REVIEW Jpn. J. Cancer Res. 83, 547–558, June 1992

Basic Design Considerations for Clinical Trials in Oncology

Steven Piantadosi,[1] Nagahiro Saijo[2] and Tomohide Tamura[2]

[1]Oncology Biostatistics, Johns Hopkins Oncology Center, 550 North Broadway, Suite 1103, Baltimore, Maryland 21205, USA and [2]Pharmacology Division and Medical Oncology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 170, Japan

1. Introduction

The purpose of this article is to outline for practicing oncologists fundamental considerations in the design of clinical trials. Although there has been a considerable amount written about this topic in the English language literature, a large amount of earlier writing may be inaccessible to some readers because of its technical nature and appearance in specialized journals or books. There is actually relatively little written about this subject in the clinical literature, especially given its importance in oncology today. A relatively recent review of some important topics in cancer clinical trials was given by several authors,[1)] and the merits of design were emphasized.[2)]

This article is not intended to be either a refresher or a comprehensive treatment of biostatistics or clinical trials methodology. To the contrary, we have made every effort to keep mathematical details out of the paper in favor of a less technical discussion. Although a mathematical discussion of some topics is easier and more efficient, we recognize that many practicing oncologists do not have facility with statistical notation. This is not to say that certain important points have been avoided. Many of these can be explained based on intuitive reasoning without having to rely on technical arguments. The reader interested in more comprehensive or technical background is referred to more detailed sources.[3–15)]

In addition, this article is not intended to serve as guidelines for the design of clinical trials that would necessarily result in the approval of cancer therapies by regulatory agencies. Our purpose is to promote scientifically valid methodology, to encourage investigators to use the most convincing and accurate clinical tools available, and to increase the sophistication with which trial designs and publications are reviewed.

First, we offer some basic definitions and logical concepts of clinical trials. Second, we consider design considerations in the development of cancer treatments following the pharmacologic model of phase I, II, and III trials. Third, we outline certain design modifications and adaptations which have been found to address certain concerns in the real world practice of clinical trials. Finally, we discuss some ethical concerns and limitations of clinical trials. Considerations for analysis and reporting of clinical trials are left to a following article. Although clinical trials have an interesting past which is helpful in understanding their present utility and many of the points made in this paper, space does not permit including details about the history of this still developing methodology. The interested reader is referred to papers dealing wholly or in part with the history of clinical trials.[16, 17)]

2. Terminology and Concepts

To be certain that the reader understands the way in which important terms are used in this article and in most of the clinical trials literature, we offer the following definitions. These definitions are intended to be instructive and not authoritative.

According to Dorland's Medical Dictionary,[18)] an experiment is "a procedure done in order to discover or to demonstrate some fact or general truth." Stedman's Medical Dictionary[19)] defines an experiment simply as "a test or trial." For the purposes of discussing clinical trials, these definitions are insufficient and we offer the following alternative:

experiment: a test in which the investigator controls the exposure or treatment assigned to the subject(s).

This different emphasis in our definition is important because it distinguishes experimental from non-experimental tests, one of the most important characteristics of clinical trials. We have also carefully chosen the word "assigned" rather than using "applied" because, at least in humans if not in all experiments, the investigator cannot guarantee the application of an intended treatment. This point will be expanded below. The remaining definitions are:

clinical trial: a planned experiment in humans (other definitions are possible)

variability: stochastic (random) error in the estimate of an effect of clinical interest

bias: systematic (non-random) error in the estimate of an effect of clinical interest

randomization: selection of experimental subjects from a larger group by chance alone

inference: drawing conclusions while accounting for random variability

protocol: the detailed plans for conducting an experiment (in humans)

type I error: wrongly concluding that a treatment effect or difference exists (false positive)

type II error: wrongly concluding that no treatment effect or difference exists (false negative)

confounder: a (prognostic) factor that is associated both with treatment and outcome and can affect both

blocking: forming groups or "blocks" of treatment assignments to induce balance.

We temporarily postpone the definitions of phase I, II, and III clinical trials, which are discussed individually below.

While not offered as part of a definition, there is one other important and distinguishing characteristic of clinical trials. Clinical trials of all types are tests of treatment policy and not of treatment received. This means that the investigator can only establish the intent or policy to treat patients in a certain way and cannot guarantee that patients on (or off) the study will actually receive the treatment planned. The failure to understand, appreciate, or accept this basic limitation of clinical trials explains many of the disputes between clinicians and their biostatistical counterparts.

Some investigators have suggested statistical methods[20, 21] to test the treatment itself (i.e., treatment received) rather than the treatment policy. However, the current best approach is to assure a high fidelity between the intended treatment and the received treatment rather than excluding registered or randomized patients from the analysis of a completed trial.[22] Nothing but bias is to be gained by excluding those patients who were unable to complete the intended treatment. Such results do not accurately estimate the effect of the treatment in those patients who actually receive it because potentially strong prognostic (confounding) factors are not controlled.

3. Purposes and Strengths of Clinical Trials

Investigators would not engage in as complicated, expensive, and unwieldy a methodology as clinical trials unless there was much to be gained. Because most biological experiments in cancer treatment are characterized by potential errors of about the same size as the treatment differences or effects being sought, there is much to be gained from careful planning and conduct of clinical trials. The major purposes of conducting clinical trials are to 1) estimate important clinical effects, 2) quantify errors, 3) reduce or eliminate bias, 4) provide high degree of credibility in results, and 5) influence clinical practice.

In an effort to meet these goals, investigators often conduct a variety of clinical studies which could loosely be called "clinical trials." However, many commonly accomplished types of medical studies are specifically excluded as clinical trials by our pointed definition above. Most medical investigations can be classified on the basis of who or what controls three essential components: exposure of the subjects, collection of the data (endpoint ascertainment), and analysis. The relative strength of evidence resulting from various medical studies is in the following order from highest to lowest:

1) confirmed or replicated clinical trials: independent verification of results,
2) clinical trials: the treatment or exposure, data collection, and analysis are all carefully planned,
3) observational study: the data collection and analysis are planned while the exposure occurs "naturally,"
4) pseudo-experiment: only the analysis is planned while the treatment is determined by other factors (physician and patient preference) and the data were collected for some other purpose (e.g., database or registry),
5) case series: minimal analysis because none of the essential components are planned,
6) case report: a demonstration only that some event is possible.

