An Examination of the Double-Blind Method as It Has Been Applied to Megavitamin Therapy

A Hoffer, M.D., Ph.D., F.A.O.P.[1]

Introduction

Under my direction, the first double-blind experiments were conducted in psychiatry. With the first one we examined the efficacy of a yeast nucleotide preparation which had been claimed to be effective in treating chronic schizophrenic patients. None of the patients improved. With the second experiment we compared the efficacy of nicotinic acid, nicotinamide, and placebo in combination with the standard treatment of that day (electroconvulsive therapy, psychotherapy, and sedatives). The clinician in charge of each case decided whether or not to use ECT. Half from each of the three groups (usually the sickest patients) received ECT. Nicotinamide was used as a hidden control group to compensate for the vasodilation (flush) produced by nicotinic acid when the medication is started. None of the nursing or medical staff were aware of the nicotinamide group. The 30 patients were admitted from the community to a psychiatric ward and had not spent long periods of time in any institution. They were mainly acute and subacute. On the average they were in hospital about two months while going through the experimental program.

1.1201 C. N. Towers, Saskatoon, Saskatchewan, Canada, S7K 2L5

They were re-evaluated every three months after discharge and a final evaluation was made one year after discharge. The follow-up worker made his evaluations blind. Three grams per day of nicotinic acid or nicotinamide was used for 33 days while in hospital.

At the end of one year after discharge, one-third of the placebo group were well. However, two-thirds of the vitamin groups were well, there being no difference between nicotinamide and nicotinic acid. After that, three more double-blind experiments were completed in Saskatchewan. In each case, the results when vitamin B3 was used we're superior to standard treatment only.

Since then, as a result of the work of physicians, now known as Orthomolecular psychiatrists, many significant improvements have been made. The treatment today contains many additional variables and produces a greater number of recoveries. Orthomolecular therapy includes megadoses of vitamins as a main component but also includes attention to nutritional therapy, to tranquilizers, antidepressants, and to ECT. The best comprehensive outline is in *Orthomolecular Psychiatry* (Hawkins and Pauling, 1973).

ORTHOMOLECULAR PSYCHIATRY, VOLUME 2,„NUMBER 3, 1973, Pp. 107 - 114

Only one other psychiatrist, Dr. H. Osmond, has had as much experience as I in conducting double-blind experiments, not in the number of trials, but in the number of years that trials have been conducted. Over the years we have become more and more aware of the inherent defects of this method, and in a series of papers we have drawn these difficulties to the attention of the medical world (Hoffer and Osmond, 1961, 1963; Hoffer, 1967). I have concluded that the double-blind method has so many imperfections that its use is limited, that it leads to a large number of serious errors, i.e., it would do so if clinicians took it seriously, and that its main function is to make it easier for government agencies to turn down new drugs. It seems to be most appreciated by groups who have access to unlimited funds and limited access to clinical curiosity and creativity. I have finally concluded that the observations and conclusions made from double-blind experiments have as little relevance to the therapy of patients as do observations on monkeys in a cage to their behavior in their native habitat. The beast seems to be the same, but there the resemblance ends.

The defects of the double-blind methodology are these:

(1) The basic assumptions behind the use of statistical analyses are ignored (Hogben, 1957). The design assumes that the comparison groups will be equivalent, will be homogeneous, and will be invariant. Over time a proper randomization may approximate equivalence but since psychiatric populations are very heterogeneous, it is impossible to obtain equivalent groups. This can be overcome by very large samples, but these are very rare in medical research. Nor can we assume the conditions are invariant. There is a natural historical drift in the nature and intensity of diseases due to factors only dimly understood. Therefore, according to Hogben, statistical analyses based upon probability theory is invalid.

(2) Little account is taken of the disease model being followed in most double-blind experiments. Diseases may be short-lived, like pneumonia, or may last most of one's life, like diabetes or schizophrenia. The same variability of models is inherent in any psychiatric population. Since chronic patients usually require chronic treatment, it is inappropriate to use short-term treatment for chronic patients. But chronic double-blind experiments are very difficult to control so they are seldom used. We therefore find that short experiments are applied to groups of patients whose duration of illness varies from several months to many years (Click and Margolis, 1962).

