THE

# JOURNAL OF MENTAL SCIENCE

*[Published by Authority of the Royal Medico-Psychological Association]*

No. 435 [NEW SERIES NO. 399] APRIL, 1958 VOL. 104

## Original Articles

# CLINICAL RESEARCH IN PSYCHIATRY*

By

**G. A. FOULDS, M.A., Ph.D.**

*Runwell Hospital, Wickford, Essex*

### INTRODUCTION

SINCE the author has been working for some fourteen years alongside psychiatric colleagues his slender assets are as heavily and inextricably invested in the future of Psychiatry as those of any psychiatrist. Such criticisms as are offered must, therefore, from this point of view be regarded as within rather than between group criticisms.

To anticipate the charge of flogging a dead horse or setting up a straw-man, the literature on the assessment of the effects of treatment was examined for the years 1951 to 1956 inclusive. Thirty-six papers were culled from the British literature (*J. Ment. Sci.*, 31; *J. Neurol. and Neuropsychiat.*, 5) and 36 from the American (*J. Nerv. and Ment. Dis.*, 7; *Amer. J. Psychiat.*, 20; *Psychiat. Qtrly.*, 9). Alternate copies of each journal were examined. If the appropriate copy was not available, the next one back was taken.

In the British literature 16, or 44 per cent., of the studies used control groups of some sort. These were not always the appropriate ones and occasionally the author seems to have lost control, as it were, somewhere between the Introduction and the Results. The author claimed that the treatment was a success in 19 per cent. of the studies in which controls were used; on the other hand, 85 per cent. of authors of uncontrolled studies claimed success. Such a comparison is not even possible with my sample of the American literature since only 4, or 11 per cent., of their studies were controlled.

Table I shows the percentage of treatments claimed as successful in 20 controlled and 52 uncontrolled British and American studies.

TABLE I

*72 Anglo-American Papers*

| | Success | Failure | Total |
|---|---|---|---|
| Controls .. .. .. .. | 5 | 15 | 20 |
| No controls .. .. .. | 43 | 9 | 52 |
| Total .. .. .. | 48 | 24 | 72 |
| n=1 | $\chi^2=21\cdot06$ | $P<\cdot001$ | |

* Based on a paper read at the Annual General Meeting of the Royal Medico-Psychological Association in Oxford in July, 1957.

Thus 28 per cent. only of the studies used some sort of controls; in uncontrolled studies success was claimed in 83 per cent. as against 25 per cent. in the controlled studies (and in some of these appropriate statistical treatment was lacking). For one degree of freedom, $\chi^2$ is $21 \cdot 06$ and $P < \cdot 001$. It would appear, therefore, from this study of the literature that one can safely conclude that claims for the success of a treatment are closely associated with absence of the means whereby these claims can be scientifically substantiated. It has been argued, in spite of the above evidence, that the author is flogging a dying, if not a dead, horse. The percentage of controlled studies in 1957 so far is little more than half that in 1956.

## Florence Nightingale and the Fairy From Out of the Wood

In a B.B.C. pantomime skit a character kept appearing with the introduction "I am the fairy from out of the wood and I goes about a-doing of good". Needless to say her good endeavours invariably resulted in the creation of chaos. This, in more technical language, is sometimes referred to as the flight into activity.

When Florence Nightingale first went to the Crimea, she found men dying in the most appallingly filthy conditions. Her assistants wanted to start nursing at once. She very courageously stopped them on the grounds that their work would be piecemeal and hindered at every turn until they had obtained the full co-operation of the medical men on the spot and of the Government at home. A few more men died; but many thousands have since been saved.

A resolute determination not to put the humanitarian cart before the scientific horse has not been a striking characteristic of psychiatry. Some years ago one of the treatments in vogue was as follows: "A drink for a fiend-sick man, to be drunk out of a church-bell: Githrife, cynoglossum, yarrow, lupin, flower-de-luce, fennel, lichen, lovage. Work up to a drink with clear ale, sing seven masses over it, add garlic and holy water, and let the possessed sing the *Beati Immaculati:* then let him drink the dose out of a church-bell, and let the priest sing over him the *Domine Sancte Pater Omnipotens.*" Doubtless control groups were then—as so often now—thought to be unethical, so that it would have been impossible to determine whether, in such cases as appeared to respond well, the therapeutic agent was the drink, the *Beati Immaculati* or the *Domine Sancte Pater Omnipotens.*

Obviously the clinician in charge of a case, with the great responsibility which this involves, must have the final decision on the question of giving or withholding treatment; but is the argument usually advanced against the withholding of treatment a valid one?

