Linking observational and genetic approaches to determine the role of C-reactive protein in heart disease risk

Aroon D. Hingorani1*, Tina Shah1, and Juan P. Casas2

1Department of Medicine, Centre for Clinical Pharmacology, BHF Laboratories, University College London, Rayne Building, 5 University Street, London WC1E 6JJ, UK; and 2Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Online publish-ahead-of-print 9 May 2006

This editorial refers to 'C-reactive protein gene haplotypes and risk of coronary heart disease. The Rotterdam Study'† by I. Kardys et al., on page 1331

In the last decade, there has been a burgeoning interest in associations between concentrations of various circulating biomarkers and later coronary disease.1 Of these, the acute phase reactant, C-reactive protein, has garnered the greatest attention. This spotlight on C-reactive protein reflects the strength and consistency of the observed associations, the fact that C-reactive protein is a stable and readily measured analyte, and the focus on the link between inflammation and atherosclerosis. The findings led initially to the proposal that its measurement might provide useful predictive information on coronary disease risk.2 Later, the proposal was made that C-reactive protein might even play a contributory role in atherogenesis.3

Associations identified between C-reactive protein and incident blood pressure, diabetes, and metabolic syndrome were used in support of this, and experimental data from studies of vascular cells and tissues in vitro and from infusions of C-reactive protein provided preliminary evidence for pro-inflammatory, pro-adhesive, and pro-thrombotic actions.3 These data are exciting because it was through a similar mix of observational and experimental work that cholesterol was first identified as a potential causal factor in atherosclerosis, a link which randomized clinical trials of cholesterol-lowering drugs subsequently confirmed to be causal. If the same is true for C-reactive protein, it would represent a new therapeutic target for coronary disease prevention.

More recently, however, evidence has emerged, which questions the role of C-reactive protein in both the prediction of clinical events and in the pathogenesis of atherosclerosis. A meta-analysis of the prospective cohort studies, including a very large new study, indicated that the strength of the C-reactive protein-coronary event association may be smaller than initially supposed and that C-reactive protein may offer limited predictive information, particularly after levels of traditional risk factors are taken into account.4

This is because C-reactive protein is itself associated with a range of established risk factors for cardiovascular disease including smoking, blood pressure, diabetes, as well as HDL-cholesterol. Once these are measured, it has been argued, there is little additional predictive information in a C-reactive protein measurement.

Associations among established risk factors and biomarkers also limit inference on causation in observational studies because associations of any one factor such as C-reactive protein with incident disease could be affected by confounding (where the link arises because of the common association of the biomarker and the disease outcome with other causal risk factors like smoking, or with an enlarging list of other biomarkers of less certain causal relevance). Statistical adjustment can limit confounding, but rarely overcomes it completely. Adequate adjustment can only be made for those risk factors or biomarkers that have been measured. In a recent review, perhaps 100 biomarkers or exposures (including C-reactive protein) were listed as having shown association with coronary risk in observational studies.1

Importantly, most of these correlate to a greater or lesser degree with one another. However, no study has adjusted for them all. Even if all were assayed, the inherent biological and analytical variability in any measure generally limits the ability to make complete adjustment. Despite this, the persisting association of C-reactive protein with disease risk, after adjustment for a generally limited range of risk factors or other biomarkers, referred to as ‘independent association’ in statistical parlance, has sometimes been over-interpreted as equating with causation. The possibility of residual confounding by unmeasured exposures remains however, and the link with the disease may not be causal. There is an additional problem. Observational studies may not well define the direction of causation, should it exist. Reverse causation (where atheroma leads to an elevation of C-reactive protein rather than vice versa) could provide an alternative explanation for the link between C-reactive protein concentration and disease risk. Although this is more of a problem in case-control studies, it may still...
affect prospective studies of initially healthy subjects, because atherosclerosis has such a long subclinical phase.

Against this setting, another prospective observational study demonstrating the association of C-reactive protein with coronary events is reported by Witteman et al. This association, which diminishes but persists after adjustment for a range of potential confounders, is insistently concordant with prior work. So what is different about this study? The beauty of the current work is that, in line with a growing number of similar studies, it takes the observational analysis, with its inherent limitations, a stage further by incorporating a genetic approach. Taking advantage of its recent resequencing (http://pga.gs.washington.edu/data/crp/ and Carlson et al.), Kardys et al. type three common genetic variants (single nucleotide polymorphisms; SNPs) in the gene that encodes C-reactive protein. Through their association with the 12 or so additional untyped variants in the gene, these three ‘tagging’ SNPs adequately capture almost all the genetic diversity at this locus in European subjects. All the SNPs, as well as the haplotypes inferred from them, are associated with differences in C-reactive protein concentration of about 0.6 mg/L between the most extreme categories, representing nearly 25% of the interquartile range of the C-reactive protein distribution in this population. Seasoned readers of the journal may have become wary of positive findings from genetic association studies with disease end-points, because of their inconsistency. Nevertheless, they can be reassured that, in this case, the association between C-reactive protein genotype and C-reactive protein concentration (an intermediate phenotype) is almost certainly correct. The genotype–C-reactive protein associations identified by the Rotterdam study are closely concordant with other studies of C-reactive protein SNPs.6–8 In general, genotype–biomarker associations appear to be more readily established than genotype–disease associations perhaps because: (a) a change in level or property of the biomarker is the immediate consequence of the genetic variation, so the ‘signal-to-noise ratio’ is greater in such studies; (b) a given biomarker is likely influenced by a smaller range of genes and exposures than the disease; (c) biomarkers are continuous traits, so power is enhanced greatly over studies where the outcome is categorical (e.g. disease event); and (d) biomarker levels are less prone to missclassification error than commonly used disease end-points such as ‘myocardial infarction’ or ‘angiographic coronary artery disease (CAD)’.

