*Eur J Cancer Clin Oncol*, Vol. 25, No. 11, pp. 1519–1523, 1989.
Printed in Great Britain

0277–5379/89$3.00 + 0.00


*Commentary*

# Bone Marrow Transplantation Versus Chemotherapy in Acute Non-Lymphocytic Leukemia: A Meta-Analytic Review

COLIN B. BEGG,* LOUISE PILOTE† and PHILIP B. McGLAVE‡

**Memorial Sloan-Kettering Cancer Center, New York, †Harvard School of Public Health, Boston, MA and ‡University of Minnisota, Minneapolis, MN, U.S.A.*

(A COMMENT ON: Hermans J, Suciu S, Stijnen T *et al.* Treatment of Acute Myelogenous Leukemia. An EBMT–EORTC retrospective analysis of chemotherapy versus allogeneic or autologous bone marrow transplantation. *Eur J Cancer Clin Oncol* 1989, **25**, 545–550.)

IN THIS JOURNAL Hermans *et al.* [1] have performed a retrospective analysis in order to compare alternative therapeutic approaches to the treatment of acute nonlymphocytic leukemia in young adults. They studied the relative efficacy of bone marrow transplantation (BMT) and conventional chemotherapy for patients in first complete remission, using registry information from the European Cooperative Group for Bone Marrow Transplantation, and data from two trials conducted by the European Organization for Research and Treatment of Cancer as controls. They used disease-free survival (DFS) from complete remission (CR) as the primary endpoint. Their analysis involves an adjustment to correct for the bias that is caused by the fact that many patients who might have received transplants are automatically excluded from the BMT series by virtue of an early relapse prior to transplantation. For this purpose they employed the Mantel–Byar procedure [2]. The analysis demonstrates significant superiority for allogeneic BMT over conventional chemotherapy ($P = 0.01$), with the results of autologous BMT lying between those of the other two groups. Further analyses using Cox regression with time-dependent covariates demonstrate that the superiority of allogeneic BMT over conventional therapy is primarily in longer-term disease-free survival, the short-term results being very similar. Less clear-cut conclusions are evident for comparisons of autologous BMT and chemotherapy.

In their summary the authors acknowledge the limitations of retrospective investigations of this nature, and point out some additional biases which might have influenced the results due to, for example, underreporting of early failures to the registry, temporal effects in patient care, and other factors affecting patient selection differentially between the comparison groups. They conclude by cautiously suggesting that allogeneic BMT produces better disease-free survival in this patient population.

We believe that the analytic methodology used by the authors is sound, and that their cautious interpretation of their results is appropriate. In the following we have endeavored to assemble results of recently published related studies to see if consistent results are emerging, and to try to provide a more confident conclusion. We restrict attention to comparisons of allogeneic BMT and chemotherapy, since very little information is yet available on results of autologous BMT. After adjusting for a variety of biases that affect the comparability of published studies, we demonstrate that there is great consistency in the results to date, with BMT displaying a 15–20% improvement in long-term DFS.

## RESULTS OF PUBLISHED STUDIES

We have assembled our studies for analysis by a variety of means including anecdotal knowledge, the follow-up of articles cited by relevant papers,

Accepted 20 June 1989.
This work was supported by grants CA-35291, CA-21737, CA-45814 and CA-39640, awarded by the National Cancer Institute, DHHS.

and a MEDLINE literature search using the keywords 'acute myelogenous leukemia', 'bone marrow transplant' and 'chemotherapy'. We were interested primarily in comparative studies of BMT and chemotherapy [1, 3–7], but we also assembled recent uncontrolled studies of BMT [8, 9], and also studies of long-term follow-up of patients treated with chemotherapy [10–15] as a potential source of further controls.

The varied nature of these studies highlights our first problem, namely which studies should be included in a meta-analysis of this nature. The ideal study would be a randomized comparison of the two treatments employing an analysis based on the intent-to-treat principle. That is, for example, all patients randomized to BMT would be included in the BMT group regardless of whether they received this treatment. This is especially important since many patients typically relapse prior to transplantation and so failure to employ the intent-to-treat principle would lead inevitably to serious bias, causing overestimation of the efficacy of BMT. Indeed it is for the purpose of correcting this bias that Hermans *et al.* [1] use the Mantel–Byar adjustments in their retrospective study. Unfortunately there have been no randomized studies to date, and so we have to make do with the next best thing, the few available prospective controlled studies, that is the studies by Appelbaum *et al.* [3], Cassileth *et al.* [4], Champlin *et al.* [5] and Conde *et al.* [6]. In each of these studies a cohort of patients in complete remission and potentially eligible for transplant was assembled, and those who consented and for whom a donor could be found were transplanted, the remaining patients being used as controls. As in the case with randomized studies it is important that the intent-to-treat principle be applied in analyzing these studies. This was not the case in all of these studies, and so we have made a number of adjustments to the published results in order to make the data suitably comparable, details of which will be discussed later.

