h

The Power of Designed Experiments

by

Nancy Flournoy
University of Missouri
Columbia, MO 65211, USA
flournoyn@missouri.edu

Abstract: Statistics in developing countries tends to focus on issues arising from government requirements for enumeration and characterization of populations. In contrast, I call attention to the power of designed experiments. I follow a very simple example with a vignette of discovery. This vignette illustrates the power of employing a continual process of hypothesis development and testing. I conclude with warnings against reliance on classical randomized trials, because they can waste time and resources without adequate attention to exploration and hypothesis generation.

keywords: Experimental design, randomized trials, exploratory trials, hypothesis generation

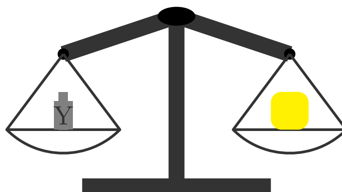
1 Introduction

Statistics in developing countries tends to focus on issues arising from government requirements for enumeration and characterization of populations. In contrast, I call attention to the power of designed experiments. I follow a very simple example with a vignette of discovery. This vignette illustrates the power of employing a continual process of hypothesis development and testing. I conclude with warnings against reliance on classical randomized trials, because they can waste time and resources if adequate attention is not given to exploration and hypothesis generation.

Consider four gold bars denoted A, B, C and D. One dollar ($1 USD) is charged each time the scale is used to weigh the gold bars. The scale is not very accurate, giving results that have $N(0, \sigma^2)$ errors. How would you weight the gold bars if you have only $4?

Three different approaches to answering this question show dramatically different reliability is obtained depending the weights and measurements that are taken.

1. Weight each gold bar individually yields the following four random variables:

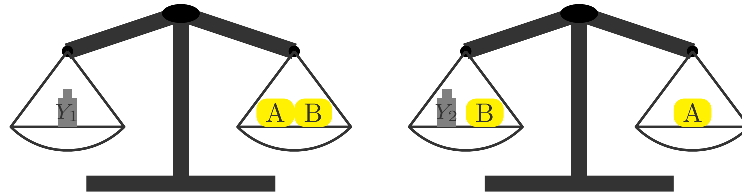


$$Y_1 = A + \varepsilon_1; \qquad Y_2 = B + \varepsilon_2;$$
$$Y_3 = C + \varepsilon_3; \qquad Y_4 = D + \varepsilon_4.$$

The precision of this method is summarized by the variances of the estimated weights:

$$\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \sigma^2.$$

2. First weight the gold bars A and B together and then weight bars C and D together to obtain their total weights. Next obtain the differences in the weights of bars A and B and then of bars C and D as shown in the figure below.

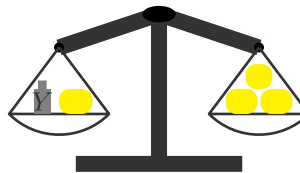


$$Y_1 = A + B + \varepsilon_1; \qquad Y_2 = A - B + \varepsilon_2;$$
$$Y_3 = C + D + \varepsilon_3; \qquad Y_4 = C - D + \varepsilon_4.$$

With this method, the variances of $A+B$, $A-B$, $C+D$, and $C-D$ are each σ^2 from which it follows that

$$\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \frac{1}{2}\sigma^2.$$

3. Weigh the difference between one gold bar and the sum of the other three to obtain the random variables given below:



$$Y_1 = A + B + C - D + \varepsilon_1; \qquad Y_2 = A + B + D - C + \varepsilon_2$$
$$Y_3 = A + C + D - B + \varepsilon_3; \qquad Y_4 = B + C + D - A + \varepsilon_4$$

Calculating the variances of A, B, C and$D$ from the variances of the Ys one finds the precise of this method:

$$\text{Var}(\widehat{A}) = \text{Var}(\widehat{B}) = \text{Var}(\widehat{C}) = \text{Var}(\widehat{D}) = \frac{1}{4}\sigma^2.$$

The three methods of weighing the gold bars correspond to three experimental designs. In summary,

✓All methods give unbiased estimates of the weights.
✓Method 3 yields the smallest variances, namely, $\frac{1}{4}\sigma^2$.
✓To achieve the same precision

Method 1 needs \$16
Method 2 needs \$8
Method 3 needs \$4

There are no designs better than Method 3. It can be shown that Method 3 is the optimal design.