(3) The three main components of the patient-doctor relationship are the status of patient and doctor tied together by the relationship. The patient is impelled to have some faith and trust that the physician will be able to help him. The greater the threat to the person, the greater is the drive to have faith. The physician, on the basis of his education and experience, will have a certain degree of confidence that he can help the patient, either to achieve a cure or to alleviate the discomfort. There is no generally accepted term for this complex relationship. Terms like placebo effect, psychotherapy, etc., have been used but do not quite get at the essential relationship. Placebo effect is generally described as a positive response in the patient due to his expectations, his faith in the doctor or in his medication. Negative placebo effects have been given less consideration.

Hoffer and Osmond (1961) presented the term *obecalp* to describe another aspect of the relationship. Inasmuch as a positive response to an inert substance is a placebo response, so a negative response to an active chemical is an obecalp reaction (placebo backward). This we defined as a reaction in which a compound of known potency produces no response in test subjects due to factors such as fear of medication, either in the subject or in the doctor and others involved in giving the drug, lack of faith, negative suggestion, and so on.

There are two phases in the placebo and obecalp reaction, the initial reaction when treatment is just underway, and the sustained reaction. Of these the sustained reaction is more rational and more important. During the initial phases, the patient is dependent upon the transaction with his doctor and has no way of directly experiencing whether he is being helped. However, once treatment has started he begins to experience changes in himself, good or bad, from which he draws conclusions which enhance his placebo or obecalp reaction. If a patient finds that his depression or his hallucinations become less troublesome, this increases his faith and trust in the medication and so enhances placebo response. A positive feedback develops which increases the level of placebo effect as more and more improvement is experienced. This is the optimum placebo effect which will remain high as long as improvement is sustained. The physician is also affected in the same way for, as he observes his patient's improvement, his faith in the treatment that he is giving goes up, which in turn is evident to his patient.

Generally there must be a sufficient placebo effect for the patient to start the treatment. This the doctor achieves by his manner of confidence in giving the diagnosis to his patient, in his explanations of it, and in describing the importance of the treatment and the outcome or prognosis. If the patient remains unconvinced, he will not even start the therapy unless he has a family who can persuade him to cooperate, and supervise the medication. If this is impossible, the patient will have to be hospitalized for treatment. The patient's initial faith may be zero, provided the family can ensure the program will be followed. Thus, patients have been given nicotinic acid or nicotinamide disguised in their food because they were violently opposed to taking any medication, and they have recovered. In the same way, a pellagrin given nicotinamide in his food will also recover whether or not he knows this is being done.

However, unless there is a sustained placebo effect, the recovery will not be maintained since the patient will discontinue medication. Fortunately, most psychiatric patients when well do not wish to relapse into their previous sick condition. Many schizophrenics who have been well for many years on the megavitamin therapy will discontinue medication but will in most cases resume medication if they become aware of a resurgence of symptoms.

In JAMA 224, 1584, 1973, Borda's report on patient evaluation of tranquilizers was reviewed. Borda studied over 15,000 patients in 10 hospitals of whom 25 percent had received at least one tranquilizer. There was a marked discrepancy between the patients' and their doctors' evaluation of efficacy. About 66 percent of the physicians reported good results but only 21 percent of the patients reported feeling better. With this major disagreement, it is simple to understand why patients are so reluctant to take them, and why there is such a development in slow-release, long-acting tranquilizers. Apparently tranquilizers, either because they are ineffective or because of undesirable side effects, do not sustain the placebo effect. A major problem in treating chronic schizophrenics is in keeping them on their tranquilizers. There must therefore be enough initial and sustained placebo reaction to ensure the medication will be taken.

