

The Scientific Basis of Clinical Trials: Statistical Aspects¹

Edmund A. Gehan²

Lombardi Cancer Center, Georgetown University Medical Center,
Washington, DC 20007

Abstract

The objective of this paper is to delineate some of the advances in clinical oncology that have occurred since the 1950s in the context of methodological principles established by Dr. Emil J Freireich and colleagues. Four statistical aspects of the methodological developments in clinical trials are considered and illustrated by real examples: a quantitative approach to the design and analysis of clinical trials; the randomized controlled trial; the nonrandomized controlled trial; and the use of regression models in clinical studies.

Introduction

The first randomized clinical trial in oncology was planned in 1954 and reported in 1958 by Frei *et al.* (1). The clinical trial involved a randomized comparison of two regimens of combination chemotherapy, 6-MP³ and either continuous or intermittent MTX in 65 patients. The trial was essentially negative, both because the results were very poor and similar on the two treatment regimens. It is indeed fortunate that this did not discourage Frei *et al.* (1). These investigators are perhaps those most responsible for the great improvements in the therapy of acute leukemia that have been made since 1958.

My objective is to put the advances in clinical oncology since the 1950s in the context of methodological principles established by Dr. Emil J Freireich and colleagues. In considering the statistical aspects of clinical trials, the most difficult part of my task is selecting from among the many advances in the methodology of clinical trials that Freireich has either established or participated in directly. I have chosen four aspects of the methodological developments in clinical trials that will be illustrated by real examples: a quantitative approach to the design and analysis of clinical trials; the randomized controlled trial; the nonrandomized controlled trial; and the use of regression models in clinical studies.

Quantitative Approach to the Design and Analysis of Clinical Trials

Two studies are considered as examples: the effect of chemotherapy on acute leukemia in the human (2), a study done at the NCI; and a randomized trial of sequential and combination 6-MP and MTX in acute leukemia. The purpose of the leukemia study at NCI was "to present certain aspects of the clinical course of acute leukemia treated with the available chemotherapeutic agents," and it was expected that "these data can be useful for comparison with data accumulated in the future when newer forms of therapy are available." The analysis was of 178 patients with an established diagnosis of acute leukemia treated at NCI between 1953 and 1958. The study demonstrated certain facts that are now very well known such as that as age increased, the number of patients with ALL decreased and the number with acute myelogenous leukemia increased. Also, there was a decreasing frequency of CR by age group up to age 20 and over age 20.

Freireich suggested certain survival curves that established that the survival of patients with acute leukemia can be attributed almost entirely to the period of time that patients spend with "hematological improvement." Three groups of patients were considered: ALL patients under age 20, acute myelogenous leukemia patients, and ALL patients age 20 or over. For each group, three survival curves were calculated: a survival curve from the start of study for patients showing hematological improvement; a survival curve for patients that had hematological improvement after subtraction of the duration of improvement; and a survival curve for patients who did not show hematological improvement. The latter two curves were nearly identical, demonstrating that the advantage in survival for patients having hematological improvement could be attributed mainly to the period of time spent in hematological improvement. In the conclusion, it was stated that "the risk of death during periods of active disease has not changed during the past ten years and the improved survival can be attributed directly to the duration of hematological responses resulting from chemotherapy."

Another concept was that the initial response of the patient is independent of the number of subsequent responses. At that time, some clinical investigators believed that patients could be classified as "responders" or "nonresponders," and the category for the patient predicted subsequent survival. However, it was demonstrated that the average number of later responses for patients having a first response of CR was almost the same as that for patients whose initial response was partial remission or even no response. The average number of later responses for patients having a first response of CR was 1.0, as was the average for patients who did not respond initially. Patients who had improvements or partial remissions initially had an average of 0.7 subsequent responses. It was concluded that "response to subsequent courses of therapy was found to be independent of response to the first course of therapy." From here, it required only one further step to conclude that combining chemothera-

¹ Presented at "Foundations of Clinical Cancer Research: Perspective for the 21st Century," a symposium to honor Emil J Freireich, M.D., on the occasion of his 70th birthday, March 14-15, 1997, M. D. Anderson Cancer Center, Houston, TX.

² To whom requests for reprints should be addressed, at Lombardi Cancer Center, Podium A, Georgetown University Medical Center, 3800 Reservoir Road, NW, Washington, DC 20007.

