Commentary: Genes as instruments for evaluation of markers and causes

Robert J Glynn

In this issue of the journal, Casas et al. provide important information on a genetic determinant of C-reactive protein (CRP) with little apparent relationship to risk of coronary heart disease (CHD). The implications of these relationships for our understanding of a possibly causal role of inflammation in CHD can shed light on the contributions of Mendelian randomization to causal inference in epidemiology. Mendelian randomization gives a focused evaluation of causality, but it relies on strong assumptions and may be of limited usefulness in the evaluation of CRP as a marker of inflammation when the genetic variant has a modest ability to predict the phenotype of interest.

The mechanics of Mendelian randomization

Mendelian randomization capitalizes on the random allocation of a genetic variant that influences an intermediate phenotype such as CRP to evaluate the possibly confounded relationship of that phenotype with an outcome such as CHD. A variety of approaches have been used in actual practice for causal inference about CRP and outcomes. The simplest approach relies on qualitative comparisons between the effects of genetic variants on CRP and the relationship of CRP with CHD. Miller et al. made such comparisons when they found four SNP variants to have consistent and statistically significant associations with higher CRP across three populations, yet these SNP variants had no apparent relationship with risk of CHD. Generally, the term Mendelian randomization is reserved for studies that provide a formal statistical comparison of the relationship of a phenotype with an outcome, based on the impact of the genetic variant on the phenotype and outcome. Often an adjusted estimate of the association of the phenotype with the outcome is obtained through use of methods of instrumental variables, and this is the approach taken by Timpson et al. to obtain adjusted estimates of the relationship of CRP with components of the metabolic syndrome. Alternatively, Casas et al. use an approximate approach to compare the observed relationship of CRP with CHD with that expected from the relationship of a genetic determinant with CHD. Differences in interpretation of similar results across these three studies of CRP relate to different evaluations of the limitations of Mendelian randomization for causal inference, including population stratification, gene–environment interactions, and canalization. As these threats to validity are discussed elsewhere, I will focus on the statistical challenges to evaluation of CRP via Mendelian randomization that arise because of the skewed distribution of CRP, its heterogeneity across populations, the modest correlation of the genetic determinant with CRP, and the uncertain shape of the relationship of CRP with risk of CHD.

Inference in Mendelian randomization requires integration of information on three relationships: the impact of the genetic determinant on the intermediate phenotype; the association between the genetic determinant and the outcome; and the observed, but possibly confounded, relationship of the intermediate phenotype with the outcome. Casas et al. found that men who are homozygous for the rarer allele of the +1444C>T polymorphism of the CRP gene have a 21% increased level of CRP, a finding consistent with the relationship seen among the men studied by Miller et al. (Figures 1 and 2). The implications of this relationship for CHD risk are complicated because of the highly significant heterogeneity in mean log-CRP levels across the populations studied by Casas et al. and the presumed multiplicative effect of the mutation on CRP levels. Specifically, a 21% higher CRP-level associated with the genetic determinant assumes a greater absolute impact of this determinant on CRP in people with higher average CRP levels. Casas et al. assume a uniform increase of 0.68 mg/l in CRP associated with the mutation, which will underestimate the variability of the impact of a mutation that acts multiplicatively on CRP. While they do not allow for the variability of the mean CRP across populations under the strong assumption of a fixed-effects model, they do not allow for within-population variability in response to CRP.

Figure 1 Probability density of log-CRP in male participants in the PRINCE trial; 802 men with 10% homozygous for the T allele of the +1444C>T polymorphism of the CRP gene
Figure 2 Probability density of log-CRP in male participants in the Physicians’ Health Study; 499 men with 10% homozygous for the T allele of the +1444C>T polymorphism of the CRP gene

Also, the impact of the mutation would be less on a person at the median CRP level of 2.01 mg/l in whom a 21% increase would imply a 0.42 mg/l higher CRP level.

