Weibull Prediction of Event Times
in Randomized Clinical Trials

Gui-shuang Ying, Ph.D.

Assistant Professor of Ophthalmology
University of Pennsylvania, School of Medicine

October 27, 2010
Introduction

- Interim analysis:
  - Data analysis performed prior to the completion of the trial
  - Monitor safety and efficacy of trial
  - Hope to stop as soon as convincing data arise

- When do interim analyses:
  - Calendar time
  - # of events

- Departure from interim analysis schedule:
  - Injure trial’s credibility
  - Inflate type I error (Proschan MA, 1992)
Example: The REMATCH Trial

- Compare left ventricular assist device to medical therapy for end-stage heart failure

- Design:
  - Enroll $N=140$ patients to get 92 deaths
  - Analyze all-cause mortality by logrank test

- Interim analysis plan:
  - Analyses after 23, 46, 69 and 92 deaths
  - O’Brien-Fleming boundary
How to plan interim analysis and schedule DSMB meetings when landmark time is random?
Real-Time Prediction

- Use the data from ongoing trial itself

- Prediction can be updated frequently as data accumulating

- Potentially more realistic and accurate

- Prediction interval available to reflect the uncertainty of prediction
Prediction Approaches

- Exponential prediction proposed by Bagiella & Heitjan (Statistics in Medicine 2001; 20:2055-63)
  - Exponential survival, constant Poisson enrollment
  - Simple, convenient, and potentially efficient

- Nonparametric prediction proposed by Ying, Heitjan & Chen (Clinical Trial 2004; 1:352-61)
  - Based on Kaplan-Meier survival estimator
  - Robust to distribution assumptions

- Weibull prediction proposed by Ying & Heitjan (Pharmaceutical Statistics, 2008; 7:107-120)
Why Weibull Prediction?

- Exponential predictions require strong distribution assumption, can be biased
- Nonparametric predictions can be less efficient than exponential prediction
- Weibull survival model
  - Widely used in survival analysis
  - Works well for the long-tailed survival data
  - Compromise approach between exponential and nonparametric prediction
Notations

- Data elements:
  - 2 treatment arms, \( j = 1, 2 \)
  - Enrollment start at calendar time 0
  - \( t_0 = \) current calendar time when we make prediction
  - \( t = \) some time in the future, \( t > t_0 \)
  - \( e_{ji} = \) enrollment time of subject \( i(j) \)
  - \( c_{ji} = \) loss follow-up time of subject \( i(j) \) from randomization
  - \( t_{ji} = \) event time of subject \( i(j) \) from randomization
  - \( t_{end} = \) enrollment end time, pre-specified or estimated

- More notations:
  - \( N_j(t) = \) # subjects enrolled in group \( j \) by time \( t \), \( N(t) = N_1(t) + N_2(t) \)
  - \( D_j(t) = \) # events in group \( j \) by time \( t \), \( D(t) = D_1(t) + D_2(t) \)
  - \( C_j(t) = \) # loss of follow-up in group \( j \) by time \( t \), \( C(t) = C_1(t) + C_2(t) \)
  - \( Y_{ji}(t) = \) indicator whether subject is at risk at time \( t \), 1=Yes
  - CDF for survival in group \( j \) is \( F_j \), density is \( f_j \)
  - CDF for loss of follow-up in group \( j \) is \( G_j \), density is \( g_j \)
3 Models for Prediction

- Model for time to enrollment
- Model for time from enrollment to event
- Model for time from enrollment to loss of follow-up
3 Components of Predicting # of Events

• First piece: \( D(t_0) = \# \) events occurred by \( t_0 \)

• Second piece: \( \# \) events expected to occur among subjects enrolled and at risk of failure
  - \( Q(t_0, t) = Q_1(t_0, t) + Q_2(t_0, t) \)

• Third piece: \( \# \) events expect to occur among subjects to be enrolled
  - \( R(t_0, t) = R_1(t_0, t) + R_2(t_0, t) \)

• Expected \( \# \) events by time \( t \) given experience to time \( t_0 \)
  - \( ED(t \mid t_0) = D(t_0) + Q(t_0, t) + R(t_0, t) \)
Point Prediction

- Let:
  - \( D^* = \) landmark event number
  - \( t^* = \) predicted landmark time

- Straightforward prediction:
  - Solution of the following equation with respect to \( t^* \)
    \[
    D^* = \hat{ED}(t_0, t^*) = D(t_0) + \hat{Q}(t_0, t^*) + \hat{R}(t_0, t^*)
    \]
General Expression for $Q$ and $R$