Using this type of hierarchy, one can see that the strength of evidence in medical studies is directly related to the amount of prospective design involved in the investigation. The more control exerted by the investigators over the essential components, the better will be the design, and the more credible will be the results. See Byar[23] for a general discussion of this topic.

4. Errors Arising from Clinical Trials

There are qualitatively different types of errors that can result when making inferences about treatment effects from clinical trials — bias and random error. Although the type of error seems to relate primarily to analysis, both types of error can be controlled through relatively simple design considerations. A review of some of these concepts, although somewhat old, has appeared in the clinical literature.[24]

Bias

Bias causes the investigator to mis-estimate consistently the treatment effect of interest. The error is systematic and neither its size nor its direction can be exactly determined. There are many sources and types of bias in clinical trials.[25] Some sources are the investigators' belief or wishes, selection of patients, post entry exclusion of

patients, informative censoring or loss of data, subjective evaluation of endpoints by unblinded investigators or patients, and improper analysis. Most sources of bias can be controlled by proper design through the appropriate use of eligibility criteria, randomization, and blinding.

One of the most serious sources of bias in clinical studies is indicated by the following quote from the physician Galen, who lived from AD 138–201 and heavily influenced western medicine:

All who drink of this remedy recover in a short time, except those whom it does not help, who all die. Therefore, it is obvious that it fails only in incurable cases.

This statement typifies a type of clinical dogmatism which can encourage us to define prognosis retrospectively, i.e., in terms of outcome. The bias that can result from retrospective looking definitions, analyses, and patient selection is well known to clinical trial methodologists but is particularly difficult to eliminate from clinical reasoning.

Random Errors: Type I

Besides bias, there are 2 other types of errors which can be made when formally testing a statistical hypothesis in the way that is commonly done in clinical trials. The type I error is a "false positive" result and occurs if there is no treatment effect or difference but the investigators wrongly conclude that there is. This type of error can be controlled by the investigator when the analysis is done (provided that the hypothesis being tested is correctly formulated). That is to say, the chance of making a type I error is almost always under the control of the investigator, even well into the analysis stage of a clinical trial. The type I error is unique in this regard because it is the only error on which the design of a simple investigation has minimal impact. One important exception is discussed below.

Clinicians often seem to emphasize the smallness of the type I error as quantified by the *P*-value. In fact, some clinical reports contain only *P*-values, apparently used as a measure of "effect." However, the size of the *P*-value is a consequence of two things: the magnitude of the estimated treatment difference and its estimated variability (which is itself partially a consequence of sample size). Thus, the *P*-value partially reflects the size of the experiment, which has no biological importance. The *P*-value also only partially reflects the size of the treatment difference, which does have major biological importance.

It is a consequence of this that causes some investigators to say things like "the effect might be statistically significant in a larger sample." This, of course, misses the point because any effect other than zero can be statistically significant in a large enough sample. What the investigators should really be talking about is the size and clinical significance of an estimated treatment effect rather than its *P*-value. In summary, *P*-values only quantify the type I error and incompletely characterize the biologically important effects in the data.

There is one circumstance in which the type I error must be carefully considered in the design phase of a clinical trial. This occurs when the investigators intend repeatedly to examine accumulating data and perform statistical tests, as is done in sequential or group sequential interim monitoring of clinical trials. Failing to account properly for the effect of such repeated hypothesis tests can greatly increase the type I error rate. This point will be expanded below in a discussion of sequential methods.

Random Errors: Type II

The type II error is a "missed effect" and occurs when investigators fail to detect a treatment effect or difference that is actually present. The chance of not making a type II error is called the power of a clinical trial. Quantification and control of this error frequently require the assistance of an experienced clinical trials biostatistician. This error can only be controlled by proper design (i.e., a sufficiently large sample size) and not by any procedures used in the analysis of clinical trials. Many investigators are either unaware of or ignore this fact and as a result perform clinical trials whose power is unknown or much lower than one might expect.

There are two other points worth emphasizing about the power or type II error of a study. The first is that all investigations have an adequate statistical power to detect treatment differences or effects of some (large) size. Thus, to discuss power, we must keep in mind the alternative treatment effect that is of clinical importance. The treatment effect against which a small study has high power may be too large to be of clinical interest. A small study can reliably detect only large differences and a large study can reliably detect small differences.

A second point about the type II error is that whenever a clinical trial concludes that there is no effect (or no treatment difference), the power of the experiment against *meaningful* clinical alternatives should be quantified. Investigators are usually quick to quantify the type I error but seldom give the same attention to the type II error. When a trial has high power against meaningful clinical alternatives but also shows no significant difference between treatments (i.e., a "negative result"), this indicates that a treatment difference is unlikely and should be as useful to the scientific community as a "positive" result.

As an example, consider a hypothetical randomized clinical trial intended to compare response rates resulting from two treatments for colorectal cancer.

Investigators decide that 50 patients per treatment group should be placed on study, based mostly on the size of the available patient population. How large a treatment effect can the clinical trial reliably detect? Using standard sample size and power calculations,[26] such a clinical trial would have 90% power to detect the difference between response rates of 66% and 36%.

Thus, investigators might be disappointed to learn that treatment differences smaller than this, but also of clinical importance, could be easily missed by the study.

The problem in the above example is that investigators have let the sample size determine the treatment difference rather than *vice versa*. The following alternative design procedure would be much better.

Before finalizing the design for the randomized trial, investigators consult an experienced clinical trial methodologist and pose the question: "We believe that a difference in response rates between these treatments of 20% (60% vs 40%) is clinically very important. How large must our clinical trial be to reliably detect such a difference (e.g., 90% of the time) if it really exists?" After some further discussion and calculation, the statistician replies that 129 patients per treatment group are required.

The investigators now know whether or not the trial is feasible, given the accrual resources at their institutions. To complete the example, suppose that the trial is completed as planned and the response rates are found to be equivalent at 50% in each treatment group. The investigators are concerned that this "negative result" may not be useful.

However, the statistician points out that the trial had 90% power to detect a difference of 60% versus 40%, if the difference really existed. The investigators are reassured that any difference between the treatments is not likely to be clinically important by their definition. The low chance of a type II error, 10%, is quantified in the published report.

When the power is high, even negative findings from such a study should be clinically very helpful.

