Prognostic Discrimination in “Good-Risk” Chronic Granulocytic Leukemia

By Joseph E. Sokal, Edwin B. Cox, Michele Baccarani, Sante Tura, German A. Gomez, Judith E. Robertson, C. Y. Tso, Thomas J. Braun, Bayard D. Clarkson, Francisco Cervantes, Ciril Rozman, and the Italian Cooperative CML Study Group

The prognostic significance of disease features recorded at the time of diagnosis was examined among 813 patients with Philadelphia chromosome-positive, nonblastic chronic granulocytic leukemia (CGL) collected from six European and American series. The survival pattern for this population was typical of “good-risk” patients, and median survival was 47 mo. There were multiple interrelationships among different disease features, which led to highly significant correlations with survival for some that had no primary prognostic significance, such as hematocrit. Multi-variable regression analysis indicated that spleen size and the percentage of circulating blasts were the most important prognostic indicators. These features, and age, behaved as continuous variables with progressively unfavorable import at higher values. The platelet count did not influence survival significantly at values below 700 x 10^9/liter but was increasingly unfavorable above this level.

Although there is variation in the survival data for chronic granulocytic leukemia (CGL) reported from different sources, this is due principally to inhomogeneities in the composition of individual series and to widely varying initial proportions of poor-risk patients. When these factors are accounted for, death rates recorded at different centers in the United States and Europe during the past 20–30 yr are seen to be quite similar, and the results of management with “effective” chemotherapy can be distinguished from those of less effective treatment. Among “good-risk” patient populations (i.e., Philadelphia chromosome-positive, not blastic) treated according to any of several therapeutic schedules, the death rate is quite low (5%–10%) during the first year after diagnosis, but increases thereafter to a relatively stable rate of 23%–28%/yr during the third through the fifth years. Median survival in such populations is about 3.5 yr from diagnosis.

There is disagreement as to whether pretreatment prognostic discrimination is possible within the “good-risk” population. Some authorities have stated that Ph'-positive patients in the chronic stage of the disease cannot be further segregated into subgroups with different life expectancies. On the other hand, several recent papers have ascribed highly significant prognostic import to a number of disease features recorded at the time of diagnosis. This question is of considerable practical as well as theoretical interest. Increasingly complex and hazardous therapeutic procedures have been introduced in recent years in attempts to improve survival in this uniformly fatal disease. Allogeneic bone marrow transplantation is now being undertaken during the chronic stage of CGL at several centers, and some investigators recommend that it be performed soon after the diagnosis of leukemia is established. Obviously, early consideration of such high-risk therapy would be more appropriate for a patient in substantial danger of blast transformation and death within 2 yr than for one who may be expected to do well with conventional chemotherapy for the next 4–5 yr.

We undertook a cooperative study to evaluate the prognostic significance of various disease features recorded prior to treatment and to develop a prognostic classification for CGL. Our findings indicate that it should be possible to divide the Ph'-positive nonblastic population into subgroups of patients with significantly different survival patterns.

MATERIALS AND METHODS

Patients

This report is based on data submitted by cooperating investigators for 813 nonblastic Ph'-positive patients. This clinical material was collected from 6 sources, over somewhat different time periods.
of CGL. The following was adopted, as a compromise among the
of chromosomal analysis at the contributing institution were

SOKAL ET AL.

patients were incorrectly classified as Ph'-positive.

Nevertheless, we believe that very few of these
techniques of chromosomal study varied somewhat among different
institutions and over different time periods, with both direct prepara-
tions from bone marrow aspirates and short-term cultures being
utilized. Banding techniques were not available when many of these
patients underwent diagnostic study and, thus, were used in rela-
tively few cases. Nevertheless, we believe that very few of these
patients were incorrectly classified as Ph'-positive.

Most patients were treated conventionally, receiving single-agent
chemotherapy (most commonly with busulfan) in a dosage that
avoided hematologic toxicity. A sizable minority, however, were
subjected to splenectomy and to repeated courses of intensive
combination chemotherapy.9 About 10% received immunothera-
py12 as an adjunct to conventional chemotherapy.

Patients whose disease had been diagnosed before the availability
of chromosomal analysis at the contributing institution were
excluded in order to avoid a selection bias favoring long survivors.
The patients excluded because Ph' status was unknown consisted principally of those in whom, for various reasons, chromosomal study
was not attempted. Comparison of the survival of these patients with
that of patients whose Ph' status was known revealed no bias toward
either long or short survival in selection for chromosome study.

Techniques of chromosomal study varied somewhat among different
institutions and over different time periods, with both direct prepara-
tions from bone marrow aspirates and short-term cultures being
utilized. Banding techniques were not available when many of these
patients underwent diagnostic study and, thus, were used in rela-
tively few cases. Nevertheless, we believe that very few of these
patients were incorrectly classified as Ph'-positive.

There is no generally accepted set of criteria for the blastic stage
of CGL. The following was adopted, as a compromise among the
different definitions of blastic disease used by the investigators
participating in this study: (a) blasts 20% or more in peripheral
blood or bone marrow, (b) blasts plus promyelocytes, 30% in blood or
50% in marrow, or (c) presence of leukemic tumor masses or tissue
infiltration with immature leukemic cells (extramedullary blastic
transformation).

