STATISTICS IN MEDICINE, VOL. 3, 401–403 (1984)

COMMENTARY: OBSERVATION VERSUS EXPERIMENT

HENRY BLACKBURN

Division of Epidemiology, School of Public Health, University of Minnesota, Stadium Gate 27, 611 Beacon Street S.E., Minneapolis, Minnesota 55455, U.S.A.

I would like to discuss how to make more explicit the indications and conditions for observational and for experimental studies in the search for causes. A better definition of these indications and conditions would reduce unnecessary controversy and help our peers in the review process. It would also recognize that there *are* differences in the strength of inference possible from different study approaches and designs, though each has a role to play; one does not necessarily replace the other.

First, observational studies. Where are they appropriate and cost effective?

We have heard a good deal today about when the randomized clinical trial (RCT) is *not* feasible, so that we are required to rely on observation. Many aspects of infeasibility are societally determined, i.e. cost, ethics and the quality of peer review available. For example, we cannot randomize babies to specified alcohol and smoking habits. But observation may also be appropriate when an experiment is infeasible because of physiological constraints. For example, one recent case series and a few case–control studies suggest that apolipoprotein A-1 may play a role in atherosclerosis independent of the high density lipoprotein (HDL) level. Despite the apparent independence of their effects, however, one subfraction probably cannot be modified independently of the other in an experiment among humans. To learn about the causal role of subfractions, we thus must turn to better observational studies.

The RCT is also not feasible when therapy cannot be delivered effectively. There is now a wide concern about 'usual', 'optimal' or 'maximal care' control groups when subjects are referred for treatment outside the purview of the experimenters. We shudder still about the 'usual care' group in the Multiple Risk Factor Intervention Trial (MRFIT).[1] Their 'usual care' was affected by rapid changes in society and medical practice, so that intervention in that control group may have differed less from the 'special intervention' group than had been originally expected.

An RCT may also be inappropriate for very complex causal issues. Today's discussions have suggested repeatedly that only very plain and simple questions about treatment for or prevention of a disease can be effectively answered by an RCT.

Can we propose other general indications and criteria for the use of observational studies along with, or instead of trials?

We may view controlled observational studies in a hierarchy of formality. Historical control studies are the least formal of the controlled designs, just 'one step up' from the case-series. They may be used to confirm a hunch rapidly, or to extend a limited clinical observation or anecdote. Historical control studies are effective in situations when clinical judgement also works, i.e. when 'what you see is what you have'. They may be appropriate to generate hypotheses for more definitive studies, when the outcome and treatment are relatively short-term, or when the outcome

0277–6715/84/040401–03$01.00

is not highly variable—e.g. when most cases die. Additionally, these observational methods are appropriate when there is good knowledge about the confounding variables, and when these confounding variables can be reliably measured. Observational methods can also be used when dealing with relatively uncommon events and when only small samples or few cases exist, and when one can afford to 'play around' by such a pre-test approach. Finally, observations are appropriate when there is simply a plethora of good ideas, possibilities and hunches available about cause, therapy or prevention.

The case–control study is more formal in control design. It is indicated and appropriate when the above criteria are present but especially desirable when there is access to cases and controls representative of the population from which they both derive. We think of the marvellous life that investigators must have in such places as Rochester, Olmstead County, Minnesota or Gothenburg, Sweden, where cases and controls represent the community, where total access to well documented records is available and there is complete ascertainment of cases.

A cohort study is yet again a step upward in formality. It reduces sources of selection bias and accounts for temporal relationships because the variable under study is identified before the onset of disease. These studies can be used when the same general conditions prevail as above for other types of observational studies, but when, in addition, there are resources available to cover the relative inefficiency of this approach. (The cost of case finding and the identification of the 'excessive' numbers of controls in cohort studies can be expensive). Perhaps the best of all observational worlds is when both *cases and controls* in a case–control study can be chosen from a *cohort* representative of the population.

