EDITORIAL

Use of Colony-Stimulating Factors in the Treatment of Acute Myeloid Leukemia

By Elihu H. Estey

In this issue of Blood, Ohno et al become the first to report a trial in which patients with acute myeloid leukemia (AML) received granulocyte-colony-stimulating factor (G-CSF) both before and during, as well as after, chemotherapy. Fifty-eight patients in relapse or resistant to 2 courses of initial therapy ($n = 50$) or with untreated disease arising out of a myelodysplastic syndrome ($n = 8$) were randomly assigned to receive G-CSF (200 μg/m$^2$ daily) or placebo beginning 2 days before, and continuing until 35 days after, a mitoxantrone, behenoyl cytosine arabinoside, etoposide regimen. The G-CSF group took less time to attain neutrophil counts of ≥500 and ≥1,000 after chemotherapy. However, there were no statistically significant differences between the two groups in marrow blast percent on day 1 of chemotherapy, persistence of blasts after chemotherapy, incidence of febrile episodes or documented infections, complete remission (CR) rate, event-free survival of responding patients, or survival of all patients. Statistical significance was assessed by $P$ values, which were <.001 for the indices of neutrophil recovery but >.3 for the other endpoints.

Clinical trials of CSFs in conjunction with chemotherapy have been ongoing in AML since the late 1980s. To date, the vast majority of trials have used granulocyte-macrophage–CSF (GM-CSF) or G-CSF. These trials have been designed either to decrease the rate of infection, particularly fatal infection, after chemotherapy, to increase the sensitivity of leukemic blasts to chemotherapy, or, as in the study of Ohno et al, to meet both objectives. To lower the infection rate, GM-CSF or G-CSF have been administered after completion of chemotherapy so as to accelerate neutrophil recovery. Given in vitro data indicating that CSFs can stimulate growth of clonogenic blasts, this approach has the potential risk of accelerating the disease. To sensitize (or “prime”) leukemic blasts to chemotherapy, GM-CSF or G-CSF have been administered 1 to 8 days before, and/or during, chemotherapy. Sensitization has been postulated to occur through recruitment of cells into the cell cycle, although non—cell-cycle–related mechanisms have been proposed. Such attempts at priming may also enhance growth of leukemia, interfere with the cytotoxic effects of antileukemia agents, and sensitize normal hematopoietic progenitors to the effects of chemotherapy, thereby producing prolonged myelosuppression. To reduce this latter risk, or to improve both supportive care and leukemia sensitization, some studies, such as the current one of Ohno et al, have administered GM-CSF or G-CSF after completion of, as well as before and during, chemotherapy.

Table 1 summarizes trials of GM-CSF or G-CSF in chemotherapy of newly diagnosed or relapsed AML with regard to the major therapeutic endpoints of neutrophil recovery, infection, CR, remission duration, and survival rates. Trials were included if a comparison, non–CSF-treated group was specified. Trials are grouped in the table by type of CSF administered and timing of CSF administration. Daily doses were 120, 200, or 250 μg/m$^2$ GM-CSF, and 200 or 400 μg/m$^2$ G-CSF. These trials, including the current study of Ohno et al, agree on several points. First, the incidence of serious nonhematologic toxicity after the use of either CSF in newly diagnosed or relapsed patients is similar to that seen in historical or concurrent controls not receiving CSFs. Second, in no trial has administration of GM-CSF or G-CSF beginning after completion of various chemotherapy regimens, in newly diagnosed or relapsed disease, been associated, on average, with persistence or regrowth of leukemia during induction, or with shorter remissions, compared with patients receiving the same chemotherapy without CSF. Third, administration of GM-CSF or G-CSF either after or before and during, or before, during, and after various chemotherapies has reduced the median time to neutrophil counts of ≥500 and ≥1,000 by an average of about 7 days, with the only exception being a relatively small study. Such reduction has occurred in both newly diagnosed and relapsed patients and during both induction and consolidation cycles. There has been no effect on platelet recovery.