³ The abbreviations used are: 6-MP, 6-mercaptopurine; MTX, methotrexate; NCI, National Cancer Institute; ALL, acute lymphocytic leukemia; CR, complete remission; PE, protected environment; HCT, historically controlled trial; AMSA, amsacrine; OAP, oncovin, ara-c, prednisone.

peutic agents into one treatment regimen could increase the chance of an individual patient responding by the independent action of the drugs in the combination.

In 1961, Frei *et al.* (3) reported a study of 6-MP *versus* MTX *versus* the combination of 6-MP and MTX in a randomized clinical trial conducted by what was then called Acute Leukemia Group B involving a collaboration among Drs. Frei, Freireich, Holland, and others. The design of the study in Phase I involved a randomized comparison among MTX *versus* 6-MP *versus* the combination of the two. In Phase II, patients failing after MTX received 6-MP, and those failing after 6-MP received MTX. A total of 462 patients were entered into study. There were no overall differences in the survival experiences on the three treatment programs; the median survival was about 9–10 months for patients on each treatment. This study developed further the concept of the independent action of 6-MP and MTX. In Phase I, 26% of the patients had CRs after 6-MP treatment, 22% had CRs after MTX, and 42% of the patients responded to the combination treatment. As stated in the paper, “if it is assumed that when combination therapy is given, 6-MP and MTX act independently, the expected number of remissions for the combination can be calculated. For example, if 26% of the children with ALL have CRs on 6-MP (the 6-MP CR rate) and in addition 22% (the MTX CR rate) of the 6-MP failures respond to MTX, the expected CR rate for the combination therapy would be 42%.” This combined rate when applied to the 40 children receiving the combination therapy resulted in 16.9 as the predicted number of CRs, and this was very close to the 17 observed CRs. The agreement was very good between observed and predicted number of remissions in children but not quite as good in adults. Quoting from the paper, “in terms of remission induction, combined 6-MP and MTX provided a therapeutic advantage over either drug used alone. The data suggest that the advantage arises in the two drugs acting independently on the patient.” This randomized comparative study confirmed and extended a concept of the independent action of drugs and presaged the development of multidrug regimens for patients in subsequent studies.

Furthermore, the study demonstrated that response in Phase II was almost the same as that in Phase I. The conclusion was that “responsiveness to the second course of anti-metabolite therapy (Phase II) was as good as that in the first course of treatment. This was true for remission rate, remission duration, and even survival when appropriate corrections were made. Thus, responsiveness to drug therapy is maintained as the disease progresses temporarily,” a concept established in this study.

The Randomized Controlled Trial

In 1963, Freireich *et al.* (4) reported another trial of Acute Leukemia Group B, a prospective, randomized, double-blind, placebo-controlled, sequential study of 6-MP *versus* placebo in the maintenance of remissions in pediatric leukemia. In the design of the study, patients were put into CR (in Phase I) with steroid treatment and were then randomized to 6-MP or placebo (in Phase II), which was administered double-blind. Patients were paired at each institution by remission status, and the primary end point was the length of CR between the two groups

of patients in Phase II. In Phase III, patients relapsing after placebo received 6-MP.

This study was analyzed several ways, each of which established that 6-MP led to substantially longer remissions than placebo. Among the 21 pairs of patients for whom a preference for 6-MP or placebo could be recorded, there were 18 preferences for 6-MP compared with 3 preferences for placebo. The preferences for 6-MP or placebo were plotted as part of a sequential treatment plan, and the path of the preferences passed through the upper boundary, which indicated that 6-MP was the preferred treatment, and the study could be stopped after entry of the 21 pairs of patients.

It is of interest that the difference between 6-MP and placebo treatment was also established in the nonrandomized portion of the study. Those patients relapsing on placebo who were later treated with 6-MP had remissions that were on the average 4 weeks longer than those on placebo. This study was the first to establish a very important concept, *i.e.*, that patients with acute leukemia should be treated in the remission phase of the disease. It was a precursor to adjuvant studies in other forms of cancer, such as breast cancer, in which treatments are administered when the patients are in a disease-free state. Furthermore, it demonstrated that a patient receiving placebo could be used as his or her own control for evaluating subsequent treatments. However, the study, contrary to the title, has not become a model for the evaluation of other potentially useful therapy. Perhaps this is because clinicians since that time have been reluctant to randomize between a potentially useful therapy and placebo in this setting.

