Resolving inconsistency in the results of genetic association studies of cardiovascular disease

Aroon HINGORANI
BHF Laboratories, Centre for Clinical Pharmacology, Department of Medicine, Rayne Building, University College London, London WC1E 6JJ, U.K.

ABSTRACT

In this issue of Clinical Science, Agema and co-workers report the results of a genetic association study of eNOS (endothelial nitric oxide synthase) gene polymorphisms (−786T→C, intron 4b→a and Glu298→Asp) in patients with angiographic CAD (coronary artery disease), and/or prior MI (myocardial infarction) and a group of healthy population-based controls. However, the findings of this study appear to contradict previous studies published on the eNOS polymorphisms, and this commentary will attempt to resolve the inconsistency in such genetic association studies.

In this issue of Clinical Science, Agema and co-workers [1] report the results of a genetic association study of eNOS (endothelial nitric oxide synthase) gene polymorphisms (−786T→C, intron 4b→a and Glu298→Asp) in angiographic CAD (coronary artery disease) and/or prior MI (myocardial infarction). The rarer −786C, Asp298 and intron 4a alleles of eNOS have been associated with an increased risk of ischaemic heart disease or MI and, in the case of the −786T→C and Glu298→Asp variants, with altered NO-dependent vascular responses in vivo [2,3]. Molecular studies have shown that the −786C allele in the gene promoter may attenuate mRNA transcription, and that the Glu298→Asp amino acid polymorphism results in enhanced proteolytic cleavage of the mature enzyme, although for both the clinical and molecular studies, the data has been inconsistent. Nevertheless, the findings of Agema et al. [1], that homozygosity for the common Glu298 allele was associated with increased risk of CAD or prior MI, are surprising; more so because, in the same study, the rare Asp298 allele was associated with longer ischaemia times in a subset of patients with data from ambulatory ECGs. The design of the study appears robust, the methods used for genotyping have been used in many prior studies and genotype frequencies in all groups conformed to the Hardy–Weinberg equilibrium. How then might these results be explained?

Substantial inconsistency in the results of the 3000 or so genetic association studies of CAD conducted since 1985 has been widely recognized and led some to question the validity of this approach to identifying CAD genes. In principle, inconsistent results could arise due to true biological variability in the association between genetic polymorphisms and disease in different data sets, arising from varying patterns of linkage disequilibrium between marker and disease variant, different causative polymorphisms (allelic heterogeneity) or from substantial variation in the modulating impact of the environment among the different populations studied [4]. Recent reviews have suggested that these explanations are unlikely to account for much of the inconsistency and that it is more likely that falsely negative and falsely positive studies both contribute substantially to the observed variability [4].

Why should there be such a high proportion of false positive and false negative studies in the genetic literature and can the available data be resolved in any meaningful way? Since there are many thousands of genetic candidates, each with many SNPs (single nucleotide polymorphisms), spurious association, arising by chance from multiple hypothesis testing, is likely to account for a substantial number of positive studies. Undetected ethnic admixture in cases or controls might also falsely distort allele frequencies in some situations and, if disease risk

Key words: cardiovascular disease, gene polymorphism, myocardial infarction, nitric oxide synthase (NOS).

Correspondence: Dr Aroon Hingorani (email a.hingorani@ucl.ac.uk).
also varies by ethnicity, a specific form of confounding due to population stratification may result. However, provided individual studies are conducted in broadly homogeneous ethnic groups, the chance of such confounding is also thought to be small [5]. The other side of the coin is that for most genetic association studies, sample sizes have been too small to reliably detect the small genotypic ORs (odds ratios) in question. Sample sizes of thousands are required to have adequate power to detect genes conferring ORs as low as 1.2, with minor allele frequencies in the range of 5–10 % [6]. Thus small studies with OR estimates nominally reported as negative, because wide confidence limits encompass the null, may be entirely consistent with real genetic effects.

