A Double-Blind Placebo-Controlled Trial of Granulocyte-Stimulating Factor in Elderly Patients With Previously Untreated Acute Myeloid Leukemia: A Southwest Oncology Group Study (9031)

By John E. Godwin, Kenneth J. Kopecky, David R. Head, Cheryl L. Willman, Catherine P. Leith, Harry E. Hynes, Stanley P. Balcerzak, and Frederick R. Appelbaum

Older age is a poor prognosis factor in acute myeloid leukemia (AML). This double-blind trial was designed to test the hypothesis that granulocyte colony-stimulating factor (G-CSF) used as supportive care could improve the treatment of elderly AML patients. Two hundred thirty-four patients 55 or more years of age with a morphologic diagnosis of de novo or secondary AML, French-American-British (FAB) M0-M7, excluding M3, were randomly assigned to a standard induction regimen (daunorubicin at 45 mg/m² intravenously [IV] on days 1 through 3 and Ara-C at 200 mg/m² IV continuous infusion on days 1 through 7) plus either placebo or G-CSF (400 μg/m² IV over 30 minutes once daily). Results are reported here for 211 centrally confirmed cases of non-M3 AML. The two groups were well balanced in demographic, clinical, and hematological parameters, with median ages of 68 years in the G-CSF and 67 years in the placebo groups. The complete response (CR) rate was not significantly better in the G-CSF group: 50% in the placebo and 41% in the G-CSF group (one-tailed P = .89). Median overall survival was also similar, 9 months (95% confidence interval [CI], 7 to 10 months) in the placebo and 6 months (95% CI, 3 to 8 months) in the G-CSF arms (P = .71). We found a significant 15% reduction in the time to neutrophil recovery in the G-CSF group (P = .014). G-CSF had no impact on recovery from thrombocytopenia (P = .80) or duration of first hospitalization (P = .27). When infection complications were evaluated, G-CSF had a beneficial effect on the duration but not on incidence of infection. G-CSF patients had fewer days with fever and shorter duration of antibiotic use. However, there was no difference in the frequency of total documented infections or in the number of fatal infections (15% placebo vs 26% G-CSF). In this study of elderly AML patients, G-CSF improved clinical parameters of duration of neutropenia and antibiotic use, but did not change CR rate or survival or shorten hospitalization.

© 1998 by The American Society of Hematology.
colony-stimulating factor (r-metHuG-CSF) administered after standard induction chemotherapy-induced marrow aplasia in elderly AML patients. Our goals were to determine the safety of this treatment and its impact on the complete response rates and duration of survival in patients 56 years of age or older. We also examined, as secondary endpoints, the duration of neutropenia and thrombocytopenia, the total number of febrile days, the number of days of antibiotic therapy, the number and type of infection episodes, and the number of hospital days.

PATIENTS AND METHODS

Eligibility

Eligibility was limited to patients 56 years of age or older with AML, French-American-British (FAB) class M2-M5, that was morphologically confirmed by bone marrow aspiration, biopsy, and cytological staining.20 Patients with acute promyelocytic leukemia, AML Mx, and blastic transformation of chronic myelogenous leukemia were excluded. Patients who received prior chemotherapy for acute leukemia were not eligible, although administration of hydroxyurea to control high cell counts was permitted. Patients with AML after a previous diagnosis of myelodysplastic syndrome (MDS) or after prior chemotherapy or radiotherapy were eligible. A history of prior treatment of MDS with low-dose cytosine arabinoside was permitted, but 30 days must have elapsed from prior treatment and all toxicities must be resolved. Patients who received prior treatment with erythropoietin, G-CSF, granulocyte-macrophage colony-stimulating factor (GM-CSF), or other myeloid growth factors were eligible. Patients had normal liver, renal, and cardiac function as indicated by (1) a bilirubin less than two times the institutional upper limits of normal, (2) serum creatinine less than two times the institutional upper limits of normal, and (3) an ejection fraction ≥50% as measured by mutigated cardiac blood pool (MUGA) scan. Patients with unstable cardiac arrhythmia or unstable angina and pregnant or lactating women were not eligible.

