

Editorial

The Public Reaction to Double-Blind Controlled Clinical Trials

The *New York Times*, October 3, 1999, detailed the trials and tribulations of research physicians who were not able to enroll enough patients to proceed with double blind studies. A double-blind controlled therapeutic trial, using treatment widely used for breast cancer and for ovarian cancer was called off because over 100 medical institutions searching for 2.5 years could not find the 285 patients needed. They found 25. The major issue is whether marrow transplants for treating breast and ovarian cancer are of greater benefit than standard chemotherapy alone. Since 1980 this treatment has been provided by many private and university hospitals even though these double-blind studies had not been completed. This is not a criticism. The vast majority of medical and surgical treatments have never been put through the double blind procedure. Medical research purists who demand that every new treatment be subjected to this test ought first to examine their own bailiwick and subject their own favorite treatments to the same test and clean up their own field. Recently five double blind experiments were completed and although the results showed that the bone marrow treatment was no better, physicians using the treatment could not admit that it should no longer be used. But the main issue was why patients with breast cancer and other forms of cancer would not allow themselves to be enrolled in these studies.

The main reason offered was that with terminal illnesses patients will grasp at any treatment which offers them some hope. This is true and I agree with these patients who make these decisions but it does not really answer the question. Is it possible that the public knows something about these double-blind experiments that the researchers do not know. Do these patients know instinctively that these double-blind experiments, the gold standard of modern

medicine, are perhaps best labeled as “fools gold” standard, that they are unethical, that they do not remove bias in the evaluation of treatments and that they remove the most essential element of any doctor-patient relationship, hope. Without hope patients have nothing, and in the treatment of depression, as soon as they lose hope they become immediate risks for suicide.

Another reason is that diseases are difficult to define and samples from the theoretical universal population are very difficult to find. By the time each patient is examined to see if they fit the parameters of the study there are very few left who can be entered into the trial. This was one of the problems with the bone marrow transplant trials. Thus in the trials testing drugs to treat AIDS it is becoming increasingly difficult to find patients who are not taking vitamins and other complementary therapies. Patients known to be taking vitamins are not enrolled because it takes away from the purity of the trial. Many patients do not tell their physicians.

I am qualified to examine seriously the double blind method but I have no vested interest in defending it. With Humphrey Osmond we were the first psychiatrists to run double blind experiments when we examined the therapeutic properties of vitamin B₃ in the treatment of schizophrenia. Between 1952 and 1960 we completed six clinical trials. They all showed that adding this vitamin doubled the recovery rates of these patients. However we knew from simple previous clinical trials (pilot studies) that this vitamin was effective and it was amply confirmed on thousands of our patients I and my colleagues treated in Canada and the United States. We did not need the double blinds to conclude that the treatment was effective, but we did need them for more practical reasons. The research statistician in Ottawa advised us that we must conduct our experiments in this way, with the implied threat that we would not get any research funds if we did

not. The method we were advised to use became the modern gold standard.¹ To declare today that the double-blind is not the gold standard is like criticizing the Emperor for not wearing any clothes; it is one of the major heresies of modern medical science. At that time we had no experience with this clinical design and hoped that it was really the best way of testing drugs. We also knew that it would be easier to receive scientific acceptance of our conclusions if the experiments were done this way. But as we continued to work with them it became clear that there were many errors in the design. We were able to complete these experiments simply because our patients were captive patients in psychiatric wards and had no say in how they were going to be treated. These experiments could not have been done with patients in the community who were free to cooperate or not. Finally it became clear that in spite of the fact we had done these experiments using the double-blind method they were not acceptable because in those years long ago psychiatrists did not believe that schizophrenia was a biochemical disease. They believed that it was psychosocial problem caused by some malign influence in the family (usually mother) or in the community. They could not swallow the idea that a psychosocial disease as complex as schizophrenia, could possibly respond to a simple vitamin, especially since they all knew that no one in North America suffered from a deficiency of vitamin B₃, or any vitamins for that matter. I will discuss only three issues:

- (1) Is the double-blind a valid measure of therapeutic response? Is it really the best we have, the gold standard?
- (2) Is it ethical?
- (3) Does it remove bias?