Numerous methods for the calculation of sample size, power, and other statistical design parameters for a variety of types of clinical trials have appeared in the statistical literature. For a statistical review of some methods, see Donner.[27] Most oncologists cannot expect to maintain a working knowledge of all of these methods. One relatively simple and comprehensive computer program is available to perform the necessary calculations for a variety of clinical trials.[28] This program is commercially available only for IBM-PC compatible computers. A review of more technical statistical, but related computer programs, is given by Goldstein.[29]

5. Phase I Studies

As defined here, phase I clinical trials have four main objectives. These are to estimate the maximal tolerated dose of a new agent, to determine the organ systems affected and the type of toxicity, to gather information on the pharmacologic properties of the drug (e.g. clearance, half-life, and "area under the curve"), and to look for evidence of treatment efficacy. These multiple objectives are typified by a recent phase I study of the experimental agent cis-diammine platinum.[30] For an excellent review of phase I methods, see Von Hoff.[31] From a clinical trial design perspective, the major issues in phase I testing are the following.

Patient Selection

Phase I testing can be performed in one of several general categories of patients. These include normal volunteers, previously untreated patients with the malignancy in question, heavily pretreated patients with a single type of cancer, or patients with one of a variety of advanced stage cancers. The best patient population in which to test a new phase I agent will depend upon the expected effects of the agent based on preclinical studies, existing treatment options for the patients, and the generalizations to be made from the study. Often, phase I testing is done in heavily pretreated patients with advanced disease, and for whom few, if any, treatment options exist. Provided such patients have similar types and extent of disease and normal organ system function, most of the goals of phase I testing can be achieved. However, reliable information on treatment efficacy may be difficult to obtain under these circumstances. In general, disease-oriented phase I trials are preferred because the patient population is more homogeneous.

Starting Dose

The starting dose for phase I testing is often chosen based almost entirely on the LD50 (or LD10) observed in animal studies. From a purely statistical design perspective, the usual starting doses are often too low, i.e., too far from the maximum tolerated dose (MTD). Greater efficiency (fewer patients treated) could be gained if doses were started close to the MTD. This is often not possible because of ethical considerations.

Patient Exclusions

Many phase I protocols contain provisions for "patient evaluability." Often, these criteria require that patients live a certain length of time or complete a fixed number of treatment courses to be considered for the study endpoints. Investigators should keep in mind that all events that occur after registration are outcomes, including failure to complete treatment. Patients with certain

outcomes cannot be excluded from consideration without biasing the results. In other words, post entry exclusion of patients because of "inevaluability" is a powerful way to create bias!

6. Phase II Studies

The major objectives of phase II clinical trials are to determine treatment feasibility, estimate response and toxicity rates, and sometimes to select the best of several new treatments for phase III testing.[32-34] Phase II testing is required for new agents[35] as well as new applications of standard agents.[36] We do not distinguish here between "early" and "late" phase II trials, but use the term broadly to include studies with the stated goals. There are many research needs which cause investigators to make greater scientific demands on phase II studies such as the need to learn about the relative treatment efficacy of new agents as quickly as possible. However, design fundamentals can be safely applied based on the objectives of the study regardless of the terminology used to describe it.

With regard to patient eligibility criteria for phase II studies, both the strictness of the criteria as well as their interpretation at an individual institution are important. Many phase II studies are performed by a single investigator at a single institution. Because of differences in patient referral patterns and investigator interpretation of eligibility criteria, the results of such studies seem not to generalize well to other centers with apparently similar patients. This is often not a serious problem because institutional comparisons are of little importance. However, the results of more rigorously performed phase III clinical trials are often disappointing based on the expectations generated by phase II results, possibly because of patient selection bias or "evaluability criteria" that often work in favor of new phase II agents.

The comments for post entry exclusions in phase I studies are equally relevant to phase II studies. In principle, all patients who meet the eligibility criteria are "evaluable" for both toxicity and response. The relevant clinical effect is the chance that the treatment will produce a given outcome in a new patient. This is not conditional on the new patient being "evaluable," which is unknown when the treatment is selected. If A patients are "evaluable," B are not, and R patients "respond," the relevant probability of a beneficial response is $R/(A+B)$ not R/A.

Randomized Phase II Trials

Also, in phase II studies we must consider the possible role of randomization to eliminate bias in treatment selection when testing several agents at the same time. This randomization is not intended to justify formal treatment comparisons like phase III trials because the sample sizes in phase II studies are too small. Instead, randomization can be used simply to eliminate bias in treatment selection. The use of randomization in this setting has been discussed by several authors.[37,38] A fuller discussion of randomization is presented below.

Early Stopping

Another important consideration in phase II trials is the use of designs which allow for early termination of studies when unexpectedly large or small response or toxicity rates are seen. Such designs are termed "sequential" or "group sequential" and have been discussed in detail by many authors.[39-43] Using these designs, the interim results of a trial can be examined to determine if the study should proceed to a larger size. The number of interim analyses and the overall properties of these designs must be planned in advance, particularly under the guidance of an experienced trial methodologist. Fully sequential designs (evaluation of results after each patient), are sometimes useful provided each patient can be evaluated before a new accrual is ready. Such designs satisfy some ethical concerns in testing new therapies, particularly if one wishes to minimize the number of patients given a potentially inferior agent. Also, when treatments are much better (or worse) than expected, these designs will result in smaller sample sizes being needed to detect the effect.

Estimation vs. Selection Designs

It is important to recognize some differences between these designs and the common fixed sample size studies. Sequential designs are really "selection" designs because they are intended to reliably select or reject treatments with a certain property. Fixed sample size designs are "estimation" designs because they are intended to reliably estimate a particular treatment effect. If a sequential design causes the investigators to terminate a trial early, the resulting effect estimates can be biased.[44,45] The direction of the bias depends upon the rule for stopping (high or low) and the size of the bias depends upon how early the study met its stopping criteria.