Data Collection and Processing

The available information regarding the following, as of the time
of diagnosis, was retrieved by cooperating investigators: sex, age,
date of diagnosis of CGL, presence or absence of symptoms presum-
ably due to leukemia and date of onset of symptoms, spleen and liver
size (centimeters below the costal margin), peripheral blood count,
bone marrow findings, results of chromosome study, leukocyte
alkaline phosphatase and serum alkaline phosphatase, lactate dehy-
drogenase, SGOT, and B12 levels. In addition, information regarding
antileukemic therapy was reported in most cases. Finally, the date of
the last follow-up and the patient's status at that time (alive, dead of
leukemia, or dead of other cause) were recorded.

This information was transmitted to the Cancer Center Database
either on one-page study forms for each patient or, in the case of
institutions that had entered these data into their computers, via
computer print-outs or copies of computer tapes. The data were
entered into the computer facility of the Cancer Center, with each
patient identified by source and by a study number at that source. Data
were subjected to automated checks for consistency and
reasonableness, and print-outs of data items were examined by one
or more of us. These checks resulted in identification of some errors
in data submission or processing. Delinquencies in patient follow-up
were easily identified by the computer, and a print-out of delinquent
cases was generated for each series. Correspondence with cooperat-
ing investigators resulted in elimination of most of these delinquen-
cies and correction of errors in data entry. For patients who
remained lost to follow-up, information was obtained regarding
disease status when last seen. Patients lost during the chronic stage
were handled statistically by conventional censoring. This was not
done for patients in the accelerated or blastic stage, however, in
order to avoid bias toward longer survival. Instead, for survival
calculations, these patients were considered to have died 90 days
(blastic patients) or 270 days (patients in the accelerated stage)
after their last follow-up date, which was usually some time after
disease progression had been recorded.

Table 1. Patients With Chronic Granulocytic Leukemia Included in This Study

<table>
<thead>
<tr>
<th>Source</th>
<th>No. of Patients</th>
<th>Dates of Diagnosis</th>
<th>Nature of Patient Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roswell Park Memorial Institute</td>
<td>172</td>
<td>1962-1979</td>
<td>All patients 12 yr or more old, except 24% who were blast at diagnosis or Ph'-negative</td>
</tr>
<tr>
<td>University of Bologna</td>
<td>167</td>
<td>1965-1980</td>
<td>All patients, except 9% who were blast at diagnosis or Ph'-negative</td>
</tr>
<tr>
<td>Italian Cooperative CML Study Group</td>
<td>167</td>
<td>1972-1977</td>
<td>Recently diagnosed patients in the chronic stage entered into two therapeutic protocols, one of which was restricted to patients 8–65 yr old; 41% excluded because Ph' unknown or negative</td>
</tr>
<tr>
<td>Memorial Sloan-Kettering Cancer Center</td>
<td>105</td>
<td>1968-1979</td>
<td>Ph'-positive patients with full work-up, followed regularly at MSKCC (approximately half of the Ph'-positive patients seen in this period)</td>
</tr>
<tr>
<td>University of Barcelona</td>
<td>101</td>
<td>1969-1981</td>
<td>All patients, except 42% who were blast at diagnosis, Ph' unknown, or Ph'-negative</td>
</tr>
<tr>
<td>Duke University</td>
<td>101</td>
<td>1967-1981</td>
<td>All patients, except 36% who were blast at diagnosis, Ph' unknown, or Ph'-negative</td>
</tr>
</tbody>
</table>
**Statistical Methods**

Survival estimates were obtained by the Kaplan-Meier product-limit method. Prognostic factor relationships were determined using Cox's proportional hazard model for covariate analysis of censored survival data. The underlying assumption of the model is that the hazard (relative failure rate) for the $i$th individual, $\lambda_i(t)$, is related to the hazard for the overall group, $\lambda_0(t)$, and the individual's $j$ covariate values, $x_{1i}, x_{2i}, \ldots, x_{ji}$, by the expression:

$$\lambda_i(t) = \lambda_0(t) \exp \left( \sum \beta_j x_{ji} \right)$$

where $\beta_j$ are regression coefficients estimated from the data using an iterative conditional likelihood method. The significance test for stepwise addition of a variable to a model is based on the approximate chi-square distribution of twice the increment in log likelihood with the entry of that variable into the model. Breslow's approximation to the risk set permutations on tied data was used.

**RESULTS**

The patient population consisted of 464 males (57%) and 349 females (43%), ranging in age at diagnosis from 5 to 84 yr. Of the 813 patients, 69% have died of leukemia, 4% have died of other causes, 25% were alive at the last reporting date, and 16 patients (2%) have been lost to follow-up. Of the latter, 10 were in the chronic stage of CGL, 2 in the accelerated phase, and 4 blastic when last seen. The actuarial survival curve for this population (Fig. 1) conforms to the pattern previously described for “good-risk” patients. The indicated death rate during the first year after diagnosis is 5% and during the second year, 14%. From 2 to 8 yr after diagnosis, annual death rates fluctuate between 20% and 31%, averaging 25%/yr. Median survival is 47 mo.

Information regarding sex, age, spleen and liver size, hemoglobin/hematocrit, platelet count, WBC count, and the percentage of circulating myeloblasts was available for almost all patients. The presence or absence of symptoms attributable to leukemia, and the percentage of basophils and eosinophils, were known in two-thirds of the cases. However, data for other disease features were available for only a minority of the patient population.

