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 requires 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
  - $ED(t \,|\, t_0) = D(t_0) + Q(t_0, t) + R(t_0, t)$

Point Prediction

- Let:
  - $D^\star$ = landmark event number
  - $t^\star$ = predicted landmark time
- Straightforward prediction:
  - Solution of the following equation with respect to $t^\star$

$$D^\star = \widehat{ED}(t_0, t^\star) = D(t_0) + \widehat{Q}(t_0, t^\star) + \widehat{R}(t_0, t^\star)$$

General Expression for Q and R

- General expression for Q_j:

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

- 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 μ
- Survival in group j is Weibull with parameters (α_j, β_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 (λ_j, γ_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 (α_j, β_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 (λ_j, γ_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:

$$
\begin{aligned}
p(\alpha_j, \beta_j) &\propto L(\alpha_j, \beta_j) \times \pi(\alpha_j, \beta_j) \\
&= (\alpha_j \beta_j)^{D_j(t_0)} \left\{ \prod_{i=1}^{D_j(t_0)} t_{ji} \right\}^{\alpha_j - 1} \exp\left\{ -\beta_j \sum_{i=1}^{N_j(t_0)} {t_{ji}}^{\alpha_j} \right\} \qquad (1) \\
&\times \alpha_j^{u_{\alpha_j} - 1} e^{-v_{\alpha_j} \alpha_j} \times \beta_j^{u_{\beta_j} - 1} e^{-v_{\beta_j} \beta_j}.
\end{aligned}
$$

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 (ν=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 μ 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 $T^\star$ corresponding $D^\star$th event

- Repeat B times to generate the distribution of $T^\star$

- 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

		Median Interval Length		Coverage Rate	
t_0	n	Weibull	Nonparametric	Weibull	Nonparametric
6	500	15.2	Inf	0.998	0.616
12	500	11.0	25.8	0.996	0.948
18	500	8.17	17.8	0.978	0.984
24	500	5.62	11.5	0.964	0.972
30	500	3.28	5.27	0.964	0.974

Results from Gamma Distributions

		Median Interval Length		Coverage Rate	
t_0	n	Weibull	Nonparametric	Weibull	Nonparametric
6	500	14.5	∞	0.996	0.850
12	500	10.2	21.1	0.984	0.972
18	500	6.76	13.8	0.982	0.976
24	500	4.17	6.89	0.960	0.988
30	339	1.61	1.81	0.956	0.968

Results from Lognormal Distributions

		Median Interval Length		Coverage Rate	
t_0	n	Weibull	Nonparametric	Weibull	Nonparametric
6	500	12.0	17.6	0.416	0.908
12	500	7.30	12.5	0.792	0.988
18	500	4.32	6.06	0.912	0.988
24	451	1.77	1.94	0.905	0.965

