Defining Response in GVHD Treatment Trials

Paul Carpenter, MB.BS.
Associate Member, FHCRC
Associate Professor, UW

- Disclosures:
  - Contract Support for clinical trials
    - Novartis: Nilotinib for relapse prophylaxis after HCT
What is Clinical Benefit in Acute GVHD Treatment Trials?

- **Direct Benefit:**
  - Prompt reversal of Grade III/IV symptoms clearly benefits the patient

- **Survival Benefit:**
  - Complete resolution of GVHD may be neither durable nor associated with long survival
  - Potent GVHD therapies that achieve high CR rates may increase infection mortality rates
  - “Minimal” residual acute GVHD activity may be compatible with long-term survival
Study Question

Can response to GVHD therapy be used as a surrogate for survival in order to demonstrate clinical benefit?
“Not too hot, not too cold, this one’s just right”: is how Goldilocks would pick GVHD therapy

What is the distribution of response rates among those destined to survive?
Is strict CR the best predictor?

Study Hypotheses

- GVHD responses that are more predictive of 6 month survival should consider:
  1. Relaxation of the requirement for CR because minimal residual GVHD activity is not clearly deleterious.
  2. Whether response durability improves prediction
  3. Whether steroid dose predicts survival

- Why choose 6-month survival?
  1. Survival analysis makes fewer assumptions than NRM
  2. Deaths < 6 months are often due to acute GVHD
  3. Deaths > 6 months have multiple causes (esp, relapse)
Study Design

- Additional data collection performed on a subgroup from a recently published large retrospective cohort (Mielcarek et al, Blood 2009)
- Detailed GVHD organ staging at Days 7, 28 and 56 after start of GVHD therapy
- Assigned complete response (CR) at each time-point
- Also defined “very good partial response” (VGPR)
- Detailed steroid dose data were available for this cohort.
Distinguishing VGPR from CR based on Organ Staging

<table>
<thead>
<tr>
<th>Stage</th>
<th>Skin BSA (%)</th>
<th>Liver Bilirubin (mg/dL)</th>
<th>Gut stool (L/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>&lt; 2.0</td>
<td>≤ 0.5</td>
</tr>
<tr>
<td>1</td>
<td>&lt;25</td>
<td>2.0 - 2.9</td>
<td>&gt; 0.51</td>
</tr>
<tr>
<td>2</td>
<td>25-50</td>
<td>3.0 – 5.9</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>3</td>
<td>&gt;50</td>
<td>6.0 -14.9</td>
<td>&gt;1.5</td>
</tr>
<tr>
<td>4</td>
<td>Bullae</td>
<td>≥ 15.0</td>
<td>&gt;2</td>
</tr>
</tbody>
</table>

VGPR included:
1. Minor rash resolving, evanescent, but not progressive
2. Minor stable or resolving elevations of bilirubin
3. Minimal but stable or resolving GI symptoms

1 – or persistent anorexia, nausea, vomiting
2 – or severe abdominal pain ileus

5/19/09 FDA AGVHD
Study Cohort

- Began with data set of Mielcarek et al (Blood 2009; 113:2888-2894)
- N = 733 adults transplanted 2000-2005
- N = 248 selected for further data collection because acute GVHD was diagnosed before Day 30
- Day 30 criterion balances the need to:
  1. Maximize probability for capturing response data at Days 7, 28 and 56.
  2. Minimize bias from including subjects whose stay in Seattle was prolonged by illness
Number of subjects evaluable for response at each time-point

<table>
<thead>
<tr>
<th>Day</th>
<th>Evaluable (N)</th>
<th>Not Evaluable (N)</th>
<th>Reason Not Evaluable</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>236</td>
<td>12</td>
<td>7 Missing data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 Relapse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 Deaths</td>
</tr>
<tr>
<td>28</td>
<td>225</td>
<td>23</td>
<td>4 Missing data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 Relapse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18 Deaths</td>
</tr>
<tr>
<td>56</td>
<td>204</td>
<td>44</td>
<td>3 Missing data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 Relapse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 Second transplant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>37 Deaths</td>
</tr>
</tbody>
</table>
## Causes of Death at each time-point