Table 2 shows the distribution of more commonly recorded features among the six series. There are many statistically significant differences among these values, but most of them would not be considered clinically significant. The male:female ratio varied from 0.8 (Barcelona) to 1.7 (Bologna), and the mean age ranged from 39 to 48 yr. There was less hepatomegaly in the Sloan-Kettering series than at the other institutions; this center also had a lower percentage of blasts and more patients with normal leukocyte alkaline phosphatase than any of the others. Serum lactate dehydrogenase (LDH) values were increased to more than twice the upper limit of the normal range in the majority of patients in all four series reporting this determination. The marked differences in the number of additional karyotypic abnormalities reported are probably artifactual, reflecting special interest in chromosome studies at some institutions (with more detailed examination of larger numbers of mitoses), differences in referral patterns (e.g., many patients seen at Duke had already been identified as Ph1-positive and were not restudied), and incomplete reporting.

There were multiple interactions among different disease features; these are summarized in Table 3. The correlation coefficients listed probably underestimate the degree of association between disease features in a number of instances because of deviations from linear relationships. An example is shown in Fig. 2. It is evident that there is a highly significant correlation between the WBC count and spleen size. However, there is substantial scatter of individual points, with about a tenfold variation in leukocyte count at each spleen size. Furthermore, the relationship of log WBC to spleen size is lost about halfway through the plot. This interaction is better described by the two dotted lines than by the conventional linear regression. In contrast, the negative correlation between hematocrit and spleen size fitted a single regression line.

**Correlations With Survival**

On univariate analysis, many of the recorded disease features had highly significant correlations with survival. However, because of the interactions described above, multivariable regression analysis was necessary in order to identify those with primary prognostic significance. To reduce the number of indeterminate cases, these analyses were undertaken in a subpopulation of patients whose diagnosis of CGL was established before 1978. This provided a minimum of 4 yr of patient follow-up for survival calculations. There were 678 such patients, 391 males (58%) and 287 females.
Table 2. Characteristics of 813 Patients According to Source

<table>
<thead>
<tr>
<th>Feature</th>
<th>Roswell Park</th>
<th>Bologna</th>
<th>Italian Coop. Group</th>
<th>Sloan-Kettering</th>
<th>Barcelona</th>
<th>Duke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/female (no.)</td>
<td>104/68</td>
<td>106/61</td>
<td>90/77</td>
<td>65/40</td>
<td>46/55</td>
<td>53/48</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>42 ± 15*</td>
<td>48 ± 15</td>
<td>42 ± 15</td>
<td>39 ± 14</td>
<td>44 ± 17</td>
<td>44 ± 18</td>
</tr>
<tr>
<td>Symptoms (% of pts)</td>
<td>63</td>
<td>85</td>
<td>69</td>
<td>86</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Spleen size (cm)</td>
<td>7.9 ± 6.7</td>
<td>8.5 ± 7.6</td>
<td>7.3 ± 5.4</td>
<td>6.2 ± 6.9</td>
<td>7.6 ± 7.3</td>
<td>5.4 ± 5.8</td>
</tr>
<tr>
<td>Liver size (cm)</td>
<td>1.9 ± 2.8</td>
<td>2.4 ± 2.5</td>
<td>3.2 ± 2.3</td>
<td>0.7 ± 1.5</td>
<td>2.3 ± 3.0</td>
<td>1.7 ± 2.8</td>
</tr>
<tr>
<td>Hematocrit (%)</td>
<td>35 ± 7</td>
<td>33 ± 6</td>
<td>35 ± 7</td>
<td>35 ± 7</td>
<td>36 ± 8</td>
<td>36 ± 7</td>
</tr>
<tr>
<td>WBC (10^9/liter)</td>
<td>459 ± 343</td>
<td>425 ± 314</td>
<td>399 ± 293</td>
<td>426 ± 259</td>
<td>426 ± 257</td>
<td>479 ± 348</td>
</tr>
<tr>
<td>Platelets in blood (%)</td>
<td>2.2 ± 2.9</td>
<td>2.0 ± 2.3</td>
<td>2.4 ± 2.8</td>
<td>1.2 ± 1.8</td>
<td>2.1 ± 2.5</td>
<td>2.1 ± 3.0</td>
</tr>
<tr>
<td>Blastsin blood (%)</td>
<td>3.5 ± 4.3</td>
<td>2.8 ± 2.9</td>
<td></td>
<td></td>
<td>4.1 ± 4.4</td>
<td>2.7 ± 3.9</td>
</tr>
<tr>
<td>Basophils and eosinophils (%)</td>
<td>6.2 ± 5.9</td>
<td>4.5 ± 3.8</td>
<td></td>
<td></td>
<td>6.6 ± 5.3</td>
<td>6.0 ± 3.4</td>
</tr>
<tr>
<td>Leukocyte alk phos low/normal/ high (no. of pts)</td>
<td>97/10/4</td>
<td>28/23/9</td>
<td>86/2/4</td>
<td>29/1/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blasts in marrow (%)</td>
<td>2.3 ± 2.6</td>
<td>2.7 ± 2.4</td>
<td>3.3 ± 3.1</td>
<td>2.9 ± 3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promyelocytes in marrow (%)</td>
<td>9.7 ± 4.8</td>
<td>5.4 ± 5.3</td>
<td>12.6 ± 7.4</td>
<td>6.2 ± 4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum alk phos normal/high (pts)</td>
<td>85/25</td>
<td>79/6</td>
<td>52/18</td>
<td>51/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum SGOT, normal/high (pts)</td>
<td>76/15</td>
<td>34/52</td>
<td>65/8</td>
<td>28/37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum LDH, less/ more than twice normal (pts)</td>
<td>39/48</td>
<td>36/51</td>
<td>28/34</td>
<td>20/51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional karyotypic abnormalities reported (pts)</td>
<td>14</td>
<td>27</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

*Mean ± SD.

