Chapter II

DESIGN FOR THE EVALUATION OF THERAPY

Joseph Zubin

Three basic questions seem to me worth your serious attention:

1. What are the minimal requirements of the design for research in evaluating the outcome of therapy?

2. What have designs which have been used up to the present time revealed about the efficacy of various therapies?

3. What future steps must be taken to improve evaluation of outcome of therapy?

The minimum essentials of a design for evaluating the outcome of any therapy, be it in the mental disorders or in the physical disorders are: 1) a homogeneous group of patients; 2) a comparable untreated control group; 3) a sufficient follow-up period and 4); specific criteria for evaluating outcome.

The need for homogeneity in age, duration of illness, diagnosis and other important relevant variables in the patient group under treatment is so fundamental in any research that it can be dismissed without further comment. The control group, however, needs more discussion. Despite some opinions to the contrary, the principle of the control group is essential to the evaluation of therapy and is as applicable to mental disorders as it is to physical disorders. Differences of opinion may arise however in the methods of selection of controls. Because of our ignorance of the etiology of many mental disorders we can not always be sure that our controls are truly comparable to our treated cases. Furthermore, since we have to deny, at least temporarily, the benefits of any specific therapy to the control group, some hesitancy arises in the selection of a given patient for such a purpose and the fact that he does not receive specific treatment may lower his morale and motivation for improvement. Despite these difficulties no better method is now available and the question of maintenance of morale can now be answered by providing the control group with "total push". If the number of patients to be treated is small, the statistical analysis of the data requires that the control group be matched patient for patient with the treated group in all variables relevant to outcome. If larger groups, say several hundred, are to be treated, random sampling from the large reservoir of patients can be made for both the treated and the control group. The tendency of placing in the control group patients for whom treatment is contraindicated is a dangerous one since these patients may differ from the treatment group also in their chances for recovery.

The period of follow-up is another much neglected but important consideration. Immediate outcome of therapy after a short follow-up period (3 months for example) is often quite different from the outcome at the end of a long follow-up period (5 years or more). The duration of the follow-up period, the method of follow-up whether by personal interview or by mail and the thoroughness of the follow-up should be recorded adequately if trustworthy comparisons are to be made with other studies. Not only the status of the surviving
patients should be obtained but also the status before death of the deceased, and the status at time of last contact of those who were subsequently lost to the study.

The criteria for the evaluation of outcome must be specific and stated in objective terms. Such terms as cured, improved, ameliorated, worsened, can vary so much in their significance that they are worthless for comparative studies. In the absence of better standards, the criterion of improvement most in vogue today is whether the patient was discharged and remained at home. This criterion is often faulty because discharge may sometimes depend merely upon the willingness of the family to receive the patient or on other incidental factors. But it is often the best criterion of improvement available today. When more specific methods for evaluating degree of improvement are available, a question often arises about the degree of improvement to be expected in mental patients. Excluding the milder non-chronic neuroses and possibly the true manic-depressive psychoses, it is doubtful whether cure in the sense of return to premorbid level is to be expected any more frequently in chronic mental disease than in chronic physical diseases like tuberculosis or cancer. One of the most crucial tests of the value of a specific therapy in the chronic physical disorders, prolongation of life, is denied the psychiatric therapist since prolongation of life of a mental patient has never been regarded as a goal in itself. Yet, the remarkable achievement of the shock therapies in reducing patient mortality due to inanition or physical exhaustion is one of the unheralded outcomes of present-day therapy. Long-term follow-up studies usually indicate a higher level of mortality for the control group even in hospitals where good care is afforded. In view of the generally higher mortality rates which characterize chronically ill mental patients, a characteristic common to all chronic illness, evaluation of outcome of therapy must eventually take into consideration the extension of life following treatment.