- General expression for $Q_j$:

$$Q_j(t_0, t) = \sum_{i=1}^{N_j(t_0)} Y_{ji}(t_0) \frac{[F_j(t - e_{ji}) - F_j(t_0 - e_{ji})] - \int_{t_0 - e_{ji}}^{t - e_{ji}} G_j(u) f_j(u) du}{[1 - F_j(t_0 - e_{ji})][1 - G_j(t_0 - e_{ji})]}$$

- General expression for $R_j$:

$$R_j(t_0, t) = \frac{\mu}{2} \int_0^{\min(t_{end}, t) - t_0} \left\{ \int_0^{t - t_0 - u} f_j(s)(1 - G_j(s)) ds \right\} du$$
Assumptions for Weibull Prediction

• Enrollment follows Poisson with rate $\mu$

• Survival in group $j$ is Weibull with parameters ($\alpha_j, \beta_j$):
  - CDFs: $F_j(t) = 1 - \exp(-\beta_j t^{\alpha_j})$
  - Densities: $f_j(t) = \alpha_j \beta_j t^{\alpha_j-1} \exp(-\beta_j t^{\alpha_j})$

• Loss of follow-up in group $j$ is Weibull with parameters ($\lambda_j, \gamma_j$):
  - CDFs: $G_j(t) = 1 - \exp(-\gamma_j t^{\lambda_j})$
  - Densities: $g_j(t) = \lambda_j \gamma_j t^{\lambda_j-1} \exp(-\gamma_j t^{\lambda_j})$
Priors

- Prior for enrollment rate:
  \[ \mu \mid (A, B) \sim \Gamma(A, B) \]

- Priors for \((\alpha_j, \beta_j)\) of Weibull event time distributions:
  - \(\alpha_j \sim \Gamma(u_{\alpha_j}, v_{\alpha_j})\)
  - \(\beta_j \sim \Gamma(u_{\beta_j}, v_{\beta_j})\)

- Priors for \((\lambda_j, \gamma_j)\) of Weibull loss time distributions:
  - \(\lambda_j \sim \Gamma(u_{\lambda_j}, v_{\lambda_j})\)
  - \(\gamma_j \sim \Gamma(u_{\gamma_j}, v_{\gamma_j})\)
Posterior Distributions

- Posterior for enrollment rate:
  $\mu \sim \Gamma(A + N(t_0), B + t_0)$

- Posterior for Weibull distributions:

  $p(\alpha_j, \beta_j) \propto L(\alpha_j, \beta_j) \times \pi(\alpha_j, \beta_j)$
  
  $= (\alpha_j \beta_j)^{D_j(t_0)} \left\{ \prod_{i=1}^{D_j(t_0)} t_{ji} \right\}^{\alpha_j - 1} \exp \left\{ -\beta_j \sum_{i=1}^{N_j(t_0)} t_{ji} \alpha_j \right\}$
  $\times \alpha_j^{u\alpha_j - 1} e^{-v\alpha_j \alpha_j} \times \beta_j^{u\beta_j - 1} e^{-v\beta_j \beta_j}$.  
  (1)
Approximation of the Posterior Distributions

- Construct a first-stage approximation to the posterior of the Weibull parameters
  - Centered at Bayesian mode
  - Dispersion matrix equal to the inverse of curvature of the log posterior at the mode

- Generate the parameter values from multivariate $t$-distribution
  - small degree of freedom ($\nu=4$)
  - location and dispersion as in first step

- Improve the approximate posterior by Sampling Importance Resampling (SIR)
  - Sampling weight $w(\alpha_j, \beta_j) = q(\alpha_j, \beta_j)/t(\alpha_j, \beta_j)$
  - $q(\alpha_j, \beta_j)$ is unnormalized posterior density
  - $t(\alpha_j, \beta_j)$ is the approximating multivariate $t$ density.
Algorithm for Weibull Prediction

- A three-step algorithm:
  1. Sample from the posterior of $\mu$ and $(\alpha_j, \beta_j, \lambda_j, \gamma_j), j = 1, 2$
  2. Given the current data and sampled parameters, complete the data:
     - Enrollment, failure and loss times for new subjects (if any)
     - Failure and loss times for subjects still in the study
  3. With each subject has time to event and time to loss:
     - determine the each subject’s status
     - rank the event times and find $T^*$ corresponding $D^*$th event

