

THERAPEUTIC TRIALS IN CANCER¹

INDEXED

by

Marvin A. Schneiderman, Ph. D.
National Cancer Institute
National Institutes of Health
Bethesda, Maryland, United States of America

The controlled clinical trial (or in the language of I. D. H. Todd, the "random trial") is a part of the process by which a new therapeutic modality comes into eventual use in cancer. The principles of the controlled therapeutic trial as set out by Hill,¹ Lasagna & Meier,² Marshall & Merrell³ are as applicable in cancer as in any other disease. The nature of much of cancer, however, requires emphasizing different portions of these authors' arguments.

The special problems of the trial of new treatments in cancer arise out of three special qualities of cancer:

1. The considerable toxicity or danger inherent in most new treatments - requiring a balancing of toxic and therapeutic effects within a very narrow range.
2. The long-term nature of cancer, driving experimenters to seek measures of response evident earlier than the ultimate measure, survival in good health.
3. The ethical proscriptions which make it difficult to mount an acceptable trial in early disease - where probability of cure or effective treatment is greatest.

¹ Working paper prepared for WHO Expert Committee on Cancer Treatment, Geneva, Switzerland, 9-15 March 1965.

Before examining the special problems of cancer, it may be wise to review some general principles of the controlled clinical trial:⁴

1. A trial is not ethically permissible unless one has strong reason to anticipate a net gain from the new treatment over the results derivable from present treatments. No trial is ethically permissible unless one has honest doubts about which of the contrasted treatments is really better.
2. In a comparative trial there must be no systematic bias tending to favour one treatment. This implies that, in so far as we are able, we must try to "ensure that these patient groups are similar in all relevant respects, except in the treatments they receive".⁵ This goal is best achieved through two steps. First: for a patient to be admitted to the trial, he must be judged able to receive any of the treatments included in the trial. A person too old or too ill to be treated surgically, but able to receive radiation therapy is not eligible for a trial contrasting radiation and surgery. The patient's eligibility for all treatments in the trial must be established prior to his assignment to any treatment. Second: having established the patient's eligibility for any of the competing treatments, the patient is assigned by a formal randomization procedure to one of the treatments.

Any one of several formal randomization procedures is acceptable. One of the most effective is to prepare a set of cards (for example, small file cards), with one card for each patient who will eventually be included in the trial. On each card is written the name or other identification of one of the contrasting treatments. As many cards are prepared as one expects individuals in each of the treatments. Then, from a table of random numbers (6, 7), following the table in some prescribed fashion, a random number of sufficient number of digits is placed on each card. (The "sufficient number" of digits is decided by total size of the experiment. If the total experiment size is 10 or less, a two digit number is usually sufficient. If the experiment size is between 10 and 100, a three digit number will do; for an experiment size between 100 and 1000, a four digit number, etc. The purpose of having more digits than the experiment size might indicate is to keep duplication of numbers low.) The cards are then arranged in increasing order, according to the random numbers. Ties are broken by flipping a coin or some similar procedure.

The cards are now placed in envelopes, numbered in increasing order. The envelopes are usually sealed. The first patient eligible for the trial receives the treatment called for on the card in the first envelope, the second patient, the treatment on the card in the second envelope, etc.

Not acceptable are the informal, "nearly random", assignment procedures. Haphazard assignment is not random in the sense intended here. Alternating assignment is not random. Assignment by day of the week is not random. Studies have shown that assignment by patient serial number, telephone number, year of birth may not be random.

Other acceptable procedures involve using an honest roulette wheel, tossing an honest coin, or an honest die (the Japanese Standard Association has prepared 20-sided dice in three colours for use as a randomization device.) The patient's birth date (date in the month) can be used - if the date is not known or available to the investigator before he has decided that the patient is eligible for the study.

Where there are elements known to have an effect on response to treatment (age, sex, institution, stage of disease, etc.) it is proper to stratify by these elements and randomize within strata. Restraint in creating strata will avoid fractionating the patient population into too many strata each with too few individuals. Billewicz⁸ has demonstrated that matching (intense stratification) may be costly in time and effort and not repay the bother.

3. Evaluation of results must be unbiased. The response measure must be an objective measure that cannot possibly be read differently by different observers (e.g. survival) - or the experiment must be designed to permit the unbiased evaluation of measures subject to varying interpretation. This is often achieved by the technique called "double-blind". Double blind is achieved by having the person evaluating the response, and the person receiving the treatment both unaware of the specific treatment the patient is receiving. If the person evaluating the response is different from the person treating the patient, the person treating the patient should also not know which specific treatment the patient received. Some people refer to this situation as "triple-blind", but this term impresses me as an abomination - to be avoided. There is no need for it. As Mainland⁹

remarks "'Double-blind' implies that no one on the physician's side (himself, assistants, nurses, secretaries, and so forth) and no one on the patient's side (himself, relatives, friends and so forth) - no one at all who can in any way influence patients' attitudes, feelings, behaviour, measurements, replies to questions, etc. - shall have any knowledge of the compounds that individual patients are receiving."