The 6-MP *versus* placebo study had a role in perhaps the most important statistical paper published in survival analysis, *i.e.*, the report on “regression models and life tables” published by David R. Cox (5). This report made the very important advance of providing a model for testing the differences in survival experience among groups of patients, adjusting for patient covariates. Cox demonstrated in the 6-MP example that a proportional hazards model yielded a good fit to the data and that the relapse rate per unit time for placebo patients was 5.2 times the relapse rate per unit time for the 6-MP patients. This report has subsequently been referred to as the Cox model and is one of the most referenced papers in the statistics and oncology literature. Consequently, this 6-MP *versus* placebo study has certainly been very important not only in the development of new concepts in clinical oncology but also in leading to a significant advance in statistical theory.

Nonrandomized Controls in Clinical Cancer Trials

The first example is a study of the PE-prophylactic antibiotic program in the chemotherapy of acute leukemia by Bodey *et al.* (6). This study had no formal statistical design. Patients with acute leukemia, 33 in this case, underwent chemotherapy in a PE and received prophylactic antibiotic regimens. No provision was made for a control group, randomized or otherwise. It was time to apply for renewal of a grant funding the project, and it became important to answer the question of whether treatment within a PE might be expected to be superior to treatment outside a PE.

The approach taken to the evaluation was to construct a

potential control group of patients for each PE patient that was as comparable as possible with respect to chemotherapy regimen, diagnosis, age, sex, infection status, WBC count, platelet count, time from diagnosis to therapy, and previous therapy. For each PE patient, a list of at least five potential pairmates was made, and Dr. Bodey ranked the patients according to his estimate of prognosis without knowing which one was in the PE. Two control patients were selected for each patient to have equivalent or superior ranks to those for the PE patient. To test the process of selecting control patients, a comparison was made between the two control groups of 33 patients; no differences were demonstrated in a paired analysis between these groups in CR rate, incidence of infection, or survival. Finding no difference between control groups provided some justification for comparing the PE *versus* the control group. When PE patients were compared with the combined control groups, there was a suggestion of a difference in CR rate (61% for PE *versus* 49% for combined control), but the main result was that there were significantly fewer episodes of both local and severe infection as related to neutrophil count in PE patients. Furthermore, there was a significant improvement in survival time (median of 40 weeks for PE patients compared to 20 weeks for the combined control patients). The report concluded that “the use of non-randomized patients as a control group has been considered unacceptable by some investigators. However, this approach was both necessary and advantageous for our study. Randomizing patients to either the PE group or control group would have resulted in the PE units being unoccupied for substantial periods of time and would have required a much longer time to complete the study.”

This was perhaps the first publication demonstrating that the PE was useful for treating patients and early evidence in clinical cancer trials that a nonrandomized study could lead to conclusions about treatments. The study was confirmed subsequently in 1978 by Rodriguez *et al.* (7) in a randomized study of a PE-prophylactic antibiotic program.

Freireich has published several papers arguing the merits of nonrandomized clinical trials in certain circumstances, the first of which was a special article in *The New England Journal of Medicine* (8). This report was written in response to one by Chalmers *et al.* (9), who wrote a report essentially stating that all nonrandomized studies are uncontrolled and that even the first patient administered a new therapy should be randomized. *The New England Journal of Medicine* article presented a positive case for the nonrandomized trial in which the control group might be selected from the literature, from a matched group of patients, such as in the PE study, or there might be a historical control group.

A second report discussed “the limitations of the randomized clinical trial” (10), and the limitations were discussed from the point of view of significance levels, one *versus* two-sided tests, ethics, the nature of evidence from randomized trials, efficiency, the capability of the cancer clinical trial to resolve conflicts, and various statistical aspects. The strengths of HCTs were summarized and are: all knowledge is historical; HCTs require smaller sample sizes and shorter length studies; there is no ethical problem in randomizing patients; it is easier to recruit patients from afar to enter a nonrandomized study; controversial questions can be resolved; there is a better record of accom-

plishment; and having HCTs as part of an evolutionary research program helps to avoid mistakes. Some examples of therapies in cancer discovered in nonrandomized studies are given by Gehan (11). Some examples of therapies for cancer discovered in nonrandomized studies include: MTX, 1- β -D-arabinofuranosylcytosine, and vincristine-amethopterin-6-MP-prednisone in acute leukemia (11–13); nitrogen mustard-vincristine-procarbazine-prednisone in Hodgkin’s disease; (14), cyclophosphamide-Adriamycin-vincristine-prednisone in non-Hodgkin’s lymphoma (15, 16); Adriamycin in soft tissue sarcoma (16); Velban, bleomycin, and cisplatin in testicular cancer (17); and combination chemotherapy in advanced and early breast cancer (18–20).