- Repeat $B$ times to generate the distribution of $T^*$
- Point prediction of landmark date is the median
- $100(1 - \alpha)$ prediction intervals are $\alpha/2$ and $1 - \alpha/2$ quantiles
Simulation Study

- Distributions for scenario 1:
  - Time to event: Treated $\sim$ Weibull(2, 3.76), control $\sim$ Weibull(2, 2.50)
  - Time to loss: Both groups $\sim$ Weibull(2, 11.3)

- Distributions for scenario 2:
  - Time to event: Treated $\sim$ Gamma(1.75, 1), control $\sim$ Gamma(3.50, 1)
  - Time to loss: Both groups $\sim$ Gamma(5.45, 1)

- Distribution for scenario 3:
  - Time to event: Treated $\sim$ Lognormal(0.70, 1), control $\sim$ Lognormal(0.30, 1)
  - Time to loss: Both groups $\sim$ Lognormal(2.70, 1)

- Predictions:
  - Landmark times of $128^{th}$
  - Prediction performed every half year since enrollment began
### Results from Weibull Distributions

<table>
<thead>
<tr>
<th>$t_0$</th>
<th>$n$</th>
<th>Median Interval Length</th>
<th>Coverage Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Weibull</td>
<td>Nonparametric</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>15.2</td>
<td>$Inf$</td>
</tr>
<tr>
<td>12</td>
<td>500</td>
<td>11.0</td>
<td>25.8</td>
</tr>
<tr>
<td>18</td>
<td>500</td>
<td>8.17</td>
<td>17.8</td>
</tr>
<tr>
<td>24</td>
<td>500</td>
<td>5.62</td>
<td>11.5</td>
</tr>
<tr>
<td>30</td>
<td>500</td>
<td>3.28</td>
<td>5.27</td>
</tr>
</tbody>
</table>
# Results from Gamma Distributions

<table>
<thead>
<tr>
<th>$t_0$</th>
<th>$n$</th>
<th>Median Interval Length</th>
<th>Coverage Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Weibull</td>
<td>Nonparametric</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>14.5</td>
<td>$\infty$</td>
</tr>
<tr>
<td>12</td>
<td>500</td>
<td>10.2</td>
<td>21.1</td>
</tr>
<tr>
<td>18</td>
<td>500</td>
<td>6.76</td>
<td>13.8</td>
</tr>
<tr>
<td>24</td>
<td>500</td>
<td>4.17</td>
<td>6.89</td>
</tr>
<tr>
<td>30</td>
<td>339</td>
<td>1.61</td>
<td>1.81</td>
</tr>
</tbody>
</table>
### Results from Lognormal Distributions

<table>
<thead>
<tr>
<th>$t_0$</th>
<th>n</th>
<th>Median Interval Length</th>
<th>Coverage Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Weibull</td>
<td>Nonparametric</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>12.0</td>
<td>17.6</td>
</tr>
<tr>
<td>12</td>
<td>500</td>
<td>7.30</td>
<td>12.5</td>
</tr>
<tr>
<td>18</td>
<td>500</td>
<td>4.32</td>
<td>6.06</td>
</tr>
<tr>
<td>24</td>
<td>451</td>
<td>1.77</td>
<td>1.94</td>
</tr>
</tbody>
</table>
Illustration: Chronic Granulomatous Disease (CGD) Study

- RCT to compare γ-IFN with placebo in treatment of CGD

- Design:
  - Outcome: time to first infection
  - Planned an interim analysis 6 months after half subjects enrolled
  - Stop if nominal p<0.0036 (O’Brien-Fleming boundary)

- History:
  - Aug. 27, 1988 - March 1989, 128 patients enrolled and randomized
  - Aug. 15, 1989: 35th events
  - 3 patients loss of follow-up
## Progress of CGD Study