Example

As an example, consider a group sequential phase II study in which the response rate is esimated after up to three groups of 20 accruals. If the response rate at an interim analysis is "acceptable," additional patients are accrued, whereas if the rate is high or low, the trial stops. Suppose the true underlying response rate is 30%, as expected by the investigators for a new treatment, and the trial will stop whenever the data reliably indicate that the response rate is greater or less than 30%. Specifically, the trial is stopped whenever the 95% confidence limits on the response rate excludes 30%. To study this design, 10,000 computer simulations of such a clinical trial were

made, with the following results. If the trial terminates after only 20 (or 40) patients because of a poor response rate, the average observed response rate will be 4% (or 14%). If the trial terminates after only 20 (or 40) patients because of a high response rate, the average observed response rate will be 57% (or 49%). If the trial continues to its fixed sample size end, the observed response rate will be 30%, consistent with the underlying true rate. Furthermore, the type I error rate will be approximately 7% rather than the 5% that one might guess from the 95% confidence limits. Thus, when a selection design terminates early, the treatment effect will be mis-estimated, although some benefits might result. The expected sample size will be nearly 60, unless the true response rate differs substantially from 30%.

Optimality

A number of investigators have suggested ways in which phase II trial designs might be optimized in the presence of certain reasonable constraints.[46-53] Unfortunately, much of this theory requires strong mathematical assumptions so that the resulting designs are "optimal" in narrowly defined ways. However, the concepts are important and the clinician should be aware that trial design can be affected both by the properties expected in the treatment under investigation as well as the goals of a sequence of related clinical trials.

7. Phase III Studies

The purposes of phase III studies, or comparative clinical trials, are to look for relative treatment and toxicity effects and sometimes to attempt to establish the equivalency of two treatments. In oncology, investigators frequently are interested in differences which are of about the same size as the biases and errors which can enter poorly planned studies. Consequently, care must be taken to control variability, bias, and loss of patients. To control bias, comparative clinical trials are frequently designed using randomization and blinding. In addition, stratification is often used to minimize imbalances between treatment groups which can occur by chance when using simple randomization. Each of these basic design fundamentals will be discussed.

Randomization

Randomization is the cornerstone of bias-reducing efforts because it guarantees that treatment selection will not be based on the patient's prognostic factors. The benefits derived from randomized treatment assignment are well known.[54, 55] Any resulting treatment differences must then be due only to a treatment effect plus random variability. Because the statistician can control variability, it is relatively simple to estimate the correct or unbiased treatment effect. It is true that many medical discoveries have been made without randomized trials. Indeed, when treatment effects/differences are large, the need for unbiased design is minimal. Some authors have argued in favor of non-randomized designs.[56] However, we have already stated the clinical oncologist's interest in modest-sized treatment effects. These can easily be obscured or even reversed by uncontrolled prognostic factors.

Another argument against randomization is that any confounders could be controlled in the analysis by the use of statistical adjustment procedures. This is true to a certain extent but relies upon two additional assumptions: 1) the confounders are known to the investigators and have been correctly measured in the experimental subjects, and 2) the assumptions of the statistical models or other adjustment procedures are known to be correct in the data under study. Randomization is a more reliable and powerful bias control method because it guarantees the expectation of a valid comparison without these assumptions and because it controls the effects of confounders whether they are known to the investigator or not. It is this latter point which provides randomized studies with their high degree of credibility but is so often overlooked by critics of the randomized method.

Blocking

In practice, randomization is often "blocked" or constrained so that exact balance between treatment groups is obtained at the end of each block or group of patients.[57, 58] This will necessitate making some assignments at the end of a block in a nonrandom fashion. The size of a block is determined by the urgency to maintain balance and the size of the trial. Small block sizes force exact balance at frequent points while large block sizes allow the imbalance to be greater. The maximum imbalance cannot exceed one-half of the block size. Even for small block sizes, the resulting block-randomized treatment assignments appear to be essentially random although they will fail sophisticated tests for randomness. On this basis, it seems that one might always prefer small block sizes. However, small block sizes have disadvantages. Investigators can more easily break the randomization if block sizes are small. Also, small block sizes produce a larger proportion of assignments which are not truly random. Consequently, the choice of a block size must be a compromise between balance and true randomization.

Blocking serves only to reduce covariate imbalances which might otherwise be partially controlled through statistical adjustment. When analyzing data arising from a blocked randomization, the blocking is often ignored because statistical methods to account explicitly for it are computationally difficult. Besides blocking, there are other methods to induce balance in treatment groups,[59-66] many of which can only be implemented

with the help of an experienced trial methodologist.

Blinding

Blinding or masking is another bias-reducing technique in which the patient (single blind), physician (double blind), and perhaps the monitors (triple blind) in a clinical trial are unaware of the individual patients' treatment assignment. As a result of blinding, treatment assessments can be made without prejudice, increasing the utility of subjective endpoints. In clinical oncology, blinding of pills and i.v. solutions is often simple to implement, particularly with the assistance of a hospital pharmacy or pharmaceutical company. It is clearly more difficult when testing other treatments like surgery.

Stratification

Stratification is a technique in which blocked randomization is performed separately within strata defined by strong prognostic factors. Stratification is not theoretically required but is helpful in reducing chance imbalances in influential prognostic factors which would otherwise have to be controlled by statistical adjustment procedures.[67] Stratification has few disadvantages and can enhance the credibility of trial results because adjustment procedures may not be needed. For this reason alone, it should always be considered.

However, stratification should not be carried to extremes. That is, if there are too many strata so that many never fill with assignments, the result can be as imbalanced as that with no stratification. For example, if each patient is in his or her own stratum, the result will be the same as simple randomization. Generally, only a few strata based on the strongest prognostic factors should be used. This will often result in 2–8 strata, but the number that can actually be supported depends on the sample size of the trial and the block size.

In multi-center trials, stratification is often done by individual institutions. This is really only necessary if there is evidence to suggest that institution is associated with prognosis. This most often happens as a result of referral patterns that tend to send to some centers patients with poorer prognostic factors. If institution is not associated with outcome, there is nothing to gain from stratifying on it.

Monitoring and Early Stopping

Plans for monitoring and early stopping of accrual are another element of good trial design that can greatly alleviate problems in starting and conducting studies. Researchers have an ethical obligation to learn about treatment differences as quickly and efficiently as possible and to minimize the number of patients who are placed on an inferior treatment. By planning for early termination of accrual when unexpectedly large differences are observed, investigators can make a clinical trial more acceptable to other researchers and perhaps even to patient participants. A full discussion of sequential and group sequential methods for use in this context is beyond the scope of this review. However, such methods are now commonplace with well described techniques in the statistical literature.[68] Furthermore, investigators should assure that such methods can be properly implemented by having the accumulating data reviewed formally at intervals by a Monitoring Committee.