2 A Vignette of Discovery

Today the nation is concerned about the purity of our blood bank system, and anxious that blood banks not become a distribution center for the AIDS virus. Such concerns seem obvious, but 40 years ago transmission of viral infection through the blood was known to occur only with hepatitis, and was not considered for other viruses. Here I review the genesis of a hypothesis that infection from cytomegalovirus could result from contaminated blood products and the experiments that were conducted to test this hypothesis. The question of cytomegalovirus infection resulting from contaminated blood products arose in the early days of bone marrow transplantation. So I begin by describing this environment and how the question came to be asked.

E. Donnell Thomas began to transplant bone marrow into patients without identical twin donors in 1969. By 1975, his Seattle transplant team had transplanted 100 patients with acute leukemia [8]. Some patients with defective blood production were also transplanted. Bone marrow transplantation is now a common treatment for childhood leukemia with a good success rate for young people with a well-matched donor. Attempts to transplant organs, such as kidneys, livers and hearts, had not been very successful until it was determined that matching a patient and donor at a few key genetic loci would substantially reduce the risk of rejection. Drugs to suppress the natural tendency of the patient's immune system to attack a foreign object further reduced the risk of rejection. In bone marrow transplantation, also, a good genetic match was needed to prevent rejection. However, because a new and foreign immune system was being transplanted with the bone marrow, drugs were not only used to reduce the risk of rejection, but to keep the transplanted marrow from deciding that the whole patient was a foreign object and mounting an auto-immune-like attack similar to Lupus.

Furthermore, in order both to destroy diseases of the blood and to prevent rejection of the new bone marrow, high doses of irradiation and drugs are given prior to the transplant. Eradicating as completely as possible all the patient's bone marrow destroys the bulk of the patient's existing immune system. Since, typically two to three weeks are required before the transplanted bone marrow's production of blood cells resumes, the Seattle team tried to anticipate problems that could result from the patient would not have a normal blood production system for an extended period of time when.

In particular, it was known that granulocytes (white blood cells) fight infection. In those days, the word 'infection' implied bacterial infection; very little was known about viruses. In order to protect the patient from infection, an elaborate and expensive system was devised to assure that the patient would have plenty of granulocytes. When the team moved into the newly built Fred Hutchinson Cancer Research Center in 1975, a large portion of one floor was dedicated to this task. On rows of beds, the bone marrow donors lay for hours each day with needles in both arms. Blood was taken from one arm, and passed through a machine that filtered off the granulocytes and returned the rest of the blood to the donor through the other arm.

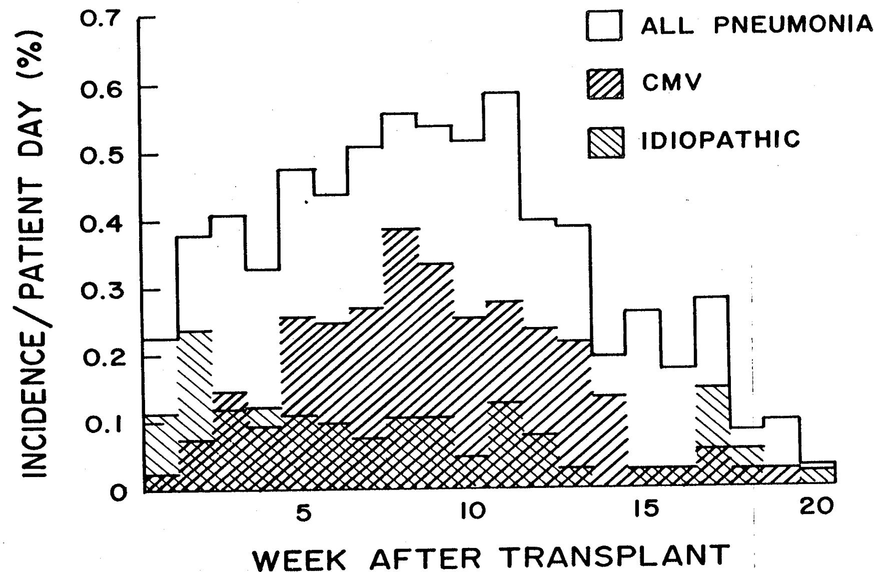


Figure 1: Incidence of CMV and all nonbacterial pneumonias expressed as percentage per patient day for each week after allogeneic marrow transplant.