Insulin Therapy was introduced before its efficacy had been established by adequate experimental methods. A large number of clinical psychiatrists came to believe, on the basis of uncontrolled observations, that insulin was in fact efficacious in certain cases. This being so, they were undoubtedly right to refuse to withhold this treatment in such cases. Their integrity cannot be questioned; but the basis for their belief can and should be. During the past few years an increasing number of clinicians have begun to doubt this belief. To some of these the faults of the ageing mistress suddenly became apparent with the advent of the beautiful young Serpasil. Soon a large number of clinicians will come to believe, on the basis of uncontrolled observations, that Serpasil . . . but perhaps this particular human characteristic has been adequately dealt with in Anatole France's *Penguin Island*, where—it may be remembered—the beginning of the epilogue describes the demolition of all skyscrapers and the end their re-erection. Meantime controlled experiments are either lacking or, as in the case of the long-term follow-up study by Penrose (1944) and the controlled study by Lehoczky *et al.* (1939) showing negative results for Insulin, so scotomatized that the recent paper by Ackner *et al.* (1957) came as a surprise.

The same story can be told of the electrical and neurological procedures, with perhaps the added weakness that the groups for which the treatment was originally thought to be most efficacious have gradually changed and always towards those groups which are thought to have the best prospects of spontaneous recovery. The drift has been with E.C.T. and leucotomy away from schizophrenia towards depression. More recently with the pharmaceutical procedures, such as largactil, and the inspirational, such as carbon-dioxide, less ambitious workers have gone straight to the group with apparently the best chance of spontaneous recovery, the anxiety states; or, if need be, to the somewhat nebulous "tension states", who will soon be definable as those who respond to a given treatment.

When a new treatment is suggested no one knows whether or not it will work. Clearly it is at this stage that it should be assessed by adequate experimental methods before ethical questions are introduced. These ethical questions arise because clinicians believe the results of reports such as those discussed below. Such reports are, however, nothing more than the dissemination of folk-lore. The withholding of treatment X cannot be unethical if it is not known to be efficacious. Continuing to give treatments with no intention of validating them would seem to be the more immoral course.

## Lack of Design in Much Psychiatric Research

A combination of confused ethical and methodological judgments has resulted in a succession of publications which are treated as research publications, but which are in fact simply the chronicling of clinical routine. Much of the methodological confusion seems to be due to the dichotomizing of clinical observation and experimental method, which are thought of as antithetical approaches to the same type of data at the same time. No scientist is likely to deny the value of the subjective observations which frequently lead to the setting up of experiments. Unfortunately many investigators consider their task complete after these uncontrolled observations have been made. The Argyll Robertson phenomenon has been cited in support of the view that such observations can stand independently of statistics; but its practical importance for psychiatry is dependent upon its occurrence in roughly two out of every three cases of GPI and its extreme rarity in other conditions. This is a statistical concept—even though the sums may not have been done. It is not even always necessary to do them. We do not require tests of significance to satisfy ourselves that the sun sets more frequently in the weşt than in the east; but, until Hume has been convincingly refuted on other than epistemological grounds, this must remain a statistical concept. Who knows what tomorrow may bring?

Concern here will be with those investigations which purport to assess the value of a psychiatric treatment. The following report illustrates one of the commonest procedures.

An adequate number of psychoneurotics were given drug 1 and then assessed as remitted, improved and unimproved. About 60 per cent. were thought to have gained substantial benefit from the treatment. Of the subgroups, onxiety rtates did the best. The conclusion is drawn that the main effect of the drug is the relief of tension. This claim is, of course, utterly unsubstantiated. This type of conclusion can only be drawn from intra-group and not from inter-group designs. To justify such a conclusion many other variables would have to be controlled. The following well-known story may bring out the logical fallacy more clearly. A conjuror, performing at a ship's concert, pointed to a parrot in a cage and said he would make it disappear. At the crucial moment the ship struck a floating mine and blew up. All were thrown into the water, including the parrot which was to be seen, with all its feathers off, pacing up and down on a floating door muttering "bloody clever". In this story it was the parrot-patient rather than the conjuror-doctor who was deceived; but it does serve to point up the fact that variables as large and efficacious as mines may intervene unremarked.