Because of the problem of repeated significance testing on accumulating data, the overall type I error must be properly controlled by prospective planning of the analysis points and the significance level for each analysis. If all interim analyses are conducted with the conventional significance level of 0.05, the resulting overall type I error will be higher than 5%. This inflation of the type I error can be very high if many interim looks at the data are performed. To control the overall type I error at 5%, each interim look should use a significance level smaller than 0.05. For example, in a clinical trial comparing response rates and using a total of three analyses after equal groups of patients are accrued, a frequently used group sequential method[43] indicates that the analyses should be conducted with significance levels of 0.0006, 0.015, and 0.048, to control the overall type I error at the conventional 5%. Using this method, note that the final analysis is conducted employing a significance level near the usual 5%, but that early in the trial, achieving statistical significance is more difficult.

Unplanned interim analyses have statistical properties which are unknown under frequentist statistical theory and can pose serious problems in interpretation for researchers and regulators. Attempts have been made to alleviate this problem by retrospectively applying group sequential methods, although this has difficulties of its own. Another alternative for monitoring is the use of Bayesian statistical methods, which have much appeal to clinicians but have not gained widespread use.[69-71] For a general discussion of monitoring alternatives, see Gail[72] and for a practical discussion, see DeMets.[73] In any case, the time to deal properly with these concerns is in the design phase of a trial.

Another reason for terminating a clinical trial early is when interim analyses demonstrate the near equivalence of the treatments and continuing the trial would be unlikely to demonstrate clinically significant differences. In this circumstance, early stopping has been based on "conditional power" calculations.[74] Although the power of such studies may be slightly reduced by lowering the sample size, they are informative and more efficient in terms of sample size. For a real example of this type of early termination, see the trial reported by Shinkai *et al.*[75]

Crossover Trials

Crossover trials are those in which patients in a comparative study receive both treatments under investigation, usually in a randomly assigned order. It is frequently stated about crossover trials that "each patient acts as his own control," which seems to explain the greater efficiency of such designs. It is true that using this design, ideally the investigator can expect to remove the within-patient variability, resulting in a more precise estimate of the treatment effect.

However, these designs cannot be applied without assurances that the assumptions guaranteeing their validity have been met. In particular, for valid inferences, crossover designs require that the disease under study has a constant intensity and that the effects of the first treatment are completely "washed out" or eliminated by the time the second treatment is applied. If the first treatment has "carryover" effects, they will confuse the assessment of the efficacy of the second treatment. One must remember that, because all patients receive both treatments, the effect of randomization is not to eliminate bias in the selection of therapy, but to allocate treatments randomly to different time periods. Under (ideal) conditions with no temporal effects, randomization would not even be required. Thus, it can be seen that valid treatment comparisons from crossover trials rely upon very different assumptions than in an independent groups design.[76, 77]

It is precisely the difficulty in satisfying these assumptions that has limited the usefulness of crossover deigns for most important clinical questions in oncology today. Endpoints such as response and survival cannot be studied using this design. Further difficulties in satisfying the assumptions of crossover designs cause the US FDA to look unfavorably on their use.[78] From an informal point of view, it is unwise to depend on the efficiency of a design whose validity relies upon doubtful assumptions. These designs should be used with great caution, if at all.

Other Design Modifications

When conducting trials in specialized circumstances or to answer certain types of clinical questions, other design modifications are necessary or desirable. For example, factorial designs are useful for studying several treatments simultaneously, especially in disease prevention trials.[79–85] Factorial design means that each of several treatments (factors) is studied simultaneously at different levels. The US the Physician's Health Study[80, 82] is an example where aspirin and beta-carotene were studied for their effect on cancer and cardiovascular disease.

Sometimes, the patient is not the experimental unit as in family (litter) studies or population interventions. In this circumstance, cluster randomization or designs might be used.[86] In these designs, groups (clusters) of subjects, instead of individuals, are allocated to the same treatment. For example, entire cities might be assigned to different interventions as for screening or smoking cessation trials.

Designs to minimize the number of patients given an inferior treatment or to address other ethical concerns have been proposed.[87–89] For some clinical constraints such as greatly differing costs for two treatments, unequal allocation of subjects can minimize cost or effort.[90, 91]

8. Ethical Concerns and Limitations

No general discussion of clinical trials would be complete without mention of limitations and ethical concerns surrounding the use of this method. Although clinical trials are conducted in all technologically advanced countries, there are frequently voiced concerns about ethics, particularly of randomized comparative studies.[92–94] It is also difficult to make statements or claims that are valid across all cultural boundaries because of differing roles of health care and physician-patient relationships. The nearly universal implementation of patient informed consent and Institutional Review Boards as ethical quality control procedures has not eliminated all concerns. In Japan, a major ethical foundation for clinical trials is the Helsinki Declaration. In a clinical trial, the idea of informed consent also usually includes explaining to the patient alternatives to participation.

Nevertheless, it is often stated that randomized trials raise ethical concerns because the physician is, or should be, an advocate only for the well being of the individual patient. Encouraging patients to accept treatment selected at random could be an abdication of this duty. Instead, the physician should become knowledgeable about treatments in other ways and optimally select the best therapy for the individual patient without resorting to chance.

This view is incorrect in two ways. First, the role of the physician as an advocate for the well being of the patient is an ideal, if it exists at all, and may well be mythical. This role is frequently "violated" in circumstances of triage, allocation of scarce and expensive technologies such as organ transplantation, and in the recommendation of vaccination (among our most cost effective and beneficial medical interventions) where the patient is knowingly placed at minor risk for the benefit of all. Perhaps if the historical role of physicians were carefully examined, we would discover that untempered advocacy for the individual patient is only a myth.

Second, and even if we accept the advocacy argument, some patients will always receive an inferior treatment regardless of our ethical rules. It is incumbent upon the scientist which exists in every physician to learn quickly

and convincingly from such occurrences to see that their frequency is minimized. Controlled experiments in the proper clinical setting are the most reliable way to accomplish this. In other words, failing to learn from one's mistakes is unethical.

In spite of our positive technological and ethical assessment, clinical trials do have definite limitations. They can only be applied in circumstances where the investigator has no opinion about the superiority of a treatment. Only physicians with this "equipoise" can participate. Physicians and patients who have preferences for a particular treatment, even irrationally based ones, are not eligible to participate in a randominized trial. Also, patients must be relatively well informed about the risks, alternatives, and possible benefits of participation in such studies. Often, patient participation is based more on altruism than expected benefit, especially in phase I and II studies.