A less well-controlled study involves the use of separate series of patients compared retrospectively. Such a study design has the additional problem that the patient populations may differ between the two comparison groups in ways that are difficult to detect. The study by Hermans *et al.* is in this category, as is the earlier study of a similar nature which we conducted [7]. The advantage of these studies is that the sample sizes are much larger than in the controlled studies. There is a potential problem of overlap, although we believe that this is a minor issue. Our earlier study [7] employed published BMT data primarily from institutions in the U.S.A., while the Hermans study involves European transplants. The controlled comparative studies [3–6], with the exception of the study by Conde *et al.*, were conducted in the U.S.A. and so would not be part of the European registry data used by Hermans *et al.* There may be some overlap involving the studies by Conde *et al.* and Hermans *et al.* but the numbers involved are so small that this dependence will have little influence on our conclusions.

Finally we have assembled some recent uncontrolled studies of BMT which would not be part of the registry [8, 9] and some reports of the long-term results of chemotherapy treatment [10–15]. In addition to the problems of the preceding studies, such uncontrolled studies are especially susceptible to publication bias since the primary rationale for their publication may be the observation of unusually favorable results [16].

We have tabulated the *reported* results of each of these studies in Table 1, using disease-free survival as the endpoint. Many of these results are abstracted from published graphics, and so there may be some inaccuracy in translation. It is clear, at the outset, that there is wide variation in the results. However, much of the variability is explainable due to differing definitions of DFS and the presence of bias in the BMT groups, in some cases due to the effect of excluding patients who failed prior to transplantation. We have systematically adjusted each of these studies in order to express the results in comparable units of measurement, and to correct bias where it is apparent. We have used DFS from time of complete response as our standard and translated each of the studies accordingly. Specific details of the corrections used are in the Appendix. The corrected estimates appear in Table 2.

Examination of Table 2 clarifies the need for the standardization of the results. The adjusted results are characterized by consistency, especially for the controlled studies, unlike Table 1 where there is great diversity. With the exception of the study by Conde *et al.*, which is very unprecise anyway due to the small sample size, the studies of allogeneic BMT consistently show 1-year disease-free survival in the region of 55% dropping to a plateau at 3 years of around 40%. By contrast the chemotherapy controls, while having similar one year DFS around 50%, experience substantially poorer long-term DFS survival in the region of 20–25%. The uncontrolled studies, not surprisingly, have somewhat more varied results, due to the more varied selection effects outlined earlier.

We have conducted a formal meta-analysis of these results, using a procedure which gives more 'weight' to the internally controlled studies depending on the extent of between-study variability, but nonetheless makes use of the uncontrolled studies. This method is analogous to the DerSimonian and Laird method [17]. (Details are available from the authors upon request.) Our analysis gives the

*Table 1. Abstracted data on disease-free survival*

| Study | % DFS (Standard Error) BMT 1 year | 2 year | 3 year | 4 year | 5 year | Chemotherapy 1 year | 2 year | 3 year | 4 year | 5 year |
|---|---|---|---|---|---|---|---|---|---|---|
| Appelbaum [3] | 55% (12%) | 55% (12%) | 52% (12%) | 48 (12%) | 48% (12%) | 54% (8%) | 25% (8%) | 23% (7%) | 23% (7%) | 23% (7%) |
| Cassileth [4] | 55% (10%) | 50% (10%) | 36% (9%) | | | 40% (8%) | 23% (7%) | 23% (7%) | 23% (7%) | |
| Champlin [5] | 81% (10%) | 60% (13%) | 60% (13%) | 60% (13%) | 60% (13%) | 48% (9%) | 31% (8%) | 30% (8%) | 30% (8%) | 30% (8%) |
| Conde [6] | 70% (23%) | 70% (23%) | 70% (23%) | 70% (23%) | 70% (23%) | 48% (17%) | 48% (17%) | 17% (13%) | | |
| Begg [7] | 68% (4%) | 57% (5%) | 53% (6%) | | | 49% (5%) | 26% (4%) | 21% (4%) | 21% (4%) | |
| Hermans [1] | 64% (2%) | 51% (3%) | 47% (3%) | 46% (3%) | | 50% (4%) | 32% (4%) | 24% (4%) | 18% (4%) | |
| Forman [8] | 61% (8%) | 52% (9%) | 52% (9%) | 52% (9%) | | | | | | |
| McGlave [9] | 65% (8%) | 62% (8%) | 62% (8%) | 62% (8%) | 62% (8%) | | | | | |
| Champlin [10] | | | | | | 60% (9%) | 48% (9%) | 32% (9%) | 32% (9%) | 32% (9%) |
| Hayat [11] | | | | | | 44% (5%) | 26% (4%) | 17% (4%) | 16% (4%) | |
| Preisler [12] | | | | | | 50% (3%) | 33% (3%) | 26% (3%) | 22% (3%) | 19% (2%) |
| Rees [13] | | | | | | 62% (3%) | 38% (3%) | 29% (3%) | 24% (3%) | 22% (3%) |
| Schwartz [14] | | | | | | 50% (10%) | 24% (8%) | 16% (7%) | 12% (6%) | |
| Wolff [15] | | | | | | 76% (7%) | 53% (8%) | 53% (8%) | 50% (8%) | 50% (8%) |