Study Design

This was a double-blind placebo controlled study. After granting informed consent, patients were randomly assigned to one of two remission induction treatment arms: cytosine arabinoside (Ara-C) and daunorubicin plus placebo or Ara-C and daunorubicin plus G-CSF. The randomization was stratified by age (56 to 64 v 65+ years of age) and by onset of leukemia (secondary v de novo) and dynamically balanced using the method of Pocock and Simon21 to assure nearly equal numbers of patients within levels of the stratifying factors. All patients received induction chemotherapy, with daunorubicin at 45 mg/m² intravenously (IV) on days 1 through 3 and Ara-C at 200 mg/m²/d as a continuous IV infusion on days 1 through 7. On day 10, a bone marrow biopsy was performed. If the day-10 marrow was hypocellular with blasts less than 5%, either G-CSF or placebo treatment was begun on day 11. Patients randomized to G-CSF received Escherichia coli-derived recombinant human G-CSF (r-metHuG-CSF; Neupogen; Filgrastim; Amgen, Inc, Thousand Oaks, CA) at 400 µg/m² IV over 30 minutes once daily. The G-CSF was continued until the absolute neutrophil count (ANC) was 1,000/µL, and then tapered over 3 days. If the day-10 marrow showed blasts ≥5%, a marrow biopsy was repeated on day 14. If the day-14 marrow showed residual leukemia, a second induction using the identical chemotherapy was administered. After the second induction, patients received either G-CSF or placebo, depending on initial randomization, when the bone marrow blasts were less than 5%. If leukemic regrowth occurred after a second induction, the patient was removed from protocol treatment. Leukemic regrowth was defined as a marrow with 30% blasts after a prior marrow with less than 5% blasts. After induction therapy, patients who achieved CR were registered to receive two courses of postremission therapy consisting of daunorubicin at 30 mg/m² IV on days 1 and 2 and Ara-C at 200 mg/m²/d as a continuous IV infusion on days 1 through 7, and patients received the same treatment assignment, G-CSF or placebo, during postremission therapy starting on day 8 after chemotherapy.

Guidelines for supportive care stipulated that neutropenic patients with a temperature of 38°C on more than one occasion within a single (24 hours) day or a single temperature greater than 38.5°C should be treated with empiric antibiotics, including an aminoglycoside and semisynthetic penicillin or cephalosporin. Fungal prophylaxis was determined by local practice, and patients with documented fungal infections or fever unresponsive to broad spectrum antibiotics after 72 hours were treated with amphotericin B. Platelet transfusions were administered for bleeding manifestations and invasive procedures, and prophylactically when the platelet count was less than 20,000/µL.

Definitions of Outcomes

Response was evaluated according to Southwest Oncology Group (SWOG) criteria, as modified from National Cancer Institute guidelines.20 CR was defined as a marrow with greater than 20% cellularity, with maturation of all cell lines, less than 5% blasts, and no Auer rods; peripheral blood with neutrophils 1,500/µL, platelets greater than 100,000/µL, and no leukemic blasts; and no extramedullary disease. Overall survival (OS) was measured from the day of randomization until death from any cause, with observation censored for patients last known alive. Relapse-free survival (RFS) was measured from the date CR was established until relapse or death from any cause with observation censored for patients last known alive without report of relapse.

Treatment failures. Patients who failed to achieve CR after induction were classified according to the type of failure (resistant disease, death during aplasia, or indeterminate).

Toxicity criteria. The criteria used to determine severity of toxicity were those of the SWOG.22

Neutropenia. The duration of neutropenia was defined as the number of days from the start of chemotherapy until the first day that the ANC was ≥500/µL for 2 consecutive days.

Thrombocytopenia. The duration of thrombocytopenia was defined as the number days from the start of chemotherapy until the first day that the platelet count remained ≥30,000/µL, and was sustained without transfusion for 7 consecutive days or more.

Febrile days. The number of febrile days was defined as the total number of days of fever (oral or core body temperature > 38°C) from the start of chemotherapy until first hospital discharge.

Antibiotic days. The total number of days of IV antibiotic and/or antifungal therapy from the start of chemotherapy until first hospital discharge was defined as antibiotic days. In addition, any use of amphotericin B was tabulated as administered or not administered (see Table 3).

Numbers of infection. Each infection was defined by the first occurrence with a culture documented pathogen and a clinically evident focus. Any culture documented infection, including oral thrush, or herpes viral infection was included in the total number of infections (see Table 3). Fever of unknown source did not constitute an infection. The presence or absence of a documented fungal infection other than oral thrush was also separately tabulated. In addition, the occurrence of pneumonia (defined as chest infiltrates and clinical diagnosis) or documented sepsis (bacteremia or fungemia) was recorded as present or absent.

Hospital days. The number of hospital days was calculated from the start of chemotherapy until first hospital discharge.

Statistical Methods

This study design called for randomization of 182 evaluable patients (91 per arm), providing statistical power of 82% to detect an increase in
the CR rate from 40% to 60% (one-tailed test at critical level of α = .05). This number of patients, accrued over 2 years and with 1 additional year of follow-up, would also provide 82% power to detect a hazard ratio (placebo:G-CSF) of 1.5 in the analysis of OS, assuming a median OS of placebo of 7 months for 56 to 64 years of age (45% of patients) and 1.4 months for 65+ years of age (55%). Demographic and clinical data for patients in this study were collected with quality control review according to standard procedures of the SWOG. This study was monitored throughout its accrual and follow-up phases by a Data Monitoring Committee of which the investigators were not members. One preplanned interim analysis was performed, but its results did not require early termination of the study.