Clinical trials are also expensive and cumbersome. They take a long time to complete and require the collaboration of many participants. Today, multi-center clinical trials are especially needed to accrue sufficient numbers of patients to test important clinical hypotheses (e.g. ref. 95), but have additional problems in design and conduct.[96, 97] When concluded, the results may not be as convincing as one might expect because of prevailing opinion, minor design, analysis, or reporting flaws, difficulty in understanding and accepting a complicated methodology, and failure to appreciate the subtle way in which new medical advances are incorporated into practice.

We also acknowledge that not all clinical trials fit neatly into one of the categories discussed above. Important clinical questions and constraints frequently require that studies be designed with characteristics of more than one "phase" of research. This can result in hybrid designs that may seem confusing, wrong, ingenious, or extra efficient. However, by following the principles of good trial design, investigators can be more comfortable with and more accurate in their assessment of clinical studies of all types.

9. Conclusions

In developing new cancer treatments, investigators are often interested in treatment effects and differences which are of about the same size as patient-to-patient variability or the bias which is a part of poorly controlled clinical studies. The only solution for making valid inferences in the face of these potential errors is to properly design and conduct clinical trials. There are a small number of important design considerations to help control bias and random errors including the use of randomization, blinding, stratification, minimizing post entry exclusions, adequate sample size, and planned interim monitoring.

Clinical trials have limitations, partly because of the rigor required to implement them. Investigators contemplating the use of these important scientific tools should focus most efforts on the design aspects of the study and concern themselves little with analysis. This is because most of the serious errors which can be made when performing clinical trials can be prevented or minimized by correct design. In this regard, consultation with an experienced clinical trial methodologist early in the design stage of an investigation will be of enormous benefit.

ACKNOWLEDGMENTS

The authors acknowledge the secretarial assistance of Helen Cromwell. Supported by a Foundation for the Promotion of Cancer Research Fellowship based on the Comprehensive 10-Year Strategy for Cancer Control.

(Received March 30, 1991/Accepted March 31, 1992)

REFERENCES

1) Leventhal, B. L. and Piantadosi, S. (ed.). Issues in clinical research. *Sem. Oncol.*, **15** (1988).
2) Piantadosi, S. Principles of clinical trial design. *Sem. Oncol.*, **15**, 423–433 (1988).
3) Meinert, C. L. "Clinical Trials" (1986). Oxford University Press, Oxford.
4) Silverman, W. A. "Human Experimentation: A Guided Step into the Unknown" (1985). Oxford University Press, Oxford.
5) Armitage, P. The design of clinical trials. *Aust. J. Stat.*, **21**, 266–281 (1979).
6) Louis, T. A., Mosteller, F. and McPeek, B. Timely topics in statistical methods for clinical trials. *Annu. Rev. Biophys. Bioeng.*, **11**, 81–104 (1982).
7) Lewis, J. A. Clinical trials: statistical developments of practical benefit to the pharmaceutical industry (with discussion). *J. R. Stat. Soc. A*, **146**, 362–393 (1983).
8) Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: I. Introduction and design. *Br. J. Cancer*, **34**, 585–612 (1976).
9) Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. Design and analysis of randomized clinical trials requiring prolonged observations of each patient: II. Analysis and examples. *Br. J. Cancer*, **35**, 1–39