Despite the well-known correlation between duration of neutropenia and risk of infection, this accelerated rate of neutrophil recovery has not consistently translated into statistically significant ($P$ value <.05) reductions in rates of infection or fatal infection (Table 1). Using the same statistical criterion, none of the studies in Table 1 found an improve...
Table 1. Comparisons of Chemotherapy With or Without GM-CSF or G-CSF

<table>
<thead>
<tr>
<th>Study</th>
<th>CSF</th>
<th>Timing of CSF Administration</th>
<th>Patients*</th>
<th>Neutrophil Recovery</th>
<th>Infection Rate</th>
<th>CR Rate</th>
<th>Remission Duration</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estey et al 1990 (H)</td>
<td>GM</td>
<td>After CHDAC induction</td>
<td>65 (ND)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Not given</td>
<td>Not given</td>
</tr>
<tr>
<td>Büchner et al 1991 (H)</td>
<td>GM</td>
<td>After TAD induction (if no leukemia in marrow on day 12)</td>
<td>92 (ND, RE)</td>
<td>Faster with GM</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>Not given</td>
</tr>
<tr>
<td>Rowe et al 1993 (R)</td>
<td>GM</td>
<td>After 3 + 7 induction (if no leukemia in marrow on day 10) and after HDAC consolidation</td>
<td>118 (ND)</td>
<td>Faster with GM</td>
<td>Lower with GM</td>
<td>NS</td>
<td>Not given</td>
<td>Longer with GM</td>
</tr>
<tr>
<td>Estey et al 1992 (H)</td>
<td>GM</td>
<td>Before and during dauno + CHDAC induction</td>
<td>232 (ND)</td>
<td>Faster with GM</td>
<td>Not given</td>
<td>Lower with GM</td>
<td>NS</td>
<td>Shorter with GM</td>
</tr>
<tr>
<td>Archimbaud et al 1993 (H)</td>
<td>GM</td>
<td>Before AE, after MA</td>
<td>58 (RE)</td>
<td>NS</td>
<td>Not given</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Bettelman et al 1992 (H)</td>
<td>GM</td>
<td>Before, during, and after 3 + 7 induction</td>
<td>57 (ND)</td>
<td>Faster with GM</td>
<td>Not given</td>
<td>Not given</td>
<td>Not given</td>
<td>Not given</td>
</tr>
<tr>
<td>Büchner et al 1993 (R)</td>
<td>GM</td>
<td>Before, during, and after induction and consolidation with TAD/HAM</td>
<td>63 (ND)</td>
<td>Faster with GM</td>
<td>Not given</td>
<td>NS</td>
<td>Longer with GM</td>
<td>Not given</td>
</tr>
<tr>
<td>Ohno et al 1990 (R)</td>
<td>G</td>
<td>After MAE induction (if no leukemia in marrow on days 8-12)</td>
<td>61 (RE, ND)</td>
<td>Faster with G</td>
<td>Lower with G</td>
<td>NS</td>
<td>NS</td>
<td>Not given</td>
</tr>
<tr>
<td>Ohno et al 1994 (R)</td>
<td>G</td>
<td>Before, during, and after MAE induction</td>
<td>58 (RE, ND)</td>
<td>Faster with G</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Estey et al 1994 (H)</td>
<td>G</td>
<td>Before, during and after FA induction</td>
<td>197 (ND)</td>
<td>Faster with G</td>
<td>NS</td>
<td>NS</td>
<td>Not given</td>
<td>NS</td>
</tr>
</tbody>
</table>

Abbreviations: H, comparisons were to a historical group not receiving CSF; R, comparisons were to a concurrently randomized group; CHDAC, continuous infusion high-dose ara-C; TAD, 6-thioguanine + ara-C + daunorubicin; HDAC, high-dose ara-C (=1.5 g/m2/d); AE, ara-C + etoposide; MA, mitoxantrone + ara-C; HAM, high-dose ara-C + mitoxantrone; MAE, mitoxantrone + ara-C + etoposide; FA, fludarabine + high-dose ara-C; ND, newly diagnosed; RE, relapsed or refractory disease; NS, no statistically significant event on the particular outcome at $P < .05$.

