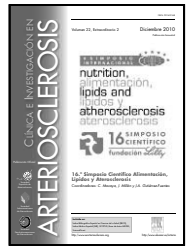
Clin Invest Arterioscl. 2010;**22(Supl 2)**:33-35

CLÍNICA E INVESTIGACIÓN EN
ARTERIOSCLEROSIS

www.elsevier.es/arterio



16.º SIMPOSIO CIENTÍFICO ALIMENTACIÓN, LÍPIDOS Y ATEROSCLEROSIS

Using large-scale epidemiological evidence to help evaluate biomarkers in cardiovascular disease

Emanuele di Angelantonio

University Lecturer in Medical Screening, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Glossary

Meta-analysis: a meta-analysis combines the results of several studies that address a set of related research hypotheses.
Bias: bias is systematic favouritism that is present in the data collection process resulting in misleading results
Prospective study: a cohort study that follows over time a group of similar individuals who differ with respect to certain factors under study, in order to determine how these factors affect rates of a certain outcome.
Random error: an error that has a random distribution and can be attributed to chance.
Regression dilution: regression dilution is a statistical phenomenon also known as "attenuation".
Polymorphism: a polymorphism is a genetic variant that appears in at least 1%of a population.

Despite the known importance of classical risk factors such as smoking, elevated blood lipids and high blood pressure to cardiovascular disease risk, the incidence of cardiovascular disease in individuals who lack these traits, together with evidence that interventions which successfully target such factors do not entirely abrogate the risk of disease has encouraged investigators to continue searching for novel biomarkers. Consequently, in recent decades, many circulating molecular biomarkers and thousands of genetic variants related to lipid, inflammatory, metabolic and haemostasic pathways have been investigated. To date, however, these novel factors have generally been investigated in individual observational studies typically involving only a few hundred cases and only few hundred controls which, due to the inherent statistical uncertainties of small sample sizes, may be prone to chance findings, selective publication biases and potentially over-stated conclusions[1]. However, the reliable identification and detailed characterisation of any such associations with cardiovascular disease is essential if ongoing research is to help improve our understanding of cardiovascular disease aetiology, better predict disease risk in individuals and suggest new targets for interventional therapies.

In the absence of individual studies of very large size, appropriate synthesis of the available reports of molecular biomarkers in coronary heart disease (CHD) by meta-analysis should provide a better preliminary indication of their relevance to CHD than can individual studies involving just a few hundred cases. This is because meta-analyses are less likely to be subject to random error than single studies, which due to their inherent statistical uncertainties may produce false-positive and false-negative results. Consequently, to enhance appropriate interpretation and to prioritise hypotheses for further investigation, there is an increasing need for systematic reviews of publications on biomarkers in CHD. Figure 1 suggests a schema for a staged approach to the evaluation of candidate biomarkers in CHD. This approach includes systematic reviews of published and unpublished data; measurement of emerging biomarkers in stored samples from existing large prospective studies; and the collaborative pooling of individual participant data from multiple studies.

Preliminary quantitative reviews ("literature-based meta-analyses") can help to prioritise research in CHD by: *a)* identifying risk markers for which the available evidence is, in aggregate, comparatively unpromising, encouraging the study of other, potentially more fruitful hypotheses; *b)* suggesting the need for new measurements in much larger studies than hitherto to achieve reliable results; *c)* indicating