Table 3. Significant Correlations Among Presenting Features*

<table>
<thead>
<tr>
<th>Feature</th>
<th>Female Sex</th>
<th>Spleen Size</th>
<th>Liver Size</th>
<th>Hematocrit</th>
<th>WBC Count</th>
<th>Platelet Count</th>
<th>Blood Blasts (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex</td>
<td></td>
<td>-0.21§</td>
<td></td>
<td>-0.13†</td>
<td></td>
<td></td>
<td>+0.22$</td>
</tr>
<tr>
<td>Age</td>
<td>-0.10†</td>
<td>+0.17$</td>
<td></td>
<td></td>
<td>-0.13†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptoms</td>
<td>+0.34§</td>
<td>+0.13†</td>
<td>-0.28§</td>
<td>+0.30§</td>
<td>+0.19†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spleen size</td>
<td>+0.33§</td>
<td></td>
<td>-0.46§</td>
<td>+0.55§</td>
<td>+0.31§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver size</td>
<td>-0.19§</td>
<td></td>
<td></td>
<td>+0.22§</td>
<td>+0.24§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hematocrit</td>
<td></td>
<td></td>
<td>-0.46§</td>
<td></td>
<td>-0.30§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBC count</td>
<td>+0.17†</td>
<td></td>
<td></td>
<td></td>
<td>+0.26§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platelet count</td>
<td>+0.11†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood promyelocytes (%)</td>
<td>+0.27§</td>
<td>+0.16†</td>
<td>-0.18†</td>
<td>+0.33§</td>
<td>+0.26§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baso + eos (%)</td>
<td>+0.12†</td>
<td></td>
<td></td>
<td></td>
<td>+0.13†</td>
<td>-0.14‡</td>
<td></td>
</tr>
<tr>
<td>Low leukocyte alk phos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marrow blasts (%)</td>
<td>0.16†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marrow promyelocytes (%)</td>
<td>0.18†</td>
<td>+0.22‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal serum alk phos</td>
<td>-0.11†</td>
<td></td>
<td></td>
<td></td>
<td>-0.18‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal serum SGOT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal serum LDH</td>
<td>-0.27§</td>
<td>-0.18†</td>
<td></td>
<td>+0.16†</td>
<td>+0.32§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional karyotypic abnormalities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlation coefficient and statistical significance. (p > 0.01 excluded.)
†p = 0.01–0.001.
‡p = 0.001–0.0001.
§p < 0.0001.
conventional linear regression over the entire range of values.

data items were missing in 79 cases (1.2%), leaving of regression analysis gave similar results, the missing items of data. Both procedures yielded of 678 by assigning the means of the known values to same data. were compared with univariate of regression analysis were used, and neuous techniques similar results. Both stepwise (step-up) and simulta-

checked in the entire group population and then were

Analyses were undertaken first in this slightly smaller population. These results

are summarized in Table 4. Both

<table>
<thead>
<tr>
<th>Feature (Direction of</th>
<th>Prognostic Significance (p Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univariate Analysis</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
</tr>
<tr>
<td>Sex (male)</td>
<td>0.5</td>
</tr>
<tr>
<td>Age (higher)</td>
<td>0.02</td>
</tr>
<tr>
<td>Spleen size (larger)</td>
<td>0.0000001</td>
</tr>
<tr>
<td>Liver size (larger)</td>
<td>0.09</td>
</tr>
<tr>
<td>Hematocrit (lower)</td>
<td>0.0004</td>
</tr>
<tr>
<td>WBC count (higher)</td>
<td>0.001</td>
</tr>
<tr>
<td>Platelet count (higher)</td>
<td>0.001</td>
</tr>
<tr>
<td>Percent blasts in blood (higher)</td>
<td>0.0000001</td>
</tr>
</tbody>
</table>

(42%). Of these patients, 529 (78%) had died of leukemia and 112 (16.5%) were alive at last follow-up, with 32 (4.7%) dead of other causes and 5 (0.7%) lost to follow-up. Median survival was 47 mo—the same as that of the entire population.

Sex and age of all patients were known. Spleen and liver size, hemoglobin/hematocrit, platelet count, WBC count, and percent of circulating blasts were known for almost all patients, but one or more of these data items were missing in 79 cases (12%), leaving 599 patients for whom all of this information was available. Analyses were undertaken first in this slightly smaller population and then were checked in the entire group of 678 by assigning the means of the known values to the missing items of data. Both procedures yielded similar results. Both stepwise (step-up) and simulta-

These results are summarized in Table 4. Both techniques of regression analysis gave similar results, but there were several major discrepancies between the findings on univariate and multivariable analysis. Percent blasts and spleen size were the features that correlated most significantly with survival. Liver size appeared highly significant on univariate analysis but had only borderline significance when considered in conjunction with other disease features. Hematocrit and WBC count also appeared to be important prognostic indicators on univariate analysis, but were found to be insignificant on multivariable regression. Shifts in the other direction were also seen. Sex achieved borderline significance on multivariable regression, and age gained in significance.

Our relatively large study population provided us with the opportunity to test the results of multivariable regression by univariate comparisons in appropriately stratified subgroups of patients. These confirmed the results of the multivariable analyses. When patients were stratified for spleen size, liver size usually lost prognostic significance. In populations stratified for percentage of blasts and spleen size, neither WBC count nor hematocrit proved to have significant prognostic value.