* Includes both CSF and control groups; numbers in the CSF group were 12, 36, 59, 56, 20, 18, 34, 30, 28, and 112 in studies 1 through 10, respectively. The proportions of newly diagnosed patients in the 1991 Büchner et al study, the 1990 Ohno et al study, and the 1994 Ohno et al study were 72%, 4%, and 14%, respectively.

ment in CR rate, and only the ECOG study found an improvement in survival rate, with this study being limited to newly diagnosed patients aged 55 to 70. One study in which GM-CSF was administered before and during chemotherapy to newly diagnosed patients reported statistically significant decreases in both CR and survival rates. Effects on CR and survival could, of course, reflect either effects on fatal infection rates or, in studies in which CSF is administered before and/or during chemotherapy, effects on the number of leukemia blasts or their sensitivity to chemotherapy. Lengthened remissions, possibly the least ambiguous test of the sensitization strategy, have been reported only by Büchner et al with GM-CSF. This is the only study in Table 1 in which patients received CSF before and during consolidation as well as induction, providing perhaps the fairest test of the sensitization approach. However, follow-up in this
study and the other studies in Table 1 is short, limiting the ability to reliably evaluate survival and especially remission duration.

There are several other pitfalls in interpreting the results of the studies in Table 1. First and simplest is the fact that the failure to demonstrate statistically significant differences between control and CSF groups does not mean that such differences do not exist or are not clinically important. Some of the studies in Table 1 have provided evidence of improved CR rates in CSF-treated patients, raising the possibility that the results of these studies might have been more convincing if more patients had been treated. Reporting a P value alone can be quite misleading. Rather, interpretation of the results would be aided by reporting confidence intervals for the true differences in rates between CSF and control groups, based on the observed data.

Ohno et al report CR rates of 11 of 30 (37%) in their control group and 14 of 28 (50%) in their G-CSF group (P = .308). Based on these data, a 95% confidence interval (CI) for the difference in the true CR rates is (−.120, .386). Although this interval contains 0, which implies that a .05 level test would accept the hypothesis of no difference, the evidence supporting this hypothesis is the same as that supporting the hypothesis that the true difference equals .264, because 0 and .264 are each equidistant from the confidence limits. With one exception, confidence intervals are not used to summarize the results of the trials in Table 1.

A second problem in interpreting these trials is the presence of unobserved ("latent") variables, such as supportive care practices, which may vary substantially between different trials, including those conducted sequentially at the same institution. Although regression methods have been used in some of these studies to adjust for differences between CSF and control groups in the distribution of known prognostic characteristics (eg, performance status for fatal infection rate, cytogenetics or antecedent hematologic disorder for CR, remission duration, or survival), these methods do not account for latent variables. Although randomization is the appropriate method for dealing with this problem, its value may be limited by small sample sizes, in which case a randomization scheme balanced within patient subgroups is more effective than simple randomization.

A more fundamental problem in interpreting the trials in Table 1 is the high level of heterogeneity among both patients and treatments. The clinical response to GM-CSF, G-CSF, or, by extension, other CSFs may depend on the precise conditions of CSF administration, the CSF used, the chemotherapy administered, other supportive care considerations, and, most of all, the specific characteristics of the patients treated. It appears likely that certain patient characteristics, such as resistance to infection, may be very different across trials as a consequence of study design rather than chance alone. For example, the current study of Ohno et al found no significant effect of G-CSF on rate of documented infection (P = .61, 95% CI for the difference in the true rates [−.186, .326]). This study thus disagrees with the earlier study of Ohno et al, in which the rate of documented infections was significantly reduced in the group receiving G-CSF after completion of chemotherapy compared with randomized controls (P = .03, 95% CI [.046, .496]). Despite the difference in timing of G-CSF administration in the two studies, the time to recovery to greater than 1,000 neutrophils was very similar (medians: 34 v 32 days for the controls, 22 v 25 days for the G-CSF groups). Given this, the differences in the effects of G-CSF on documented infections observed in the two studies very possibly reflects the markedly different incidences of febrile episodes in the two control groups: 32 of 50 (64%) in the earlier and 27 of 30 (90%) in the current study (P = .01, 95% CI [.089, .431]). In turn, this discrepancy may reflect the possibility that the earlier patients of Ohno et al were "selected" to be unlikely to develop infections because they had to survive the initial 8 to 12 days of chemotherapy before they were randomized, because randomization took place only when marrow aplasia was documented. In contrast, in the current study, patients were randomized before receiving chemotherapy. Similar considerations limit the validity of comparisons of CR and survival rates between the two studies.