In the normal situation the patient's personal physician attempts to set placebo effect at a very high level. However, in a double-blind experiment, if it remains truly blind, it is impossible for the physician to do so. Since he knows that the patient may receive an inert substance, he cannot honestly advise the patient that this will be helpful to him. The patient will detect his physician's reluctance to be a real physician to him. This is even more apt to occur in situations where double-blind experiments are run, i.e., in institutions, university wards, and so on, for in many cases the patient does

not have a physician or does not know who he is. I have seen large numbers of patients after discharge from these institutions who did not know who had been their doctor. They had been seen by medical students, internes, and residents, but very infrequently by the physician under whose name they had been admitted and who was finally responsible for treatment. By setting the placebo effect at a minimum, the double-blind destroys one of the essential components of the treatment. This is like measuring the rate of a chemical reaction, say an enzyme, which works best at 37°C, by setting the reaction at 0°C. There may be a reaction but it will not help the patient much. This is the main reason why double-blind results have little relevance to the real therapeutic effects of chemicals as commonly used by physicians.

(4) Perhaps the most serious criticism of the double-blind method is that it has not been established empirically. There is no data which shows that the double-blind method does control those factors which theoretically it was designed to control. We are therefore swept up in a technique which has received overwhelming approval by institutions and universities but which has never been subjected to the experimental test of whether it works. The discussion that I have presented so far indicates that when such a test is finally performed, it will show the double-blind not to be a useful method.

In science, a new technique is not used until it has been calibrated, i.e., compared to a current method. If it is more accurate, and more sensitive, it will supercede the older method. If it has not these advantages, it may still take over if it is more economical, quicker, and so on. But in the double-blind we find a method which is not established by experiment, not proven to be better, more difficult to run, and much more expensive, which has displaced usual clinical trials. Credit must be given to those research workers who have carried the field with enthusiasm and dedication for they had little else to bolster their position. They assumed that, because chemical treatments replaced each other, e.g., sulfonamides by penicillin, etc., and because double-blind experiments generally showed drugs considered active to be no better than placebo, this proved (1) double-blinds were superior, and (2) caused active drugs to replace inactive ones. This is a major fallacy. Better drugs have always displaced weaker drugs even before double-blinds became fashionable. If penicillin hadn't been developed we would still be using sulfonamides. Finally, the fact that double-blinds generally show active drugs to be inactive may be interpreted as showing that the technique is no good since it fails to detect therapeutic activity in compounds known to be active. Double-blinds enhance obecalp reactions.

Hogben and Wrighton, 1952, summarize their point of view in this way: (1) Hitherto it has been customary to assess the claims of therapeutic and prophylactic measures in statistical terms by recourse to tests which invoke a unique and so-called null hypothesis, namely that the procedures compared are equally efficacious. (2) This procedure has no bearing on the operational intention of the trial, viz., to find out how much advantage accrues from substituting one treatment for another. (3) Within its more restricted domain, the credentials of any significance test which takes within its scope only one hypothesis have now to meet the criticism that it takes into account only one sort of error, viz., that of rejecting the hypothesis when it is true. (4) A procedure which justifies assertions of so limited and conditional a scope may be a useful self-disciplinary convention; but its claims to rank as an instrument of statistical inference are no longer acceptable.

In spite of the mounting concern over the heavy reliance placed on results of double-blind experiments by men like Glick and Margolis (1962); Freyhan (1963) in his discussion of their paper; Bellak and Chasson (1964); Chasson (1957, 1959, 1960, 1961);

Plutchik, Platman, and Fieve (1969); Baird (1964, 1968); Cromie, (1963); Dalen (1969) (who questions the power of statistical tests. They are always used in double-blind experiments where the null hypothesis is used. This, according to Popper, is only a test of the null hypothesis for only that hypothesis is in danger of being refuted); Lasagna (1972); Feinstein (1970, 1971, 1972), Cotzias (1972). Many investigators will use no other technique and depend upon these questionable methods to resolve therapeutic controversy. They seem to be unaware that clinicians (doctors who work directly with patients) generally remain unconvinced by double-blind experiments. Double-blind experiments generally tend to prove active drugs to be inactive, i.e., yield obecalp responses. On the other hand, clinical trials in the older fashion may too often show inactive or slightly active drugs to be too active. I consider the latter error safer for medicine. If every drug used today had been forced to be double-blinded at its inception I doubt that we would have insulin, thyroid, aspirin, antihistamines, and other very valuable drugs. It has been claimed that the need to impress government agencies in the U.S.A. has substantially reduced the flow of new drugs onto the market. The government agencies will only accept double-blind studies.

