General Practice Observed

Clinical Research in General Practice*

A. M. W. PORTER

British Medical Journal, 1971, 1, 600-602

Clinical research is full of difficulties, and in general practice we have our full share, perhaps more than our share. Thus one welcomes the opportunity to meet from time to time, to learn from each other, and to discuss our research problems. I want to start by considering a problem which is fundamental to research in general practice.

The Eremitical Life

In our branch of research at the moment there are many communities and few hermits. Joint studies and multicentre studies are commonplace; studies by one or two doctors are relatively rare. As Cardinal Heenan has observed, before writing to *The Times* these days you must first collect a cluster of signatures. It seems no longer to suffice that one or two men should design and execute a study. In 1971 an army must take the field, with a general to lead the troops, staff officers to co-ordinate manoeuvres, and the usual supporting tail to wag away; Richard Bright would not have been able to squeeze them all into the sluice. Is this trend towards co-operative studies and away from individual effort either necessary or desirable? I would like now to consider some of the advantages and disadvantages of the one-man versus the communal study.

Every study, however humble, is a formidable undertaking; it requires enthusiasm and exertion. Here is the first advantage which the lone researcher has—he is highly motivated to complete and publish his study. By contrast the doctor fulfilling someone else's protocol is, in many ways, little more than a technician; a sense of obligation may be present but rarely enthusiasm.

If enthusiasm is the first problem uniformity of assessment is the second. It is not difficult to determine whether a patient is dead or alive after a treatment has been given. It may be very difficult indeed to decide in a psychiatric study whether a patient is improved or not, and, if improved, the extent of the improvement. Put more precisely "interobserver errors" may be a problem when many observers try to assess the progress of different patients. Provided the protocol is sound the lone observer has only to worry about changes in attitude with the passing of time—that is, "intraobserver errors." Here the hermit probably has the advantage.

Bound up with this question of assessment is that of communication. It is not just a question of consistency in diag-

nosis criteria, in the use of rating scales, and in other methods of assessment. All sorts of things may—and do—go wrong during a study. Let us suppose, for example, that the protocol of a trial requires that the patient provides an early morning specimen of sputum for culture. Does it matter if he expectorates after cleaning his teeth? Toothpaste may not be very good for *Haemophilus influenzae*. Somehow the point was overlooked during the pilot study; the protocol must be amended at once. Should one discard the early admissions to the study, define the matter by retrospective questioning, or just carry on? One man can clear up the muddle, but imagine the difficulties of communication if 10 investigators are involved. Thus the single-handed investigator has the advantage in terms of motivation, uniformity of assessment, and communication.

But, as always, there is a snag. It may be, and often is, very difficult to achieve samples of adequate size to show meaningful differences. It would be foolish, for example, for one general practitioner to attempt to study the morbidity and mortality associated with the contraceptive pill; the study of the Royal College of General Practitioners is an example of the legitimate use of many participants. But there are other studies in general practice which may well be done by one or two doctors; we may have to go on a little longer, but that need not necessarily matter. We should cultivate the perseverance of the old man with a gong who bumped at it all day long.

The Medical Research Council's trial of treatments in depressive illnesses is an example of a study where too many participants were used.¹ Altogether 269 patients were admitted by 55 physicians, which is an average of just under five patients per doctor. From time to time these doctors may have been absent from their outpatient departments, and it seems likely that the ratio was even smaller than this. How can uniformity of assessment be achieved in the presence of such fragmentation? How can the right hand know what the left hand is doing? Depression is not a rare disorder; would it not have been preferable for a handful of enthusiasts to have persevered until they had achieved adequate sample sizes, even if it did take longer? Difficulties multiply with size.

He Follows His Nose Wherever He Goes

My next problem is a more profound one. Before starting a study we should always sit down and formulate a hypothesis to be accepted or rejected by the experimental evidence. What are we doing and why are we doing it? Better still, we should turn the hypothesis around into the negative form, the null hypothesis, if only to remind ourselves that the onus is on us to refute the negative rather than affirm the positive. The experimental evidence cannot affirm that measures A, B, and C do improve the expectation of life for cancer patients, but the experimental evidence may be inconsistent with the

*Paper read at the sixth meeting of the General Practitioner Research Club on 30 January 1971.

null hypothesis that measures A, B, and C *do not* improve the expectation of life for cancer patients, and thus the possibility emerges that measures A, B, and C *may* improve the expectation of life for cancer patients. If the null hypothesis is consistently rejected by different experiments then the possibility can become a probability.