Conducting a double-blind trial lets one use non-objective evidence in an unbiased way. Since measures of pain, feelings of well-being, ability to return to work or to carry out usual functions may be strongly subjectively influenced, one is hardly ever able to use them except in a double-blind situation. Sometimes these are the most useful measures, so to conduct a non-blind study requiring that one not use these measures could be wasteful and misleading.

The proscription that the treating physician not know the specific treatment (drug, usually) the patient is receiving provides no ethical barrier, since he must know all the treatments the patient could possibly be receiving in the trial, and therefore can be expected to be on the look-out for the side effects or toxic reactions from all the treatments. There are many recorded instances of "side-effects" seen in placebo-treated patients. One wonders if placebo effects could be discovered in possible-to-bias unblinded experiments.

4. The trial must be neither too short to develop a convincing argument (for or against any of the contrasted treatments), nor so long that an excessive number of patients are treated with an inferior treatment. This is not easy to achieve, even with expert statistical assistance.¹⁰ Without such assistance it is even harder.

5. The last step in the clinical trial is the presentation of the data; clearly, fully and honestly. At present most editors seem to look for statements of "statistical significance" or appropriate probability values, or something of the sort. Unfortunately, there has too often been misuse of statistical significance tests, so that again professional statistical help is worthwhile. The misuse has been so great that Cromie¹¹ warns "There is certainly a case for accepting a claim of 'statistical significance' with caution."

In addition to these tribulations of every trial, the assessment of a new treatment in cancer poses several others. In chemotherapy evaluation at least, there appear to be two different kinds of trials, conducted for distinctly different purposes. There are the patient-oriented trial, and the drug-oriented trial. Patient-oriented trials are designed to give answers to the question "How shall I treat the next patient with cancer who comes into my care?" The drug-oriented trials attempt to answer the questions "Has this drug enough promise that I can bring it into a patient-oriented trial?" and "If I were to bring it to a patient-oriented trial, how is it best to give it?"

The patient-oriented trial must consider the response of every patient placed on the treatment - no matter for how short a period. Since one wants to know what happens to all patients (of a certain category) from the moment treatment starts, once a patient is placed on treatment there is no exclusion from follow-up. Of course, the patient can (or must) be taken off the treatment if it appears that a continuation is likely to result in his injury, disability or death. But he must then be recorded as a "failure".

The drug-oriented trial is concerned with the performance of the drug under the best possible conditions. This permits the dropping of patients from analysis if they have been "inadequately" treated, and for similar reasons. Permission to exclude some patients from final analysis derives from the argument that the drug must be able to do well when all the elements are biased in its favour - otherwise the drug is most unlikely to do well in the patient-oriented (hopefully unbiased) trial. There are three immediate consequences of this permissiveness. First, the drug-oriented trials do not provide unbiased estimates of what a drug will do in a general population of patients. Second, one must be prepared to find that the same drug in a patient-oriented trial will not do as well as the results of the drug-oriented trial promised. Third, it is a disservice to suggest the introduction of a drug into general use following only a drug-oriented trial.

Since most new treatments for cancer carry the threat of considerable toxicity one wishes to give them at a dose that is the result of a judicious weighting of positive (response) and negative (toxicity) elements - an optimal dose. The determination of an optimal dose is not easy. Cromie¹¹ remarks "the optimum

dose is usually determined by reference to previous trials. If this does not exist, then a pilot trial must be carried out in similar patients." This "piloting" is the major function of the drug-oriented trials, but even here their inherent bias in patient selection and reporting can be misleading.

Optimum implies a balancing of the negative and the positive. Early dose-finding trials in cancer have often been carried out in rather ill patients - where the prospect for response was small. These trials have usually provided only estimates of doses at which toxicity will occur, i.e. estimates only of the "negative" - hence a tolerable rather than an optimal dose. Since it is quite likely that ill patients show toxicity at low doses, the further strong possibility exists that the tolerable toxic doses found by early trials will be too low. Working at a non-optimal dose could bias later patient-oriented trials against the new drug. There seem to be two opposite forces at work, however, that may inadvertently be balancing each other. Finding a dose in ill patients may give too low a dose. Taking a maximum tolerated dose (see below) may give too high a dose. A maximum tolerated dose in very ill patients might be just right.