(1977).
10) Shapiro, S. H. and Louis, T. A. "Clinical Trials: Issues and Approaches" (1983). Marcel Dekker, New York.
11) Pocock, S. J. "Clinical Trials: A Practical Approach" (1983). John Wiley & Sons, New York.
12) Buyse, M. E., Staquet, M. J. and Sylvester, R. J. (ed.). "Cancer Clinical Trials. Methods and Practice" (1984). Oxford University Press, Oxford.
13) Leventhal, B. G. and Wittes, R. E. "Research Methods in Clinical Oncology" (1988). Raven Press, New York.
14) Byar, D. P., Simon, R. M., Friedewald, W. T., Schlesselman, J. J., DeMets, D. L., Ellenberg, J. H., Gail, M. H. and Ware, J. H. Randomized clinical trials: perspectives on some recent ideas. *N. Engl. J. Med.*, **295**, 74–80 (1976).
15) Freidman, L. M., Furberg, C. D. and DeMets, D. L. "Fundamentals of Clinical Trials" (1981). John Wright, Boston.
16) Office of Technology Assessment, U.S. Congress. The Impact of Randomized Clinical Trials on Health Policy and Medical Practice: Background Paper. OTA-BP-H-22 (1983). US Government Printing Office, Washington, D.C.
17) Bull, J. P. The historical development of clinical therapeutic trials. *J. Chronic Dis.*, **10**, 218–248 (1959).
18) "Dorland's Medical Dictionary, 27th Edition" (1988). W. B. Saunders Co., Philadelphia.
19) "Stedman's Medical Dictionary, 25th Edition" (1990). Williams and Wilkins, Baltimore.
20) Sommer, A. and Zeger, S. L. On estimating efficacy from clinical trials. *Stat. Med.*, **10**, 45–52 (1991).
21) Newcombe, R. G. Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Stat. Med.*, **7**, 1179–1186 (1988).
22) Gail, M. H. Eligibility exclusions, losses to follow-up, removal of randomized patients, and uncounted events in cancer clinical trials. *Cancer Treat. Rep.*, **69**, 1107–1112 (1985).
23) Byar, D. P. Why data bases should not replace randomized clinical trials. *Biometrics*, **36**, 337–342 (1980).
24) Freiman, J. A., Chalmers, T. C., Smith, H., Jr. and Kuebler, R. R. The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. Survey of 71 "negative" trials. *N. Engl. J. Med.*, **299**, 690–694 (1976).
25) Sackett, D. L. Bias in analytic research. *J. Chronic Dis.*, **32**, 51–63 (1979).
26) Fleiss, J. L., Tytun, A. and Ury, H. K. A simple approximation for calculating sample size for comparing independent proportions. *Biometrics*, **36**, 343–346 (1980).
27) Donner, A. Approaches to sample size estimation in the design of clinical trials — a review. *Stat. Med.*, **3**, 199–214 (1984).
28) Piantadosi, S. "Clinical Trials Design Program" (1990). BIOSOFT, Cambridge.
29) Goldstein, R. Power and sample size via MS/PC-DOS computers. *Am. Stat.*, **43**, 253–260 (1989).
30) Sasaki, Y., Totani, T., Shinkai, T., Eguchi, K., Tamura, T., Ohe, Y., Nakagawa, K. and Saijo, N. A phase I study of cis-diammine (glycolato) platinum (II) (254-S; NSC 375101D) administered by 5-day continuous intravenous infusion: pharmacokinetic and pharmacodynamic analysis. *Cancer Res.*, **51**, 1472–1477 (1991).
31) Von Hoff, D. D., Kuhn, J. and Clark, G. M. Design and conduct of phase I trials. *In* "Cancer Clinical Trials: Methods and Practice," ed. M.E. Buyse, M. J. Staquet and R. J. Sylvester, pp. 210–222 (1984). Oxford Univ. Press, Oxford.
32) Carter, S. K. Clinical trials in cancer chemotherapy. *Cancer*, **40**, 544–557 (1977).
33) Muggia, F. M., Rosenzweig, M., Staquet, M. J. and McGuire, W. P., Jr. Methodology of phase II clinical trials in cancer. *Recent Results Cancer Res.*, 53–60 (1980).
34) Herson, J. Statistical aspects in the design and analysis of phase II clinical trials. *In* "Cancer Clinical Trials: Methods and Practice," ed. M. E. Buyse, M. J. Staquet and R. J. Sylvester, 239–260 (1984). Oxford Univ. Press, Oxford.
35) Fukuda, M., Shinkai, T,. Eguchi, K., Sasaki, Y., Tamura, T., Ohe, Y., Kojima, A., Oshita, F., Hara, K. and Saijo, N. Phase II study of (glycolate-O,O′) diammineplatinum (II), a novel platinum complex, in the treatment of non-small-cell lung cancer. *Cancer Chemother. Pharmacol.*, **26**, 393–396 (1990).
36) Ohe, Y., Shinkai, T., Eguchi, K., Sasaki, Y., Tamura, T., Nakagawa, K., Kojima, A., Yamada, K., Oshita, F., Miya, T. and Saijo, N. Negative phase II study of 5-fluorouracil with high-dose lecovorin in non-small cell lung cancer. *Jpn. J. Clin. Oncol.*, **20**, 364–368 (1990).
37) Piantadosi, S. and Dalesio, O. The randomized phase II clinical trial. *Eur. Cancer News*, **1**, 3–4 (1988).
38) Simon, R., Wittes, R. E. and Ellenberg, S. Randomized phase II clinical trials. *Cancer Treat. Rep.*, **69**, 1375–1381 (1985).
39) Anscombe, F. J. Sequential medical trials. *J. Am. Stat. Assoc.*, **58**, 365–383 (1963).
40) Armitage, P., McPherson, C. K. and Rowe, B. C. Repeated significance tests on accumulating data. *J. R. Stat. Soc. A*, **132**, 235–244 (1969).
41) Wald, A. "Sequential Analysis" (1947). John Wiley & Sons, New York.
42) Pocock, S. J. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191–199 (1977).
43) O'Brien, P. C. and Fleming, T. R. A multiple testing procedure for clinical trials. *Biometrics*, **35**, 549–556 (1979).
44) Hughes, M. D. and Pocock, S. J. Stopping rules and estimation problems in clinical trials. *Stat. Med.*, **7**, 1231–1242 (1988).
45) Chang, M. N., Wieand, H. S. and Chang, V. T. The bias of the sample proportion following a group sequential phase II clinical trial. *Stat. Med.*, **8**, 563–570 (1989).
46) Whitehead, J. Designing phase II studies in the context of a programme of clinical research. *Biometrics*, **41**, 373–383