Information regarding other disease features was available for smaller numbers of patients, and these were unequally distributed among the different series. For example, leukocyte alkaline phosphatase was known in only 30% of the cases, and most of these came from two institutions (see Table 2). Therefore, we did not include such features in the regression analyses summarized in Table 4. As there was no suitable subpopulation in which values for most of these other features were known, separate analyses were performed for each feature, restricted to the patients for which information regarding that feature was available, and adding it as a fifth variable in regression analysis to percent of blasts, spleen size, platelets, and age. In the case of serum LDH activity, which had been found to be a highly significant prognostic indicator at one institution, analysis was also performed after excluding patients from that institution. The results are presented in Table 5. The presence of karyotypic abnormalities in addition to the Ph' chro-
mosome (54 patients), marrow blasts above 5% (26 patients), and basophils plus eosinophils above 15% (19 patients) were statistically significant unfavorable features. Major elevation in serum LDH activity lost its prognostic significance when data from the Roswell Park Memorial Institute were excluded.

Of the features with statistically significant prognostic import, some had progressive influence on risk status over almost their entire range of values. These included spleen size, percent of circulating blasts, and, to a lesser extent, age. Other features appeared to have little or no prognostic significance until a boundary well beyond their median value was reached. The percentage of basophils and eosinophils, and of marrow blasts, fell into this category, as did the platelet count. The latter did not influence survival significantly at values below $700 \times 10^9$/liter; beyond this boundary, it behaved as a continuous variable, with increasingly unfavorable import at higher values.

Effects of Combining Prognostic Criteria

When survival curves were plotted for patients segregated into approximately equal groups according to percentage of circulating blasts, spleen size, or age, median survivals differed by 17, 14, and 6 mo, respectively. In the case of the platelet count, segregation according to the prognostically significant boundary of $700 \times 10^9$/liter produced grossly unequal groups, with a 10-mo difference in median survival. As might be expected, combining these prognostic criteria resulted in the identification of patient groups with larger differences in survival. Figure 3 shows the effect of combining blasts and spleen size. Because there was a significant positive correlation between these features (see Table 3), there were more patients in the “fewer blasts—smaller spleen” and “more blasts—larger spleen” groups than in either of the other two. This pair of criteria identified a good-risk group containing almost one-third of the patients, with a median survival of 61 mo, as well as a poor-risk group (28% of patients) with a median survival of 34 mo. Only one of the intermediate groups (fewer blasts—larger spleen) was significantly different from both of these.

Figure 4 shows the effect of adding age and platelet count as additional criteria. A consistent trend is evident: among patients stratified for blasts and spleen size, age and platelet count provide additional prognostic discrimination. However, patient groups are now quite small and the confidence limits for most of the survival curves would overlap. These would not be very useful for predictive purposes.

Cox Model Analysis

The Cox model provides a quantitative technique for examining the effects of various combinations of prognostic variables. An important advantage of this mathematical model is that a prognostic feature can be treated as a continuous variable, in contrast to classification systems that require assignment to discrete categories, such as “spleen greater than 10 cm.” To determine whether the Cox model could provide a

### Table 5. Prognostic Significance of Disease Features Reported Only in Part of the Study Population

<table>
<thead>
<tr>
<th>Feature (Worst Prognosis)</th>
<th>No. of Patients Studied (No. of Series*)</th>
<th>Prognostic Significance ($p$ Value)</th>
<th>Univariate Analysis</th>
<th>Multivariable Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms</td>
<td>409(5)</td>
<td>0.15</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Leukocyte alk. phosphatase (abnormal)</td>
<td>206(4)</td>
<td>0.07</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Blood promyelocytes</td>
<td>381(4)</td>
<td>0.3</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Basophils + eosinophils (&gt;15%)</td>
<td>461(5)</td>
<td>0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Marrow blasts (&gt;5%)</td>
<td>232(4)</td>
<td>0.001</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Marrow promyelocytes</td>
<td>216(4)</td>
<td>0.9</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Serum alk. phosphatase</td>
<td>246(4)</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Serum SGOT (abnormal)</td>
<td>236(4)</td>
<td>0.6</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Serum LDH (over twice normal)</td>
<td>228(4)</td>
<td>0.0001</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Serum LDH, excluding RPMI</td>
<td>148(3)</td>
<td>0.03</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Additional karyotypic abnormalities</td>
<td>306(5)</td>
<td>0.001</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

*Of the six series listed in Table 1.

---

![Fig. 3. Survival according to spleen size and percentage of circulating blasts among 624 patients diagnosed before 1978.](image)
useful representation of risk status in CGL, it was constructed with variables representing the disease features regularly recorded in all six series and found to have unequivocal prognostic significance, namely: age, spleen size, platelet count, and percent of circulating blasts. Age, spleen size, and percent of blasts were used as continuous variables without transformation. As the platelet count had little prognostic significance over a wide range of values, but became an increasingly significant factor at abnormally high values, a simple linear expression for this variable did not seem appropriate. Several different formulations were tested, and a function of the square of the platelet count was selected; this provided an expression utilizing a single variable, which gave increasing weight to higher values. In order to include all patients represented in Table 4, mean values were assigned for the missing items of data (age, none missing; spleen size, 5% missing; platelets, 6% missing; percent blasts, 5% missing). This procedure eliminates the regression term for the missing variable and, hence, consideration of that feature. It has a neutral effect if the variables segregate independently, but may introduce some bias if there are significant associations among variables. In the present case, there were statistically significant correlations; two were concordant (percent blasts and spleen size, percent blasts and square of the platelet count) and one was discordant (age and spleen size). Thus, on balance, there was modest positive association of prognostic features, and the use of mean values may have introduced a slight bias against prognostic discrimination.