<table>
<thead>
<tr>
<th>Time (t0)</th>
<th>Number of Enrollments</th>
<th>Number of Events</th>
<th>LogRank Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo Treatment</td>
<td>Placebo Treatment</td>
<td></td>
</tr>
<tr>
<td>10/26/88</td>
<td>9 9</td>
<td>2 0</td>
<td>0.1063</td>
</tr>
<tr>
<td>11/25/88</td>
<td>19 17</td>
<td>3 0</td>
<td>0.0630</td>
</tr>
<tr>
<td>12/25/88</td>
<td>32 35</td>
<td>4 0</td>
<td>0.0319</td>
</tr>
<tr>
<td>01/24/89</td>
<td>44 45</td>
<td>4 0</td>
<td>0.0281</td>
</tr>
<tr>
<td>02/23/89</td>
<td>50 57</td>
<td>10 1</td>
<td>0.0027</td>
</tr>
<tr>
<td>03/25/89</td>
<td>65 63</td>
<td>11 2</td>
<td>0.0054</td>
</tr>
<tr>
<td>04/24/89</td>
<td>65 63</td>
<td>13 3</td>
<td>0.0017</td>
</tr>
<tr>
<td>05/05/89</td>
<td>65 63</td>
<td>14 4</td>
<td>0.0045</td>
</tr>
<tr>
<td>05/24/89</td>
<td>65 63</td>
<td>16 5</td>
<td>0.0042</td>
</tr>
<tr>
<td>06/23/89</td>
<td>65 63</td>
<td>18 6</td>
<td>0.0037</td>
</tr>
<tr>
<td>07/23/89</td>
<td>65 63</td>
<td>21 6</td>
<td>0.0006</td>
</tr>
<tr>
<td>08/15/89</td>
<td>65 63</td>
<td>24 11</td>
<td>0.0027</td>
</tr>
</tbody>
</table>
Predictions for CGD Study

- Prediction plan:
  - Monthly prediction of landmark times of 18th and 35th event

- Priors:
  - Enrollment rate: $\mu \sim \Gamma(30, 15)$
  - Event in placebo arm: $\alpha_0 \sim \Gamma(1.5, 1), \beta_0 \sim \Gamma(2426, 1)$
  - Event in $\gamma$-IFN arm: $\alpha_1 \sim \Gamma(1.5, 1), \beta_1 \sim \Gamma(808, 1)$
  - Loss of follow-up in both arms: $\gamma \sim \Gamma(1.5, 1), \lambda \sim \Gamma(4043, 1)$
Predictions of Final Analysis Date

Date of Prediction
02/23/89
05/24/89
08/22/89
11/20/89
02/18/90
05/19/90
08/17/90
11/15/90
02/13/91
05/14/91
08/12/91
11/10/91
02/08/92
05/08/92
08/06/92
11/04/92

Projection Method:
Weibull
Nonparametric
<table>
<thead>
<tr>
<th>Date of Prediction</th>
<th>Predicted Landmark Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/26/88</td>
<td></td>
</tr>
<tr>
<td>11/25/88</td>
<td></td>
</tr>
<tr>
<td>12/25/88</td>
<td></td>
</tr>
<tr>
<td>01/24/89</td>
<td></td>
</tr>
<tr>
<td>02/23/89</td>
<td></td>
</tr>
<tr>
<td>03/25/89</td>
<td></td>
</tr>
<tr>
<td>04/24/89</td>
<td></td>
</tr>
<tr>
<td>05/24/89</td>
<td></td>
</tr>
<tr>
<td>06/23/89</td>
<td></td>
</tr>
<tr>
<td>07/23/89</td>
<td></td>
</tr>
<tr>
<td>08/22/89</td>
<td></td>
</tr>
</tbody>
</table>

Predictions of Interim Analysis Date

Date of Prediction

01/19/90
01/24/89
02/23/89
03/25/89
04/24/89
05/24/89
06/23/89
07/23/89
08/22/89

- **Weaker Prior**
- **Current Prior**
- **Stronger Prior**
Predictions of Final Analysis Date

Date of Prediction
Predicted Landmark Date

02/23/89
05/24/89
08/22/89
11/20/89
02/18/90
05/19/90
08/17/90
11/15/90
02/13/91
05/14/91
10/26/88
11/25/88
12/25/88
01/24/89
02/23/89
03/25/89
04/24/89
05/24/89
06/23/89
07/23/89
08/22/89

Weaker Prior
Current Prior
Stronger Prior
Conclusion

• Weibull Prediction:
  - Involve simulating future course of trial on enrollment, occurrence of events and losses to follow-up
  - Use both prior information and accumulated data from trial itself
  - Predict accurately and efficiently in Weibull and Gamma distributions
  - Potentially has greater application

• Predict other outcomes:
  - # of events at specific time
  - Predictive power
  - Optimal combination of enrollment and study length
References


- Qiang J, Stangl DK, George S. A Weibull model for survival data: Using prediction to decide when to stop a clinical trial. *Bayesian Biostatistics*, Edited by Berry DA, Stangl DK; 1996.