The bone marrow donor was used in hopes that the genetic match would prevent the patient from becoming allergic to the granulocyte transfusions. The bone marrow donor was expected to stay in town for at least six weeks and lie quietly with needles in both arms every day so that the patient could fight off threats of infection.

Viral Infection

Early in the development of the bone marrow transplant procedure, it was clear that patients whose donors were not identical twins were at high risk of death due to viral pneumonia [7]. Figure 1 (displaying ten years of data from Meyers, Flournoy and Thomas [4]) shows the incidence of pneumonia as a function of time following transplantation. The incidence distribution, expressed as the percentage per patient-day each week, is slightly skewed toward the time of transplant with a mode at about 7 weeks. Of 525 patients, 215 (38%) had nonbacterial pneumonia with cytomegalovirus (CMV) isolated in 40% of the cases and other viruses identified in 29% of the cases.

Eight-four percent of the 215 cases were fatal. In contrast, CMV had not been identified among patients whose bone marrow donors were identical twins [1]. At this time, we speculated that the differences might be due to the fact that patients without identical twin donors received more drugs to suppress the

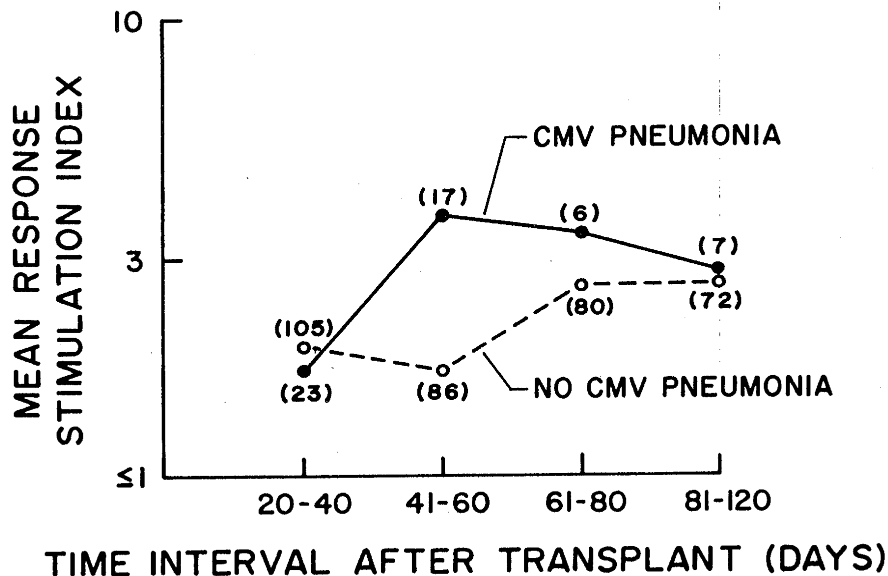


Figure 2: Mean response of lymphocytes to cytomegalovirus antigen. Numbers in parentheses represent the sample size in each group.

immune system than did patients with identical twin donors, but we failed to recognize that their reduced drug therapy was linked to reduced blood transfusion support.

Because CMV was isolated most frequently, and associated with the highest fatality rate, the race was on to characterize the course of illness, identify prognostic factors and find therapies. In order to standardize diagnostic procedures so that the course of the risk period could be established and to identify cases of viral infection early so that intervention trials might be feasible, we instituted a program of blood and urine surveillance testing. Between October 1977 and August 1979, antibody to CMV was measured in 158 patients and their donors prior to transplantation and periodically following transplant. The incidence of CMV infection was approximately the same regardless of the presence or absence of antibody to CMV before transplant in either the donor or the recipient [3]. However, average antibody titers increased by 41-60 days after transplant among patients who contracted CMV pneumonia (see Figure 2). Among patients whose pretransplant CMV titers were positive pretransplant (seropositive), average titers remained high. Whereas, among patients whose pretransplant titers were negative (seronegative), average titers remained low until about 60 days after transplant and then began to rise without regard to the marrow donor's pretransplant titer (see Figure 3).