The following hazard ratio function was derived from these data:

$$\lambda(t)/\lambda_0(t) = \exp 0.0116(Age-43.4)$$

$$+ 0.0345(Spln-7.51) + 0.188 \left(\frac{Plt}{700}\right)^2 - 0.563$$

$$+ 0.0887(Blasts-2.10)$$

The hazard ratio for each patient was calculated from this expression. These ranged from 0.41 to 5.68 for 677 patients. There was one “outlier” value (17.87, for a patient who died of leukemia at 18 mo), which was excluded. To test the model, 5 groups of 60–75 patients were selected: from each extreme of the risk ratio distribution, from the center, and from two intermediate bands of values. The risk ratios for these groups fell within relatively narrow ranges, except for the highest risk group, which contained a threefold range of hazard ratios. Survival curves for these 5 groups of patients are shown in Fig. 5. Median survival ranged from 28 to 67 mo. Although the differences between adjacent curves are not statistically significant, all other differences are highly significant (p < 0.003).

To test the goodness of fit of this mathematical model, a predicted survival curve was generated for each of the five patient groups by applying the median hazard ratio of that group to the actuarial survival calculated for the entire population of 678 patients. Figure 6 shows these predicted curves and the observed survival for the lowest risk, highest risk, and central groups. Similar results were obtained for the two intermediate groups. We concluded that this mathematical model was suitable for analysis of survival data in CGL.

To determine whether the prognostic significance of individual disease features was similar in different
patient populations, we used part of our patient population to generate a prognostic function and the remainder to test it. For the "training group," we used patients from the Roswell Park, Bologna, and Italian Cooperative Study Group CML/73 series. These series had already been examined for prognostic features, with positive and generally similar results. For the "test group," we used patients from the Barcelona, Duke, Memorial Sloan-Kettering, and Italian Cooperative Study Group CML/74 series. Independent analysis of the training group (assigning mean values for missing items of data) again indicated that spleen size, percentage of circulating blasts, platelet count, and age were the features with unequivocal prognostic significance. These features were used as variables in a manner similar to that described above, and the Cox model was generated, using stepwise regression to determine the coefficients for these four variables. The hazard ratio for each of the 361 patients in the training group was calculated, and this population was divided into 3 subgroups of roughly similar size, using hazard ratios of 0.8 and 1.2 as boundaries. Survival curves for these groups are shown in Fig. 7. Median survival for the low-risk group, containing 32% of the patients, is 60 mo and that of the high-risk group (28% of the patients) is 32 mo.

The same formula was then used to calculate the hazard ratios in the test population of 317 patients and again, these patients were divided into three subgroups using hazard ratios of 0.8 and 1.2 as boundaries. The survival curves for these groups are shown in Fig. 8. The distribution of patients among risk groups is not as symmetrical as in the training population. However, each survival curve is similar to the corresponding curve in Fig. 7, and these curves also differ significantly from each other and diverge progressively with time. Again, median survival of the low-risk group is almost twice that of the high-risk group (58 versus 31 mo).

Taken together, the low-risk subgroups of the training and test populations contain 263 patients (39% of the total), with a 2-yr survival of 90% and an average death rate during the next 5 yr of 19%/yr. The high-risk subgroups contain 158 patients (23%), with a 2-yr survival of 65%, followed by an average death rate of 35%/yr.

Finally, the training and test groups were reversed. The Cox model was generated from the 317 patients in the Barcelona, Duke, Memorial Sloan-Kettering, and Italian Cooperative Study Group CML/74 series. Again, percentage of blasts, spleen size, platelet count, age, and other features were used as variables in a manner similar to that described above, and the Cox model was generated, using stepwise regression to determine the coefficients for these four variables. The hazard ratio for each of the 361 patients in the training group was calculated, and this population was divided into 3 subgroups of roughly similar size, using hazard ratios of 0.8 and 1.2 as boundaries. Survival curves for these groups are shown in Fig. 7. Median survival for the low-risk group, containing 32% of the patients, is 60 mo and that of the high-risk group (28% of the patients) is 32 mo.

The same formula was then used to calculate the hazard ratios in the test population of 317 patients and again, these patients were divided into three subgroups using hazard ratios of 0.8 and 1.2 as boundaries. The survival curves for these groups are shown in Fig. 8. The distribution of patients among risk groups is not as symmetrical as in the training population. However, each survival curve is similar to the corresponding curve in Fig. 7, and these curves also differ significantly from each other and diverge progressively with time. Again, median survival of the low-risk group is almost twice that of the high-risk group (58 versus 31 mo).

Taken together, the low-risk subgroups of the training and test populations contain 263 patients (39% of the total), with a 2-yr survival of 90% and an average death rate during the next 5 yr of 19%/yr. The high-risk subgroups contain 158 patients (23%), with a 2-yr survival of 65%, followed by an average death rate of 35%/yr.

Finally, the training and test groups were reversed. The Cox model was generated from the 317 patients in the Barcelona, Duke, Memorial Sloan-Kettering, and Italian Cooperative Study Group CML/74 series. Again, percentage of blasts, spleen size, platelet count, age, and other features were used as variables in a manner similar to that described above, and the Cox model was generated, using stepwise regression to determine the coefficients for these four variables. The hazard ratio for each of the 361 patients in the training group was calculated, and this population was divided into 3 subgroups of roughly similar size, using hazard ratios of 0.8 and 1.2 as boundaries. Survival curves for these groups are shown in Fig. 7. Median survival for the low-risk group, containing 32% of the patients, is 60 mo and that of the high-risk group (28% of the patients) is 32 mo.