<table>
<thead>
<tr>
<th>Day</th>
<th>Deaths, N</th>
<th>Cause</th>
</tr>
</thead>
</table>
| 7   | 4         | 3 Conditioning toxicity ± Infection  
1 GVHD + conditioning toxicity |
| 28  | 18        | 6 Conditioning toxicity ± Infection  
1 GVHD  
1 GVHD + conditioning toxicity  
1 Relapse  
8 Infection  
1 Other (Cerebral hemorrhage) |
| 56  | 37        | 7 Conditioning toxicity ± Infection  
6 GVHD  
2 GVHD + conditioning toxicity  
3 GVHD + infection  
13 Infection  
4 Relapse  
2 Other (Cerebral hemorrhage, Idiopathic pneumonitis) |
We compared subjects with either CR or CR+VGPR to subjects who were non-responders or less than very good partial responders.
A 2 x 2 type analysis was performed.

**Classic 2 x 2**

<table>
<thead>
<tr>
<th>POS test</th>
<th>With Disease</th>
<th>Without Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease &amp; POS test</td>
<td>Disease &amp; POS test</td>
<td>No Disease &amp; POS test</td>
</tr>
<tr>
<td>a = true +ve</td>
<td>b = false +ve</td>
<td></td>
</tr>
<tr>
<td>NEG test</td>
<td>Disease &amp; NEG test</td>
<td>No Disease &amp; NEG test</td>
</tr>
<tr>
<td>c = false -ve</td>
<td>d = true -ve</td>
<td></td>
</tr>
</tbody>
</table>

**GVHD Response “Test” 2 x 2**

<table>
<thead>
<tr>
<th>CR</th>
<th>Dead</th>
<th>Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not in CR</td>
<td>Dead &amp; not in CR</td>
<td>Alive &amp; not in CR</td>
</tr>
<tr>
<td></td>
<td>a = true +ve</td>
<td>b = false +ve</td>
</tr>
<tr>
<td>in CR</td>
<td>Dead &amp; in CR</td>
<td>Alive &amp; in CR</td>
</tr>
<tr>
<td></td>
<td>c = false -ve</td>
<td>d = true -ve</td>
</tr>
</tbody>
</table>

*A positive “not in CR” test predicts for death at 6 months.*
GVHD Response “ROC” plot

**Sensitivity**

\[ \text{Sensitivity} = \frac{a}{a+c} \]

= proportion of dead without CR
= “Correctly Predicted Dead”

**1 minus Specificity**

\[ 1 - \text{Specificity} = \frac{b}{b+d} \]

= proportion of survivors without CR
= “Falsely Predicted Dead”

<table>
<thead>
<tr>
<th>Dead &amp; not in CR</th>
<th>Alive &amp; not in CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Dead &amp; in CR</td>
<td>Alive &amp; in CR</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>
Day 7 CR or CR/VGPR “Tests” Lack Sensitivity and Specificity for Predicting Death at 6 Months

All patients
(includes deaths before endpoint)

Survivors
CR Test at Day 28 or 56 Has Sensitivity but Lacks Specificity

All patients

Survivors

Day 56 – red    Day 28 – blue    Day 7 - black
VGPR/CR Test at Day 28 or 56 Improve Specificity with Loss of Sensitivity

All patients

Survivors

Day 56 – red  Day 28 – blue  Day 7 – black
Durable Day 28/56 VGPR/CR Tests Improve Sensitivity with Little Loss of Specificity

All patients

Survivors

Day 56 – red   Day 28 – blue   Day 7 - black
Steroid Dose or % Dose Reduction at Day 28 or 56 was not Predictive of Death at 6 Months
Conclusions

- Poor surrogates for 6 month survival include:
  - Day 7 CR or VGPR response rates
  - Day 28 or 56 CR or durable CR rates
- Durable VGPR/CR response has marginal potential as a surrogate for survival
- A critical unknown is how a treatment that changes response will change 6 month mortality via other downstream effects
- Difference between Minnesota and Seattle results might relate to definition of PR.
- Where do PRs best fit in the analysis?
Future Directions

- Prospective evaluate durable VGPR/CR as an endpoint in clinical trials because it may be more useful than strict CR.
- Need to better understand the relationship between beneficial effects (response induction) of GVHD treatments and harmful effects that can occur downstream of CR/VGPR (opportunistic infection, relapse, chronic GVHD).
Acknowledgements

- Barry Storer, PhD
- Marco Mielcarek, MD
- Paul Martin, MD