required to settle the question grew and grew. My contribution consisted of encouraging my scientific staff members to keep going until they had enough patients to settle the matter—in this case 23,000 patients. The study showed definitively that the 677 mutation is associated with increased risk of CHD.9

So what have the past 18 years taught me about Mendelian Randomization studies? Anything I could say about the topic has already been said better in the excellent review by Davey Smith and Ebrahim,6 but I would still like to stress a few points.

The first thing I learned is that you need a simple well-defined phenotype. In the case of apoE the three genetic variants produced clearly different cholesterol levels, via a logical mechanism (although logic can be deceptive; the apoE2 variant was originally discovered as the cause of a rare type of hyperlipidaemia, and it took a while to establish that most of the homozygous E2/E2 subjects have low, rather than high, plasma cholesterol). I have to admit that the phenotype which I chased for over 20 years, namely an exaggerated versus a reduced response of blood cholesterol to diet, is too fuzzy and hard too measure to allow proper genetic analyses.

The second thing is that you need large numbers to get a result that will stick. That implies that only the urgent issues are worth pursuing: if the question that you are trying to solve is not sufficiently inspiring and challenging you will be tempted to give up when the first 100 or even 1000 subjects fail to give you a clear answer, which is often the case.

Finally, the apoE episode and all that followed taught me that my early switch from chemistry10 and biochemistry11 into nutrition science did not mean that I had wasted my time. The matrix algebra that I struggled with in trying to understand quantum chemistry later eased my path into biostatistics, and my PhD studies in molecular biology taught me the genetics that I needed to understand the apoE polymorphism. A training in the basic sciences can come in handy in unexpected ways.

References
clinical trials confirming that associations that have been found in observational studies are causal—by showing that treatment of particular risk factors using suitably specific therapeutic agents diminishes the risks of developing disease. The third important strand of evidence confirming (albeit indirectly) causality of a risk factor consists of studies in animal models of disease; the technology to create transgenic and gene-targeted animals has resulted in an explosion of activity in this field. Although such models often confirm the importance of risk factors in humans (for example, the development of stroke in certain hypertensive rat models, or of atheromatous disease in genetically engineered hyperlipidaemic mice), importantly they do not in all cases.

Investigators studying genetic susceptibility to complex traits such as CHD have looked on somewhat enviously at the steady flow of scientifically robust data originating from epidemiology and clinical trials groups, in contrast to the extreme difficulties that have been encountered attempting to identify genes contributing significantly to the population burden of conditions such as atherosclerosis, cancer, and obesity. Of course, the principal reason for this dichotomy is that ‘classical’ epidemiologists have so far been studying effects that are much larger than the effect of any single genetic locus is likely to be. For example, in the large International Study of Infarct Survival (ISIS) case-control study of myocardial infarction (MI), current smoking was associated with a relative risk for MI of 4.6 (95% CI: 4.1, 5.3) among 2554 cases of premature MI (males <55 and females <65) and 4831 controls, whereas in that same study the relative risk for MI among those with the ε3/ε4 genotype at the apolipoprotein E ε2/ε3/ε4 polymorphism relative to the ε2/ε3 genotype was only 1.17 (95% CI: 1.09, 1.25). The apolipoprotein E ε2/ε3/ε4 polymorphism is the only common genetic variant for which convincing large-scale evidence of an association with MI risk exists to date.

Despite the outstanding successes of observational epidemiology in recent decades, however, there is some concern that much of the ‘low-hanging fruit’ has now been picked, and that the identification of novel causal risk factors using classical methodology will become exponentially more difficult. In a recent International Journal of Epidemiology review, Davey Smith and Ebrahim drew attention to several instances (mostly to do with vitamin intake and cardiovascular or cancer risk) where the findings of observational epidemiology had not been confirmed by subsequent clinical trials; these examples should be added those hypothesized causal associations which essentially cannot be validated in humans by classical means at present because no suitably specific agent has yet been developed for clinical trial purposes (a good example being the association between plasma C-reactive protein and cardiovascular risk). There are three principal reasons why further such difficulties might lie ahead. Firstly, the sizes of effect that are being claimed for novel risk factors are smaller than for the classical risk factors; this means that even larger studies will be required to produce robust results. Secondly, the associations between novel risk factors and disease might be confounded by other inaccurately measured or unmeasured factors which are themselves related to the risk of disease. For example, plasma fibrinogen, a hypothesized novel risk factor for CHD, is very strongly associated with smoking, a causal factor: while statistical correction for smoking in the assessment of the relationship between fibrinogen and CHD risk is possible, measurement error would render the correction likely to underestimate the effect of smoking (Figure 1). Thirdly, for some diseases, ‘reverse causality’ may be a problem—in the case of atherosclerosis, it is known that the process begins in early life, and pathological studies clearly show its inflammatory component, so those with baseline higher levels of high-sensitivity C-reactive protein or other inflammatory markers may have subclinical disease causing their inflammatory marker profile rather than the other way round (Figure 2).