So why bother to study gene variants when one can measure C-reactive protein itself? Even if genotype contributes to differences in C-reactive protein concentration, it is only one of the many factors to do so, and the genetic contribution to the variance in C-reactive protein is likely to be much smaller than that due to the combined effect of acquired exposures such as abdominal obesity, smoking, blood pressure, and so on. The answer is that, there is one particular property of genotype that distinguishes it from all the acquired exposures that influence C-reactive protein. Genotype is determined at random at conception according to Mendel’s Laws, because of the independent assortment of alleles during meiotic cell division and the reconstitution of the diploid genome during fusion of gametes. In effect, allocation to a low-C-reactive protein genotype (or haplotype) at conception is analogous to being randomized to a specific C-reactive protein-lowering drug in a clinical trial. This natural Mendelian randomization should ensure that factors that could confound associations of C-reactive protein itself with coronary disease should be distributed evenly among individuals with differing C-reactive protein genotype (or haplotype).9,10 It is clear from Kardys’s study that this is indeed the case. Although a whole range of measured (and by extrapolation unmeasured) exposures relevant to coronary risk varied by quartile of measured C-reactive protein, none differed among the four common haplotypes or genotypes of the C-reactive protein gene, despite a substantial genetic influence on C-reactive protein concentration. Moreover, as genotype/haplotype is determined prior to the disease, reverse causation should also be overcome. Thus, studying the effect of C-reactive protein indirectly through its gene, counter-intuitively, provides a less biased and confounded assessment of its influence on disease risk than studying C-reactive protein itself. If a high C-reactive protein actually contributed causally to coronary disease risk, individuals with genotypes that raise C-reactive protein concentration should have a higher risk of coronary disease, commensurate with the effect of genotype on C-reactive protein level. In the current study, however, no such increase in coronary risk was observed.

So is the case now closed? Has a causal role for C-reactive protein in atherosclerosis been unequivocally refuted? The answer is no, but evidence such as this is forcing a more cautious and critical approach. The null point estimate from the current genetic study is concordant with genetic risk estimates from other work2,8 and consistent with recent experimental studies in vitro, which have indicated that the apparently pro-atherogenic actions of C-reactive protein observed in earlier studies may have been mediated by bacterial peptides and sodium azide present in commercial recombinant C-reactive protein preparations rather than by C-reactive protein itself.11 Because the genetic effect on C-reactive protein is small, a very large genetic association study with about 10 000 cases would be required to exclude a small but, in population terms, extremely important causal effect of C-reactive protein on coronary outcomes. Accumulating this number of cases may be beyond the capacity of individual research groups, but could be achieved through a collaborative effort to pool studies of the type reported here. The development of networks of genetic investigators may help facilitate this and, in the absence of a selective C-reactive protein-lowering drug, may provide the only currently practical way of providing randomized evidence on the effect of C-reactive protein-lowering on disease risk.12 Although statins lower C-reactive protein, this is neither their primary nor main action and it is unclear if this effect is achieved independently of or through cholesterol reduction.

What of C-reactive protein and disease prediction? C-reactive protein might still provide useful information on coronary risk even if it does not play a causal role in coronary disease, because its concentration might capture information on the extent of exposure to orthodox risk factors, inflammatory signals from adipose, or the burden or inflammatory state of atherosomatous plaques. However, the finding that common variation in the C-reactive protein gene influences C-reactive protein concentration both in health and during acute inflammation has potentially
important implications for risk prediction. Two individuals with contrasting C-reactive protein genotype but the same measured C-reactive protein concentration might have different underlying levels of risk factor exposure or inflammatory activity. If it is these that influence disease outcome, the prognostic value of a given C-reactive protein measure may need to be considered in conjunction with genotype or haplotype. Additional work will be required to assess if the predictive utility of a C-reactive protein measurement is modified by genotype.

Combining genetic and observational approaches as Kardys et al. have done, offers the possibility of better distinguishing causal from confounded associations of biomarkers and coronary disease and may, in future, help prioritize new therapeutic targets. Rich and expanding databases of human genome variation are providing the genetic tools for this type of work, but of no lesser importance are the existence of well-phenotyped cohorts, of which the Rotterdam study is an excellent example.

Acknowledgement

A.D.H. acknowledges the support of the British Heart Foundation Senior Research Fellowship.

Conflict of interest: No financial conflicts of interest.

References