Phase II and III Study Designs for aGVHD

Brent R. Logan, Ph.D.
Division of Biostatistics
Medical College of Wisconsin
Financial Conflicts of Interest

- None
Prophylaxis studies

- Competing risks
- Selection of endpoint
- Consideration in study design
- Statistical Analysis
- Phase II trials
- Phase III trials
Competing Risks

- Definition: Occurrence of one type of event precludes observation of the other type of event.

- Development of acute GVHD grade II-IV vs. Death (in absence of aGVHD).

- Problem: GVHD Prophylaxis targets development of aGVHD.

- May or may not impact mortality positively or adversely.
Summarizing competing risks data

- Cumulative incidence function: Probability an individual has experienced the event of interest by time \( t \), and before the competing event.

- Not appropriate to use the estimator \( 1 - KM(t) \):
  - Assumes independent competing risks.
  - Estimates incidence of one event if the other event could not occur.
Competing Risks Example 1

- PB vs. BM in URD TX (Eapen et al. BBMT 2007)
competing risks example 1

Cumulative Incidence of death without grade II–IV aGVHD by stem cell source

Reasons for lower non-GVHD mortality in PB group: fewer people at risk, other factors??
Competing Risks Example 1

Difference in aGVHD-free survival is slightly less pronounced than difference in incidence of aGVHD
Competing Risks Example 1

Little to no difference in OS despite greater incidence of aGVHD
Competing Risks Example 2

- Ratanatharathorn et al. (Blood, 1998): FK506+MTX vs. CSA+MTX in HLA-identical sibling BM transplants (n=329)
- Stratified on patient age (<40 vs. >=40) and donor/recipient sex match
Incidence of aGVHD grade II-IV

Worse survival for FK-506 group despite lower aGVHD incidence


Copyright ©1998 American Society of Hematology. Copyright restrictions may apply.
Advanced Disease

CSA+MTX (n=48)

FK-506+MTX (n=68)

Ratanatharathorn, V. et al.
Blood 1998;92:2303-2314

Early Disease

Copyright ©1998 American Society of Hematology. Copyright restrictions may apply.
Further investigation

- Horowitz et al. (BBMT, 1999): Comparison of matched IBMTR controls for advanced disease patients in above trial \( (n=116) \)
- Matched on diagnosis, pre-tx disease status, age \( \text{(within 5 years)} \)
- Controls received CSA+MTX
# Two-year survival

<table>
<thead>
<tr>
<th>Group</th>
<th>Probability</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FK-treated trial pts</td>
<td>27%</td>
<td>0.51</td>
</tr>
<tr>
<td>FK-matched controls</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>CSA-treated trial pts</td>
<td>42%</td>
<td>0.66</td>
</tr>
<tr>
<td>CSA-matched controls</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>FK-matched controls</td>
<td>24%</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>CSA-matched controls</td>
<td>45%</td>
<td></td>
</tr>
</tbody>
</table>
Possible endpoints

- **Survival**
  - May not be sensitive enough to treatment (e.g., PB vs. BM study) – see also Martin and Nash (BBMT, 2006)
  - Once patient develops aGVHD, may receive additional treatment

- **Incidence of acute GVHD grade II-IV**
  - Does not include the impact of treatment on mortality (e.g., FK506 study)

- **Composite endpoint: acute GVHD free survival**
  - Event = Death or development of acute GVHD grade II-IV
  - Will a composite endpoint mask a treatment effect?