Regression Models and Clinical Trials

Freireich has been a strong proponent of the quantitative approach in clinical trials and was an early advocate of the use of regression models relating patient outcomes to patient covariates as a means of testing differences between treatments, adjusting for prognostic factors. Examples of the use of regression models in breast cancer and acute leukemia are given in Budzar *et al.* (20) and Smith *et al.* (21).

Perhaps the ultimate use of regression models was a study entitled “A strategy for evaluation of new treatments in untreated patients: application to a clinical trial of AMSA for acute leukemia” (22). The design of this trial involved the use of regression models for calculating the predicted probability of response and the predicted probability of being in remission for 1 year. Patients were assigned to treatment depending upon their predicted probabilities of response from the regression model. Basically, this was a large nonrandomized study designed to overcome a major problem in cancer research, *i.e.*, how to evaluate new treatments when there are already highly effective established treatments.

The design of this study, proposed by Freireich, involved deciding who were the most favorable patients by their regression model and administering them the current standard therapy, which was Adriamycin-OAP in this study. Those patients who had unfavorable chances of CR received the proposed new therapy, in this case AMSA-OAP. Hence, the strategy was that patients with unfavorable prognosis received new treatments, whereas patients with favorable prognosis received the current standard treatments, both for remission induction and maintenance. The response rate for all patients receiving AMSA-OAP was 52% compared with a 48% rate for matched patients receiving Adriamycin OAP in a previous study. Also, 35% of AMSA-OAP patients survived 1 year compared with 31% for the control group, not a statistically significant difference. Although the data for CR rate and survival are not very good, it should be remembered that these patients had unfavorable prognosis and a major objective was to evaluate the new therapy.

The conclusion of the study was that “it is concluded that AMSA-OAP is equivalent to Adriamycin-OAP in the induction of CRs (estimated CR rate, 61%) and that the assignment of patients for treatment based on predicted prognosis is an ethical and efficient strategy for the evaluation of new therapies in previously untreated patients with acute leukemia.”

Summary and Potential Future Advances

It has been a great privilege to work closely with J Freireich and colleagues beginning at the National Cancer Institute and subsequently at M. D. Anderson Hospital on many important studies in clinical cancer research. It is indeed a credit to his very fine career to have contributed substantially to new knowledge by taking a quantitative approach to the design and analysis of clinical trials. He has been a leader and involved in important advances in the therapy of cancer, both in randomized and nonrandomized clinical trials. He has been a strong advocate of the use of statistical regression models both for the evaluation of therapies and even for the assignment of patients to treatment in some circumstances. In the future, I expect that the concepts developed by Freireich and colleagues will be developed and refined further. It must be said, however, that improved methodology for the design and analysis of clinical trials will not result in the cure for cancer. Rather, approaches to prevention of cancer and innovative treatments are needed.

From the point of view of methodology, the basic strategies for randomized and nonrandomized studies will continue to be used, with perhaps a deeper understanding of their advantages and limitations. There will be refinements in the prediction of patient outcomes using improved statistical models and more formal ways of using prior data, such as via Bayesian methods. Finally, improved methods are needed for dealing with noncompliance in clinical trials by patients and physicians, and I hope J Freireich will put this topic on his agenda for future work in cancer research.

Acknowledgments

I thank Olga Cobbs for assistance in preparation of the manuscript.