1. The Gold Standard

In essence the double blind methods consists of two components. It is a method of selecting patients to be given the active

compounds to be tested and the patients to be used as a control. I do agree that we have to compare the outcome of treatment of any new therapy against the best results that are found with the previous orthodox or standard therapy. There are many ways of doing so including using concurrent groups, or using historical controls. But it is important to have identical groups so that proper comparisons can be made. I agree the groups ought to be randomized so that we can approach a more or less condition of equality. I do not agree that they have to be double-blind. The second major element is the mathematical analysis of the results which is based upon probability theory. I will consider the mathematical procedure first.

According to probability theory a phenomenon to be investigated has to be invariant i.e. provided the same conditions are present the results will always be the same. This applies to gravity for example which appears to be invariant. It can be applied to dice which are manufactured according to stringent standards so that out of six tosses any one number will on the average come up once. If these strict manufacturing methods are not used the dice are not trustworthy. But this criterion can not be applied to any biological phenomenon. A disease is not invariant. Epilepsy described 2000 years ago is not described in the same way today and uses diagnostic techniques unheard of in the past. Diseases change, the causes change and the way they respond to treatment changes. Thus one of the essential elements of probability theory is not present in clinical studies. The second element is that the groups (they are called samples) truly represents the total population from which they are drawn. This only approximates being true in human studies when thousands of cases are used and may not be true even then. Thus the two main elements for the theoretical basis for the double-blind controlled methods are not present in dou-

ble-blind controlled therapeutic trials. Sir Lancelot Hogben,² analyzed the double-blind in detail and it was his book which brought to my attention these basic two serious flaws in the theoretical basis of the double blind.

In my opinion it may have some value for certain conditions. It is not required for diseases for which we have no effective treatment. If no patients recover or very few do, to the point that it is considered a miracle when they do, then why bother? Any treatment that will do better than that should and will be used. For severe infectious encephalitis I doubt that any physician would demand double-blind controlled experiments when any treatment that appears to be effective is available. It may be most needed for conditions which do not threaten life, which are relatively mild and for which in most cases the natural recovery rate is so high they are merely inconvenient, for example, the common cold. But for diseases like terminal cancer not only do double-blinds not serve any useful purpose they are in fact detrimental since they remove hope and it is now well known that hope is an important element in the treatment of terminal cancer.

In 1970 the head of the M.D. Anderson hospital in Texas invited me to present my views about the use of double blinds to the annual meeting of the National Cancer Institute, USA, in Puerto Rico.³ He told me that he and his statistical advisors had studied my criticism of the method⁴ as applied to cancer and had concluded that I was right. He was under pressure from the National Cancer Institute to conduct all their trials with this method. I hope my criticism of the method helped him resist the pressure.

Today double-blinds are demanded even for diseases for which there is no known effective treatment. An example is idiopathic thrombocytopenic purpura (ITP). A few years ago, a university professor of medicine discovered that vitamin C

cured this condition.⁵ Of the first eight he treated all recovered. There was no known cure, but when he submitted his article to the *New England Journal of Medicine* it was rejected because it had not been double blind. I have treated two cases and both recovered. Since there is no treatment why would anyone demand that a placebo study be done? The placebo recovery rate is well known to be zero.

Do we have better methods? We do and they go back to the dawn of medicine. They are accurate clinical observations made by ethical physicians who have no vested interest in commercially developing their compound. They promote their treatment by reporting as accurately and as frequently as possible what they have found, keeping nothing secret and offering no secret remedies. The next step is for physicians of equal stature to repeat these open clinical trials. If they find similar results than the treatment is established. If they do not, they must then get together and find out exactly what each did so that they establish out why they obtained divergent results. When this is done frequently enough the true value of their treatment will become clear. Any new treatment for the same disease will then have to be tested in exactly the same way. If there is preliminary evidence that the new treatment is better and certainly no worse than the standard treatment, then a randomized, controlled, open experiment can be conducted and physicians can honestly advise their patients what they are going to do and assure them that they will be no worse off and may in fact be much better off. Hope is accentuated, not washed out by the dishonesty of the double blind. Many patients will still refuse to cooperate and they must not be pressured.