Treatment comparisons were based on assigned treatments, ie, G-CSF patients who did not in fact receive G-CSF were retained in the G-CSF arm. CR rates were analyzed using logistic regression models.25 Distributions of OS and RFS were estimated by the method of Kaplan and Meier.24 The effects of treatment, patient, and disease characteristics on OS and RFS were analyzed using the proportional hazards regression model of Cox.25 Because neutropenia or thrombocytopenia may increase the risk of death, the primary statistical analyses of neutrophil and platelet recovery were not based on simple Kaplan-Meier estimates. It is well known that, when the outcome of interest (eg, neutrophil recovery) is subject to censoring by a nonindependent competing outcome (eg, death), the simple Kaplan-Meier estimate is biased, and effects of a covariate, such as treatment assignment, on the two outcomes are nonidentifiable.26 The primary comparisons of endpoints for neutrophil and platelet recovery were based on the procedure described by Lin et al27 to control for dependent censoring. If OS is similar in the two treatment arms, the problem of dependent censoring may have little impact on the comparison between arms. For this reason and for the comparability with prior studies that did not account for dependent censoring, Kaplan-Meier plots were also calculated for time to neutrophil and thrombocyte recovery. Numbers of febrile days, antibiotic days, numbers of infection, and hospital days were compared using the Wilcoxon rank sum test. Quantitative factors were treated as continuous variables in regression analyses, but grouped for time to neutrophil and platelet recovery. Numbers of infection and hospital days were compared using the Wilcoxon rank sum test. Quantitative factors were treated as continuous variables in regression analyses, but grouped when necessary for descriptive tables or figures. In all analyses, statistical significance of differences between treatment arms is expressed in terms of one-tailed P values for improvement with G-CSF. All other test results are reported as two-tailed P values. Results are based on data available September 23, 1996.

**RESULTS**

**Patient and Disease Characteristics**

A total of 234 patients (116 randomized to G-CSF and 118 to placebo) were registered on SWOG-9031 from 68 institutions between January 1992 and February 1994. Diagnoses of AML (non-M3) were confirmed by central morphology review for 211 patients (90%): 106 G-CSF and 105 placebo. The other 23 patients included 10 with central review diagnoses other than non-M3 AML: 4 RAEB, 1 M4 AML, 2 M5/M6, and 3 with too few blasts to meet the protocol definition of AML. For the other 13 patients, adequate materials were not submitted for eligibility review. The following analyses are based on intent to treat and include the 211 patients with centrally confirmed diagnoses of AML. Nine of these 211 cases failed to meet other eligibility criteria. Results from all 234 registered patients or the 202 fully eligible patients were separately analyzed and found to be similar to these results. Infection complications, duration of hospitalization, and hematologic recovery analysis are based on the numbers of patients evaluable for these outcomes (see below). The 211 centrally con-

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Placebo (N = 105)</th>
<th>G-CSF (N = 106)</th>
<th>Total (N = 211)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (yr)</td>
<td>67</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Age 56-64</td>
<td>41 (39%)</td>
<td>37 (35%)</td>
<td>78 (37%)</td>
</tr>
<tr>
<td>Age 65+</td>
<td>64 (61%)</td>
<td>69 (65%)</td>
<td>133 (63%)</td>
</tr>
<tr>
<td>Range</td>
<td>56-84</td>
<td>56-88</td>
<td>56-88</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>66 (63%)</td>
<td>56 (53%)</td>
<td>122 (58%)</td>
</tr>
<tr>
<td>Female</td>
<td>39 (37%)</td>
<td>50 (47%)</td>
<td>89 (42%)</td>
</tr>
<tr>
<td>Disease status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>24 (23%)</td>
<td>26 (25%)</td>
<td>50 (24%)</td>
</tr>
<tr>
<td>De novo</td>
<td>81 (77%)</td>
<td>80 (75%)</td>
<td>161 (76%)</td>
</tr>
<tr>
<td>Performance status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>76 (72%)</td>
<td>73 (69%)</td>
<td>149 (71%)</td>
</tr>
<tr>
<td>2-3</td>
<td>29 (28%)</td>
<td>33 (31%)</td>
<td>62 (29%)</td>
</tr>
<tr>
<td>WBC count (&gt;10³/ ml), median (range)</td>
<td>13.4 (1.0-274)</td>
<td>14.3 (0.6-298)</td>
<td>13.5 (0.6-298)</td>
</tr>
<tr>
<td>Peripheral blasts (&gt;10³/mL), median (range)</td>
<td>4.1 (0-299)</td>
<td>1.8 (0.292)</td>
<td>3.1 (0-292)</td>
</tr>
<tr>
<td>Platelet count (&gt;10³/ ml), median (range)</td>
<td>59 (11-317)</td>
<td>59 (3-537)</td>
<td>59 (3-537)</td>
</tr>
<tr>
<td>Cytogenetic status*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable</td>
<td>5 (5%)</td>
<td>4 (4%)</td>
<td>9 (4%)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>53 (50%)</td>
<td>50 (47%)</td>
<td>103 (49%)</td>
</tr>
<tr>
<td>Unfavorable</td>
<td>26 (25%)</td>
<td>26 (25%)</td>
<td>52 (25%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>21 (20%)</td>
<td>26 (25%)</td>
<td>47 (22%)</td>
</tr>
<tr>
<td>MDR1 expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bright</td>
<td>37 (39%)</td>
<td>32 (34%)</td>
<td>69 (37%)</td>
</tr>
<tr>
<td>Moderate</td>
<td>18 (19%)</td>
<td>15 (16%)</td>
<td>33 (17%)</td>
</tr>
<tr>
<td>Dim</td>
<td>16 (17%)</td>
<td>17 (18%)</td>
<td>33 (17%)</td>
</tr>
<tr>
<td>Negative</td>
<td>25 (26%)</td>
<td>29 (31%)</td>
<td>54 (29%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>FAB classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>12 (11%)</td>
<td>12 (11%)</td>
<td>24 (11%)</td>
</tr>
<tr>
<td>M1</td>
<td>31 (30%)</td>
<td>27 (25%)</td>
<td>58 (27%)</td>
</tr>
<tr>
<td>M2</td>
<td>31 (30%)</td>
<td>39 (37%)</td>
<td>70 (33%)</td>
</tr>
<tr>
<td>M4</td>
<td>10 (10%)</td>
<td>11 (10%)</td>
<td>21 (10%)</td>
</tr>
<tr>
<td>M5</td>
<td>9 (9%)</td>
<td>10 (9%)</td>
<td>19 (9%)</td>
</tr>
<tr>
<td>M6</td>
<td>7 (7%)</td>
<td>1 (1%)</td>
<td>8 (4%)</td>
</tr>
<tr>
<td>M7</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Myeloid, NOS</td>
<td>4 (4%)</td>
<td>6 (6%)</td>
<td>9 (4%)</td>
</tr>
</tbody>
</table>