The megadose vitamin B3 controversy is caught in the middle of this methodology controversy. It was established on the basis of the first double-blind experiments in psychiatry on acute and subacute schizophrenics treated with ECT, sedatives, and psychotherapy. The groups were randomized. One group received placebo. So far no one has attempted to reproduce this experiment using similar patients and similar treatment. We also reported that nicotinic acid alone did not benefit chronic patients such as are found in a mental hospital (see O'Reilly, 1955). This we reiterated in many of our reports.

However, in a series of studies over a five-year period, investigators seemed unaware of these conclusions and used subacute and chronic patients in mental hospitals without ECT. They concluded that nicotinic acid was not therapeutic for schizophrenia. Hoffer (1971) reviewed the reasons for their failure to obtain positive results.

In order to illustrate the inherent defects of the double-blind design, I will review several reports by Wittenborn, Weber, and Brown (1973), Wittenborn (1973), and DeLiz (1973), all bearing upon the same double-blind therapeutic trial.

Patients newly admitted to a mental hospital but ill at least 4.8 years on the average were randomized into two groups. The randomization was imperfect since the nicotinic acid group had been HI 4.8 years on the average and the control group had been ill for three years. The experimental group were given 3 grams per day of nicotinic acid with or without tranquilizers, depending upon the usual indications for giving tranquilizers. No ECT was used. The experiment was described as double-blind but no evidence is given that it was double-blind, nor were there suggestions that the code had been broken. Out of an initial group of 140, 75 completed two years of treatment. Out of 36 patients who dropped out of the experimental group, 20 or 58 percent were uncooperative. Out of 29 who dropped from the control group, 21 or 72 percent were uncooperative. Out of 83 who started on nicotinic acid, 24 percent dropped out while out of 57 on placebo, 37 percent were dropped because they did not cooperate. This suggests that more placebo patients were suffering from negative placebo effects and might be explained by the fact that some of them were aware of the fact that they were on placebo (DeLiz, 1973). Their general conclusion was that there was no therapeutic effect from the use of nicotinic acid.

In a more recent report, Wittenborn (1973) concluded that patients with certain descriptive indices did respond well to nicotinic acid. Each patient had been examined carefully before treatment was started. From

social data so obtained 12 factors (items) discriminated between placebo and nicotinic acid as the study progressed. He concluded, "These items appear to have a common implication for the description of the premorbid personality of patients who responded relatively well to the high dosage of niacin. Thus niacin was most effective for those patients for whom some features of definite interpersonal participation was found in the premorbid background."

The 12 items were combined into a predictor scale ranging from 0.00 to 1.00. The 75 patients who scored over 0.50 indicated easily evident pathology. The average scores indicated the experimental group had more pathology than the control group (consistent with the finding that they had been sick longer). Patients with high predictor scores generally responded better to therapy.

When high-score patients (over 0.6) were compared, the group on niacin responded better than the placebo group. For example, at 24 months the niacin group had half the depression of the placebo group. For schizophrenic excitement, the incidence of significant pathology was more than twice as great in the placebo group. Paranoia and hebephrenia were twice as prevalent in the placebo group. During the discussion after presenting this paper, Wittenborn reported that two-thirds of the niacin group were well compared to one-third of the placebo group. He concluded, "A high positive predictor score was associated with a clinical significant advantage for those patients who were treated with the high level of niacin medication."

"The items of the follow-up inquiry which showed the required consistency of relationship with the predictor score in the experimental group appear to be mutually consistent and describe the kind of person who participates in ordinary interpersonal interactions and who is approaching at least some of the tasks and challenges of his life constructively. The respective correlations based on the control sample displayed in Table 3 do not show a comparable or even an inverse set of relationships with the predictor scores. Thus it would appear that among the patients treated with high-dosage niacin those patients with a high positive predictor score have resumed the constructive quality of adjustment which the high positive predictor score implied for the premorbid status of these patients.