Since—to return to the investigation—psychoneurotics were apparently taken at random, one would expect that they were being regarded as a group and that an intra-group design would be used; but the stated aim was to assess the effect of previous personality on the results. In such a case psychoneurotics, regardless of diagnosis, with pre-breakdown psychoneurotic traits of personality and those without such traits should be compared. Had the aim been to assess the value of drug 1 with psychoneurotics in general, matched groups on drug 1 and placebo 2 should have been compared. Or again, had the aim been to discover whether personality or the syndrome were the decisive factor, anxiety states and hysterics (the only groups of any size in the study) should have been divided into those with and without pre-illness neurotic traits and the design could have been: $AX_1$; $AX_2$; $AY_1$; $AY_2$; $HX_1$; $HX_2$; $HY_1$; $HY_2$—where A is anxiety, H is hysteria, X having neurotic traits, Y not having them, 1 is the drug and 2 the placebo.

By comparing $(AX_2+AY_2)-(AX_1+AY_1)$, the effectiveness of the drug with anxiety states could be determined. By $[(AX_1+AY_1)-(AX_2+AY_2)]-[(HX_1+HY_1)-(HX_2+HY_2)]$ it could be determined whether favourable response to drug 1 was a characteristic of anxiety states which distinguished them from hysterics.

By $[(AX_1+HX_1)-(AX_2+HX_2)]-[(AY_1+HY_1)-(AY_2+HY_2)]$ it could be determined whether favourable response to drug 1 was a characteristic of patients with pre-illness neurotic personality traits which distinguished them from those who do not have such traits. Characteristics of these groups are all that can be determined by such inter-group comparison.

In a second study an adequate number of depressives received drug 2 and were then assessed as remitted, improved or unimproved. No evidence of clinical improvement attributable to drug 2 was found. One cannot, however, arrive at negative conclusions any more happily than at positive ones in the absence of intra-group controls. Without drug 2 these patients might have got worse.

A third study is concerned with treatment Z in mental disorders. The total number of subjects (25) was subdivided into 6 groups. Even so, a list, admittedly tentative, of indications and contraindications was given. Sixty per cent. were again improved. Three pages were devoted to theories of action—that is to how it works before it is known whether it works. Such discussions are irrelevant to inter-group designs and can claim no superiority over armchair discussions, rather the contrary since they mislead many people into thinking that they are based on facts. One of the principal arguments put forward in favour of treatment Z is that it works with patients who have failed to respond to anything else. It is admitted that only 60 per cent. of the cases were treated successfully and that this is no better than many other treatments. One would imagine that by now, with all the treatments that are moderately successful with cases that have not responded to all other forms of moderately successful treatment there could scarcely be anyone left who is not at least much improved; yet the chronic wards seem to remain remarkably full.

## Comparisons Among and Between Groups

Most of the points in this section have been more fully and ably presented by Kline (1953), but some repetition is necessary for the main argument of this paper.

*Intra-individual comparisons:* This method applies when the procedure to be investigated is reversible within a short time, when the number of subjects is unavoidably small and when the variables are either largely unknown or numerous. The effect of sodium amytal on catatonic stupor may serve as an example. Let it be supposed that 12 stuporose patients have been collected and that the degree of stupor can be measured on a five-point scale. The procedure would be to divide the 12 into two groups as well matched as possible. Group A would be measured for stupor, then given sodium amytal, again measured for stupor, allowed time for the effects of the sodium amytal to wear off, measured for stupor, given a placebo, measured for stupor. Group B's procedure would be: measured, given placebo, measured, interval, measured, sodium amytal, measured.

If some such results as the following were obtained: A: 5—sodium amytal—1—interval—5—placebo—5 and B: 5—placebo—5—interval—5—sodium amytal—1, it might be concluded that sodium amytal tends to bring catatonics out of stupor.

If the following results were obtained: A: 5—sodium amytal—4—interval—3—placebo—2 and B: 5—placebo—5—interval—5—sodium amytal—4, it might be concluded that sodium amytal was not so reversible after all. Or again, something like this might be found—A: 5—sodium amytal—4—interval—3—placebo—2 and B: 5—placebo—4—interval—3—sodium amytal—2, in which case additional attention or suggestion might be accounting for the results.