The clinically perceptible improvement in behavior, feeling tone, ambition, sociability etc., which accompanies successful therapy is one of the most difficult areas of evaluation. The subjective clinical evaluation has been supplemented by objective rating scales and behavior charts and sometimes by tests. At the present time, however, no generally acceptable yardstick of improvement is available, though the basis for such a measuring rod has already been delineated in the literature.

The results of the application of these four criteria in the design for the evaluation of therapy have yielded some unexpected findings. A survey of the literature indicates that not more than a dozen adequately designed long-term follow-up studies are available. Nearly all the immediate outcomes in these studies indicate a definite advantage for the specific treatment as compared to the control group. Patients are discharged home much sooner and remain out of the hospital longer. In diseases for which rational therapies have been discovered, e.g., a general paresis and pellagra, the degree of improvement and its permanence is limited only by the amount of residual damage produced by the disease. In the case of the still irrational shock therapies, the degree of improvement and its duration in long term follow-up studies seem to show no advantage over the non-specific therapies given the control groups with
few exceptions. This does not lessen the important contribution made by the irrational therapies to the saving of human misery and of public funds by the early discharge of the treated patients, and by the apparent increase in their longevity. But the fact that the long-run comparison shows no advantage for the specific therapies must lead both the clinician and the research worker to ponder. (It may be pointed out parenthetically that the shock therapies are not alone in their failure to show an advantage in long-term follow-up studies. Such data as are available for psychoanalytic therapy and psychotherapy lead to the same conclusion. The data on psychosurgery are still too meager, but even for this therapy there are already some indications that the advantage is not too striking.) Why this should be so is one of the most baffling questions facing the research worker. One possible answer is that the specific therapies simply hasten the improvement of those who would eventually improve anyhow. A more hopeful note is struck by the observation that the pretreatment characteristics of those who eventually improve are often significantly different from those who fail to improve. Whether these differences are related to outcome regardless of therapy or whether some therapies will be more suitable for patients with certain characteristics can not be answered definitely at this time.

In order to improve the evaluation of outcome of therapy several important steps seem necessary. First, some easier method must be found for establishing control groups. The work with the treated patients alone is sufficient to tax even our best equipped hospitals. The provision of additional staff, "total push" methods etc. for the control group is often the straw that breaks the camel's back. Ethical problems also arise in the relegation of a given patient to a control group, as will be discussed by Drs. Jenkins and Holsopple. There is no reason why each hospital wishing to conduct a research in a specific therapy must provide its own control group. If we could provide standard control groups for the various types of mental patients to whom the various therapies are applicable, these could then become the basis for the selection of patients who are to undergo treatment by a specific therapy. Thus by inverting the procedure of matching controls to treated groups and instead matching treated groups to available standard control groups, it may become possible to hasten the process of the evaluation of therapy in all its aspects. Instead of selecting patients at random, future investigators for whom standard control groups are available will try to select for treatment patients who match the already available controls. The patient populations which would ordinarily not receive specific therapies because of various reasons not related to their illness could be utilized for establishing these standard control groups. By providing these selected populations with the best levels of care including "total push" methods, the base-lines so sorely needed in the evaluation of therapy could be established. The five-year follow-up on such a population will then make available standards of expected improvement which the matched specific therapy groups must meet or beat. The data now available in the literature on long-term follow-up studies are not adequate for reasons which have been indicated elsewhere (5). These may be summarized as follows:
1. The early pre-1930 data are not trustworthy as a guide because they are based on the total available patient population of that period, and not on such selected groups as those through which therapy is administered today.

2. There have been shifts in the character of the patient population as a result of the widespread use of therapies, and mental hygiene education.

3. The diagnoses of the previous decades are not comparable with those of today.

4. The interval of follow-up is not always indicated in previous studies and when indicated, it is not always uniform for all the patients.

5. Even when control groups are utilized, they often consist of the patients in whom some contraindication was found, or for whom permission could not be obtained for therapy. When control groups are selected they are rarely afforded the same level of "total push" as is given the treated group.