"The present post hoc treatment of the data reveals that persons whose premorbid history suggested a participatory life style tend to return to a participatory pattern of living after a year or more treatment with high levels of niacin. No such reconstructive trend was indicated for the control patients, however."

"There is a conceivable relationship between the fact that in the present sample patients with a high predictive score responded well to niacin and the fact that Hoffer and Osmond had claimed that niacin was more effective in relatively acute patients than in chronic patients. It is probable that patients who, in the present sample, had a high positive predictor score would have been classified by Hoffer and Osmond as acute schizophrenics. Perhaps in this way the differential effect observed by them could be in part explained."

"Why should a pretreatment disposition which has a favorable significance for patients treated with niacin have an unfavorable significance for patients not treated with niacin? One possible explanation for the paradoxical worsening in the control group draws upon observations that many patients with a favorable premorbid history are harmed by phenothiazine treatment in the sense that their remission is burdened."

Here is illustrated one of the main defects of the double-blind method — its inapplicability to heterogeneous groups. Since we had many times pointed out that early cases responded better and without need for as many other chemotherapies, there was no

need to use a mixture of early and well-developed cases. As a result, the double-blind by Wittenborn et al. yielded no significant difference. However, when the groups were purified (made more homogeneous) by using indications which sorted the early from late cases, the differences in the early cases became highly significant. The chronic cases should have been given ECT in combination with the vitamin if it had been desired to repeat our original double-blind experiments.

However, this is not the only problem. Recently Dr. DeLiz submitted a critique of the double-blind experiment run by Wittenborn. This is published in this Journal. The most serious charge is that the experiment was not double-blind. Patients on placebo discovered this and felt they were being deceived (as they were), and at least one and perhaps others tried to get niacin on their own. If the patients knew, so must have some of the staff. Since the preponderant feeling in most institutions is violently opposed to megavitamins, it is easy to understand that there would be immense negative pressure against niacin and positive pressure in favor of placebo. Obecalp would be strongly favored.

The findings from the second study are therefore even more significant, emerging from a study where a double-blind design was not adhered to and where obecalp reactions were favored by the staff. It is apparent that when treatment was matched to the right patients, i.e. nicotinic acid and tranquilizers to acute patients (phase one), the superiority of nicotinic acid over placebo was amply evident in a setting favoring obecalp responses. Two of the psychiatrists involved in the study eventually began to practice Orthomolecular psychiatry after leaving the study.

The main opposition to the Orthomolecular approach has come from these so-called double-blind experiments conducted by psychiatrists who accept only double-blinds as scientific evidence. The debate is not a debate between the physicians who have used similar treatment methods and have obtained conflicting results. This is the normal kind of scientific debate. It is between two sets of methodologies. It is clear that each group can reproduce each other's work. The Orthomolecular psychiatrist using any method as well as double-blind gets similar results. The double-blind methodologists also confirm each other since they all use essentially the same method.

Fortunately, double-blinds are inherently unconvincing to clinicians and to their patients and will not long stand in the way of good clinical observations made by Orthomolecular physicians, by their patients, their families, and by others who work with them.

In a personal communication, Dr. DeLiz reported that when relatives of patients on placebo complained, they were assured that this was untrue and that the patients were having fantasies. This would alienate patient from family. DeLiz wrote, "It is rather easy to see how the secondary symptoms, real reactions to a contaminated social and psychological environment as regards to niacin, were pervasive and might in their turn distort the totality of the psychiatric, social, and psychological rating procedures."

It is evident that the double-blind method does not solve the problem of proper trials and in fact is probably worse than usual clinical observations made by interested clinicians. Orthomolecular psychiatrists who value the welfare of their patients will be wise not to expose them to experiments of this kind. It is possible to run comparison experiments where one program is compared against another. The procedure is simple, relatively inexpensive, and the results would be decisive and convincing. But so far no one in any University or research setting has shown much interest in this kind of study. We will therefore have to depend on our clinical observations bolstered by psychological tests such as MMPI, HOD, and EWI, and by clinical tests of hair, blood, and urine (kryptopyrrole, for example).

REFERENCES

BAIRD, K. A.: Assessment of Drug Trials. Can. Med. Ass. J. 90,1279 only, 1964.