All the above considerations should be weighed in a context of the near-future capacity of proteomic technologies to measure not just a few but many thousands of plasma proteins on stored blood samples. Imminently, epidemiologists will be confronted by a very substantial number of disease-factor associations many of which will be weak, confounded, or due to reverse causality but some of which are very likely to reflect hitherto unsuspected causal pathways. The existing supporting strands of evidence for causality (clinical trials and animal models) will not provide much help.

Katan was among the first to recognize that genetics could potentially contribute importantly to the debate regarding causality. His brief contribution to the Lancet correspondence pages in 1986 directly addresses a topic of increasing debate among geneticists and epidemiologists in 2003. The problem he addresses is that of the association between low serum cholesterol levels and cancer and the main obstacle he addresses is that of ‘reverse causality’—in this case, that pre-existing occult tumour might cause lower cholesterol levels, rather than lower cholesterol levels causing cancer. In the early 1980s, the central role of the apolipoprotein E molecule in cholesterol metabolism was discovered, and the association between the E2, E3, and E4 isoforms of that molecule (determined by the ε2, ε3, and ε4 alleles of the apolipoprotein E gene) and blood levels of low density lipoprotein cholesterol were observed in a number of populations. Katan reasoned that, since apolipoprotein E genotypes were determined at conception, they would determine long-term differences in blood cholesterol between individuals and could not be altered by the subsequent development of disease. Thus, if the causal arrow pointed from low cholesterol to cancer, there would be a higher frequency of the allele predisposing to lower cholesterol (ε2) and a correspondingly lower frequency of the allele predisposing to higher cholesterol (ε4) among cancer cases, whereas if it pointed in the other direction genotypes would be randomly distributed.
among cases and controls (Figure 3). So far as I am aware Katan's hypothesis was never tested in the way that he proposed it. Subsequent to Katan's Lancet letter, Gray and Wheatley in 1991 proposed a similar use of genetic data to avoid bias when comparing bone marrow transplantation (BMT) with chemotherapy, and coined the term ‘Mendelian randomization’.4,5 If anything, Gray and Wheatley's approach was even more ingenious than Katan's. In the area of leukaemia treatment, by the time their article was written, it was already thought to be unethical to withhold BMT from those who did have human leukocyte antigen (HLA)-matched donors (a minority of patients). So, a randomized trial comparing allogeneic BMT with no BMT (or extra chemotherapy) might not be possible on ethical grounds, and comparing these treatments using observational data would be subject to major biases, chief among which would be selection bias. Gray and Wheatley proposed that comparing the survival of those patients who had HLA-compatible siblings with those who did not would constitute an unbiased assessment of the value of allogeneic BMT. This is because the ‘allocation’ to either group would have been made years before the onset of disease, and therefore no selection bias could occur. This approach assumed that complete data on HLA typing would, in general, be available, and that most suitable patients with donors would go on to receive BMT. Here, the random segregation of HLA alleles in the meioses producing the affected individual and any siblings produces a de facto randomization entirely analogous to that employed in the clinical trial setting. A number of subsequent studies have adopted Gray and Wheatley's approach in acute myeloid leukemia, and consistently better survival among those with HLA-matched donors has been observed.5,6

