Evolution of genetic analysis strategies in coronary heart disease: a case of unnatural selection?

The recent news that a rough draft of the sequence of the human genome has been completed has been interpreted as ushering in a new era of enhanced understanding of complex disease processes, individualized risk prediction and therapy, and novel rationally-designed therapeutic agents. For any of these promises to be fulfilled it will be necessary to document, in large studies comprising both family and case-control collections, robust associations between genotypes at single nucleotide polymorphisms (SNPs) of candidate genes and the diseases under study. By far the commonest study design has been comparison of the allele frequency of polymorphisms of candidate genes that are associated with changes in the level of a plasma risk factor (such as fibrinogen or cholesterol) between cases of disease and controls. In the majority of instances, however, the sequence of events has been the publication of a small positive study (of the order of a few hundred cases) typing a novel polymorphism, followed by a large number of studies of similar or even smaller size either confirming or failing to confirm the initial finding. Not surprisingly, the technique of meta-analysis has been applied to groups of such studies. Perhaps the most extensively studied polymorphism is the insertion/deletion polymorphism of the angiotensin-converting enzyme gene, in which case there has been clear evidence of publication bias towards small positive studies, and in which — some 8 years after the initial report of an association of the DD genotype with myocardial infarction[1] — a large study has now ruled out the presence of any substantial association[2]. The paper by Kluijtmans and Whitehead in this issue[3] raises important issues in the interpretation of case-control genetic association studies and introduces a novel method of analysis (‘restricted’ meta-analysis) which attempts to use data from a selected subset of studies to address mechanistic questions regarding genotype, plasma risk factor levels and risk of atherothrombotic disease.

Genetic studies provide a potential alternative to far more costly intervention trials in the evaluation of novel plasma risk factors for coronary disease, such as homocysteine level as studied by Kluijtmans and Whitehead[3]. If a genetic variant influences the plasma level of a potential risk factor, and has probably done so in any given individual since early life, then if the risk factor is causal, a similar size and direction of association would be expected to be observed between the variant and disease, as between the plasma level and disease. In the case of homocysteine, disruption of the MTHFR gene leading to severe hyperhomocysteinemia causes premature atherosclerosis[4]. Observational epidemiology suggests that homocysteine variability within the normal range is also associated with differences in coronary risk, but the causal role of the plasma homocysteine level in this range is unproven. A common C to T substitution at position 677 of the MTHFR gene (with an allele frequency of approximately 0.35) results in an alanine to valine amino acid substitution in the protein. Homozygosity for the rarer T allele (TT genotype) is associated with an elevation of some 25% in plasma homocysteine, although this elevation only seems to occur in those who are relatively deficient in folate. A large number of studies have tested for association between this polymorphism and cardiovascular disease, and a previous meta-analysis of 23 studies encompassing about 6000 cases and 6000 controls did not find any evidence of an association between genotype and disease[5]. Even with these numbers, however, it is possible that the relatively small increase in risk predicted by the epidemiology might not have been detected. This seems to indicate the need for large studies of the order of several thousands of cases and controls and for studies in susceptible populations (for example where folate deficiency is prevalent) or selected patient groups enriched for genetic contribution (for example early onset disease).

Kluijtmans and Whitehead[3] take an alternative approach and actually restrict the studies in their meta-analysis to the first ten published studies that reported positive association between the polymorphism and disease. In this way they include only five of the 23 studies in the previous
meta-analysis, together with five published since that meta-analysis, and thus only 1800 cases and 3000 controls are included in their paper. Since the sine qua non in meta-analysis of clinical trial data has been the rigorous search for and inclusion of all studies, regardless of size, place of publication or indeed publication at all, this approach will raise eyebrows. Indeed it might have been wise to avoid the term meta-analysis as the two approaches have quite different objectives. The authors seek to justify their approach as the selection of ‘genetically vulnerable’ populations, but it results, for example, in the inclusion of some studies in both European and Japanese cases and controls, and the exclusion of other studies in these same populations, which does not argue strongly in favour of the approach increasing population homogeneity. Kluittmans and Whitehead then use this abstracted data to address an extremely important issue — where is the heterozygote in terms of risk? This question was not addressed in the previous meta-analysis, but is clearly critical both in assessing the mechanism of any genetic influence on risk and the population attributable risk of the genotype. They conclude that there is a difference in risk between CC and CT genotypes, with CC having least risk, CT intermediate risk, and, stated as one of their assumptions, TT highest risk.

Is this finding surprising considering the methodology used? Overall, the answer is no. In fact this finding is largely predicted by the study design, whether or not the CT genotype is genuinely associated with risk, and so the result is, unfortunately, not very informative. Certainly, the graded genotypic effect on risk observed could be the result of a graded biological effect and plausible mechanisms can be constructed. But the important issue, given the conflicting data regarding the MTHFR TT genotype, is that the various artefacts that confound case-control association studies will persist in the ‘restricted’ meta-analysis.