BAIRD, K. A.: Medicines Domination by the Control Worshippers. Can. Doctor, 27 only, 1968.

BELLAK, L. and CHASSAN, J. B.: An Approach to the Evaluation of Drug Effect During Psychotherapy: A Double Blind Study of a Single Case. J. Nerv. Mental Dis. 139, 20-30,1964.

CHASSAN, J. B.: On the Unreliability of Reliability and Some Other Consequences of the Assumption of Probabilistic Patient States. Psychiatry 20, 163-171, 1957.

CHASSAN, J. B: On the Development of Clinical Statistical Symptoms for Psychiatry. Biometrics 15, 396-404, 1959.

CHASSAN, J. B.: Statistical Inference and the Single Case in Clinical Design. Psychiatry 23, 173-184, 1960.

CHASSAN, J. B.: Stochastic Models of the Single Case as the Basis of Clinical Research Design. Behavioral Science 6, 42-50,1961.

COTZIAS, G. C: Limitations of Controlled Double Blind Studies of Drugs. New Eng. J. Med. 287, 937 only, 1972.

CROMIE, B. W.: The Feet of Clay of the Double Blind Trial. The Lancet 2, 994-997,1963.

DALEN, PER.: Causal Explanations in Psychiatry: A Critique of Some Current Concepts. Brit . J. Psychiatry 115, 129-137, 1969.

DeLIZ, A. J: A Note of Criticism Concerning Wittenborn's Paper on an Experimental Double-Blind Research Design Dealing with the Action of Nicotinic Acid on Schizophrenia. J. Orthomolecular Psychiatry 2: 3, 1973.

FEINSTEIN, A. R.: What Kind of Basic Science for Clinical Medicine? New Eng. J. Med. 283, 847-852, 1970.

FEINSTEIN, A. R.: Clinical Biostatistics IX. How Do We Measure Safety and Efficacy? Clin. Pharm. and Ther. 12, 544-558, 1971.

FEINSTEIN, A. R.: The Need for Humanized Science in Evaluating Medication. The Lancet 2, 421-423, 1972.

FREYHAN, F.: Discussion of Paper by Click and Margolis.

CLICK, B. S. and MARCOLIS, R: A Study of the Influence '.'of Experimental Design on Clinical Outcome in Drug Research. Am. J. Psychiatry 118, 1087-10%, 1962.

HAWKINS, D. R. and PAULING, L: Orthomolecular Psychiatry. W. H. Freeman and Co., San Francisco, 1973.

HOFFER, A.: A Theoretical Examination of Double Blind Design. Can. Med. Ass. J. 97, 123-127,1967.

HOFFER, A.: Megavitamin Therapy for Schizophrenia. Can. Psychiat. Ass. J. 16, 499-504, 1971.

HOFFER, A. and OSMOND, H.: Double Blind Clinical Trials, j. Neuropsychiatry 2, 221-227,1961.

HOFFER, A. and OSMOND, H.: Some Problems of Stochastic Psychiatry. J. Neuropsychiatry 5, 97-111, 1963.

HOGBEN, L: Statistical Theory: The Relationship of Probability, Credibility and Error. George, Allen and Unwin, Ltd., London, 1957.

HOGBEN, L. and WRICHTON, R.: Quoted in Hoffer and Osmond (1963).

LASAGNA, L.: The Nature of Evidence: Triangle 11, 145-152,1972.

O'REILLY, P. O: Nicotinic Acid Therapy and the Chronic Schizophrenic. Dis. Nerv. System 16, 67-72, 1955.

PLUTCHIK, R., PLATMAN, M.B., and FIEVE, R. R.: Three Alternatives to the Double Blind. Arch. Gen. Psychiatry 20, 428-432, 1969.

WITTENBORN, J. R.: The Selective Efficacy of Niacin in the Treatment of Schizophrenia. N.I.M.H. Conference, Washington, 1973.

WITTENBORN, J. R., WEBER, E. S. P., and BROWN, M.: Niacin in the Long-Term Treatment of Schizophrenia, Arch. Gen. Psychiatry 28, 308-315, 1973.