More recently, the phrase ‘Mendelian randomization’ has been applied by our group to an approach which can be used in large case-control studies to address the causality of hypothesized plasma risk factors whose association with disease might be due to confounding with other, causal risk factors (for example, the hypothesized factor plasma fibrinogen concentration, the causal factors smoking, diabetes and higher levels of blood lipids, and CHD).7 The approach consists of identifying a regulatory genetic polymorphism which is associated with differences in the plasma levels of a hypothesized risk factor, measurement of the risk factor in a large number of cases and controls, and determination of the strength of three associations: between plasma factor and disease risk, between genotypes and plasma factor, and between genotypes and disease. In this case, because ‘Mendelian randomization’ takes place at conception, the genotype-determined differences in risk factor levels should not be affected by the presence of other potentially confounding factors (in the absence of significant gene-environment or gene-gene interactions) and thus if association between the plasma risk factor and disease is causal, it should be reflected by an association between genotype and risk which is to some degree commensurate with the ‘composite’ associations between genotype and plasma factor, and between plasma factor and disease (Figure 4). This assumes no direct effect of the polymorphism on disease risk other than through its effect on the risk factor (by a pleiotropic effect on another causal factor) and also assumes that genetically determined and environmentally determined differences in plasma levels of the factor will be equal with regard to any influence on risk. Katan's (and Gray and Wheatley’s) argument can only be used in this novel way in studies of many thousands of cases and controls, not only because of the smaller size of the novel hypothesized associations now under study, but also because the influence of any individual common genetic polymorphism (with a few exceptions such as the angiotensin-I converting enzyme I/D polymorphism)8 on a plasma factor is, in most cases so far described, small. This means that to detect definite differences between the ‘expected’ relationship between genotype and disease given the strengths of the composite associations and the actual genotype–disease relationship, large numbers are needed.

This novel use of ‘Mendelian randomization’ has already generated some interest from both geneticists and epidemiologists—Clayton and McKeigue identified the approach as potentially more useful at detecting causal pathways amenable to modification than approaches focusing on gene–environment interaction,9 while Davey Smith and Ebrahim considered the approach as likely to be more robust than much conventional observational epidemiology. As both these authors, and Little and Khoury have observed, however, a number of methodological issues remain, in particular our very incomplete current knowledge of haplotype structure throughout the human genome.10 One way in which ‘Mendelian randomization’ could be confounded is by the presence of polymorphisms having pleiotropic effects on disease risk in association (linkage disequilibrium) with the marker polymorphism typed in a study, or indeed by unsuspected pleiotropic effects of a marker polymorphism thought to be only regulatory in its effects. For some plasma factors (including fibrinogen), linkage disequilibrium relationships between polymorphisms influencing plasma levels are well defined, and genotype-factor associations...
are supported by large numbers of studies. However, for other factors (such as C-reactive protein), such relationships are at a much earlier stage of definition.\textsuperscript{11} Identification of regulatory polymorphisms and characterization of haplotypes which influence plasma levels of hypothesized risk factors must be a priority if this approach is to succeed—this will be facilitated by data from the international haplotype mapping consortium,\textsuperscript{12,13} and by novel technological approaches to rapidly identify regulatory polymorphisms in both coding and non-coding genome sequence.\textsuperscript{14,15}

As proteomic technology becomes progressively more applicable to large sample sizes, it will eventually be possible to study the expression of many thousands of proteins in many thousands of individuals. Establishing the causality of any of the thousands of associations likely to emerge will be extremely challenging, since the development of the number of animal models or therapeutic agents necessary to test this number of associations will be impossible. A potential next step might be the early investigation of these associations by ‘Mendelian randomization’, to focus attention on those with more substantial evidence of causality. The development of animal models and/or the conduct of intervention trials could be restricted to those proteins that have passed through the genetic ‘sieve’.

Nearly 20 years on, we now have the genetic technology to make widespread use of the insights of Katan, Gray, and Wheatley. Such is the promise of this approach that, although I doubt that either Katan’s letter or Gray and Wheatley’s paper is currently a ‘citation classic’, I would be willing to bet they will be in 5 years.

References
\textsuperscript{1} Keavney B, Parish S, Palmer A et al. Large-scale evidence that the cardiotoxicity of smoking is not significantly modified by the apolipoprotein E epsilon2/epsilon3/epsilon4 genotype. \textit{Lancet} 2003; \textbf{361}:396–98.
\textsuperscript{4} Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. \textit{Bone Marrow Transplant} 1991; \textbf{7}(Suppl.3):9–12.
\textsuperscript{12} Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. \textit{Trends Genet} 2003; \textbf{19}:135–40.
\textsuperscript{13} Cousin J. Human genome. HapMap launched with pledges of $100 million. \textit{Science} 2002; \textbf{298}:941–42.
\textsuperscript{14} Ding C, Cantor CR. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. \textit{Proc Natl Acad Sci USA} 2003; \textbf{100}:3059–64.
\textsuperscript{15} Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. \textit{Nat Genet} 2003; \textbf{33}:469–73.