EDITORIAL

Whatever happened to the notion of controlled experiments?

I was asked recently to review a book that comprised an excellent set of research reviews on subjects of great topical interest relating to the roles and physiological effects of fat in the human diet. Research reviews are often disappointing in that they present a catalogue of information on a topic in a 'take-it-or-leave-it' fashion without much attempt critically to examine the original publications to assess for the reader the validity of their conclusions. Such catalogues can be useful in that they save the reader the burden of conducting a literature search but it is only when the reviewer applies his critical faculties that we appreciate the poor quality of much published literature.

The book of which I speak contained at least two of its twelve chapters in which this critical faculty had been exercised to a praiseworthy degree. Both highlighted fundamental deficiencies of many of the publications that had been reviewed; publications that had often been instrumental in developing widely held views of the influence of fats in the diet. The most outstanding faults had been in the design of nutritional and epidemiological experiments and, in particular, the quality of the controls employed One reviewer stated that 'about one-third of the studies [that he had reviewed] were poorly controlled and *one-third completely lacked a control group*' (my emphasis). The second author began an early section in his chapter thus: 'Evaluation of the results of many published experiments on the effects of diet on serum lipoprotein metabolism is hampered by poor methodology'.

Disturbing though these assertions are, they did not come in any way as a surprise, since I myself had come fresh from completing a review of a similar topic and had needed to revise my former views of certain phenomena when I discovered that some of the original experiments did not possess the impeccable quality I had fondly imagined. Herein lies another disadvantage of the 'catalogue' style of review. Certain publications are continually being cited as providing evidence for such-and-such a phenomenon with the result that they are thereafter automatically quoted as doing so and the number of people who have ever examined the original publication grows ever smaller. We become content with 'reviews of reviews'.

Apart from the inadequacy, or worse, complete lack of a control group, the following were cited by the review authors quoted above as being faults in design and methodology: trends over a period of time are not taken into account and dietary periods are rarely sufficient to do so anyway; the number of subjects (I am here considering only 'human' experiments) is frequently insufficient for biochemical variables, that fluctuate within an individual and are not related to diet, to be averaged out and for type II errors to be minimized. They go on to say that 'obviously, food intake of experimental subjects has to be carefully controlled as lack of adherence makes the data useless'. Yet, how often is this done? Very infrequently, I suggest. Quite often the authors claim that the dietary component being studied is X (for example, to continue the lipid analogy, a particular fatty acid or type of fatty acid), whereas in reality the test diets differ in many other respects, such as their content of many minor components that might well have potent biological effects (e.g. cholesterol or minor plant sterols, fat-soluble vitamins and many others). The conclusion that an observed effect is due solely to component X cannot rigorously be sustained, yet the result is quoted by a reviewer anxious to add support to a particular argument that is being developed, and from that day on enters the literature as a proven fact that will rarely again be challenged. Even the most critical reviewers fall into the trap of quoting dubious evidence when they are hard put to it to find enough material to support a particular line of argument. This poses several problems for the critical scientist and his conscience.

Few of us are involved in pure research for its own sake: most of us are concerned in some way or other with the practical application of research results. We may be on committees concerned with making recommendations for 'healthy eating', updating recommended dietary standards, organizing relief programmes, advising the food industry on appropriate strategies, recommending research programmes or approving research grants. We may be attempting to inform students through our lectures, or the public through articles in the newspapers, magazines, radio or television programmes. All these activities require, in some degree or other, a review of the scientific literature. The first problem is, therefore, how rigorous should we be? If the reviewer that I quoted above is correct that about two-thirds of the papers reviewed should be disregarded, our conclusions may well be rather different than if we had accepted most at face value. I constantly hear the plea: 'We cannot wait until we have absolute proof: we should accept the consensus of the evidence that we do have and act on it'. While I can sympathize with the need to take positive action on what may seem to be sufficient evidence. I am inclined to think that, on the basis of the arguments outlined here, we do so at our peril. I will go further and suggest that most of our conclusions and advice are based on insufficient or dubious evidence.

The second problem is one for the journals and their editors. Most (but not all) of the papers that are reviewed will have been subjected to peer review before publication. Does this mean that for two-thirds of the time, editors are passing papers that are unfit? Does this apply to our Journal in which we have taken a certain amount of pride (justifiably?) in editorial rigour? Refereeing can never be entirely based on absolute scientific standards: it always involves an element of human judgement. What one referee will find acceptable, another will reject as unsound. Where there is total conflict about publication (a fairly rare occurrence in my limited experience with this Journal) the chairman holds the balance of power but he, too, is human.

It can be argued whether there is, indeed, a good case for accepting only those papers that are demonstrably flawless. Nutrition research, especially where it involves human subjects, is difficult. Ethical problems abound. The biggest single problem is (in my view) the measurement of what people actually eat. So-called rigorous methods for measuring food intake almost certainly distort people's normal eating habits. On the other hand, many apparently widely accepted epidemiological relationships (e.g. between dietary fat intake and cancer) are based on intake estimates such as food balance sheets using trade statistics that provide unacceptably crude estimates of nutrient intakes. Should we reject all work that may have needed to included some 'soft data' because that was the best that could be achieved in the circumstances? Is it better to have a flawed study than no study at all? There may well be a case for accepting work that is generally good but that has a number of weaknesses as long as we improve the mechanisms for making the reader aware of those weaknesses. Greater awareness by authors themselves of the weaknesses of their own work and the ability to express the reservations in the discussion should be encouraged. Editors, too, may need to appreciate the contribution a paper (including, importantly, those with 'negative' results) makes to a broad area of research as well as assessing its attributes in isolation. But is a study with no control one that we should not tolerate under any circumstances?

MICHAEL I. GURR

Printed in Great Britain