*Favorable: t(8;21), inv(16), t(16;16), +14. Unfavorable: complex, >3 abnl, inv(3q), -5/5q-, -7/7q-, 11q or 17p, del(20q),dmins/hsrs, +13, t(9;22).†MDR1 expression was evaluated by the MDR1 specific antibody MRK16 and quantified by the KS statistic.
the G-CSF arm), and prior exposure to myeloid growth factors was reported for 4 patients (1 G-CSF and 3 placebo).

**Treatment**

Three of the 211 patients (1 G-CSF and 2 placebo) did not receive induction therapy according to the protocol: 1 patient refused, 1 received doxorubicin rather than daunorubicin, and 1 received induction therapy for ALL. Including these 3, 51 patients (24%) did not receive blinded drug: 10 (6 G-CSF and 4 placebo) died before the earliest possible start of drug and 41 others (22 G-CSF and 19 placebo) never received the study drug for other reasons, most frequently because they had residual blasts in the marrow and never met the treatment requirement of a marrow with less than 5% blasts (15 in each arm). Three other patients, all randomized to G-CSF, received open label G-CSF rather than blinded drug. Among the 160 patients (78 G-CSF and 82 placebo) who received blinded drug or open label G-CSF, 145 (91%) completed all possible treatment: 116 (57 G-CSF and 59 placebo) were treated until ANC recovery and 29 (18 G-CSF and 11 placebo) died or had progression of leukemia while on blinded drug. The remaining 15 patients stopped receiving blinded drug early for a variety of reasons. Reasons identified for more than 1 patient included prolonged neutropenia (1 G-CSF and 3 placebo), patient refusal (2 placebo), and decision to attempt other induction therapy (2 placebo).

**Response to Induction Therapy**

Ninety-five (45%) of the 211 patients achieved CR. Eighty-six patients (41%) achieved CR after the first induction attempt. Of the remaining 125 patients, 48 received the second induction attempt, and 9 of these 48 patients (19%) achieved CR. Of the 116 patients who did not achieve CR, the responses of 5 patients were not evaluated: these included 3 patients who did not receive induction therapy and 2 who were removed from study by their institutions. The remaining 111 patients who did not achieve CR had treatment failure classified according to standard criteria as follows: 73 (35% of the 211) due to resistant disease; 19 (9%) due to complications of aplasia; and 19 (9%) due to treatment failure of indeterminate cause.20