Should the experiment require the comparison of three quickly reversible drugs and a placebo, 16 cases all of one diagnostic category might be collected and the following design adopted:

A: 1—2—4—3
B: 2—3—1—4
C: 4—1—3—2
D: 3—4—2—1

This would control the effects of position and sequence and the value of the treatments could be assessed by Analysis of Variance.*

*Inter-group comparison:* This method may be used to ascertain characteristics of groups which distinguish them from other groups. Suppose that the hypothesis to be tested is that, relative to anxiety states, depressives tend to go off to sleep more quickly, wake sooner and are then unable to get to sleep again; whilst the anxiety states, relative to depressives, are slow to get to sleep, but wake later and more bemused. Two groups are selected, matched on some at least of the variables likely to be relevant, and the necessary observations and recordings are made. If the hypothesis were supported by the observations, it could be concluded that such and such are the characteristic

* Since the presentation of this paper just such an experiment has been designed by Fraser Roberts and reported by Raymond *et al.* (1957).

sleep habits of depressives as contrasted with anxiety states and conversely. Nothing can be said about processes. It cannot, for example, be concluded that depression causes people to wake up in the middle of the night. It might be found subsequently that most people other than anxiety states tend to wake up in this way. If it could be established that this particular sleep rhythm occurred with greater frequency among those who were depressed than among those who were not, it would still be necessary to show that this sleep rhythm occurred only when these particular individuals were depressed and not when they were well.

*Intra-group comparison:* This method may be used for testing either reversible or irreversible procedures. If it be desired to assess the effect of brief stimulus therapy on anxiety states, criteria must be set up to enable the group to be made as homogeneous as possible. When as many variables as may be have been matched, randomization should be used in the hope that any unknown variables will break even between the groups. Pre-treatment measures are then taken for each group and one given BST, the other a dummy treatment. The pre- and post-treatment differences of the two groups can then be compared. Something of this sort has, in fact, been done by Montagu and Davies (1955).

There is no doubt that the difficulty of obtaining control groups is often very great in psychiatry, perhaps particularly so in irreversible procedures such as leucotomy. The answer is not, however, to do studies without them and then draw conclusions which could only be drawn had they been used. Once we have found in controlled experiments a treatment that works we can, of course, use that treatment as our yard-stick for evaluating subsequent treatments, and control groups in the sense of untreated patients would no longer be necessary. If the findings of Raymond *et al.* (supra) are confirmed, such a yard-stick for drug trials will be available in amylobarbitone. A second best method is available and has been used, for example, by Petrie (1952) when she compared two different surgical insults; but all this tells us is that one is a cut above another. It does not tell us whether the better is better than nothing.

## In Conclusion

A conference on Mental Health in Oxford, attended by eminent psychiatrists and interested scientists, discussed research for several days. The proceedings were reported in a book entitled *Prospects in Psychiatric Research* (1953). No mention was made of the need to train research workers in methodology. Without some such training the prospects seem to me to be disturbing. Particularly at this time when new treatments are being produced monthly together with attractive brochures, more attention must be paid to the design of experiments if the danger is to be avoided of slipping back into something akin to Mediaeval Alchemy—dose for a fiend-sick man by courtesy of Messrs. A and B. A Special Medical Correspondent in *The Times* defended the Medical Research Council's relative lack of support of research in Mental Health chiefly on the grounds that insufficient people capable of doing such work were available. Must we then sit back and wait for some Minervaesque research workers to spring fully armed from the head of Jove? If, when we think of research, we think of the cause of schizophrenia, perhaps we should. If, on the other hand, we think as well of such comparatively mundane tasks as the one discussed here—the evaluation of the work of our "drug addicts", or of the somatotherapies recently so carefully reviewed by Staudt and Zubin (1957), then surely the answer is "no". Here, if anywhere, is a field in which even the guiding hand of a genius is unnecessary. Here, if anywhere, the less gifted can do a useful job of work. It should be someone's responsibility to see that there are more of us.

### References


1. Ackner, B., Harris, A., and Oldham, A. J., *Lancet*, 1957, *i*, 607–611. 23 March.
2. Kline, N. S., *Psychiat. Quart.*, 1953, **27**, 3, 474–495.
3. Lehoczky, T., Eszenyi, M., Horanyi-Hechst, B., and Bak, R., *Z. ges. Neurol. Psychiat.*, 1939, **160**, 24.
4. Montagu, J. D., and Davies, L. S., *J. Ment. Sci.*, 1955, **101**, 424, 577–592.
5. Penrose, L. S., *Ontario Dept. of Health*, 1944.
6. Petrie, A., *Personality and the Frontal Lobes*, 1952. London.
7. Raymond, M. J., Lucas, C. J., Beesley, M. L., O'Connor, B. A., and Fraser Roberts, J. A., *Brit. Med. J.*, 1957, 63–66. 13 July.
8. Staudt, V. M., and Zubin, J., *Psychol. Bull.*, 1957, **54**, 3, 171–196.
9. Tanner, J. M. (Ed.), *Prospects in Psychiatric Research*, 1953. Oxford.