We appear to have agreed here that the RCT is the gold standard when the question to be studied can be stated clearly, therapy is simple, the disease is not complex, one can afford to pay for the greater knowledge gained and there is a need to determine relatively *small* effects in a phenomenon that affects masses of people. It may particularly obtain when the phenomenon is a long-term chronic disease process, the outcome is delayed, treatment effect lags, and there are inadequate ideas or few options available for treatment or prevention.

The 'public health trial' has not been mentioned today. I will mention it. The community-based, population-wide trial, in contrast to randomized trials in high risk persons or units, may best be considered when (a) the disease and treatment are more complex than can be handled in a randomized clinical trial, (b) we are dealing with a mass condition and large disease burden not being handled adequately by the medical system (i.e. life-style modification), (c) the major causes of the mass phenomenon appear importantly extrinsic (i.e. the result of powerful environmental-cultural factors rather than factors intrinsic to the individual), (d) there is a clear public need for development and testing of practical methods for the transfer of technology or for evidence of effect preparatory to wide health action and (e) it is necessary to generalize to the entire population as in public health policy. In these circumstances, the development and testing of the feasibility, cost and effect of population-wide strategies are needed. I believe that it is possible to enhance the design of such community or public health trials to strengthen the causal inference of the preventive effect (over and above the background 'noise' of rapid and wide societal change) by using these design elements: control (comparison towns), repetition (staged entry), sensitive measurement of risk and disease trends (cohort surveys for individuals and cross-sectional surveys for whole community change), measurement of dose-effect (individual change according to exposure intensity), and measurement of linkage between the health programme and change in risk factors and disease rates. The NHLBI is now supporting a few such population trials in cardiovascular disease prevention.

I sense that many of us here think that observational approaches are also strengthened when the investigators can synthesize from a wide base of evidence, and can bring to bear wide experience,

several skills, and a knowledge of mechanisms from the clinic and from the laboratory to the population situation and the larger public health issue.

We seem to accept that observational studies are appropriate, even adequate, when random variation is minor compared to the effect (one thinks of treating scurvy). Observation is also the particularly important mode when there is more than one possible answer. Dr. Feinstein's point about the likelihood of misinterpretation when opposite effects actually occur within the treated population is relevant here. Also, we seem more inclined to accept observational approaches when there is great confidence that time is not a strong factor affecting the issue, when the quality of medical records is good, when complete ascertainment of cases is possible and one can deal with representative populations. In fact, we asked ourselves this question, with an obvious answer: are not *all* medical and public health decisions based ultimately on observation and logic? Surely the interpretation and application of outcomes from RCTs are based on logical deduction (or otherwise). For example, some people have inferred, erroneously I believe, from the inconclusive MRFIT Trial results in high risk, middle-aged men, that it makes no difference whether most people stop smoking, or have their high blood pressure controlled or modify eating patterns and blood lipoprotein levels. Some have erroneously, I believe, extended the absence of clear effect in MRFIT to the idea that a policy *to prevent high risk in the first place* or in the entire population is inappropriate. Some of us also think it reasonable to extrapolate results of trials of one type of hypoglycemic agent or beta blocking drug or contraceptive or aspirin-like drug to another—in the same general family—rather than require new trials for each. Some also think that the Coronary Artery Surgery Study[2] results can reasonably be extrapolated to current by-pass techniques and the LRC Coronary Prevention Trial extrapolated to cholesterol lowering in general. Others would quarrel with such extrapolations.

We have certainly reached consensus here that *all* methods of study and inference are enhanced by (a) clarifying the questions, (b) reducing variation in measurement and (c) reducing sources of bias. We agree that whereas observational studies can be good science, the randomized clinical trial, if well done, is the best that science can offer.

REFERENCES

1. Multiple Risk Factor Intervention Trial Research Group. 'Multiple Risk Factor Intervention Trial', *Journal of the American Medical Association*, **248**, 1465–1477.
2. CASS Principal Investigators and their Associates. 'Coronary Artery Surgery Study', *Circulation*, **68**, 939–950.