Table 2 summarizes CR rates by the stratification factors, AML onset and age, and treatment assignment. The CR rate was not significantly better in the G-CSF arm (41% [43/106] v 38% [14/37] for placebo; one-tailed \(P = .09\), with adjustment for stratification). The estimated regression coefficient for the effect of G-CSF relative to placebo was \(-0.36\) (95% confidence interval [CI], \(-0.92\) to 0.21). In univariate analyses, the CR rate did not vary significantly between age strata 65 years of age or older versus 56 to 64 years of age (two-tailed \(P = .59\)), but did differ between the two AML onset groups (two-tailed \(P = .0005\)). The CR rate was significantly lower for patients with secondary AML 24% (12/50) compared with de novo 52% (83/161). The CR rate appeared particularly low among secondary AML patients in the G-CSF arm (12% [3/26]; Table 2). However, estimation of the CR rate based on this small number of patients is imprecise: the corresponding 95% CI is wide (2% to 30%) and substantially overlaps that for the CR rate of 38% observed in the placebo patients (95% CI, 19% to 59%). In a multivariate analysis of response (reported by Leith et al28 ), three independent prognostic factors were identified: secondary AML, cytogenetic status, and MDR1 expression. The CR rate was significantly worse for patients with secondary AML (two-tailed \(P = .0035\)) or unfavorable cytogenetic status (two-tailed \(P = .0031\)). MDR1 expression was evaluated by the MDR1 specific antibody MRK16 and quantified by the Kolmogorov-Smirnov (KS) statistic. The CR rate decreased with increasing expression of MDR1:MRK16 negative (67% CR); dim (45% CR); and bright/moderate (34% CR) (two-tailed \(P = .0041\)). (For more details, see Leith et al.28) After accounting for these effects, none of the other factors considered retained significant independent prognostic association with CR, including age (treated as a continuous variable, two-tailed \(P = .51\)). There was no evidence of higher CR rate in the G-CSF arm in the multivariate analysis (one-tailed \(P = .97\)).

**OS and RFS**

Of the 211 patients, 180 have died. The remaining 31 patients were last known to be alive between 18 and 53 months (median, 33 months) after entering the study. Figure 1 shows a Kaplan-Meier plot of the OS experience for all patients by randomization arm. There was no significant difference in survival between the placebo and G-CSF arms (one-tailed \(P = .71\)). The estimated relative risk of death (G-CSF relative to placebo) was 1.09 (95% CI, 0.81 to 1.46). The median survivals were 9 months (95% CI, 7 to 10 months) in the placebo and 6 months (95% CI, 3 to 8 months) in the G-CSF arm. Comparing survival by disease onset, the median OS was 8 months (95% CI, 6 to 9

---

**Table 2. CR in Elderly AML by Age, Treatment, and Disease Status**

<table>
<thead>
<tr>
<th>Disease Status</th>
<th>Placebo</th>
<th>G-CSF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>De novo</td>
<td>53%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Secondary</td>
<td>38%</td>
<td>40%</td>
<td>40%</td>
</tr>
</tbody>
</table>

---

**Fig 1.** Kaplan-Meier estimates of the distributions of survival from day of study entry, by treatment arm, based on 211 patients with centrally confirmed diagnoses of non-M2 AML. Tickmarks indicate censored data.
months) for patients with de novo AML and 7 months (95% CI, 3 to 8 months) for those with secondary AML. Despite these similar median survival times, OS appeared markedly poorer among patients with secondary AML by 12 to 18 months (Fig 2). Consequently, in univariate analysis, OS was significantly associated with disease onset (two-tailed \( P = .030 \)) as well as age (treated as a continuous variable \( P = .0003 \)). Although secondary AML was significantly associated with a lower CR rate and with shorter OS in the univariate model, AML onset was not an independent prognostic factor for OS. In the multivariate analyses reported by Leith et al.,\(^{28}\) three factors were found to have independent prognostic significance for OS. OS was significantly worse for patients with unfavorable cytogenetics (\( P < .0001 \)) and with increasing age (\( P = .014 \)) and increasing white blood cell count (\( P = .029 \)). After accounting for these effects, none of the other factors considered in this study was significantly associated with OS, including AML onset (two-tailed \( P = .29 \)) and treatment assignment (one-tailed \( P = .80 \)).

Of the 95 patients who achieved CR (52 placebo and 43 G-CSF), 41 of 52 (79%) and 33 of 43 (77%) of the placebo and G-CSF arms, respectively, received protocol postremission therapy. Reasons for failure to receive postremission therapy on study included impaired cardiac function (7), other medical reasons (7), correct or mistaken diagnosis of progression of AML (6), and patient refusal (1). Of the patients who achieved CR, 77 have relapsed (44 placebo and 33 G-CSF), and 5 others (2 placebo and 3 G-CSF) have died without report of relapse, all due to consolidation toxicities: infection, CNS hemorrhage, and complications of surgery. RFS was not significantly better in the G-CSF arm (one-tailed \( P = .38 \)). The estimated relative risk of relapse or death (G-CSF relative to placebo) was 0.93 (95% CI, 0.59 to 1.47). The median RFS was 9 months for the placebo group (95% CI, 7 to 10 months) and 8 months for the G-CSF group (95% CI, 4 to 10 months; Fig 3). In multivariate analyses, RFS was significantly worse for patients with unfavorable cytogenetics (\( P = .028 \)), but was not significantly associated with any other factors, including AML onset (\( P = .42 \)), age (\( P = .75 \)), or treatment assignment (\( P = .38 \)).