E-mail: ed303@medschl.cam.ac.uk

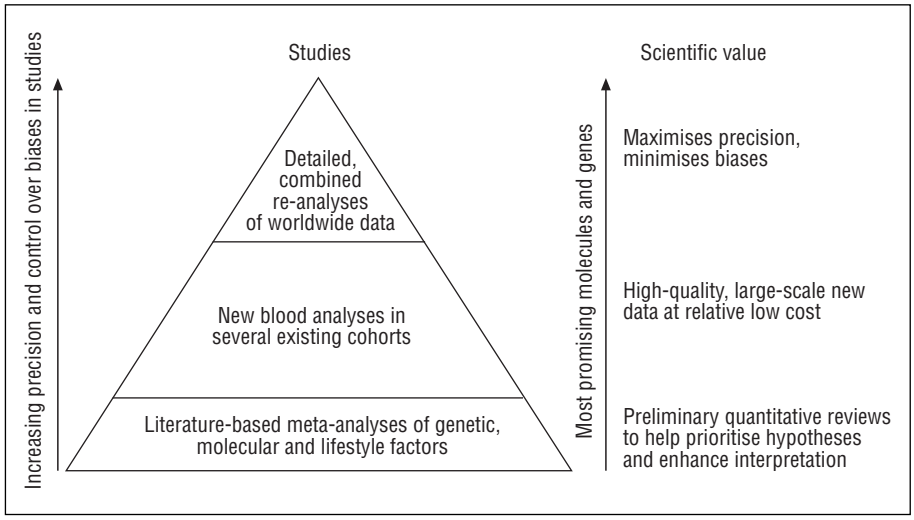


Figure 1 Outline of a staged approach to prioritise and evaluate novel and emerging markers in cardiovascular diseases.

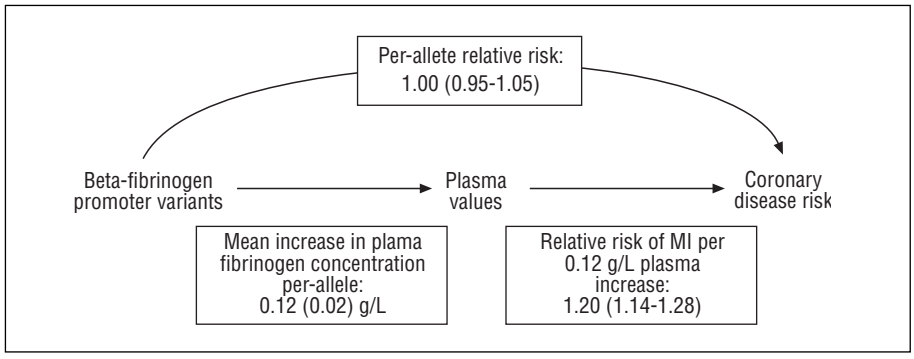


Figure 2 Integration of data involving gene → phenotype → CHD risk.

that existing data would, if properly brought together into a detailed synthesis, be sufficient to yield reliable results. However, such literature-based reviews cannot provide: *a)* precise estimates of risk marker-disease associations under a range of different circumstances, such as at different ages, in women and men, at different levels of established risk factors; *b)* reliable characterisation of the shape of any dose-response relationships; *c)* consistent approaches to adjustment for possible confounding factors; and *d)* detailed investigation of potential sources of heterogeneity. Moreover, most available assessments of emerging risk markers have related CHD risk solely to baseline measurements (which can lead to substantial underestimation of any associations due to regression dilution bias[2]) and, have based statistical adjustment for possible confounding factors only on baseline values (which can lead to residual biases). But, if a risk marker is of potential aetiological relevance, it may also be important to characterise in detail its degree of within-person variability, both to understand the sources of this variability and to enable appropriate correction for regression dilution[2].

Such uncertainties can be addressed by analyses of individual data from a comprehensive set of relevant prospective studies of cardiovascular outcomes (i.e., individual participant data meta-analysis). The value of this approach has been demonstrated by the Emerging Risk Factors Collaboration (ERFC)[3], an analysis of individual data on over 1.2 million participants in 110 long-term prospective studies in approximately general populations, including 75,000 incident major cardiovascular outcomes. The ERFC has, for example, demonstrated approximately log-linear associations for each of total and high density lipoprotein (HDL) cholesterol with CHD and other vascular outcomes[4]. These findings are of considerable public health importance, refuting earlier suggestions of threshold levels at which these established risk factors cease to be relevant, and demonstrating the importance of total and HDL cholesterol to CHD outcomes under a wide range of circumstances, notably in men and women and in the elderly for whom these risk factors were previously regarded by some authorities as unimportant.

Despite their advantages over individual studies of customary size, individual participant meta-analyses of several prospective studies of emerging risk markers may not distinguish reliably whether associations of particular biomarkers with CHD reflect a causal relationship or mainly a marker of established cardiovascular risk factors to which the biomarker is correlated, or mainly a marker of subclinical disease, or some combination of these possibilities. For example, the Fibrinogen Studies Collaboration has reported approximately log-linear associations of fibrinogen with CHD risk under a wide range of different circumstances[5]. The

magnitude of this association, however, reduced considerably following adjustment for several established cardiovascular risk factors[5], as could be expected given the large number of established and emerging risk factors to which plasma fibrinogen is correlated[6]. The existence of these many correlates makes it difficult, therefore, to determine to what extent the observed associations of fibrinogen with CHD risk are independent from these markers.