A strength of the analysis by Casas et al. is their consideration of a range of estimates from 1.45 to 1.92 for the odds of CHD in men above the second tertile vs those below the first tertile of CRP, based on findings from previous observational studies and a meta-analysis. Application of these odds ratios to calculate the expected impact of the mutation if CRP were causally related to CHD requires estimation of the difference in CRP levels between people above vs below these two tertiles. For this difference, Casas et al. use a value of 1.4 mg/l based on the observed difference between tertiles from one study. However, the first tertile is the maximum value for the lower risk group while the second tertile is the minimum value for the higher risk group, the difference between tertiles will substantially underestimate the difference in mean CRP levels between those above the second vs below the first tertile, and the underestimation is particularly large because of the skewed distribution of CRP. If log-CRP follows a normal distribution with a mean of 0.7 and standard deviation of –1.0, as indicated by their data, then the actual difference in mean CRP levels between these two groups is 6.3 mg/l, a 4.5-fold greater difference from the estimate of Casas et al. If we use 0.42 mg/l as the difference in CRP between TT subjects and C allele carriers, and 6.3 mg/l as the mean difference between those below the first tertile vs those above the second tertile, then the expected OR for TT homozygous subjects according to the approach of Casas et al. applied to odds ratios between 1.45 and 1.92 would range from 1.03 to 1.04, in good agreement with the observed relationship of the risk of CHD associated with the genetic determinant.

Strength of instruments, markers, and causes

Genetic determinants are often weakly related to an intermediate phenotype and the strength of this association has different implications for evaluation of markers vs causes. In each of the two populations shown in Figures 1 and 2, homozygosity for the rarer allele of the +1444C>T polymorphism explained <1% of the variance in log-CRP. Appropriate causal inference can still follow from instrumental variable approaches with low explained variance if distributional assumptions are met.

However, a low explained variance implies that an apparently null relationship between a genetic determinant and the outcome says little about the possibility that the intermediate phenotype is a useful marker of risk of the outcome.

Traditional observational epidemiological studies have limited ability to distinguish markers from causes. The criteria for causality proposed by Hill provide only a rough guideline for causal inference in the context of observational data. Often it is only possible to conclude that an exposure of interest such as CRP serves as a marker for an outcome such as CHD. It may well be that some underlying process such as inflammation affects CRP and it is inflammation that contributes independently to CHD. Indeed, as Figure 3 illustrates, genetic determinants can influence CRP, but they may be independent of inflammation, which might causally influence CHD and is marked by CRP. Recent evidence from the Framingham Offspring Study indicates that the available genetic determinants of CRP together account for <1% of the variance in CRP. Further, this study also found that all available lifestyle and biomarker correlates of CRP together account for 26% of the variance of CRP. The unexplained component of CRP may reflect a powerful independent marker of CHD risk.

Clayton and McKeigue suggested that Mendelian randomization may be most useful for evaluation of intermediate phenotypes that are the targets of specific interventions for their modification. Currently no commonly used therapies are specifically directed towards CRP reduction. However, the value of CRP as a marker does have potential clinical relevance as data suggest that statin therapy directed towards those with elevated CRP has potential benefits through joint reduction of lipid levels and CRP. Indeed, a large, ongoing randomized trial (the JUPITER trial) is evaluating the potential benefits and risks of statin therapy among those with normal LDL cholesterol but raised CRP.

Figure 3 Possible causal diagram for genetic determinants, traditional risk factors, inflammation, CRP, and CHD. Mendelian randomization lowers the likelihood that CRP is causally related to CHD but says little about the possible importance of CRP as a marker of inflammation.
depend on a causal relationship of CRP with CHD, but only whether CRP is a useful indicator for people at increased risk not identified by traditional risk factors. Thus, traditional epidemiological methods can identify markers of potential value for broad-based interventions, and a null result from a Mendelian randomization study will say little about the value of such interventions when genetic determinants explain little of the variability in the markers.

Disclosure
The author has a contract from Astra Zeneca to serve as the independent statistical monitor of its ongoing JUPITER trial.

References