I do not know of anything more difficult than the effort of conceptual thinking required to formulate a concise hypothesis before a study begins. It asks that we condense an idea out of the vapours of vagueness—out of a primeval wasteland of thought—and that we then turn it into a precisely framed hypothesis to be presented to the judgement of sound experimental method. But, however difficult, it is vital. It is, for example, all too easy to announce that you are going to do a study to examine the effect of a district nurse attachment on the demand for home visiting; one could easily and cosily drift into this sort of study. We might emerge at the end with a whole mass of data which suggest certain conclusions ("it was found that the coffee consumption of the practice increased in geometrical progression with the number of ancillary staff employed . . ."). These conclusions are all too often presented to the reader in a manner which implies that they represent acceptance of hypotheses defined before the work began. But do they? "Jack, Jack Joe, Bent his bow, Shot at a pigeon And killed a crow." That Jack is a nice chap of best intentions does not change the fact that the corpse is a horrible scientific mess.

An investigator who formulates or changes his hypothesis after or during the implementation of the study is, in fact, doing a retrospective study, with all the disadvantages that attach to such a study. Two things are best committed to paper before one begins and thereafter should remain unaltered and inviolable: the first is the null hypothesis to be accepted or rejected by the study, and the second is your bias towards the likely outcome. It is, I think, courteous to the reader of the final report to reveal what was in your mind before you actually started; he may then better judge the possible influence of your bias on the outcome of the study. The proper sequence may therefore be summarized as follows: background knowledge, vague idea, precise null hypothesis, experimental method, rejection or acceptance of null hypothesis, conclusion, result properly joins background knowledge. The improper sequence is: vague idea, failure to formulate a null hypothesis, experimental method, formulation of hypotheses compliant with experimental data, scientific chaos, conclusion improperly joins background knowledge.

Every study, however precisely defined and framed, has in fact an element of vagueness and is *inter alia* an observational study from which the most surprising and unexpected observations may spring; the classical example of this was the discovery of penicillin by Sir Alexander Fleming. Such incidental observations should, however, be extracted and then embodied in the background knowledge from which a fresh hypothesis and a fresh experiment can emerge. It may well be that the new observation is of such importance that we abandon our original experiment and start afresh with a different one. What we must not do is confuse the two and modify the original hypothesis half-way through the experiment. If I may borrow two military expressions, we must look to the front and maintain our objective.

Let us, therefore, always define the contour of our thoughts before we begin a study. Pussy come down. Come down from that tree—are you tabby or tom?—are you Victor or Jonathan, George or Bill Bailey?—or are you perhaps a Jellicle Cat? Come down from the clouds and let's have a look at you.

And Live Laborious Days

What are we doing and why are we doing it? There is a third question we should always ask. Is it worth while? As general

practitioners we must always be amateurs in this whole rummy business. Research is a hobby to be fitted into our day. We shall never have enough time; we will often be short of resources, money, energy, and that elemental human necessity—encouragement. It behoves us, therefore, to harbour our resources, to be resentful of their deployment, and to reflect carefully before we expend any of our limited energies on a project. I am reminded of T. S. Eliot's prayer: "Teach us to sit still, even among these rocks."

Is my proposed study worth while? Is the return likely to justify the effort? Will my study lead to a determinant—a recommendation for executive action, that one will do something as a result of the study one would not otherwise have done.² If I demonstrate in the research study I mentioned earlier that a district nurse attachment leads to a reduction in the demand for home visiting, then does this matter? Is it of any importance? Will I then have some recommendation to make? If the honest answer to these questions asked before the study begins is "no," then we would have to ponder whether our time could not be better spent. "Come on Amanda, be positive," as our practice midwife so nicely put it at a critical moment. Not every study, of course, need have a determinant. We may wish, for example, to test an idea that shoe size divided by the square root of shirt size is a discriminant which distinguishes asthma from bronchitis. Support for this hypothesis might be of considerable theoretical interest. Which of these two studies should we embark on?—or is neither worth while? We must always do a most careful assessment before making such decisions.

I must take this opportunity to suggest that general practice research at the moment is too much concerned with medical and practice organization. I am not decrying this sort of study and some excellent work has been done. But what has become of clinical research? Why are we general practitioners much keener to look for links between appointment systems and the mean consulting time than we are to look for genetical markers in spina bifida, asthma, or congenital pyloric stenosis? We are uniquely fortunate. We live and work in the world as it is. Our unfortunate colleagues who work solely in hospitals must usually carry out their research on highly selected patients. If you wish to find out the mean duration of symptoms in acute prolapse of a lumbar disc you are better placed to do it from a general practice than from an orthopaedic hospital.

There is, of course, room for both types of research in general practice, but the prevailing fashion for organizational research has produced an imbalance. We should surely concentrate on trying to solve more of the clinical problems, which can be solved only from general practice, and worry less about such questions as the optimum floor area of a consulting room—consulting rooms have four walls and four walls limit the horizon.

Triple Pillar

The methodology of the modern drug trial relies on two requirements: patients must be allocated randomly into the different treatment groups and the clinician and patient must remain ignorant of the allocation. These are the pillars on which the integrity of the study rests. If we ignore them we shall be rightly ignored in turn by editor and colleague alike. It is therefore, nothing short of astonishing that an equally important requirement for the proper execution of a drug trial is consistently overlooked and omitted.