Focused genetic studies may help to overcome some of these potential limitations of observational epidemiology[7]. "Mendelian randomisation" experiments attempt to minimise confounding and avoid reverse association bias by measurement of common polymorphisms in regulatory regions of genes that have been reliably associated with differences in circulating biomarker concentration (but not with any known change in biomarker function). According to Mendel's second law, the inheritance of genetic variants should be subject to the random assortment of maternal and paternal alleles at the time of gamete formation. So, if the levels of a particular biomarker actually increase the risk of CHD, then carriage of alleles that expose individuals to a long-term elevation of that biomarker should confer an increased risk of CHD in proportion to the difference in biomarker levels attributable to the allele. Because of the randomised allocation of alleles from parents to offspring, potential confounders should be distributed evenly among the genotypic classes, and any bias due to reverse causation should be avoided because genotypes are fixed at conception and are unlikely to be modified by the onset of disease[7]. Hence, by helping to judge the likelihood of any causal associations in CHD and estimating their magnitude, such focused genetic analyses should help to prioritise biomarkers for further study (eg, as therapeutic targets) and elucidate disease pathways.

This approach has been applied to the study of plasma levels of fibrinogen (Fig. 2). A report of a null association of fibrinogen genotypes with CHD risk in a total of about 12,000 CHD cases and 18,000 controls, has decreased the likelihood of a major causal role for fibrinogen levels[8], but even larger numbers would be needed to exclude the possibility of a modest but still potentially important effect. For example, it has been estimated that greater than 15,000 cases and greater than 15,000 controls would be needed to confirm or exclude 5-10% increases in CHD risk per 1 SD increase in blood levels of C-reactive protein (CRP)[9]. The CRP-CHD Genetics Collaboration is, therefore, generating data and conducting pooled analyses of known relevant CRP genetic variants in about 46,000 CHD cases and about 150,000 controls from 47 contributing studies. This approach is being extended to the study of several other candidate biomarkers, including HDL-C, lipoprotein(a) and interleukin-6.

The potential limitations of Mendelian randomisation analyses include: the need for very large sample sizes because most genotypes have only modest effects on concentrations of biomarkers; the scope for residual confounding by unrecognised pleiotropic effects of genotypes; and the potential obscuring of causal associations by processes related to developmental adaptation ("canalisation")[9]. Furthermore, ideal Mendelian randomisation analyses should probably involve information on genotypes, biomarker levels and CHD status derived from the same individuals in a single very large prospective study (which, for clinical CHD outcomes, may require upwards of 20,000 incident CHD cases). In the current absence of any such studies, however, it has been necessary to combine information from several different studies, only relatively few of which may involve concomitant assessment of genotype, biomarkers and CHD status (indeed, in the case of fibrinogen, studies focusing on biomarker-CHD and gene-CHD associations have largely been non-overlapping).

Conclusions

Approaches that enable study of the separate and combined effects of genetic, biochemical and lifestyle factors should yield new scientific insights that contribute importantly to the prediction and prevention of CHD.

Conflict of interest

The author declares he has not any conflict of interest.

References

1. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.
2. Wood AM, White IR, Thompson SG, Lewington S, Danesh J; Fibrinogen Studies Collaboration. Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27,247 adults in 15 prospective studies. Int J Epidemiol. 2006;35:1570-8.
3. Danesh J, Erqou S, Walker M, Thompson SG; The Emerging Risk Factors Collaboration. Analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. Eur J Epidemiol. 2007;22:839-69.
4. Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, et al; Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. JAMA. 2009; 302:1993-2000.
5. Danesh J, Lewington S, Thompson SG, Lowe GD, Collins R; Fibrinogen Studies Collaboration. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. JAMA. 2005;294:1799-809.
6. Kaptoge S, White IR, Thompson SG, Wood AM, Lewington S, Lowe GD, et al; Fibrinogen Studies Collaboration. Associations of plasma fibrinogen levels with established cardiovascular disease risk factors, inflammatory markers, and other characteristics: individual participant meta-analysis of 154,211 adults in 31 prospective studies: the fibrinogen studies collaboration. Am J Epidemiol. 2007;166:867-79.
7. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1-22.
8. Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, et al. Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. Int J Epidemiol. 2006;35:935-43.
9. Danesh J, Hingorani A, Wensley F, Casas JP, Smeeth L; CRP CHD Genetics Collaboration. Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomisation. Eur J Epidemiol. 2008; 23:531-40.