The same formula was then used to calculate the hazard ratios in the test population of 317 patients and again, these patients were divided into three subgroups using hazard ratios of 0.8 and 1.2 as boundaries. The survival curves for these groups are shown in Fig. 8. The distribution of patients among risk groups is not as symmetrical as in the training population. However, each survival curve is similar to the corresponding curve in Fig. 7, and these curves also differ significantly from each other and diverge progressively with time. Again, median survival of the low-risk group is almost twice that of the high-risk group (58 versus 31 mo).

Taken together, the low-risk subgroups of the training and test populations contain 263 patients (39% of the total), with a 2-yr survival of 90% and an average death rate during the next 5 yr of 19%/yr. The high-risk subgroups contain 158 patients (23%), with a 2-yr survival of 65%, followed by an average death rate of 35%/yr.

Finally, the training and test groups were reversed. The Cox model was generated from the 317 patients in the Barcelona, Duke, Memorial Sloan-Kettering, and Italian Cooperative Study Group CML/74 series. Again, percentage of blasts, spleen size, platelet count, age, and other features were used as variables in a manner similar to that described above, and the Cox model was generated, using stepwise regression to determine the coefficients for these four variables. The hazard ratio for each of the 361 patients in the training group was calculated, and this population was divided into 3 subgroups of roughly similar size, using hazard ratios of 0.8 and 1.2 as boundaries. Survival curves for these groups are shown in Fig. 7. Median survival for the low-risk group, containing 32% of the patients, is 60 mo and that of the high-risk group (28% of the patients) is 32 mo.
and age proved to be the significant prognostic features. The coefficient for blasts was almost identical to that obtained in the previous analysis and that for the platelet variable differed by about 20%; however, there were almost twofold differences in the coefficients for spleen size and age. Risk ratios for each patient were recalculated according to this hazard function and, again, patients in the training and test populations were divided into 3 groups using hazard ratios of 0.8 and 1.2 as boundaries. The results were similar to those described above. As before, a more symmetrical distribution of patients and somewhat better separation of survival curves was obtained in the population used to generate the model than in the test population, but the survival curves in each differed significantly and diverged progressively.

Other Observations

We tested several alternative formulations of the Cox model in addition to those described above. This included limited trials of other mathematical expressions for the variables, such as a quadratic rather than a linear function for percent of blasts (Blasts²); these did not improve the results. Use of a simpler hazard ratio function, based on percentage of blasts and spleen size as the only variables, proved feasible and passed the type of test described above; a function generated from half of the patient population was able to segregate the remainder into three groups with significantly different survival patterns. However, the fit between predicted and observed survival data was not as good as that obtained with the four-variable model.

In examining various subgroups of patients, we observed that when the number included in a multivariable regression was less than 200, the results were much less consistent. Although spleen size and percentage of circulating blasts were the most consistently significant prognostic indicators found in this study, one or the other often lost statistical significance when groups of 50–150 patients were subjected to analysis. Sometimes, no disease feature achieved statistical significance as a prognostic indicator. However, when a hazard ratio function generated from a large population was tested in such subgroups, statistically significant prognostic discrimination was almost always achieved.

DISCUSSION

The patient population assembled for this study appears to constitute a reasonable sample of “good-risk” patients with CGL. On univariate analysis, many disease features seemed to have prognostic significance. This has been the finding in most recent studies of this subject.3–5,8,10 In contrast, most negative reports are based on older series that included many poor-risk patients and less effective treatment. An example of this is seen in an earlier report from one of the institutions participating in the present study, in which 178 cases of CGL diagnosed during 1948–1967 were reviewed.17 A search for prognostic indicators gave essentially negative results, in striking contrast to findings in the current series of patients. However, in the earlier study, chromosomal analysis was performed in only 14 cases, treatment consisted of radiation alone for some patients and inferior chemotherapy for others, and survival was significantly poorer than among the patients registered in the present study. It is not surprising that the prognostic influence of initial disease characteristics was not evident in such a heterogeneous population.

Most previous studies do not seem to have examined the interactions among different disease features, but rather, to have assumed that these have independent prognostic significance. This is clearly not the case. Undoubtedly, the multiple interrelationships summarized in Table 3 are responsible for the apparent prognostic significance on univariate analysis of disease features that are found to have no predictive value by multivariable regression, such as the hematocrit. We were rather surprised by the extent of these associations. Only the platelet count and the presence of additional karyotypic abnormalities appeared to be essentially independent of other disease features.

A simple binary classification, using the two most important prognostic features, identified low- and high-risk groups of patients with almost a twofold difference in median survival, but did not characterize the intermediate groups well (Fig. 3). Adding platelet count and age provided additional prognostic information (Fig. 4). Combined use of all four criteria, with a point-scoring system such as that tested by Tura et al.,6 would probably provide better definition of risk categories, including those for the intermediate-risk patients. Devising a scoring system that would reflect the relative importance of these criteria might pose some problems, but the principal disadvantage of this type of classification system is that features that behave biologically as continuous variables are scored in discrete categories. This forces a number of gross inconsistencies. For example, in the schema illustrated in Fig. 3, the score for 2% blasts would be the same as that for 15% blasts, but different from that for 1% blasts. In contrast, the Cox model can accommodate both continuous and discrete variables, and the weight assigned to a particular prognostic feature is determined by its relative biologic importance in the population used to generate the model. The only disadvantage to the use of such a model is that it requires a relatively sophisti-
cated computer program. In view of the rapidly increasing use of computers to process biomedical data, this will probably not constitute a significant problem at most centers.