Perhaps the major concern in the interpretation of case-control studies is that false-positive results may occur due to unsuspected genetic inhomogeneity between case and control groups. Thus, in an extreme case, if the cases and controls came from two genetically distinct populations who had great differences in the prevalence of culturally or genetically transmitted risk factors for coronary disease and, incidentally, a difference in allele frequency at a polymorphism that was completely disease-neutral, an association of genotype with disease would be found, but it would not be due to any effect of the polymorphism or a near neighbour on disease risk. Lander has used the example of a hypothetical allele that appears to convey the ability to eat with chopsticks existing at the HLA-A locus, which has very different allele frequencies in Orientals and Caucasians. This is not, however, a merely theoretical or facetious issue, as well-known examples in the genetic literature illustrate. If the 10 studies, a minority of those published, abstracted by Kluittmans and Whitehead were false-positives because of such inhomogeneity, then their finding would be very much expected without the need for any genetic effect on risk — if TT homozygotes are commoner in the case ‘population’ then, by the Hardy–Weinberg law, CT heterozygotes will be so also. Unfortunately the presence of Hardy–Weinberg equilibrium (the conformation of genotypes to expected proportions given the allele frequencies) in the abstracted studies is no safeguard against such an occurrence.

A second concern is that the selected studies were not all restricted to the possible role of the TT homozygotes, but rather, some examined the role of the T genotype per se. This is appropriate given the lack of biological data that might mandate testing only a recessive, as opposed to a co-dominant or even dominant, model. In some of the studies the frequencies of the T allele in patients and control subjects were the primary findings and odds ratios for the CT genotype were reported together with those for the CC genotype. Thus the positive findings of these analyses depended in part on the contribution of the CT heterozygous cases. As a result, the established impact of publication bias in small association studies would have been carried over into the current analysis.

A further issue is the possibility of systematic genotyping error in the selected studies leading to false-positive results. Data regarding the specifics of genotyping methodology are often difficult to discern from individual studies but it is prudent to genotype subjects blinded to case-control status, randomly allocated among plates, with the presence of both positive and negative controls. In the example of the C677T MTHFR polymorphism, genotyping is usually by HinfI restriction enzyme digestion of PCR product; this assay is known to be susceptible to partial digests and this may vary from batch to batch and according to experience. If cases and controls were grouped together, spurious differences between genotype frequencies might arise. In this case, the methodology selected by Kluittmans and Whitehead would again return differences in the genotype frequency of heterozygotes without biological foundation.

This is not to denigrate the importance of the question regarding heterozygote risk that Kluittmans and Whitehead have proposed, nor the very real probability that certain alleles may only be associated with risk in certain populations. However, to address these issues, an alternative to ‘restricted’
meta-analysis, that is, the study of families, may be preferable. In such a study design, an affected individual and other members of the family (parents if available, otherwise siblings) are ascertained and the probability of transmission of the different alleles at the locus under study from parents to the affected individual calculated. Because transmission is being studied, there is no problem with population substructure between cases and controls, although extreme care still needs to be paid to the technical issues of genotyping. We would suggest that any positive association discovered in a case-control study, however strongly supported by meta-analyses and appropriately sized studies, be regarded as preliminary until family studies confirm that the association is due to identity or proximity of the genotyped polymorphism to a causative genetic variant on the chromosome. Until such time, it seems prudent to adhere to the epidemiological principles of large size in individual studies and inclusivity in meta-analysis. 

B. KEAVNEY
H. WATKINS
Department of Cardiovascular Medicine, University of Oxford, Oxford, U.K.

References


The post myocardial infarction exercise test: still worthy after all of these years

See page 300 for the article to which this Editorial refers

While the prognostic value of exercise testing post-myocardial infarction has been established by meta-analysis, both pre[1] and post[2] the thrombolytic era, there are other benefits of performing the test. Hospital discharge can be optimized and expediated. The patients’ response to exercise, their work capacity, and limiting factors at the time of discharge can be assessed. Guidelines for exercise at home can be formulated and reassurance given of physical status, and risk of complications. The test provides a safe basis for advice on return to work, and can demonstrate to the patient, relatives, or employer the effect of the myocardial infarction on the capacity for physical performance. It can cause an improvement in the patients’ self-confidence by making them less anxious about daily physical activities[3]. The test has been helpful in reassuring spouses of post-myocardial infarction patients of their physical capabilities[4]. The psychological impact of performing well on the exercise test is impressive and in fact many patients increase their activity and actually rehabilitate themselves after being encouraged and reassured by their response to this test.

Exercise testing remains useful after hospital discharge. It is an important tool for activity counselling and in exercise training, as part of comprehensive cardiac rehabilitation, where it can be used to develop and modify the exercise prescription, and assess the patient’s progress. For all these reasons, national guidelines call for exercise testing post-myocardial infarction[5].

In this issue, Domínguez et al.[6] heighten our understanding of the post-myocardial infarction