Another advantage of the standard population is the possibility of providing prognostic indicators for each patient based on his status, characteristics, test performance, etc., at the time he became a member of the standard control group. In this way, the prognosis under control condition can be compared with actual outcome under the specific therapies for each type of patient. Thus, a probability chart can be provided indicating the chances of recovery for each type of patient under control conditions of non-specific therapy ("total push"). ECT, insulin, psychoanalysis, psychotherapy, psychosurgery, and any other therapy that may be tried.

To be sure, this is a somewhat idealistic plan which will have to compromise with many realistic difficulties arising from our lack of knowledge about the true comparability of various individual patients who are to be matched with the controls. Special travelling teams of expert psychiatrists, psychologists, social workers and biometricians will have to be provided for keeping the research on standard populations up to date and for deciding comparability of patients in various parts of the country. But without such a method for evaluating outcome of therapy, we are reduced to accepting the judgment of outcome made by three of the most biased persons connected with the therapy—the clinician, the patient and his family.

In dealing with long-term follow-up studies a new consideration arises which has thus far not received any attention in the evaluation of outcome of mental disorders. When a 5-year follow-up is undertaken, the group of improved and discharged patients under study may show the following changes in status: (a) remain in that category to the end of the follow-up, (b) relapse (c) die of causes unconnected with their disease and, (d) be lost sight of. The unimproved and undischarged patients may similarly (a) remain in their status quo, (b) improve and be discharged, or undergo one of the other two changes in status described above. Furthermore, the patients may go through relapse and recovery several times in the course of the follow-up. Statistical methods for evaluating the net outcome regardless of the temporary alterations are a sine qua non in future evaluation of outcome of therapy. The need for such statistical evaluation of outcome has been forcibly brought to our attention by a recent controversy in the Journal of Mental Science regarding Karagulla's study on the outcome of convulsive treatment in depression. Both Karagulla (3) and her critic Slater (4) used crude rates of improvement, the first eliminating deaths and the
second including deaths in his calculations. Neither argument is conclusive and only the application of the mathematical model devised by Jerzy Neyman (1,2) can provide a definitive answer through the computation of net improvement rates which are freed from the influence of deaths, relapses, etc. A tentative analysis indicates that the net improvement rate of the treated group is generally not different from the corresponding rate in the control group in the above study.

Summary

The basic elements in the design for the evaluation of therapy are: 1) a homogeneously selected group of patients; 2) a comparable control group; 3) a sufficiently long follow-up period; and 4) specific objective criteria for evaluating outcome. A review of the literature indicates that only a small number of studies satisfy these criteria. The results of these studies indicate that specific therapies as contrasted with non-specific therapies yield better immediate outcome, but in the long run, no better eventual outcome. Such a result can be explained either by assuming that the present-day therapies do not alter the basic course of the disease but only hasten the improvement of patients who would have eventually improved anyhow, or that insufficient knowledge is now available for selecting the most suitable therapy for each patient.

In order to provide prognostic criteria and generally available control groups for future studies, a reversal in ordinary procedure is suggested. Instead of first selecting the group for treatment and then searching for controls, it is proposed to select standard control groups which will be carefully evaluated and studied in long-term follow-up. Investigators will then select their treatment groups to parallel the already established standard control groups. These standard control groups can then serve as a gauge for evaluating the results of a specific therapy. In this way it may be possible to determine quickly and directly the efficacy of each of the established specific therapies and of a newly-proposed therapy. The provision of such standard control groups and their supervision as well as the decision on the comparability of treatment groups will entail many difficulties and involve the creation of a special commission of experts. Without such standard procedures, however, there is little hope for attaining scientifically valid evaluation of the outcome of therapy. This is the strategic moment for developing such standard populations since the general public is aroused to the need of more humane care in all the states. In another decade, it will be impossible to find states in which the natural history of mental disease, uncomplicated by specific therapies, can be studied.

References

1. Fix, E. and Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. Human Biology, 23, 205-241, 1951