Overall Toxicity and Infection

Toxicity of induction therapy was evaluated in 207 patients (104 G-CSF and 103 placebo). The remaining 4 include 3 who did not receive protocol induction therapy and 1 who was removed from study due to pre-existing insufficient cardiac function. Nonhematologic toxicities were comparable in the two treatment groups. Bone pain was reported for only 1 G-CSF patient, compared with 5 placebo patients. Fatal induction toxicities occurred in 20% (21/104) of the G-CSF group and 19% (20/103) of the placebo arm. There was no apparent increase in leukemic relapse in the G-CSF arm, because, as indicated above, the RFS did not differ by treatment assignment.

To compare the infectious complications, several outcomes were measured during the first hospitalization: the number of febrile days; the duration of IV antibiotic therapy; the total number of documented infectious episodes; the proportion of patients with fungal infections other than oral Candida; the proportion requiring Amphotericin-B; the proportion with pneumonia; the proportion with positive blood cultures; and the frequency of infection-related deaths (Table 3). The number of culture confirmed infections during the first hospitalization was evaluated for 206 patients (103 G-CSF and 103 placebo). Twenty-eight (27%) of the G-CSF patients had no documented infections during their first hospitalization, whereas the other 75 (73%) had a total of 163 such infections. In comparison, 37 (36%) of the placebo patients had no documented infections, and the remaining 66 (64%) had a total of 141. The distribution of number of infections per patient was not significantly lower on the G-CSF arm (\( P = .82 \)), and there was no evidence that G-CSF was associated with a reduced proportion of patients with fungal infection, pneumonia, positive blood cultures; or treatment with Amphotericin-B (Table 3). Accounting for multiple infections and varying lengths of hospitalization, it was noted that the average rate of documented infection in the G-CSF arm (163 in 3,285 patient-days, or 1.5 per 30 patient-days) was not significantly different than the rate in the placebo.
Infection was the most frequent cause of induction death. However, the risk of fatal infections was not significantly lower in the G-CSF arm, with 20 (19%) infection deaths, compared with 14 (14%) in the placebo arm \((P = .90)\).

However, some improvement was seen in the measures of infection duration in the G-CSF arm. The number of febrile days during the first hospitalization tended to be shorter for the G-CSF patients, with a median of 8 days (range, 0 to 79 days), compared with 10 days (range, 0 to 34 days) for placebo patients (one-tailed \(P = .091\)), and the number of days on IV antibiotics was decreased in the G-CSF arm, 22 days (range, 0 to 128 days), compared with 26 days (range, 0 to 69 days) for placebo patients, a marginally significant improvement (one-tailed \(P = .053\); Table 3).

**Hematologic Recovery and Hospital Duration**

The duration of neutropenia in the first induction course was evaluated for 207 patients (104 G-CSF and 103 placebo). Neutrophil recovery was observed in 149 patients (75 G-CSF and 74 placebo). Another 47 patients (25 G-CSF and 22 placebo) died or relapsed without neutrophil recovery. For the remaining 11 patients, hematologic follow-up was incomplete and observation was therefore censored at the date of each patient’s last reported neutrophil count. On average, the duration of neutropenia was 15% shorter with G-CSF, which is a significant decrease \((P = .014)\). For comparison with results of previous studies, Kaplan-Meier curves were calculated for time to neutrophil recovery. For reasons previously discussed (see Statistical Methods), these curves should not be viewed as unbiased estimates of the distributions of time to neutrophil recovery. Nevertheless, it is apparent from Fig 4 that the curves from our study are similar to those reported previously. In particular, the median, ie, the point at which the curve reaches 50%, was 24 days for the G-CSF patients, 3 days less than the median for the placebo group. The fact that our analysis, which is not based on the problematic assumptions underlying the interpretation of the Kaplan-Meier curves and corresponding log rank tests, found an effect consistent with those reported earlier lends credence to the previous findings.

Recovery from thrombocytopenia was observed in 137 patients (64 G-CSF and 73 placebo), whereas another 52 (29 G-CSF and 23 placebo) died without platelet recovery. Observation was censored for the remaining 18 patients. There was no significant difference in the duration of thrombocytopenia by treatment assignment \((P = .80); Fig 5\). On average, duration of thrombocytopenia was 7% longer with G-CSF; however, the 95% CI for this difference included both longer and shorter
durations of thrombocytopenia, ranging from 12% shorter to 33% longer.