A classic approach to the dose-finding problem has been through the "therapeutic-index" route. Following Ehrlich & Hata,¹³ experimenters have attempted to use information on both response and toxicity through finding a minimum effective dose (MED) and a maximum tolerated dose (MTD). The ratio of MTD/MED is the "therapeutic index". This has many uses, related to the relative safety of different drugs - but it has unfortunately little to do with finding an optimal dose. What usually happens in cancer is that the two doses MED and MTD are found, and if the material has an acceptable therapeutic index, a dose near the MTD is chosen as the suitable dose for further work. Two deficiencies cloud this procedure. First, the dose chosen is not optimal (except by chance) under any definition of optimal. Second, the Ehrlich-Hata procedure is wasteful of data, using only the relative positions of dose response curves for toxicity and response, to determine the MED and MTD. It has been shown¹⁴ that the slopes of the curves must be considered in finding an optimal dose and are of considerable consequence in choosing between two drugs with the same therapeutic index. New experiment

strategies, capable of eliciting both response and toxicity information in an unbiased way, and thus leading to reasonable determination of optimal doses, still need to be developed.

One further problem exists. It is sometimes difficult to separate toxicity due to the disease from toxicity due to drug. Recognizing this, at least one group has attempted to do dose-finding in a randomized, blind trial - with different doses of the drug assigned to patients through a formal randomization scheme. These workers¹² recommend "the use of randomized assignment of patients, fixed non-escalating doses and the inclusion of a low dose for separation of drug toxicity from natural disease progression." If this could be done as Cromie suggests in "similar" patients, similar to the ones who will eventually be treated by the drug, then an evaluation of both response and toxicity, (and hence optimality) might possibly be made.

The treating physician's goal in treatment of cancer is to return the patient to the full, active life that he experienced before his illness. It is extremely difficult to measure what is a "full, active life", and to determine whether after treatment it is as full or as active as it was before the illness. An immediate second choice is to measure survival, since long survival is more likely to be associated with a full active life than is short survival, though not always. Paradoxically, if one is lucky, one might have to wait a long time before deciding between the merits of competing treatments based on survival.

If survival is only a substitute measure - and one for which one has to wait a long time, perhaps another substitute measure which gives an answer quicker might be used, if it could do almost as well. The desire to get an earlier answer has led experimenters to a variety of substitutes, from measuring tumours, or various chemicals in the blood or urine, to such directly disease-related measures as cell counts in bone marrow specimens in leukemia. Usually the substitute measures are also the so-called objective measures.

These substitute measures spawn three major problems. First, do they really correlate with the ultimate desired achievement of the full active life? Second, if one's studies are confined to patients with "measurable" disease, can one logically extend the results from these patients to patients who do not have measurable disease? Third, can the measures themselves distort or warp the trials?

Most attempts to show the correlation of substitute measures to some other desired end have tried to correlate "response" with survival. This is usually done by showing that "responders" have longer survival than non-responders. If it takes some time for response to occur, this can be misleading. If one cannot be sure that a response has occurred until (say) two weeks of treatment have elapsed, then all deaths before two weeks must fall into the non-responder group, and thus lower the average survival of the non-responders. (This fallacy is similar to the one which purports to show that judges on high courts tend to have much longer life expectations than ordinary people. Since persons rarely become judges on high courts before some rather advanced age, all people who die before this age fall into the non-judge group and therefore lower the average life expectancy for non-judges.) In a paper²³ reporting the best responses in acute leukemia in childhood (99.4 per cent. remissions; median survival of about 18 months) that I have seen, the author remarks ". . . but it must be remembered that by definition only patients surviving at least 30 days after diagnosis were included."

For some objective measures such as tumour-shrinkage in lung cancer, it has sometimes been difficult to show any relation between "response" and survival.¹⁵ While a convincing argument can be developed that a patient whose measurable tumour shrinks ought to be better off than a patient whose tumour does not shrink, empirical verification is necessary. Asher¹⁷ wrote "One of the most important things about treatment is that it should be effective - not merely that it ought to be effective." The same can be said of measures of response.

There is some evidence from the study of advanced breast cancer that slightly longer survival (median survivals 7-1/2 months versus 6 months) is found in patients treated with a drug which produced relatively high (about 19 per cent.) response compared with drugs producing low (about 2.5 per cent.) responses.¹⁶

The second and third problems - the problems of generalization and of distortion probably belong together. It is clear that in order to generalize from them, patients with "measurable" disease should differ only in the detail of measurability from the patients without measurable disease. Generalizing to other

patients, even when the early substitute measure correlates closely with the desired return to full active life could be in error if there were other differences. Thus, the pressure to find patients with "measurable" diseases not only has the effect of reducing the patient population on whom trials can be conducted but it also may have the effect of creating misleading extrapolations.