The Cox model appears suitable for analysis of survival in CGL. The four-variable function used to generate the data illustrated in Figs. 5–8 yielded quite satisfactory results. The model could probably be improved by including terms for basophils and eosinophils, bone marrow blasts, and additional karyotypic abnormalities. However, because of the large number of patients for whom we had no information regarding these features, we could not test this question. Our experience suggests that relatively large numbers of patients may be required in order to obtain consistent results with the Cox model. Because of this, we cannot be certain whether the loss of significance of high LDH values after exclusion of the Roswell Park data (Table 5) reflects a real difference among institutions or simply the effect of reducing the number of cases subjected to regression analysis below a critical level. Such analyses have often been applied to clinical series containing fewer than 150 patients, but we are not aware of any definitive studies of their dependability in these circumstances. At any rate, although we could not use the smaller subgroups to generate a prognostic model, we could apply a hazard function derived from a large population to these same subgroups quite successfully.

A number of quite different therapeutic approaches to CGL are represented in this study. If some of these treatment schedules were significantly superior to others, this would produce variations in survival data that would make it more difficult to identify disease features with prognostic significance. However, it would not detract from the significance of criteria that were identified, unless there was a systematic association of better or worse treatment with particular disease features. For example, if patients with smaller spleens and fewer blasts had received more effective treatment than those with larger spleens and more blasts, spleen size and percentage of blasts would have been artifactually identified as significant prognostic criteria. Such an explanation of our results is not tenable. Although we do not have detailed information regarding treatment for all patients in the study, we know that at no center was it determined by such criteria.

One assumption implicit in this study was that treatment would not affect the prognostic significance of the disease characteristics being evaluated, e.g., that the significance of spleen size is similar among patients receiving conventional therapy with busulfan and those subjected to splenectomy plus combination chemotherapy with agents used in acute leukemia. The positive results obtained despite our pooling of patients subjected to different treatment schedules constitute indirect evidence that this assumption is correct. However, they do not exclude the possibility that a treatment received by a minority of patients nullified the prognostic significance of one or more disease features, but that this was not apparent when these patients were commingled with larger groups treated in other ways. A definitive answer to this question will require evaluation of prognostic features and hazard ratio functions in patient populations segregated according to treatment.

A question of some interest is whether the prognostic differences identified in this study reflect intrinsic differences in the course of CGL among different patients or simply, earlier versus later diagnosis. That is, do the high-risk patients have more aggressive disease or more advanced disease? We favor the former explanation. Survival curves for the lower and higher risk patient groups diverge progressively (Figs. 5–8). The lower risk patients still had a relatively low death rate 5 yr after diagnosis when, presumably, their disease should have been “advanced.” In contrast, the higher risk population reached a stable, significantly higher death rate within 2 yr of diagnosis. If the differences in prognosis were due simply to earlier versus later diagnosis, one would expect initial differences in death rates, followed by parallel survival curves.1

If the validity of this approach to prognostic classification of CGL is confirmed, it should prove useful in evaluating the effects of different treatment schedules on survival. This has proven difficult in the past, because CGL is an uncommon neoplasm and treatment comparisons have usually involved relatively small numbers of patients. Little has been done in therapeutic trials to assure comparability of patients in different groups, except to restrict studies to Ph1-positive, nonblastic patients. However, our results indicate that substantial differences in expected survival may be found among such “good-risk” patients. The ability to calculate hazard ratios for individual patients will make it possible to examine patient groups for initial risk status and to stratify patients according to prognostic classification. The effects of treatment on survival could then be evaluated with greater precision, both in retrospective reviews of completed studies and in prospective therapeutic trials.

ACKNOWLEDGMENT

Ospedale S. Matteo, Pavia; A. Cajozzo. Cattedra di Patologia Medica dell'Università, Palermo; G. Perona, Cattedra di Ematologia dell'Università, Verona; P. Leoni, Istituto di Patologia Medica dell'Università, Ancona; G. Torlontano, Cattedra di Ematologia, Università "G. D'Annunzio," Chieti; A. Porcellini, Divisione di Ematologia, Ospedale S. Salvatore, Pesaro; R. Battista, Divisione di Ematologia Ospedale Regionale, Vicenza; F. Leoni, Divisione di Ematologia, Arcispedale Santa Maria Nuova, Firenze; R. Giustolisi, Cattedra di Ematologia dell'Università, Catania; F. De Cataldo, Divisione di Ematologia, Ospedale Niguarda, Milano; V. Rizzioli, Cattedra di Ematologia dell'Università, Parma; C. Carnevali, Istituto di Patologia Medica II dell'Università, Parma; V. Liso, Istituto di Clinica Medica II dell'Università, Bari; E. Bajetta, Istituto Nazionale Tumori, Milano; G. Broccia, Ospedale A. Businco, Cagliari; L. Bruzzese, Istituto di Clinica Medica dell'Università (I Facoltà), Napoli; E. Lanzi, Ospedale Generale Regionale, Monza; A. Alberti, Servizio di Ematologia, Ospedale Civile, Catanzaro; A. Neri, Divisione di Ematologia, Ospedali Riuniti, Reggio Calabria; M. Tonazo, Divisione di Oncologia Medica, Perugia; V. Grifoni, Cattedra di Clinica Medica RR dell'Università, Genova.

REFERENCES

Prognostic discrimination in "good-risk" chronic granulocytic leukemia

JE Sokal, EB Cox, M Baccarani, S Tura, GA Gomez, JE Robertson, CY Tso, TJ Braun, BD Clarkson and F Cervantes