanding of cellular pathways, then test the hypotheses? By declaring that ''priority will be given to studies that consider biologically plausible interactions of multiple genes in a pathway as well as interactions of environmental exposures with genetic variants that are involved in the metabolism of those exposures,'' the importance of interactions may become falsely established as fact through publication bias.

Cellular pathways are far from understood. Information on current databases such as KEGG can be based on as little as observed coexpression in RNA microarray assays (6). It seems likely that these laboratory data-derived resources carry as great a proportion of false-positive association as the genetic epidemiology literature. Environmental exposures are also problematic: the editors note that many are poorly measured if observed at all. In many cases, this is unavoidable. For example, exposure to infectious disease (including subclinical infection) may play a role in the development of type 1 diabetes. Even with the benefit of large-scale experimental breeding, plant geneticists struggle to discover beneficial gene-environment interactions (7). Meanwhile, many diseases may be explained, at least in part, by the simple additive effect of several rare but highly penetrant alleles. In all cases, the suggestion by Cordell (8) and others to first consider main effects is compelling.

Solutions have been proposed in other fields: the CONSORT statement (9) aims to ''encourage(s) transparency with reporting of the methods and results so that reports of RCTs can be interpreted readily and accurately.'' There is also an international registry of Randomized Controlled Trials, the Standard Randomized Controlled Trial Number, to guard against post hoc hypothesis ''discovery.'' Lodging of data from multiple studies by collaborating groups, followed by planned, impartial statistical analysis, can allow the truth to emerge from the data (10). Rather than defining a niche for the Journal as a publisher of interaction reports, would it be more profitable to concentrate on methods to ensure more rigorous review and to enhance sharing and pooling of data?

References

1. Rebbeck TR, Martínez ME, Sellers TA, Shields PG, Wild CP, Potter JD. Genetic variation and cancer: improving the environment for publication of association studies. Clin Epidemiol Biomarkers Prev 2004;13:1985–6.
2. Sterne JAC, Davey-Smith G. Sifting the evidence—what's wrong with significance tests? BMJ 2001;322:226–31.
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
4. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS 2003;100:9440–5.
5. Altman DG. The scandal of poor medical research. BMJ 1994;308:283–4.
6. Nakao M, Bono H, Kawashima S, Kamiya T, Sato K, Goto S, Kanehisa M. Genome-scale gene expression and pathway reconstruction in KEGG. Genome Inform Ser Workshop Genome Inform 1999;10:94–103.
7. Chaves MM, Oliviera MM. Mechanisms underlying plant resilience to water deficits: prospects for water-saving agriculture. J Exp Biol 2004;55:2365–84.
8. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 2002;11:2463–8.
9. Altman DG. Schulz KF, Maher D, et al. for the CONSORT group. The revised CONSORT statement for reporting randomized controlled trials: explanation and collaboration. Ann Intern Med 2001;134:663–94.
10. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83000 women with breast cancer from 16 countries. Lancet 2004;363:1007–16.

Cancer Epidemiology, Biomarkers & Prevention

Publication Policy or Publication Bias?

Graham Byrnes, Lyle Gurrin, James Dowty, et al.

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/14/6/1363

Cited articles	This article cites 10 articles, 3 of which you can access for free at: http://cebp.aacrjournals.org/content/14/6/1363.full#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/14/6/1363.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/14/6/1363. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.