Prognostic factors were sought among patients transplanted between 1979

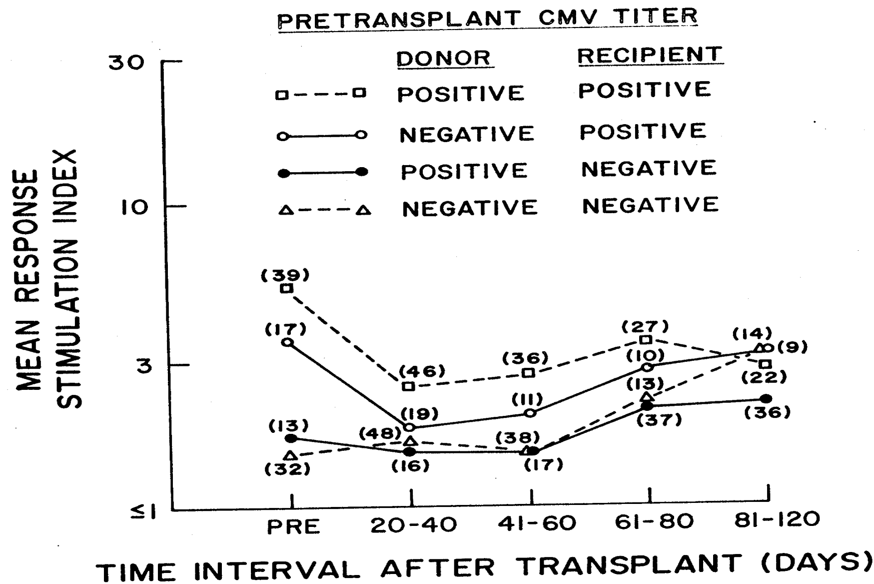


Figure 3: Mean response of lymphocytes to cytomegalovirus. Numbers in parentheses represent the sample size in each group.

and 1982 who had at least four surveillance cultures [5]. The surveillance data showed that just over half (51.5%) of the 545 recipients of marrow transplants without an identical twin donor became infected with CMV. CMV was cultured from 280 (51.4%) of the patients; 168 (30.8%) had at least a 4 fold rise in titers (seroconverted). Much attention in this study focused on the relationship between the surveillance test results and the subsequent development of pneumonia. Also, the relationship between surveillance results, pneumonia and a complication of the transplant procedure called graft-versus-host disease (GVH) was investigated. An association between GVH and CMV clearly existed, suggesting that fatalities due to CMV would be reduced by eliminating GVH. Our studies of GVH are not further discussed here.

Among patients who had CMV found in their blood prior to transplant, 69% subsequently became infected (i.e., their either seroconverted and/or began to excrete CMV in their urine). Among patients without CMV in their blood prior to transplant, 57% of those whose donors did and 28% of those whose donors did not have CMV in their blood subsequently became infected. These observations suggested that patients having CMV in their blood prior to transplant were at high risk of infections, whereas among patients without CMV in their blood prior to transplant, the donor might be passing infection to the patient, either through the marrow transplant itself or through the blood transfusions given after transplant.

A proportional hazards regression analysis was performed separately for patients with and without CMV in their blood prior to transplant. Among seropositive patients, all the significant covariates were demographic variables, disease characteristics or treatment complications for which no known control was possible. Thus the models did not suggest possible interventions. However, among seronegative patients, the relative rate of CMV infection was 2.3 times greater (p=0.0006) if the granulocyte transfusions were also found to be positive for CMV.

Interventions

In light of these results, we focused on the patients whose pretransplant titers were negative pretransplant. Thinking that prophylactic CMV immune globulin might prevent CMV infection from developing, eligible patients were randomized to receive globulin or nothing with stratifications for the use of prophylactic granulocyte transfusions and the donor's titer to CMV. At the onset of the CMV immune globulin study, we took the association seriously enough to stratify for it but not so seriously as to study it directly. To be eligible for this study [6], a patient had to be seronegative for antibody to CMV prior to transplant and not excrete the virus into the urine for the first 2 weeks after transplantation. Patients excreting virus during this period were presumed to have been infected with CMV before transplantation and were excluded from final analysis. CMV infection rates are shown in Table 1.

Table 1: Incidence of CMV Infection

Patients receiving	**GLOBULIN**	NO **GLOBULIN**
Granulocytes from seropositive donors	7/8 (88.5%)	6/7 (85.7%)
Granulocytes from seronegative donors	1/5 (20.0%)	0/6 (00.0%)
No granulocytes	2/17 (11.8%)	8/19 (42.1%)

The overall difference in infection rates between globulin recipients and controls was not significant, but since sample sizes were small within strata, hope that globulin might be effective was sustained by the difference observed among patients receiving no granulocytes.