- Must examine both competing risks to understand effect of treatment
aGVHD-free survival vs. incidence of acute GVHD

- Change in incidence of aGVHD may induce a change in incidence of non-GVHD mortality in the opposite direction
  - Fewer people at risk
  - Depends on relationship between competing events
  - This could reduce the “treatment effect” and hence the power, but is difficult to measure
  - More impact as the rate of non-GVHD mortality increases

- Power also depends on where the relevant proportions are on the interval [0,1]
  - Closer to edges generally results in greater power to detect the same absolute difference
  - Incidence of aGVHD vs. aGVHD-free survival may be at different points on this continuum
Consideration of competing risks in study design

- Eligibility: patients with low likelihood of competing risk
- Stratification: consider factors affecting incidence of aGVHD as well as those affecting mortality
Statistical Analysis of GVHD-free survival

- Fix time point of interest (e.g. 100 days).
  - In absence of censoring prior to 100 days, reduces to binary data
  - Proportion who experience acute GVHD or death in first 100 days
  - Minimal/No censoring early in clinical trials of GVHD
  - Alternatively compare Kaplan-Meier estimates

Advantages:
- Simple, clinically interpretable outcome
- Simple analyses: Chi-squared tests, logistic regression
- Minimal assumptions (e.g. proportional hazards)

Disadvantages: Possibly less powerful
Statistical Analysis of GVHD-free survival

- **Time to Event Analysis**
  - Usual survival techniques apply
    - log-rank test
    - Cox proportional hazards model
  - Modestly more powerful
  - Assumption of proportional hazards

- Implications for study design
  - Most of events occur in first 100 days (non-constant hazard rate)
  - Exponential assumption with staggered entry and follow-up to end of trial leads to incorrect power calculations
Statistical Analysis of incidence of aGVHD

- Fix time point of interest (e.g. 100 days).
  - In absence of censoring prior to 100 days, reduces to binary data
  - Proportion who developed aGVHD grade II-IV out of all patients
  - If censoring can compare cumulative incidence estimates

- Advantages:
  - Simple, clinically interpretable outcome
  - Simple analyses: Chi-squared tests, logistic regression
  - Minimal assumptions (e.g. proportional hazards)

- Disadvantages: Possibly less powerful
Statistical Analysis of incidence of aGVHD

- **Time to Event Analysis**
  - Log-rank test, Cox proportional hazards model
    - compare cause-specific hazard functions of two groups
    - Given that you are alive without aGVHD today, what is the likelihood that you will develop aGVHD tomorrow?
  - Models for cause-specific hazard function may not correspond directly to cumulative incidence (Logan et al, BBMT, 2007)
    - Related to differences in other competing risk
  - Direct comparisons or models for cumulative incidence function
    - Gray’s test (Ann Stat, 1988)
    - Fine and Gray regression models (JASA, 1999)
    - Pseudovalue regression models (Klein and Andersen, 2005)
Other design considerations for prophylaxis studies

- Group sequential monitoring: appropriate due to short-term outcome
- Treatment discontinuation/crossover: Intention-to-treat analysis
- Initiation of treatment for grade I aGVHD (e.g. steroids)
  - Event/Failure
Phase II Prophylaxis Trials

- Cumulative incidence of aGVHD may be more appropriate as a primary endpoint
- Key toxicity endpoints, including mortality should be monitored
- Two-stage or multi-stage designs to monitor futility may be appropriate
- Single-arm study vs. randomized control arm
  - Good historical data on incidence of acute GVHD
  - Publications, CIBMTR, etc.
aGVHD Treatment Trials

- Multi-state Model
- Defining an appropriate endpoint
- Treatment Discontinuation/Crossover
- Statistical Analysis
- Design issues
Multistate Model: What happens after treatment for aGVHD?

- Additional states overlaid on these
  - Flares which may or may not require systemic therapy
  - Development of chronic GVHD
Selection of primary endpoint

- **Dichotomization of states:**
  - **Success:** Alive in CR at day 56, no addition of systemic agents, no cGVHD

- **Fixed point vs. continuous time**
Statistical Analysis

- **Binary outcome**
  - Little to no censoring in early time period; data collection on GVHD response is feasible
  - Chi-square tests, logistic regression

- **If censoring:**
  - Multi-state model estimation (Aalen-Johansen estimate)
  - Pseudovalue techniques can be used to model state probabilities in the presence of censoring (Andersen et al., Biometrika 2003)
Treatment
Discontinuation/Crossover