2. Is It ethical?

In my opinion it is not for it introduces into the therapeutic equation dishonesty. Patients are sophisticated enough today to know that they are not being advised hon-

estly and physicians must feel very uncomfortable when they know they must lie. How can they impart any sense of hope and optimism when they know that the compound they are giving those patients is a useless placebo or some drug in which they have no confidence? Many years ago I tried to do a study on my private patients. I wanted to compare the efficacy of a new anti-depressant which was slightly different from another one but was claimed to be freer of side effects. I explained to each patient what I hoped to do, advised them they would not have to pay for the drugs and cautioned them that if they had the slightest reservation they must tell me so that they would not be included in the trial. I told them that they would be no worse off and that they would have fewer side effects. The required number agreed. But over the next six months I discovered to my surprise that the drop-out rate was much greater than I had expected. These patients voted with their feet. They lost confidence in me because I was offering them a new product and did not know which one they would get and they refused to participate. The drop-out rate from the previous drug was not nearly as high as it was in this attempt to do the double-blind.

3. Does it Remove Bias?

This is the most serious criticism, for the main reason for using this method is to remove bias on the part of the evaluator. This is well known to investigators but they turn a blind eye to it. Many have demanded that no double-blind is conducted properly unless a method is worked in to prove that bias did not exist. Very few studies included these controls on the blindness of the study. It is almost impossible to prevent bias. Many years ago an investigator told me about his double-blind controlled experiment to test the efficacy of intravenous valium for patients who were drunk from alcohol. He told me that the entire procedure was double-blind. Then he added that

the nurses giving the injection knew what they were giving since the valium in the fluid gave it slightly different visual properties and they could tell when what was given by the way it swirled around in the syringe. Yet this experiment was considered double blind. The Lehmann Ban studies in Montreal on the efficacy of niacin did not use a hidden control. They called their study double blind but it is impossible to avoid knowing that when you flush you must have gotten niacin and when you don't flush you probably did not get it.

It is possible that a test may be very good and accurate even though it does not conform to certain rules. The double blind test is not one of these. The double blind has never been validated. It is a method which became the gold standard with no examination of whether it is valid, whether it does what it is supposed to do. If it is the gold standard, which standard are we talking about—the gold standard or the fools gold standard?

Over the years I challenged many supporters of the double blind to explain why these validation experiments were not conducted. They simply ignore the question. The gold standard for modern clinical testing of drugs is based upon a series of hypothesis, none of which have been supported by valid research data. Why do we still use it?

I think the women who refused to participate in the breast cancer trial were correct and ought to be congratulated. Rather than questioning their motives and education to determine why they refused, I suggest that the investigators examine their own values and beliefs carefully and objectively and leave aside all matters such as getting papers accepted for publication or getting research grants. Perhaps one day the *New York Times* will give equal space to the question Why do investigators persist in using a clinical design that is so prone to error as the double-blind?

Discussion

A major criticism of the double-blind, which is usually not discussed, is that it often shows compounds to be not effective when in fact long usage has shown that they are effective. The best examples are the early double-blind experiments which showed that L-dopa was not therapeutic for Parkinson's disease. A more recent example just appeared in which it is shown that imipramine is no more effective than placebo in treating depression. Physicians and psychiatrists know from long usage that this anti depressant does have anti-depressant properties.