Illustration: Chronic Granulomatous Disease (CGD) Study

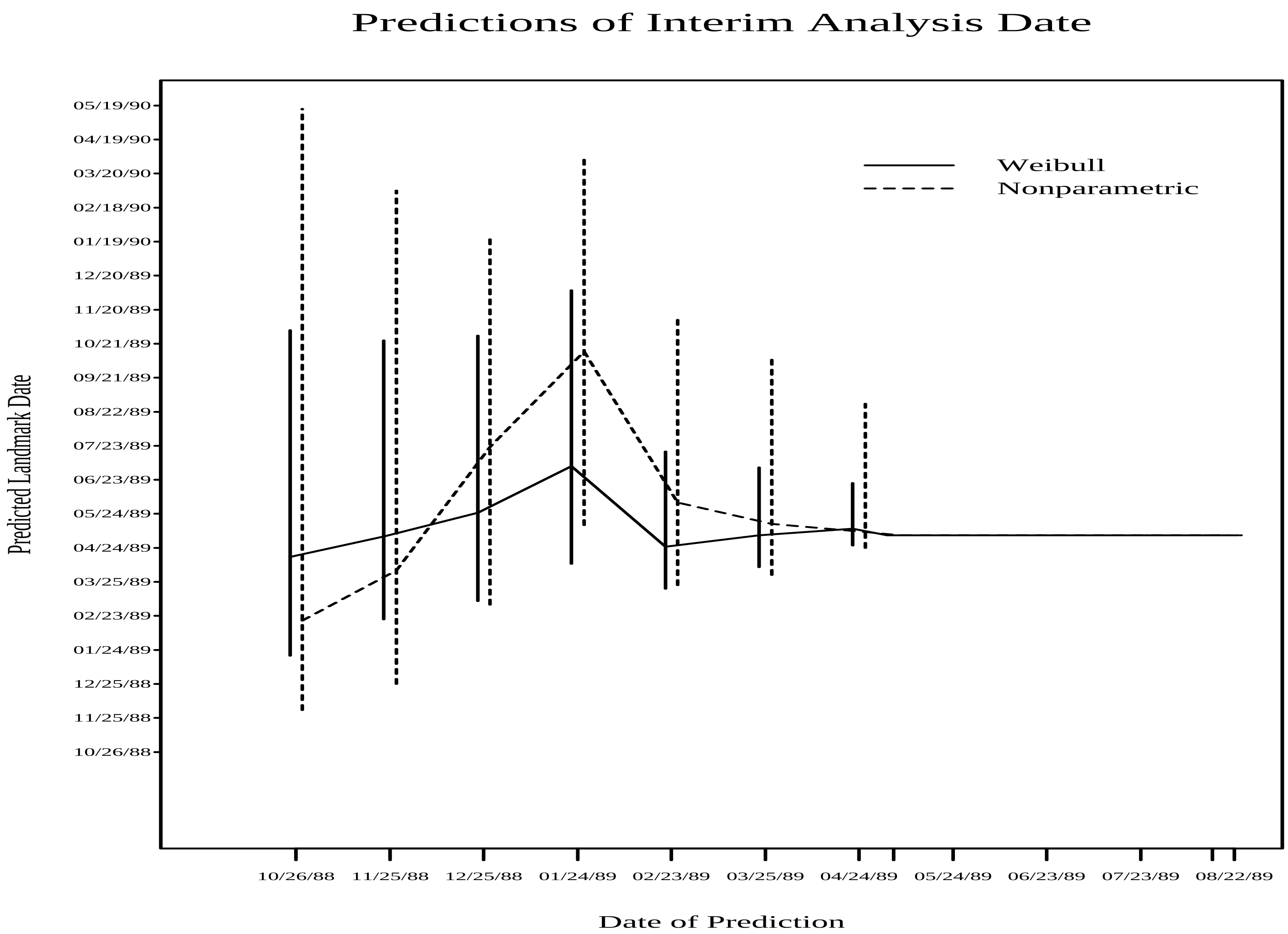
- RCT to compare γ-IFN with placebo in treatment of CGD
- Design:
  - Outcome: time to first infection
  - Planned an interim analysis 6 months after half subjects enrolled
  - Stop if nominal $p<0.0036$ (O'Brien-Fleming boundary)
- History:
  - Aug. 27, 1988 - March 1989, 128 patients enrolled and randomized
  - Aug. 15, 1989: 35^{th} events
  - 3 patients loss of follow-up