- Discontinuation due to toxicity
  - Should this be considered a failure?
- Discontinuation due to ineffectiveness to treat GVHD
  - Failure
- Failure contingent on addition of new systemic therapy
- May be difficult to assess reason for discontinuation or addition of new therapy
Impact of Treatment Discontinuation/Addition of New Therapy

- If Discontinuation/Crossover not considered a failure
  - Treatment effect is diminished through ITT analysis
  - Greater sample size required to maintain power (n=532 vs. 340 for 80% power)
- Differences in toxicity rates not explicitly penalized

<table>
<thead>
<tr>
<th>Group</th>
<th>% CR (if no crossover)</th>
<th>% crossover</th>
<th>% CR overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>50%</td>
<td>10%</td>
<td>48.5%</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td></td>
<td>47.0%</td>
</tr>
<tr>
<td>Control</td>
<td>35%</td>
<td>10%</td>
<td>36.5%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td></td>
<td>35.7%</td>
</tr>
</tbody>
</table>
Impact of Treatment Discontinuation/Addition of New Therapy

- If Discontinuation/Addition of new therapy is considered a failure
- Net effect of treatment on GVHD and toxicity
- Power for n=340 patients drops from 80% to 25% in scenario below with different toxicity rates

<table>
<thead>
<tr>
<th>Group</th>
<th>% CR (if not crossover)</th>
<th>% crossover</th>
<th>% Success (CR, No new trt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>50%</td>
<td>10%</td>
<td>45.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>35%</td>
<td>10%</td>
<td>31.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>33.2%</td>
</tr>
</tbody>
</table>
Phase III Design Issues

- Group sequential monitoring: Appropriate due to short-term outcome

- Blinding
  - Many investigator decisions on GVHD treatment not related to treatment of interest
  - Difficult to constrain these across multiple centers
    - center heterogeneity
    - limited numbers of eligible patients
Phase II Design Issues

- Two-stage or multiple stage design for futility
- Historical control rate
  - No consensus outcome or time point in literature
  - May be difficult to get for complex outcome
- Use of concurrent randomized control arm
- Seamless Phase II/III design
Randomized control arm

- Problems with use of historical control rate (Vickers et al., Clin Can Res 2007)
  - Different patient population
  - Poor or no estimate of historical outcome
  - Poor documentation of historical outcome in publication
- Account for sampling variation in historical data: Bayesian or frequentist
- Consider adjusting phase II results to account for differences in case mix
Randomized control arm

- Randomized control arm vs. historical control
- Tradeoff between bias and variability
  - Randomized control arm: No selection bias, but greater sampling variability (2 sample test)
  - Historical control arm: Possible selection bias, but less sampling variability (1 sample test)
**Randomized control arm: Example**

- Assume 35% d28 CR rate as primary outcome
- Margins of error from confidence intervals get wider

<table>
<thead>
<tr>
<th>n/group</th>
<th>Fixed Historical control</th>
<th>Randomized control</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>+/-15%</td>
<td>+/-21%</td>
</tr>
<tr>
<td>60</td>
<td>+/-13%</td>
<td>+/-17%</td>
</tr>
<tr>
<td>80</td>
<td>+/-11%</td>
<td>+/-15%</td>
</tr>
</tbody>
</table>
In Summary

- There are important competing risks to be considered in GVHD prophylaxis trials.
  - Cumulative incidence of aGVHD is statistically more acceptable than 1-KM
  - Alternative may be composite endpoint (GVHD-free survival)
  - Limit impact of competing risk by eligibility and stratification

- For Phase II prophylaxis studies
  - Consider multistage designs to minimize futility
  - Use a control arm only in the absence of good historical data

- Considerations for GVHD treatment trials
  - How does drug toxicity enter into the endpoint
  - How does post failure treatment enter into the endpoint
  - How do later flares enter into the endpoint
  - How does chronic GVHD enter into the endpoint
  - When is a control arm not needed