The duration of hospitalization was evaluated for 207 patients (103 in G-CSF and 104 placebo). Fifty-two patients died during their first hospitalization (28 G-CSF and 24 placebo), and the remaining 155 were discharged alive. The median length of the first hospitalization, whether the patient was discharged alive or died in the hospital, was 29 days in each treatment arm (range, 4 to 155 days for G-CSF and 3 to 106 days for placebo). There was no difference in the length of hospital stay in the G-CSF arm compared with placebo (one-tailed \( P = .27 \)). It must be stressed that there was no uniform protocol for discharge criteria and that each investigation site had different capabilities to provide G-CSF treatment or transfusions in the outpatient setting.

**DISCUSSION**

This study demonstrates that G-CSF administered during induction therapy for untreated AML in elderly patients, 56 years of age or greater, significantly shortens the duration of neutropenia but does not improve the rate of CR, OS, or disease-free survival. The use of G-CSF in this elderly population reduced the duration of infection but not its incidence, as evidenced by the decrease in febrile days and days on intravenous antibiotics, with no difference in the number or type of documented infections or the number of fatal infections. The use of G-CSF as a supportive care measure in this population was not associated with an increase in toxicity or obvious evidence of leukemic regrowth.

Our overall CR rate of 52% in the de novo AML group is similar to recent results in other randomized trials of myeloid growth factors in elderly AML patients. The study design differs from these trials in that we included secondary AML patients. This group made up approximately one fourth of our study population (24%; Table 1). Secondary AML is commonly seen in the elderly, and our study population therefore may more closely reflect the overall population pattern of AML in the elderly patient. Secondary AML patients are known to have a significantly worse prognosis, and our study confirms this observation, with only a 24% CR rate in this group. Multivariate analysis of these data showed that secondary AML, unfavorable cytogenetics, and increasing MDR expression are independently associated with decrease in CR. Thus disease characteristics play a predominant role in achieving CR in a population already consisting of elderly AML patients. Evaluating this multiple logistic regression analysis of CR rates, a referee has suggested that the one-tailed \( P \) value of .97 observed in the G-CSF arm, if converted to a two-sided test with the resulting \( P = .064 \), might be viewed as marginally significant evidence of a decreased CR rate with G-CSF. However, such a result must be interpreted with particular caution, because it was obtained in an exploratory post hoc analysis. The evidence for such a detrimental effect of the G-CSF is not at all significant in the stratified comparison for which this study was designed: the one-tailed \( P \) value of .89 obtained for this designed comparison corresponds to a stratified analysis two-tailed \( P = .21 \).

The survival duration and RFS were similar in our trial to other recent AML studies in the elderly and remain significantly worse than those seen in patients less than 50 years of age. Despite the negative effect on CR rate, secondary AML status was not an independent predictor of OS in our multivariate analysis. In this analysis, the disease factor unfavorable cytogenetics was associated with decreased survival; in addition, age emerged as an independent poor prognosis factor for survival.

The effect of G-CSF on myeloid recovery in our study is notably consistent with that reported in other recent randomized trials of G-CSF in AML patients. We estimated a 15% reduction (95% CI, 3% to 27%) in the average time to neutrophil recovery with G-CSF, using an analysis that does not rely on a questionable assumption of independent censoring. Other investigators have reported the effect on neutrophil recovery in terms of changes in median time to recovery, as estimated using the product-limit method of Kaplan and Meier. Reported reductions in the estimated median times to recovery from neutropenia with G-CSF have included 6 days in elderly AML (Dombret et al), 5 days in adult AML patients (Heil et al), and 8 days in relapsed or refractory adult AML patients (Ohno et al). Our results lend credence to the previous claims that G-CSF reduces the duration of neutropenia, because we obtained results similar to those in the earlier reports if we use the same analytic methods. The Kaplan-Meier estimate of the time to median recovery was 3 days shorter with G-CSF compared with placebo in our study. This consistency of results in different studies was seen despite the fact that the proportion of patients who actually received drug differed in each study and the timing of the initiation of G-CSF was different in each study: day 8 in the study of Heil et al, day 9 in the study of Dombret et al, day 10 to 14 in the study of Ohno et al, and day 11 to 15 in our own study. The present study methodology is similar to the CALGB study reported by Stone et al in that patients were randomized to receive growth factor before starting induction chemotherapy. Therefore, some patients did not receive blinded drug, which began no sooner than day 8 for the CALGB and day 11 for the present study. In the CALGB study, 27 to 388 (7%) did not receive blinded drug; the corresponding figure is 51 to 211 (24%) for the present study. In the studies of Ohno et al, Dombret et al, and Heil et al, patients were randomized to growth factor shortly before starting drug treatment and thus all patients received the drug. However, there is no evidence that the timing of randomization influenced the treatment comparisons of myeloid recovery in the present study. For example, we can attempt to simulate the effect of delaying randomization until the start of blinded drug by considering only the 160 patients who actually received blinded drug. This of course risks the introduction of bias from the uncontrolled effects that determined whether patients actually received blinded drugs. However, if we ignore this possible bias, the Kaplan-Meier estimate of the median time to neutrophil recovery was 4 days shorter in the G-CSF arm, essentially the same as the difference of 3 days based on all 211 eligible patients.

