Lecture Outline

• Case Study
  – Scientific Background
  – Materials and Methods
    • Source of Data
    • Statistical Methods
  – Results
  – Discussion

Case Study

• “Use of cytosine arabinoside and total body irradiation as conditioning for allogeneic marrow transplantation in patients with acute lymphoblastic leukemia: a multicenter survey”
Observational Study

- Compare the disease free survival in patients treated with a particular conditioning regimen to historical reports of other treatments
  - Major analysis issues
    - Summary measures of survival distribution to report
    - Selection of groups to use as strata for estimation of outcomes
    - Comparison to historical reports of other therapies

Goals of Case Study

- Illustrate approach to a data analysis problem in which data driven analyses play a major role
  - Approach to the data
  - Issues to address during analysis of time to event data
  - Data driven identification of groups for analysis
  - Presentation of results

Scientific Background

- Bone marrow transplantation in acute lymphoblastic leukemia
  - Patients who had relapsed or were at high risk of relapse were commonly recipients of bone marrow transplantation
    - Eliminate leukemia from the bone marrow
      - high doses of chemotherapy and total body irradiation
    - Infuse healthy bone marrow from suitable donor
Scientific Background

• Conditioning regimens
  – Previous standard conditioning regimen was cyclophosphamide and total body irradiations
  – Interest in cytosine arabinoside (ara-C) due to its penetrance into central nervous system and efficacy in reinducing patients whose leukemia had relapsed

Scientific Background

• Question of interest
  – Was the experience with ara-C any better than that previously reported with the cytoxan conditioning regimen?

Materials and Methods

Source of Data

• Survey of 14 centers using ara-C as a conditioning regimen
• 222 patients treated between 1981 and 1989
Scientific Classification of Data

- (Classification by statistician)
  - Demographics (age, sex, race)
  - Presentation of ALL at diagnosis (date, WBC, immunophenotype)
  - Conditioning regimen (ara-C dose, other agents, irradiation)
  - Bone marrow transplantation (date, remission status, donor information)
  - GvHD prophylaxis
  - Outcome (relapse, cause of death)

Materials and Methods: Table 1

- Descriptive statistics on available data to provide information on materials and methods
  - Missing data (pervasive in such studies)
    - Especially race, cell count, duration of disease
    - ?Quality of cooperation among centers
  - Location: Mean, median, percentages of binary data
  - Spread: Standard deviation, frequency tables, range

Table 1: Patient characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Available cases</th>
<th>Frequency (%)</th>
<th>Mean (SD)</th>
<th>Median range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>210</td>
<td>24.9</td>
<td>(3.4/9.6)</td>
<td>10.7/0.5-40.7</td>
</tr>
<tr>
<td>3-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male sex</td>
<td>241</td>
<td>60.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>159</td>
<td>39.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis date</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-10</td>
<td>225</td>
<td>51.5</td>
<td>10.6/14.2</td>
<td>16.4/14-146</td>
</tr>
<tr>
<td>11-20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone marrow transplantation</td>
<td>380</td>
<td>95.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location (centers)</td>
<td>120</td>
<td>26.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistical Task

- Provide estimates of distribution of clinical outcomes
  - Essentially a one sample problem: All subjects received ara-C
  - However, may want to make description within groups defined by important prognostic variables
    - Readers will have to compare results to those reported in the literature or in their practice
Summary Measure

- Choice of summary measure for outcome
  - Interested in relapse and death
    - Disease free survival
  - Variable time of follow-up: Censored data
  - Choices:
    - Quantiles (e.g., median)
    - Survival probabilities at fixed points in time
      - 100 days indicative of toxic death from BMT
      - 1, 3, or 5 year survival probabilities for clinical relevance, comparisons

Predictors

- Definition of groups for presentation of estimates
  - Identify strongest predictors of survival
    - Allows more precision of estimates
    - Allows comparability with other studies
  - Methods
    - Most statistically significant predictors
    - Form of models to allow comparability

Statistical Methods

- Description of statistical methods
  - Methods for descriptive statistics: Kaplan-Meier
    - Necessary because nonstandard at the time
  - Methods for model building
    - Analysis methods: Proportional hazards
      - Interpretation
    - Model building
      - Selection of variables
      - Validity of assumptions
  - Missing data
Organization of Results

- Presentation builds to conclusions
  - Dispense with potential nuisance covariate
  - Calendar year
- Finding important predictors of survival
  - Univariately
  - Multiple regression model
- Presentation of estimates within major strata
- Comparison to previously reported results for other regimens

Model Building: Univariate First

- Table II: Present statistically significant univariate predictions
  - Overall test for trend based on continuous model
  - Descriptive estimates within strata
    - Strata chosen independent of outcome
    - Reference group chosen to allow sufficient precision for comparisons
  - Space constraints suggest that nonsignificant variables could be examined only in Table III
- Figures display Kaplan-Meier curves by strata
  - Note depiction of censored data
Model Building: Multivariate

- Table III: Estimates of association adjusted for “final model”
  - Selection of variables (and form of variables) for final model based on
    - Statistical significance
    - Need for relatively few strata
    - Ability to compare with the literature
  - Obvious data-driven aspect to choice of final form of variables
    - Should be careful about believing any thresholds dividing the strata
Discussion

- Table IV: Direct comparison with previously reported results for cyclophosphamide insofar as possible from literature
  - Provide estimates of survival and relapse to match patient populations and time frames

Discussion

- Limitations of study
  - Observational aspect of study
  - Confounding by treatment center
  - (Data driven aspect of selecting strata)