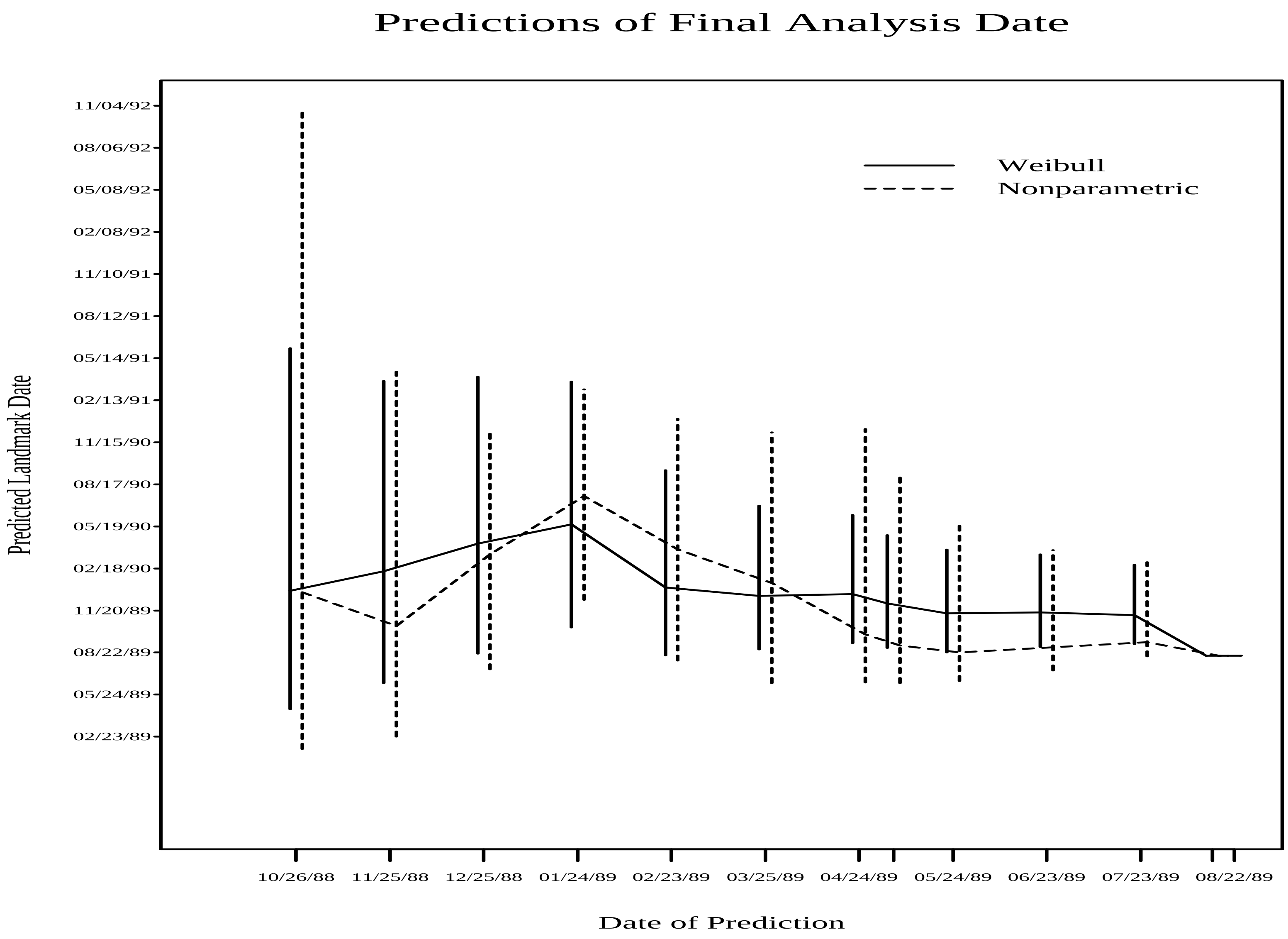
Progress of CGD Study

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

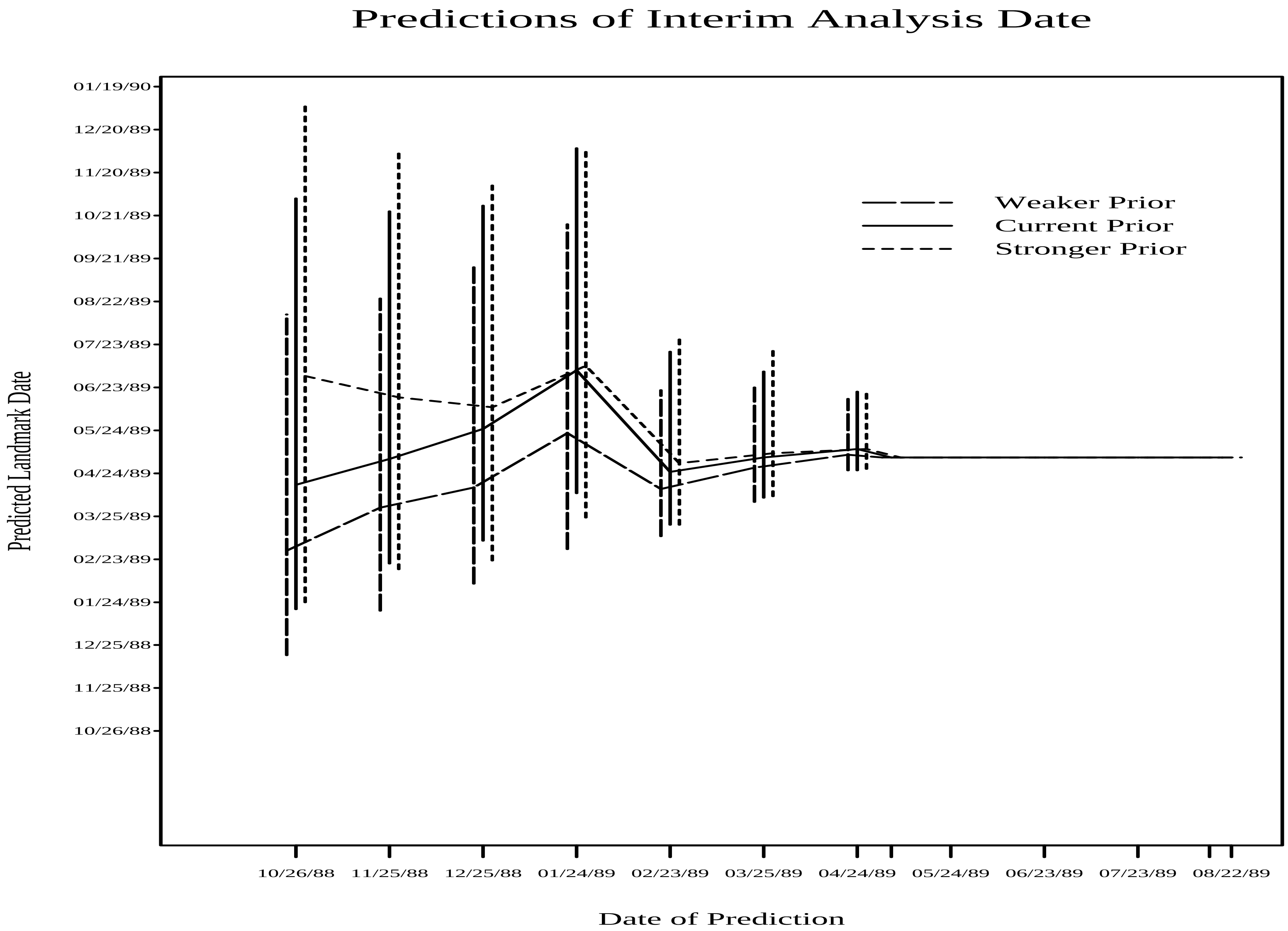
Predictions for CGD Study

- Prediction plan:
  - Monthly prediction of landmark times of 18^{th} and 35^{th} 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 γ-IFN arm: $\alpha_1 \sim \Gamma\ (1.5, 1)$, $\beta_1 \sim \Gamma(808, 1)$
  - Loss of follow-up in both arms: $\gamma \sim \Gamma(1.5, 1)$, $\lambda \sim \Gamma(4043, 1)$