In any trial conducted from an outpatient department or from a general practice we rely on the patient to remember to take his treatment as directed; the actual act of swallowing the right tablet at the right time is unsupervised. There is now ample evidence that patients often default, and this will come as no surprise to most of us. If an appreciable number of patients do not take their tablets, then, whatever else you are

doing, you are not doing a drug trial. If, in addition, the number who default from taking the active drug differs from the number who default from taking the placebo—because the active drug has more side effects—then one is in even deeper trouble. There is an absolute obligation on the part of the researcher to demonstrate to the reader that the prescribed treatments were taken. Techniques exist to do this.³ Asking the patient to bring the container back and counting the number of returned tablets is not enough; it lacks the necessary subtlety. To neglect this aspect of a trial is every bit as serious as to neglect to randomize or to neglect to conceal the identity of the treatments. It is true that the oversight operates in a “fail-safe” direction. If no one takes any of the tablets at all you are hardly likely to falsely reject the null hypothesis of no difference between active drug and placebo.

My contention is, none the less, that with a few exceptions all the drug trials which have been published from outpatient departments and from general practice in the past are suspect. I submit that logically the whole lot should be repeated with a proper and adequate methodology. The onus is on the investigator to show that the tablets were taken; why should we the readers have to make assumptions? Can anyone advance a reasoned argument to deny the inadequacy of so much past research?

Final Misgiving

As a profession we must admit that we are liable to uncritical enthusiasms after a new treatment is introduced; often it takes a long time for the sober truth to emerge. What, for example, has happened to the physical treatments in psychiatry? What has become of continuous sleep treatment, insulin coma, and prefrontal leucotomy? Can we be sure that electric convulsion therapy will not similarly pass into oblivion—is it the ceremony or the shock which works? It is a startling experience to take part in a placebo-controlled trial and learn at first hand—albeit retrospectively—the power of the placebo. Every doctor and certainly every psychiatrist should undergo this experience. Can we really be sure that drugs and treatments such as disodium cromoglycate, indomethacin, monoamine oxidase inhibitors, imipramine, and electric convulsion therapy are really superior to a placebo? The trouble is that they may have prematurely assumed a mantle of respectability. After a time interest in assessing a new treatment wanes, the number of published trials declines, and the drug becomes “established.” The number of good studies is often woefully inadequate, and invariably few have originated from general practice. The fault for that, of course, is ours. It brings me back to my earlier point that there is a need to concentrate much more on the clinical aspect of research; there is an obligation on every clinician to take part frequently in well-planned drug trials.

Finally, I should like to speculate and place this tentative hypothesis before you for your consideration: after a new treatment or drug is introduced a train of events is set in motion which is inevitable and inexorable and occurs independently of the efficacy or otherwise of the drug or treatment. I suspect that if a histogram is constructed in order to show the number of trials published each year after introduction, then the distribution is always positively skewed

and the median year is about the second or third. The studies which return an adverse verdict—I am still voicing suspicions—are not scattered randomly throughout the distribution but are to be found in the second half, in the tail: the wrinkle is stamped on the brow only with the passing of time.

The only evidence I can advance in support of this hypothesis is admittedly trifling and relates to imipramine. Twenty-five placebo-controlled trials of the drug have been published since its introduction in 1958 (see Chart). The

Year of publication of the 25 controlled trials of imipramine.

median year for publication was 1961. Four of the adverse trials were published after the median year and only two before. Of the last five trials three have returned adverse verdicts. Despite this, tricyclic drugs are now firmly established in psychiatry. A possible explanation for the clustering of adverse reports in the second half of the distribution is that the populations studied might change with time. Hospital doctors, for example, might be experimentally quicker off the mark than general practitioners, and the treatment might be more efficacious in the hospital setting. Inspection of the imipramine papers lends no support for this. It is also possible that the doctors have changed and have adopted much more rigorous standards over the past decade. Again, inspection of the various trials lends no support for this explanation. The other explanation is much more ominous. Is it possible that the placebo-controlled trial provides little or no protection against our initial enthusiasms? In those heady early days do type I errors evade the strenuous and honest efforts of investigators to achieve objectivity?

It would be interesting to look up the pattern for the other treatments mentioned earlier; all that one needs is a place in the library and three months away from the practice. Personally I would rather not do it, for such study would have no determinant and might possibly lead to an outlook of therapeutic nihilism. Would that help our patients? And helping our patients is in the end what research in general practice is all about.

References

- ¹ Medical Research Council, *British Medical Journal*, 1965, 1, 881.
- ² Atkins, H., *Guy's Hospital Reports*, 1960, 109, 81.
- ³ Porter, A. M. W., *British Medical Journal*, 1969, 1, 218.