(1985).
47) Whitehead, J. Sample sizes for phase II and phase III clinical trials: an integrated approach. *Stat. Med.*, **5**, 459–464 (1986).
48) Thall, P. F., Simon, R. and Ellenberg, S. S. Optimal two stage designs for clinical trials with binary response. *Stat. Med.*, **7**, 571–579 (1988).
49) Weiand, S. and Therneau, T. A two stage design for randomized trials with binary outcomes. *Controlled Clin. Trials*, **8**, 20–28 (1987).
50) Ellenberg, S. S. and Eisenberger, M. A. An efficient design for phase III studies of combination chemotherapies. *Cancer Treat. Rep.*, **69**, 1147–1152 (1985).
51) Herson, J. and Carter, S. K. Calibrated phase II clinical trials in oncology. *Stat. Med.*, **5**, 441–447 (1986).
52) Armitage, P. The search for optimality in clinical trials. *Int. Stat. Rev.*, **53**, 15–24 (1985).
53) Peto, R. Discussion of 'On the allocation of treatments in sequential medical trials' by J. A. Bather and 'The search for optimality in clinical trials' by P. Armitage. *Int. Stat. Rev.*, **53**, 31–34 (1985).
54) Chalmers, T. C. The control of bias in clinical trials. *In* "Clinical Trials: Issues and Approaches," ed. S. H. Shapiro and T. A. Louis, pp. 115–127 (1983). Marcel Dekker, New York.
55) Green, S. B. Randomized clinical trials: design and analysis. *Sem. Oncol.*, **8**, 417–423 (1981).
56) Gehan, E. A. and Freireich, E. J. Non-randomized controls in cancer clinical trials. *N. Engl. J. Med.*, **290**, 198–203 (1974).
57) Green, S. B. and Byar, D. P. The effect of stratified randomization on size and power of statistical tests in clinical trials. *J. Chronic Dis.*, **31**, 445–454 (1978).
58) Palta, M. Investigating maximum power losses in survival studies with nonstratified randomization. *Biometrics*, **41**, 497–504 (1985).
59) Bailar, J. C. Patient assignment algorithms: an overview. *In* "Proceedings of the 9th International Biometric Conference. Vol. I," pp. 189–206 (1976). The Biometric Society, Raleigh, N.C.
60) Simon, R. Adaptive treatment assignment methods and clinical trials. *Biometrics*, **33**, 743–749 (1977).
61) Kalish, L. A. and Begg, C. B. Treatment allocation methods in clinical trials: a review. *Stat. Med.*, **4**, 129–144 (1985).
62) Efron, B. Forcing a sequential experiment to be balanced. *Biometrika*, **58**, 403–417 (1971).
63) Pocock, S. J. and Simon, R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **31**, 103–115 (1975).
64) Freedman, L. A. and White, S. J. On the use of Pocock and Simon's method for balancing treatment numbers over prognostic factors in the controlled clinical trial. *Biometrics*, **32**, 691–694 (1976).
65) Begg, C.D. and Iglewicz, B. A. A treatment allocation procedure for sequential clinical trials. *Biometrics*, **36**, 81–90 (1980).
66) Simon, R. Restricted randomization designs in clinical trials. *Biometrics*, **35**, 503–512 (1979).
67) Senn, S. J. Covariate imbalance and random allocation in clinical trials. *Stat. Med.*, **8**, 467–475 (1989).
68) O'Fallon, J. R. Policies for interim analysis and interim reporting of results. *Cancer Treat. Rep.*, **69**, 1101–1106 (1985).
69) Cornfield, J. Recent methodologic contributions to clinical trials. *Am. J. Epidemiol.*, **104**, 408–421 (1976).
70) Berry, D. A. Interim analyses in clinical trials: classical versus Bayesian approaches. *Stat. Med.*, **4**, 521–526 (1985).
71) Berry, D. A. Interim analyses in clinical trials: the role of the likelihood principle. *Am. Stat.*, **41**, 117–122 (1987).
72) Gail, M. H. Monitoring and stopping clinical trials. *In* "Statistics in Medical Research," ed. V. Mike and K. E. Stanley, pp. 455–484 (1982). Wiley, New York.
73) DeMets, D. L. Practical aspects in data monitoring: a brief review. *Stat. Med.*, **6**, 753–760 (1987).
74) Lan, K. K. G., Simon, R. and Halperin, M. Stochastically curtailed tests in long-term clinical trials. *Comm. Stat.*, **1**, 207–219 (1982).
75) Shinkai, T., Eguchi, K., Sasaki, Y., Tamura, T., Ohe, Y., Kojima, A., Oshita, F. and Saijo, N. A randomized clinical trial of vindesine plus cisplatin versus mitomycin plus vindesine and cisplatin in advanced non-small cell lung cancer. *Eur. J. Cancer*, **27**, 571–575 (1991).
76) Brown, B. W., Jr. Statistical controversies in the design of clinical trials. *Controlled Clin. Trials*, **1**, 13–27 (1980).
77) Senn, S. J. Cross-over trials, carry-over effects and the art of self-delusion (letter). *Stat. Med.*, **7**, 1099–1101 (1988).
78) U.S. Food and Drug Administration. "General Considerations for the Clinical Evaluation of Drugs," Pub No. HEW (FDA) 77-3040 (1977).
79) Byar, D. P. and Piantadosi, S. Factorial designs for randomized clinical trials. *Cancer Treat. Rep.*, **69**, 1055–1063 (1985).
80) Hennekens, C. H. and Eberlein, K. A. for the Physician's Health Study Research Group. A randomized trial of aspirin and beta-carotene among U.S. physicians. *Prev. Med.*, **14**, 165–168 (1985).
81) Stampfer, M. J., Burning, J. G., Willet, W., Rosner, B., Eberlein, K. and Hennekens, C. H. The 2×2 factorial design: its application to a randomized trial of aspirin and carotene in U.S. physicians. *Stat. Med.*, **4**, 111–116 (1985).
82) Physicians' Health Study Research Group. Preliminary report: findings from the aspirin component of the ongoing physician's health study. *N. Engl. J. Med.*, **318**, 262–264 (1988).
83) Blot, W. J. and Li, J. Y. Some considerations in the design of a nutrition intervention trial in Linxian, People's Republic of China. *Natl. Cancer Inst. Monogr.*, **69** (1985).
84) Hennekens, C. H. Issues in the design and conduct of clinical trials. *J. Natl. Cancer Inst.*, **73**, 1473–1479 (1984).
85) Kornitzer, M., DeBacker, G., Dramaix, M., Kittel, F.,

Thilly, C., Graffar, M. and Vuylsteek, K. Belgian heart disease prevention project: incidence and mortality results. *Lancet,* **i,** 1066–1070 (1983).
86) Donner, A., Birkett, N. and Buck, C. Randomization by cluster. *Am. J. Epidemiol.,* **114,** 906–914 (1981).
87) Zelen, M. Play-the-winner rule and the controlled clinical trial. *J. Am. Stat. Assoc.,* **64,** 131–146 (1969).
88) Robbins, H. Some aspects of sequential design of experiments. *Bull. Am. Math. Soc.,* **58,** 527–535 (1972).
89) Taves, D. R. A new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.,* **15,** 443–453 (1974).
90) Brittain, E. and Schlesselman, J. J. Optimal allocation for the comparison of proportions. *Biometrics,* **38,** 1003–1009 (1982).
91) Sposto, R. and Krailo, M. D. Use of unequal allocation in survival trials. *Stat. Med.,* **6,** 119–125 (1987).
92) Schafer, A. The ethics of the randomized clinical trial. *N. Engl. J. Med.,* **307,** 719–724 (1977).
93) Burkhardt, R. and Kienle, G. Basic problems in controlled trials. *J. Med. Ethics,* **9,** 80–84 (1983).
94) Vere, D. W. Problems in controlled trials — a critical response. *J. Med. Ethics,* **9,** 85–89 (1983).
95) Fukuoka, M., Furuse, K., Saijo, N., Yutaka, N., Ikegami, H., Tamura, T., Shimoyama, M. and Suemasu, K. Randomized trial of cyclophosphamide, doxorubicin and vincristine (CAV) versus cisplatin and etoposide (PE) versus alternation of CAV and PE in small cell lung cancer. *J. Natl. Cancer Inst.,* **83,** 855–861 (1991).
96) Machin, D., Staquet, M. J. and Sylvester, R. J. Advantages and defects of single-center and multicenter clinical trials. *In* "Controversies in Cancer," ed. H. J. Tagnon and M. J. Staquet, pp. 7–22 (1979). Masson Publishing, New York.
97) Weiss, D. G., Williford, W. O., Collins, J. F. and Bingham, S. F. Planning multicenter clinical trials: a biostatistician's perspective. *Controlled Clin. Trials,* **4,** 53–64 (1983).