As another example, an early study administering GM-CSF to elderly (median age, 62), newly diagnosed patients after completion of chemotherapy failed to detect improved rates of infection or survival, contrasting with the more recent study of Rowe et al administering GM-CSF or placebo to newly diagnosed patients aged 55 to 70 upon documentation of hypoplasia, with all patients randomized before treatment and included in the results. Again, the disagreements between the two studies may reflect different incidences of mortality from infection in the control groups (32% in the earlier versus 8% in the later study). However, here the differences in the controls may result from the different chemotherapies administered (the more toxic high-dose ara-C in the earlier but not the later study), or from the fact that patient entry criteria in the former study were based not only on age but also on poor performance status (33% of the patients) and abnormal organ function. This study, as well as a study using G-CSF by the same investigators, and the two studies of Ohno et al suggest that the degree of acceleration of neutrophil recovery by G-CSF or GM-CSF will not be sufficient to benefit the relatively old and infirm patient who is likely to die before earliest neutrophil recovery. In some hospitals, such patients may constitute a significant proportion of the AML population.

The issue of timing of CSF administration is particularly relevant with regard to a study of Estey et al. This is the only study to date to find that pretreatment with a CSF, specifically GM-CSF, lowered CR and survival rates. GM-CSF has been reported to protect clonogenic leukemic blasts from the in vitro cytotoxic effect of ara-C. The same has not been found for G-CSF. Because neither CSF protects blasts from thymidine suicide, the protective effect of GM-CSF has been hypothesized to occur through a non-cell-cycle-related mechanism, eg, through effects on ara-CTP formation or DNA repair. Given these in vitro findings, the lower CR rate observed in newly diagnosed patients treated with GM-CSF before and during daunorubicin and continuous-infusion high-dose ara-C, which contrasts with
the effect of G-CSF administered before, during, and after fludarabine + ara-C in newly diagnosed patients in a study from the same investigators, 17 could be caused by use of GM-CSF rather than G-CSF. However, the different CR and survival rates could also reflect the different durations of CSF administration before chemotherapy (1 day in the G-CSF study versus a median of 4 days in the GM-CSF study). This is especially true because in studies reporting that pre-treatment with GM-CSF does not affect CR rate in newly diagnosed AML, the growth factor was administered for only 1 to 2 days before chemotherapy. 14,15 Lengthening prechemotherapy administration of some CSFs may increase leukemia burden to an extent that cannot be compensated for by an increase in chemosensitivity. In turn, effects on chemosensitivity are likely to be different in different patients because of the great interpatient variability observed in the in vitro response of clonogenic blasts to many CSFs. This high level of heterogeneity, in addition to the highly variable outcomes of either newly diagnosed or relapsed patients receiving the same chemotherapy without CSFs, suggests that even consistent observations in clinical trials of CSFs plus chemotherapy may be inferred to apply only on average in the type of patients treated, and may not apply to patient populations seen elsewhere. Finally, it should be stressed that an effect on a clinical event such as remission duration, which may seem a relatively unambiguous measure of sensitization of blasts to chemotherapy by CSFs, may be a sum of many different biologic effects of CSFs, eg, on cell cycle, on pharmacokinetics of chemotherapeutic agents, or on apoptosis after the administration of chemotherapy. In any patient each of these parameters may be affected differently by CSFs.

The data reviewed here do not support the general use of GM-CSF or G-CSF outside the context of a clinical trial. Outside such trials, these CSFs should be used according to benefit/risk ratios in individual patients, eg, as suggested by Ohno et al 17 in their concluding paragraph. Currently, it appears that the clinical effect of these CSFs will, on average, be insufficient to materially improve prognosis. In addition to trials of new CSFs, CSF combinations, or perhaps CSF receptor antagonists, 24 there may also be a need for large, randomized trials of GM-CSF or G-CSF in a patient group representative of all patients with AML to more precisely estimate response rates in different patient groups. There is also a need to search for in vitro, or ex vivo, correlates of in vivo response to a given CSF plus a given chemotherapy so as to identify those patients likely to materially benefit, not benefit, or perhaps be harmed by prospective administration of that CSF with that chemotherapy.

REFERENCES
5. Yang GS, Minden MD, McCulloch EA: Influence of schedule on regulated sensitivity of AML blasts to cytosine arabinoside. Leukemia 7:1012, 1993


20. Therneau TM: How many stratification factors are “too many” to use in a randomization plan? Controlled Clin Trials 14:98, 1993


Use of colony-stimulating factors in the treatment of acute myeloid leukemia [editorial; comment]

EH Estey