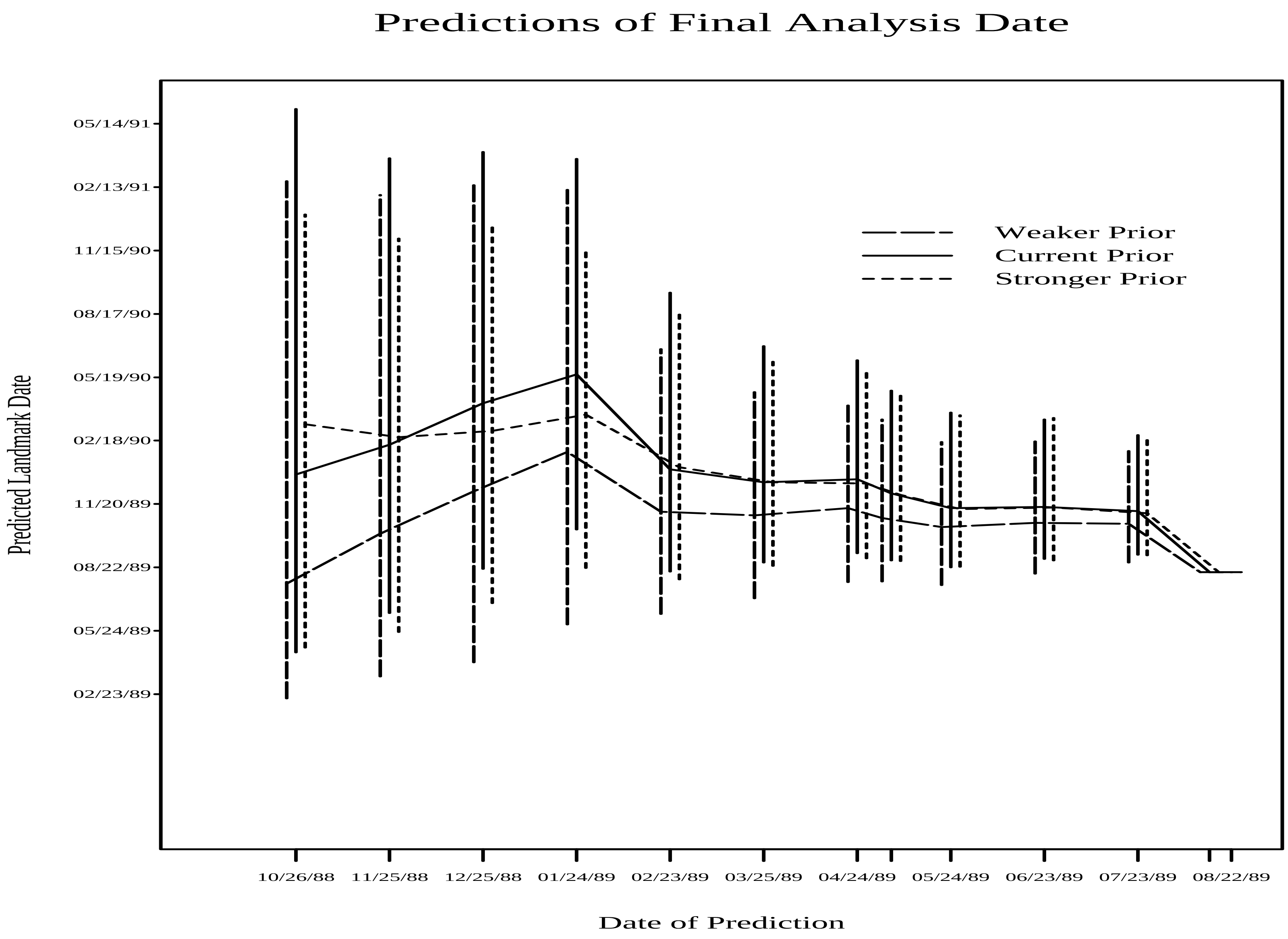

Predictions of Interim Analysis Date
Weibull
Nonparametric
Predicted Landmark Date
05/19/90
04/19/90
03/20/90
02/18/90
01/19/90
12/20/89
11/20/89
10/21/89
09/21/89
08/22/89
07/23/89
06/23/89
05/24/89
04/24/89
03/25/89
02/23/89
01/24/89
12/25/88
11/25/88
10/26/88
10/26/88
11/25/88
12/25/88
01/24/89
02/23/89
03/25/89
04/24/89
05/24/89
06/23/89
07/23/89
08/22/89
Date of Prediction



Predictions of Final Analysis Date
Weibull
Nonparametric
Predicted Landmark Date
11/04/92
08/06/92
05/08/92
02/08/92
11/10/91
08/12/91
05/14/91
02/13/91
11/15/90
08/17/90
05/19/90
02/18/90
11/20/89
08/22/89
05/24/89
02/23/89
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
Date of Prediction



Predictions of Interim Analysis Date
Predicted Landmark Date
01/19/90
12/20/89
11/20/89
10/21/89
09/21/89
08/22/89
07/23/89
06/23/89
05/24/89
04/24/89
03/25/89
02/23/89
01/24/89
12/25/88
11/25/88
10/26/88
Weaker Prior
Current Prior
Stronger Prior
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
Date of Prediction



Predictions of Final Analysis Date
05/14/91
02/13/91
11/15/90
08/17/90
05/19/90
02/18/90
11/20/89
08/22/89
05/24/89
02/23/89
Predicted Landmark Date
Weaker Prior
Current Prior
Stronger Prior
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
Date of Prediction

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

- Bagiella E, Heitjan DF. Predicting analysis times in randomized clinical trials. *Statistics in Medicine* 2001; 20:2055-63.
- Ying GS, Heitjan DF, Chen TT. Nonparametric prediction of event times in randmoized clinical trials. *Clinical Trials* 2004; 4:352-361.
- Ying GS, Heitjan DF. Weibull prediction of event times in randmoized clinical trials. *Pharmaceutical Statistics* 2008; 7:107-120.
- Proschan MA, Follmann DA, Waclawiw MA. Effects of assumption violations on type I error rate in group sequential monitoring. *Biometrics* 1992; 48:1131-1143.
- 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.
- Carroll KJ. On the use and utility of the Weibull model in analysis of survival data. *Controlled Clinical Trials* 2003; 24: 682-701.
- Rose EA et al. The REMATCH Trial: Rationale, design and endpoints. *Annals of Thoracic Surgery* 1999; 67:723-30.