Commentary: Mendelian randomization, 18 years on

Martijn B Katan

I have no notes left regarding writing ‘Apolipoprotein E isoforms, serum cholesterol, and cancer’, but I think I thought it up in Hawaii. I passed through Hawaii on my way to the US from Melbourne, where I had given a talk on diet, low cholesterol, and cancer at the 7th International Atherosclerosis Symposium in October 1985. At that time the dangers of a low serum cholesterol level was a hot topic; several scientists thought that a low cholesterol increased your risk of violent death or cancer. This rested both on observational associations and on the outcomes of early cholesterol-lowering trials. The idea became less plausible after the big statin trials showed no relation between lowering of cholesterol and the rates of violent death or cancer, but those studies came later.

I myself was convinced that the association was spurious or due to reverse causality, i.e. to occult tumours causing a lowering of cholesterol in future cancer patients. I must confess that that conviction rested partly on data and partly on emotion. Cholesterol had been a subject of fierce controversy for decades, and the scientific debate was frequently distorted by commercial interests, with dairy, meat, and egg producers on one side and margarine and oil producers on the other. Scientists were also divided, and each side distrusted whatever data the other side came up with. I had attempted to stay in the middle but I had always felt that the data in favour of cholesterol lowering were strong, and I was wary of theories that a high cholesterol level might be better than a low level. But what type of data could resolve this issue?

This is where genetics, with its clear causal pathways, should be able to help out. In my talk in Melbourne I had already argued that patients with the rare genetic disease, a-betalipoproteinaemia, do not get premature cancer even though they have almost zero plasma cholesterol levels. However, when I thought the matter over after the Melbourne meeting I realized that there were too few patients with this disease worldwide to settle the issue. When I tried to think of other subjects with low cholesterol levels, the apolipoprotein E polymorphism came to mind. I had first heard about this polymorphism at the European Lipoprotein Club meetings in Tutzing in Bavaria; these were informal gatherings where I had learned a lot about lipids. The apoE-2/3/4 polymorphism had been frequently discussed there both by its discoverer Gerd Utermann and by Gerd Assmann, who from early on had an interest in applying the new genetics to population studies. After several years of confusion it had become clear that cholesterol levels increased from the apoE2 to the apoE3 to the apoE4 phenotype. I was a molecular biologist by training, and although I was new to the lipid field I felt at ease with genetics. So it was obvious to me that there were plenty of people who carried one or two copies of the apoE2 gene, and that the large majority of them had relatively low cholesterol levels but were in other respects quite comparable with people who only carried copies of the apoE3 and apoE4 variants of the gene. Importantly, none of these people knew about their genetic status or their cholesterol levels, because this was before the era of large-scale cholesterol testing. So the E2 subjects lived their lives in exactly the same way as the E3s and E4s; a perfect experiment of nature. Each had had these genes and these low cholesterol levels from birth, so there was no need for a prospective study; measuring the apo E phenotype in cancer patients and matched controls was just as good as measuring it in a large number of newborns and following them to see who developed the disease and who did not. It had recently become feasible to determine apoE phenotypes in thousands of patients by means of protein electrophoresis (this was before the days of DNA sequencing), so I thought that a case-control study of apoE phenotype in cancer patients versus controls could settle the question of low cholesterol and cancer. As I was not an epidemiologist and did not have access to patients myself I put the idea in a letter to the Lancet, hoping that someone else would take it up. Although no one did I still felt pleased with this clean and logical approach to an important problem. However, I never felt that I had invented something new, because genetic experiments of nature had been invoked many times before to explain the aetiology of diseases; thus in the days when the role of low density lipoprotein (LDL) cholesterol in heart disease was controversial, the high rates of coronary heart disease (CHD) in familial hypercholesterolaemia had been a strong argument in favour of causality.

I remained interested in applying molecular genetics to the study of diet and heart disease, but most of our studies turned out negative, especially those of the genetic basis of hypo- and hyperresponsiveness to dietary cholesterol. My most successful involvement in a Mendelian randomization study was in a meta-analysis of the methylene tetrahydrofolate reductase (MTHFR) C677T mutation, a study in which I played a minor but, I like to think, significant role. The MTHFR enzyme converts folate into its biologically active form, and the mutation in base pair 677 causes a reduction in the activity of the enzyme which results in reduced plasma folate and increased homocysteine levels. The meta-analysis was set up to decide whether folate, and by implication homocysteine, was causally involved in CHD. As the data collection progressed it became clear that the relative risk associated with an increase in homocysteine was lower than originally thought, and therefore the number of patients
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