Figure 4 compares Kaplan-Meier estimates of the probability of infection as a function of week after transplant for globulin recipients and patients who did not receive granulocyte transfusions. The difference in rates depending upon whether or not the granulocyte donor was seronegative or seropositive finally led us to question seriously the role of granulocyte transfusions in CMV infection.

Although this study was designed to evaluate globulin, we were thunderstruck by the possibility that we were transmitting CMV through the blood. The impact should this observation be confirmed in a controlled randomized study is described by Meyers, et al. [6]

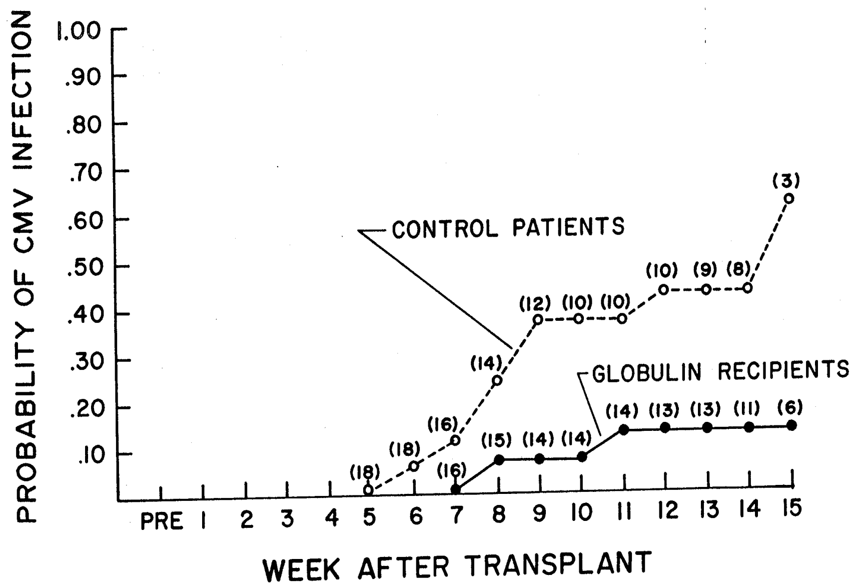


Figure 4: Probability of acquiring cytomegalovirus infection among patients who did not receive prophylactic granulocyte transfusions. The numbers in parentheses indicate the sample size of patients till at risk of infection at the beginning of each interval. The risk is different for globulin recipients and controls at p=0.03 by the Mantel-Cox test.

> screening blood products for antibody to cytomegalovirus, or more appropriately for virus or viral antigens (techniques that are not yet available), increases the burden on blood-banking facilities, decreases the pool of blood donors, and, most importantly, decreases the rapid availability of of fresh blood products such as platelets. The use of an immune globulin is therefore an attractive practical alternative among patients who need large amounts of fresh blood products such as platelets and for whom screening of blood products is less practical.

The New England Journal of Medicine rejected our paper [6] because it includes a discussion of findings external to those postulated in the initial experimental design. Such concern is compounded by the small sample sizes. However, observations concerning granulocyte transfusions led to new hypotheses and the next randomized study. Thus, while it is extremely important to distinguish between observations obtained by controlled randomized intervention and those obtained otherwise, hypothesis generation is a very important task.

To followup on the question of whether or not granulocyte transfusions carried CMV infection, we spent a year working with King County Blood Bank to develop screening procedures, set up laboratory equipment and train technicians in order to conduct a randomized clinical trial. Although we restricted the study to patients who were seronegative for CMV in 2 consecutive tests and who had not received any unscreened blood recently, more patients were available for study than the blood bank could handle. Therefore, we we studied the prophylactic capability of immune globulin at the same time in a randomized 2X2 factorial design. CMV immune globulin had no effect on the rate of CMV infection [2]. The effect of giving only CMV negative blood is summarized in Table 2.