References

- Frei, E., III, Holland, J. F., Schneiderman, M. A., Pinkel, D., Selkirk, G., Freireich, E. J., Silver, R. T., Gold, G. L., and Regelson, W. A comparative study of two regimens of combination chemotherapy in acute leukemia. *Blood*, *13*: 1126–1148, 1958.
- Freireich, E. J., Gehan, E. A., Sulman, D., Boggs, D. R., and Frei, E. The effect of chemotherapy on acute leukemia in the human. *J. Chronic Dis.*, *14*: 593–608, 1961.
- Frei, E. III, Freireich, E. J., Gehan, E. A., Pinkel, D., Holland, J. F., Selawry, O., Haurani, F., Spurr, C. L., Hayes, D. M., James, G. W., Rothberg, H., Sodee, D. B., Rundles, R. W., Schroeder, L. R., Hoogstraten, B., Wolman, I. J., Traggis, D. G., Cooper, T., Gendel, B. R., Ebaugh, F., and Taylor, R. Studies of sequential and combination antimetabolite therapy in acute leukemia: 6-mercaptopurine and methotrexate. *Blood*, *18*: 431–454, 1961.
- Freireich, E. J., Gehan, E. A., Frei, E., III, Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, E. O., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B., and Lee, S. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: a model for evaluation of other potentially useful therapy. *Blood*, *21*: 699–716, 1963.
- Cox, D. R. Regression models and life tables. *J. Royal Stat. Soc. B*, *34*: 187–220, 1972.
- Bodey, G. P., Gehan, E. A., Freireich, E. J., and Frei, F. Protected environment-prophylactic antibiotic program in the chemotherapy of acute leukemia. *Am. J. Med. Sci.*, *262*: 138–151, 1971.
- Rodriguez, F., Bodey, G. P., Freireich, E. J., McCredie, K. B., Gutterman, J. U., Keating, M. J., Smith, T. L., and Gehan, E. A. Randomized trial of protected environment–prophylactic antibiotics in 145 adults with acute leukemia. *Medicine*, *57*: 253–266, 1978.
- Gehan, E. A., and Freireich, E. J. Non-randomized controls in cancer clinical trials. *N. Engl. J. Med.*, *290*: 198–203, 1974.
- Chalmers, T. C., Block, J. B., and Lee, S. Controlled studies in clinical cancer research. *N. Engl. J. Med.*, *287*: 75–78, 1972.
- Freireich, E. J., and Gehan, E. A. The limitations of the randomized clinical trial. In: H. Busch and V. DeVita (eds.), *Methods in Cancer Research: Cancer Drug Development*, Vol. 17, pp. 277–310. New York: Academic Press, 1974.
- Gehan, E. A. The evaluation of therapies: historical control studies. *Stat. Med.*, *3*: 315–324, 1984.
- Farber, S., Diamond, K. L., Mercer, R. D., Sylvester, R. F., Jr., and Wolff, J. A. Temporary remissions in acute leukemia in children produced by folic acid antagonist, *r*-aminopteroyl-glutamic acid (aminopterin). *N. Engl. J. Med.*, *238*: 787–793, 1948.
- Freireich, E. J., and Frei, E., III. Recent advances in acute leukemia. In: C. V. Moore and E. B. Brown (eds.), *Progress in Hematology*, pp. 189–202. New York: Grune & Stratton, 1964.
- Freireich, E. J., Bodey, G. P., Harris, J. E., and Hart, J. S. Therapy for acute granulocytic leukemia. *Cancer Res.*, *27*: 2573–2577, 1967.
- De Vita, V., Serpick, A., and Carbone, P. Combination chemotherapy in the treatment of advanced Hodgkin's disease. *Ann. Intern. Med.*, *73*: 881–895, 1970.
- Gottlieb, J. A., Baker, L. H., Quagliana, J. M., Luce, J. K., Whitecar, J. P., Jr., Sinkovics, J. G., Rivkin, S. E., Brownlee, R., and Frei, E., III. Chemotherapy of sarcomas with a combination of Adriamycin and dimethyl triazeno imidazole carboxamide. *Cancer (Phila.)*, *30*: 1632–1638, 1972.
- Einhorn, L. H., and Donohue, M. *cis*-Diamminedichloroplatinum, vinblastine, and bleomycin combination chemotherapy in disseminated testicular cancer. *Ann. Intern. Med.*, *87*: 293–298, 1977.
- Greenspan, E. Combination cytotoxic chemotherapy in advanced disseminated breast cancer. *J. Mt. Sinai Hosp. NY*, *33*: 1–27, 1966.
- Cooper, R. Combination chemotherapy in hormone resistant breast cancer. *Proc. Am. Assoc. Cancer Res.*, *10*: 15, 1969.
- Budzar, A. U., Gutterman, J. U., and Blumenschein, G. R. Intensive post-operative chemoimmunotherapy for patients with stage II and III breast cancer. *Cancer (Phila.)*, *41*: 1064–1075, 1978.
- Smith, T. L., Gehan, E. A., Keating, M., and Freireich, E. J. Prediction of remission in acute leukemia: development and testing of predictive models. *Cancer (Phila.)*, *50*: 466–472, 1982.
- Keating, M. J., Gehan, E. A., Smith, T. L., Estey, E. H., Walters, R. S., Kantarjian, H. M., McCredie, K. B., and Freireich, E. J. A strategy for evaluation of new treatments in untreated patients: application to a clinical trial of AMSA for acute leukemia. *J. Clin. Oncol.*, *5*: 710–721, 1987.