Kapser, Moller, Montgomery and Zondag⁶ reported on a well designed double-blind, therapeutic trial comparing imipramine, fluvoxamine and placebo on 338 depressed patients with five North American centers cooperating. Their conclusion was "Fluvoxamine but not imipramine was significantly superior to placebo in severely depressed patients" and "No significant improvement was observed with imipramine"

Modern attempts are being made to bypass the double-blind method by using what is called meta analysis. With this technique single studies which are considered inadequate alone are simply lumped together into a large statistical table using mathematical tricks and then conclusions are drawn from these.

The powerful adherence to this flawed methodology reminds me of a meeting I attended in Prague about 30 years ago. This was the first international meeting held in Czechoslovakia during the communist regime. Most speakers were from the communist controlled countries. As each speaker presented his report I was struck by a uniformity in the structure of their presentation. Each began with a statement that communism was of course the superior system. Then they would report their medical or scientific findings. Finally after their conclusion they would finish by stating that their conclusions proved the su-

periority of the communistic system. I turned to one of my neighbors and asked him bluntly what was going on. He answered cynically, "They have to live." Do our modern users of the double blind have the same fear that they will not survive?

Communism was overthrown and I am certain the same scientist today would present papers that are very much like the ones presented by western scientists. Isn't it about time we forced all the defenders of the double-blind to stop bowing to the double-blind methodology, to seriously have another look, to let us know why they insist on using a method that has never been validated and which contains so many inherent defects that it is probably one of the worst ways of doing clinical experiments? I must make clear I am not against controlled experiments. These are essential but they do not have to be double-blind and they must take into account the many errors generated by sampling techniques. An example of a recent major trial which in my opinion should never have been published is the Finnish double-blind controlled experiments using as subjects heavy drinkers and smokers. The effect of beta carotene and vitamin A was studied using randomized groups. The randomization was not good and the group given beta carotene had been smoking on average one more year than any of the other three groups. The authors found that the beta carotene group had slightly more patients with lung cancer but that these results were not statistically significant. In fact, the fact they had been smoking one year more probably did make a difference since lung cancer is progressive and may take years to fully develop. They should have concluded only that the beta carotene did not inhibit the development of cancer. Yet they allowed the conclusion to be used widely that beta carotene increased the incidence of lung cancer. There were many other points about their design which have been severely criticized by others. The public not being aware

of the niceties of this statistical technique assumed simply that beta carotene increased the incidence of lung cancer.

In my opinion the double blind method best serves not the interests of science or medicine, but the interests of editors of journals, scientists who know their papers will be rejected if they are not double-blinded, drug companies who are forced to use them at enormous cost, those who evaluate research grant applications, and civil servants who have to rule on whether new drugs can be released to market, for the double blind removes the need to really think about the clinical results. All is left to the holy $P < 0.05$. The $P < 0.05$ actually decides whether drugs will be released. And $P < 0.05$ can be obtained if the samples are increased in size until these statistical significant but clinical artifacts are obtained.

References

1. Clancy J, Hoffer A, Lucy J, Osmond H, Smythies J, Stefaniak B: Design and planning in psychiatric research as illustrated by the Weyburn Chronic Nucleotide Project. *Bull Men Clinic*, 18:147-153, 1954.
2. Hogben L. *Statistical theory. The relationship of probability, credibility and error*. Allen & Unwin Ltd. London, 1957
3. Hoffer A: Symposium on statistical aspects of protocol design. Discussion. *Cancer Clinical Investigation Review Committee*, San Juan, Puerto Rico, 224-229, Dec. 9-10, 1970.
4. Hoffer A: A theoretical examination of double-blind design. *Can Med Assoc J*, 97: 123-127, 1967.
5. Hoffer A: Vitamine C En purpura thrombocytopenica. *Orthomoleculaire*, Holland, 259-261, Dec. 1989.
6. Kasper, S. Moller HJ, Montgomery SA, Zondag E: Antidepressant efficacy in relation to item analysis and severity of depression: a placebo controlled trial of fluvoxamine versus imipramine. *Int Clin Psychopharmacol* 90, Spp 4, 3-12, 1995