Table 2: Incidence of CMV Infection among 85 patients studied for at least 62 days after transplantation

	BLOOD TRANSFUSIONS	
Donor's CMV status	**SERONEGATIVE**	**SEROPOSITIVE**
Seropositive	1/22 (04.5%)	8/25 (32.0%)
Seronegative	3/12 (25.0%)	5/16 (31.3%)

The one patient with a seronegative donor who was assigned to receive seronegative blood products and subsequently became infected with CMV actually mistakenly received several seropositive transfusions.

As the study proceeded the blood bank personnel became increasingly agitated as they considered the ramifications of a significant finding. Blood banks all over the country would have to set up screening programs; the cost would be enormous they warned. The study results went out to blood banks, however, and viral screening procedures were put into place. The timing was fortuitous

because the AIDS crisis was building. Today the idea that viral infections can be transmitted through the blood is taken for granted.

3 Randomized Two Arm Designs Used Prematurely Waste Resources

Often interventions are complex combinations of treatments defined by doses or stress levels. When two points from a high dimensional space are defined to be "treatments" to be compared in a randomized study, information about the dose-response function is lost. For example, to define a "treatment" from a radiation schedule one must select the total amount, the number of fractions in which the radiation will be delivered, the time interval between fractions and the wave length of the radiation. After, say, 3 experiments comparing two treatments at a time one may know the precision of the treatments selected very well but still have almost no information about the response surface. Hence, the optimal "point" or combination therapy may be far from any of those studied and the experiments conducted will provide little insight as to where this optimal treatment lies. It is like knowing a lot about a few trees and nothing about the forest.

This frustration with the classical two arm clinical trial led me to study adaptive treatment allocation procedures, focusing on those that aid exploration while clustering subjects in regions of interest. This is the subject of other talks.

4 Conclusions

There are three key steps to successful experimental design. First, develop a simple, clear and important question. Pay attention and take the time to formulate a meaningful hypothesis. This period of hypothesis development may take a long time; it may utilize information that is conveniently available and/or information generated by carefully constructed adaptive allocation procedures. Second, develop a few interventions (perhaps including a control) and randomize subjects to the interventions. Third, replicate the experiments and encourage others to do likewise.

References

[1] F.R. Appelbaum, J.D. Meyers, A. Fefer, N. Flournoy, N Cheever, M.A. Greenberg, P.D. Greenberg, R. Hackman, and E.D. Thomas. Nonbacterial nonfungal pneumonia following marrow transplantation in 100 identical twins. *Transplantation*, 33:265–268, 1982.

[2] R.A. Bowden, M. Sayers, N. Flournoy, B. Newton, M. Banaji, E.D. Thomas, and J.D. Meyers. Cytomegalovirus immune globulin and seronegative blood

products to prevent primary cytomegalovirus infection after marrow transplant. *N Engl J Med*, 314:1006–1010, 1986.

[3] J.D. Meyers, N. Flournoy, and E.D. Thomas. Cytomegalovirus infection and specific cell-mediated immunity after marrow transplant. *J Infect Dis*, 142:816–824, 1980.

[4] J.D. Meyers, N. Flournoy, and E.D. Thomas. Nonbacterial pneumonia after allogeneic marrow transplantation. *Rev Infect Dis*, 4:1119–1132, 1982.

[5] J.D. Meyers, N. Flournoy, and E.D. Thomas. Risk factors for cytomegalovirus infection after human marrow transplant. *J Infect Dis*, 153:478–488, 1986.

[6] J.D. Meyers, J Leszczynski, J.A. Zaia, N. Flournoy, , B. Newton, D.R. Snydman, G.G. Wright, M.J. Levin, and E.D. Thomas. Risk factors for cytomegalovirus infection after human marrow transplant. *Ann Intern Med*, 98:442–446, 1983.

[7] P.E. Neiman, W. Reeves, G. Ray, N. Flournoy, K.G. Lerner, G.E. Sale, and E.D. Thomas. A prospective analysis of interstitial pneumonia and opportunistic viral infection among recipients of allogeneic bone marrow graphs. *J Infect Dis*, 136:754–767, 1977.

[8] E.D. Thomas, C.D. Buckner, M. Banaji, R.A. Clift, A. Fefer, N. Flournoy, B.W. Goodell, R.O. Hickman, K.G. Lerner, P.E. Neiman, G.E. Sale, J.E. Sanders, J. Singer, M. Stevens, R. Storb, and P.L. Weiden. One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation. *Blood*, 49:511–533, 1977.