Is it then possible to identify real genetic associations from the available data? Small effects might be detected reliably if studies could be pooled appropriately to enhance power. The techniques of systematic review and meta-analysis, developed for health services research to pool data from individually underpowered randomized controlled intervention trials, are now being applied to genetic studies [6,7]. For some polymorphisms, the published data set is very large allowing the prospect of some robust conclusions. The process of pooling data might also reveal outlying studies with large effect sizes that are likely to have arisen by chance. If it were shown, by appropriate synthesis of the published data, that real genetic effects existed, or that plausible candidates were unlikely to have a major role, much unnecessary repetition of research effort might be prevented, and the available data would be utilized to its best effect.

To illustrate, recent meta-analyses have shown that common variants of the MTHFR (methylene tetrahydrofolate reductase) and apoE (apolipoprotein E) genes, which influence plasma concentrations of homocysteine and cholesterol respectively, confer a moderately increased risk of CAD (OR approx. 1.25) [8,9]. The relative risk difference of approx. 25 % arising from carriage of the variant genotypes in these cases is of similar magnitude to the risk reduction associated with therapies commonly used in primary or secondary prevention of cardiovascular disease, e.g. aspirin, statins or antihypertensives. Although genotypic relative risks of this size are unlikely to impact on disease prediction in individual patients, since CAD is common, genetic effects of this size, if real, could contribute to a substantial number of clinical events in the population as a whole.

In a recently completed meta-analysis that included about up to 6000 cases and 6000 controls for each polymorphism, we found that the intron 4a and Asp298 genotypes of the eNOS genes were also associated with moderately elevated risk of CAD (OR approx. 1.3) [10,11]. With one or two exceptions, the studies contributing to the meta-analysis were small (several hundred rather than several thousand cases), and some individual studies reported null or even protective effects (like the study of Agema et al. [1]). A cumulative meta-analysis to assess the evolution of the genetic odds ratio as more studies have been reported, an approach that gives an indication of the stability of the risk estimate to additional data, showed that for the Asp298 polymorphism, at least, the risk estimate has been stable since about 2001 and that the confidence intervals have narrowed [11]. Nevertheless, all meta-analyses are limited by the quality of the available data and there is still a chance that genetic estimates from the small studies included lead to a somewhat biased estimate of risk partly, although not exclusively, due to the preferential publication of positive studies (publication bias). Formal statistical tests for small-study bias were, however, negative in the case of the Asp298 variant [10].

Current uncertainty in the field of cardiovascular genetics recalls similar controversies in cardiovascular therapeutics prior to the publication of large-scale randomized controlled intervention trials (mega-trials). For example, in the pre-mega-trial era, uncertainty existed about the efficacy of thrombolysis in acute MI because most RCTs (randomized controlled trials) to that point had been underpowered to detect the small, but important, 30 % relative mortality reduction we now know to result from treatment. The ISIS-2 and GISSI mega-trials subsequently demonstrated this treatment effect unequivocally [12,13], but a retrospective meta-analysis of the small RCTs of thrombolysis that predated ISIS-2 and GISSI showed that the summary estimate for the treatment effect was almost precisely that detected in the subsequent mega-trials [14], indicating the potential for meta-analysis of individually underpowered studies (be they genetic or interventional) to provide valid risk estimates. Very-large-scale genetic evidence from thousands of cases and controls will now be required to confirm or refute associations identified in smaller studies or even meta-analyses. Although the appropriate sample collections and prospective studies have been initiated in a number of countries, results will not become available for some years. In the meantime, several authors have proposed that data from well-conducted, technically rigorous, smaller studies, such as that from Agema et al. [1] (whether nominally positive or negative), should enter the public domain to allow pooling of results by meta-analyses that provide continually updated estimates of genetic risks for individual polymorphisms, and to reduce the possibility of publication bias.

ACKNOWLEDGMENTS

I would like to thank Professor Steve Humphries, Dr Liam Smeeth, Dr Pankaj Sharma and Dr Juan Pablo Casas for their invaluable expertise and discussions about the ideas elaborated in this commentary. I would like to
thank the British Heart Foundation for their support in the form of a Senior Research Fellowship.

REFERENCES


Received 16 July 2004; accepted 30 July 2004
Published as Immediate Publication 30 July 2004, DOI 10.1042/CS20040211