The last difficulty - the one of potential distortion - derives from a possible misunderstanding of why early objective measures are sought. Some investigators seem to have felt that given "objective" measures, it became no longer necessary to conduct controlled trials with a contra-test, or other treatment group. Since it is clear that the probability of response in cancer is strongly related to what one might call the patient's general condition,^{18,19,23} it becomes important that patient condition be taken into account in conducting a clinical trial. It is also one of the reasons that it is so difficult to compare the results from one institution to another.^{20,21} Even with agreement on the objective measures, an uncontrolled trial in one institution, on one kind of patient, can produce a substantially different set of results from a similar uncontrolled trial in another institution. Objectivity is not a guarantee against bias. The presence of a standard treatment as a control, in a properly randomized trial, would at least give different institutions a common bench mark. One of the advantages of the co-operative controlled clinical trial is that internal replications exist between institutions, so that one may be able to contrast the results on competing treatments at several centres, each possibly with patients with different clinical characteristics.

Finally, because of the toxicity of many anti-cancer agents, a special ethical problem exists in studies in which adjuvants (usually chemicals, and sometimes radiation) are added to a standard "curative" treatment. Since one has hopes that the standard treatment (i.e. surgery in early lung cancer) will often be curative of itself, great care must be taken that the adjuvant treatment (i.e. HN_2 , Thiotepa) does not pose an added risk for the patient. The Lancet²² concludes a leading article with "It would be a great pity if the over-enthusiastic and unselective use of combined therapy were to revive old fears of the treatment being worse than the disease." The adjuvant is the added push. One must take care that it pushes the disease more than it does the patient.

Tailored to the special needs of cancer, the controlled clinical trial is the ethical and effective tool to test potential advances in therapy.

REFERENCES

1. Hill, A. B. (1951) The clinical trial, Brit. med. Bull., 7, 278
2. Lasagna, L. & Meier, P. (1959) Experimental design and statistical problems in clinical evaluation of new drugs, ed. Waife, S. O. & Shapiro, A. P., Hoeber-Harper, New York
3. Marshall, E. K. & Merrell, M. (1949) Clinical trial of a new drug, Bull. Johns Hopk. Hosp., 85, No. 3, 221
4. Hill, A. B., ed. (1960) Controlled clinical trials, Blackwell Scientific Publications, Ltd., Oxford
5. Armitage, P. (1960) The construction of comparable groups. In: Hill, A. B., Controlled clinical trials (reference 4), p. 14
6. Rand Corporation (1955) A million random digits, The Free Press Glencoe, Illinois
7. Kendall, M. G. & Smith, B. B. Tables of random sampling numbers, Tracts for Computers, No. 24, Cambridge University Press
8. Billewicz, W. Z. (1964) Matched samples in medical investigations, Brit. J. prev. soc. Med., 18, No. 4, 167
9. Mainland, D. (1965) Notes from a laboratory of medical statistics, Note 98, p. 3
10. Schneiderman, M. A. (1964) The proper size of a clinical trial: Grandma's strudel method, J. New Drugs, 4, No. 1, 3
11. Cromie, B. W. (1963) The feet of clay of the double blind trial, Lancet, p 994 (9.11.63)
12. Carbone, P. et al. (1965) The feasibility of using randomization schemes early in clinical trials of new chemotherapeutic agents: hydroxyurea (NSC-32065), Clin. Pharmacol. Ther., 6, No. 1, 17
13. Ehrlich, P. & Hata, S. (1910) Die experimentelle chemotherapie der spirillosen, Springer, Berlin
14. Schneiderman, M., Myers, M., Sathe, Y. S. & Koffsky, P. (1964) Toxicity, the therapeutic index, and the ranking of drugs, Science, 144, No. 3623, 1212
15. Zubrod, C. G. et al. (1960) Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide, J. chron. Dis., 11, No. 1, 7
16. Mohberg, N. (1965) Tabulations from the cooperative breast group, (Personal communication)

17. Asher, R. (1961) A priority, Lancet, p. 1403 (23.12.61)
18. Sathe, Y. S., Zelen, M. & Sweifel, J. (1965) On the interpretation of clinical trials, J. chron. Dis. (In press)
19. Gold, G. L. (1965) The comparative effectiveness of three alkylating agents in the treatment of advanced cancer. (In preparation)
20. Ellison, R. R. (1961) Clinical applications of the fluorinated pyrimidines, Med. Clin. N. Amer., 45, No. 3, 677
21. Schneiderman, M. (1962) Clinical excursion into 5-fluorouracil, J. chron. Dis., 15, 283
22. Lancet (1964) Leader entitled Combined therapy in malignant disease, p. 797 (10.10.64)
23. Zuelzer, W. W. (1964) Implications of long term survival in acute stem cell leukemia of childhood treated with composite cyclic therapy, Blood, 24, No. 5, 477