*Table 2. Adjusted data on disease-free survival**

| Study | % DFS BMT 1 year | 2 year | 3 year | 4 year | 5 year | Chemotherapy 1 year | 2 year | 3 year | 4 year | 5 year |
|---|---|---|---|---|---|---|---|---|---|---|
| Appelbaum [3] | 49% | 46% | 42% | 40% | 40% | 54% | 25% | 23% | 23% | 23% |
| Cassileth [4] | 55% | 50% | 36% | | | 40% | 23% | 23% | 23% | |
| Champlin [5] | 54% | 47% | 40% | 40% | | 54% | 42% | 28% | 28% | |
| Conde [6] | 70% | 70% | 70% | 70% | | 48% | 48% | 17% | | |
| Begg [7] | 54% | 46% | 42% | | | 40% | 21% | 16% | 16% | |
| Hermans [1] | 54% | 43% | 40% | 39% | | 50% | 32% | 24% | 18% | |
| Forman [8] | 59% | 49% | 47% | 47% | 47% | | | | | |
| McGlave [9] | 61% | 53% | 53% | 53% | 53% | | | | | |
| Champlin [10] | | | | | | 60% | 48% | 32% | 32% | 32% |
| Hayat [11] | | | | | | 44% | 26% | 17% | 16% | |
| Preisler [12] | | | | | | 50% | 33% | 26% | 22% | 19% |
| Rees [13] | | | | | | 62% | 38% | 29% | 24% | 22% |
| Schwartz [14] | | | | | | 50% | 24% | 16% | 12% | |
| Wolff [15] | | | | | | 76% | 53% | 53% | 50% | 50% |
| Summary estimates (standard error) | 55% (2.5%) | 46% (2.5%) | 41% (2.6%) | 42% (2.7%) | — — | 53% (1.6%) | 32% (1.5%) | 25% (1.4%) | 21% (1.3%) | — — |

*DFS measured from CR. The standard errors have been omitted to simplify the display.

following summary estimates of the improvement in DFS of BMT over chemotherapy, with standard errors of the estimates in parentheses: 2% (3%) at 1 year; 13% (3%) at 2 years; 16% (3%) at 3 years; 21% (3%) at 4 years.

## CONCLUSIONS

The continued uncertainty about the relative benefits of bone marrow transplantation in acute nonlymphocytic leukemia highlights the crucial importance of randomization in the scientific study of alternative medical therapies. Even at this late stage we do not have results from any randomized trials. This is to some extent due to the ethical difficulties of randomizing after very promising results have been published, although logistical difficulties are also a factor in this case. The controlled studies that have been conducted are clearly not as reliable as randomized studies analyzed using the intent-to-treat philosophy would be, and the wide variation in the tentative results supports our general skepticism about the reliability of non-randomized investigations. Nonetheless, our efforts to homogenize the results by expressing them in commensurate units of measurement, and by making suitable bias corrections, have produced results of remarkable consistency, at least for the internally controlled studies. Consequently, we feel somewhat confident that the conclusions of Hermans' article are correct, namely that the short-term effects of BMT and chemotherapy are essentially equivalent, but that the probability of long-term disease-free survival is substantially higher for allogeneic BMT, at about 40% vs. 20%. However, we also await with great interest the results of the controlled prospective studies currently in progress.

## APPENDIX: DATA ADJUSTMENTS

We adjusted the reported results (Table 1) in order to make them directly comparable (Table 2) as follows:

1. Appelbaum [3]: We used the intent-to-treat analysis rather than the comparison which excludes patients refusing BMT. (Note: a further concern is that in patients relapsing prior to BMT the time to 2nd relapse may have been used rather than time to 1st relapse. The paper is ambiguous on this point.)
2. Cassileth [4]: No adjustments necessary.
3. Champlin [5]: We used the authors' overall survival results, since non-leukemia deaths were excluded from their BMT results, leading to serious bias.
4. Conde [6]: We used the age-stratified control group (<45).
5. Begg [7]: We used the original (unadjusted) control group and compensatingly deflated the BMT results by 20% to adjust for the patients who would have failed prior to transplantation.
6. Hermans [1]: We needed to impute the DFS from CR for the BMT group. This was accomplished by estimating the distribution of times from CR to BMT, and then convoluting this with the observed data to estimate the missing censoring distribution. The resulting curve is then deflated in similar manner to Begg above, but deflating by 15% rather than 20% since the median time to BMT is 3 months rather than 4 months.
7. Forman [8]: The DFS from CR is imputed (approximately) from the observed DFS from BMT by using the DFS at 10 months, 22 months, etc. rather than 1 year, 2 years, etc., since the median time from CR to BMT is 2 months. The resulting estimates are then deflated by 10% to adjust for the patients who relapsed prior to transplantation.
8. McGlave [9]: We used similar adjustments to Forman (above), except that we used the estimates at 9 months, 21 months, etc. and deflated by 15%. Also, we restricted the analysis to adults (>18).
9. Champlin [10]: No adjustments made. (Note: we would have excluded patients >45 years if this had been possible.)
10. Hayat [11]: As in Champlin above.
11. Preisler [12]: We restricted attention to the 20–49 age group.
12. Rees [13]: We restricted attention to the 14–39 age group (in both Tables 1 and 2).
13. Schwartz [14]: As in Champlin (9) above.
14. Wolff [15]: As in Champlin (9) above.

## REFERENCES

1. Hermans J, Suciu S, Stijnen T *et al.* Treatment of acute myelogenous leukemia: an EBMT–EORTC retrospective analysis of chemotherapy versus allogeneic or autologous bone marrow transplantation. *Eur J Cancer Clin Oncol* 1989, **25**, 545–550.
2. Mantel N, Byar DP. Evaluation of response–time data including transient states: an illustration using heart-transplant data. *J Am Stat Assoc* 1974, **69**, 81–86.
3. Appelbaum FR, Fisher LD, Thomas ED *et al.* Chemotherapy v. marrow transplantation for adults with acute nonlymphocytic leukemia: a five-year follow-up. *Blood* 1988, **72**, 179–184.
4. Cassileth PA, McGlave PB, Harrington DP *et al.* Comparison of post-remission therapy in AML: maintenance versus intensive consolidation therapy versus allogeneic bone marrow

transplantation. *Proc Am Soc Clin Oncol* 1989, **8**, 197.
5. Champlin RE, Winston GH, Gale RP *et al.* A prospective controlled trial of bone marrow transplantation versus consolidation chemotherapy. *Ann Intern Med* 1985, **102**, 285–291.
6. Conde E, Iriondo A, Rayon C *et al.* Allogeneic bone marrow transplantation versus intensification chemotherapy for acute myelogenous leukemia in first remission: a prospective controlled trial. *Br J Haematol* 1988, **68**, 219–226.
7. Begg CB, McGlave PB, Bennett JM *et al.* A critical comparison of allogeneic bone marrow transplantation and conventional chemotherapy as treatment for acute nonlymphocytic leukemia. *J Clin Oncol* 1984, **2**, 369–378.
8. Forman SJ, Krance RA, O'Donnell MR *et al.* Bone marrow transplantation for acute nonlymphoblastic leukemia during first complete remission. *Transplantation* 1987, **43**, 650–653.
9. McGlave PB, Haake RJ, Bostrom BC *et al.* Allogeneic bone marrow transplantation for acute nonlymphocytic leukemia in first remission. *Blood* 1988, **72**, 1512–1517.
10. Champlin R, Gale RP, Elashoff R *et al.* Prolonged survival in acute myelogenous leukemia without maintenance chemotherapy. *Lancet* 1984, **1**, 894–895.
11. Hayat M, Jehn U, Willemze R *et al.* A randomized comparison of maintenance treatment with androgens, immunotherapy and chemotherapy in adult myelogenous leukemia. *Cancer* 1986, **58**, 617–623.
12. Preisler HD, Anderson K, Rai K *et al.* The frequency of long-term remission in patients with acute myelogenous leukemia treated with conventional maintenance chemotherapy: a study of 760 patients with a minimal follow-up time of 6 years. *Br J Haematol* 1989, **71**, 189–194.
13. Rees JKH, Swirsky D, Gray RG *et al.* Principal results of the Medical Research Council's 8th acute myeloid leukemia trial. *Lancet* 1986, **2**, 1236–1241.
14. Schwartz RS, Mackintosh FR, Halpern J *et al.* Multivariate analysis of factors associated with outcome of treatment for adults with acute myelogenous leukemia. *Cancer* 1984, **54**, 1672–1681.
15. Wolff SN, Herzig RH, Fay JW *et al.* High-dose cytosine arabinoside and daunorubicin as consolidation therapy for acute myeloid leukemia in first remission: long-term follow-up and results. Unpublished manuscript, Barnard Cancer Center, Washington University, St. Louis, Missouri.
16. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data (with discussion). *J Royal Stat Soc A* 1988, **151**, 419–463.
17. DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Cont Clin Trials* 1986, **7**, 177–188.