Two prior randomized studies report no difference in the overall incidence of infection in AML with the use of G-CSF as supportive care after chemotherapy. However, the duration of infection as measured by the surrogates of fever and IV antibiotic use were decreased in the study by Heil et al as well as by our own study. Duration of infection was not evaluated in the trial by Dombret et al. These data are consistent with the
known myeloablative effect of AML therapy and the period of absolute neutropenia during which the majority of these infections occur. Myeloid growth factors cannot reverse absolute neutropenia in AML, but they can shorten its duration. Another variable to consider when comparing different myeloid growth factor trials, and especially when considering the possible economic impact of the study, is the dose of growth factor. Our study used a relatively high dose of G-CSF (400 µg/m² IV), corresponding roughly to 10 to 12 µg/kg. In the studies of Dombret et al and Heil et al, G-CSF was administered at a lower dose (5 µg/kg), with similar results regarding acceleration of myeloid recovery. Prior studies and our own report indicate no effect on recovery from thrombocytopenia with the use of G-CSF. These studies as well as our own suggest no increase in overall toxicity or leukemia regrowth.

There is no significant qualitative difference in these results when GM-CSF is studied in AML. In the studies reported by Stone et al and Rowe et al in elderly patients with AML, there was no difference in CR rate or overall incidence of infection with the use of GM-CSF. Whether there is any effect with the use of either G-CSF or GM-CSF on particular subgroups of patients with infectious complications in AML is still unclear. Rowe et al reported a decrease in pneumonia associated death in the GM-CSF arm compared with placebo. Heil et al reported a reduction in the total number of patients requiring amphotericin B use in the G-CSF arm but no difference in the total incidence of fungal infections. The present study shows no difference between G-CSF and placebo arms in the frequency of pneumonia, documented fungal infections, amphotericin B use, frequency of septicemia, or infection-related mortality.

Of the recently published randomized trials of myeloid growth factors in elderly AML, only the study reported by Rowe et al showed a difference in overall median survival with the use of yeast-derived GM-CSF. In that study, the observed median survival of the placebo arm was shorter than that seen in all the other trials of elderly AML; the median OS survival give by Rowe et al in the placebo arm was 4.8 months; that of the placebo arm in Stone et al was about 11 months; that of the placebo arm in Dombret et al was approximately 7.5 months; and our report indicates a median OS in the placebo arm of 9 months. The survival in the GM-CSF–treated arm reported by Rowe et al was similar to growth factor treated arms in the other randomized trials. The 7+3 chemotherapy induction regimen in the trial reported by Rowe et al administered daunorubicin at a dose of 60 mg/m², whereas the other trials in the elderly, including our own, used daunorubicin at 45 mg/m². It is unclear what role this difference in treatment played in the demonstrated effect of GM-CSF on median survival.

Our study was designed to have statistical power at a level conventionally sought in phase III cancer studies, i.e., at least 80% power to detect the design alternative hypothesis of a difference of 20 percentage points in the response rate or a hazard ratio of 1.5 in the analysis of survival. Therefore, this study cannot by itself conclusively rule out the possibility that G-CSF is associated with improvements in response or survival that, although smaller than the design alternatives, might nevertheless be of clinical interest. However, the accumulating evidence from this and the other reported trials of G-CSF and GM-CSF in AML suggest that the myeloid growth factors are not strongly beneficial with regard to the endpoints of CR rate or survival when administered as supportive care.

In summary, our results indicate that the use of G-CSF after induction chemotherapy is safe and can shorten myeloid recovery in the elderly patient with AML, but does not increase the CR rate or confer a survival benefit. The use of G-CSF or other myeloid growth factor as supportive care in elderly AML can positively affect infection duration and related outcomes. The use of G-CSF may impact the costs of AML treatment, because some measures of infection duration are significantly shortened and this could affect the economics of antibiotic use or hospitalization. However, although these are positive clinical benefits, the recommendation for the routine use of myeloid growth factors in AML therapy should await the outcome of careful analyses of the economic impact of this therapy in AML.

ACKNOWLEDGMENT

The authors are indebted to the efforts of all the institutions who participated in the trial and their data management teams; to Dr D.Y. Lin for his insightful suggestions; and to the personnel in the Statistical Center and Operations Office of the SWOG who made this study possible.

REFERENCES

A Double-Blind Placebo-Controlled Trial of Granulocyte Colony-Stimulating Factor in Elderly Patients With Previously Untreated Acute Myeloid Leukemia: A Southwest Oncology Group Study (9031)

John E. Godwin, Kenneth J. Kopecky, David R. Head, Cheryl L. Willman, Catherine P. Leith, Harry E. Hynes, Stanley P. Balcerzak and Frederick R. Appelbaum