Sample Size and Power Determination in A Joint Model of Longitudinal and Survival Data Supplement

Liddy M. Chen[1], Joseph G. Ibrahim[1,2], Haitao Chu[3]

[1]*Department of Biostatistics*
University of North Carolina
Chapel Hill, North Carolina 27599, U.S.A.
[3]*Division of Biostatistics*
University of Minnesota
Minneapolis, Minnesota 55455, U.S.A.

November 21, 2010

[2]To whom correspondence should be addressed.

Supplement I. Derivation of Sample Size Formula for Testing the Trajectory Effect

The sample size formula was derived from the score test following Schoenfeld's [1] framework. Let D denote the number of subjects who had the event in the clinical trial, and let N denote the number of subjects in the trial. Let T_i and C_i denote the event and censoring times, respectively; $S_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. Let Z_i be a treatment indicator, and let $X_i(u)$ be the longitudinal process (also referred to as the trajectory in the paper) at time $u \geq 0$. Define

$$e_i\{(X_k(S_i))^q\} = \frac{\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\beta X_k(S_i) + \hat{\alpha} Z_k\}(X_k(S_i))^q}{\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\beta X_k(S_i) + \hat{\alpha} Z_k\}}$$

and

$$G_i\{(X_k(S_i))^q\} = \frac{\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\hat{\alpha} Z_k\}(X_k(S_i))^q}{\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\hat{\alpha} Z_k\}},$$

where $X_k(u) = \theta_{0k} + \theta_{1k}u + \theta_{2k}u^2 + \cdots + \theta_{pk}u^p + \gamma Z_k$, and $q = 1, 2, \ldots$. For the hazard function $h(S) = \lambda_0(S)\exp\{\beta X(S) + \alpha Z\}$, the partial likelihood is given by

$$L_i = \left\{ \frac{\exp\{\beta X_i(S_i) + \alpha Z_i\}}{\sum_{k=1}^{N} I(S_k \leq S_i)\exp\{\beta X_k(S_i) + \alpha Z_k\}} \right\}^{\Delta_i}.$$

The score statistic for Cox's partial likelihood can be expressed as

$$S_{score} = \frac{N^{-\frac{1}{2}} \sum_{i \in D} X_i(S_i) - G_i\{X_k(S_i)\}}{\left\{N^{-1} \sum_{i \in D} G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2\right\}^{\frac{1}{2}}}.$$

Now, rewrite the score statistic as

$$\begin{aligned} S_{score} \quad = \quad & \frac{N^{-\frac{1}{2}} \sum_{i \in D} (X_i(S_i) - e_i\{X_k(S_i)\})}{\left\{N^{-1} \sum_{i \in D} G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2\right\}^{\frac{1}{2}}} \\ & + \frac{N^{-\frac{1}{2}} \sum_{i \in D} (e_i\{X_k(S_i)\} - G_i\{X_k(S_i)\})}{\left\{N^{-1} \sum_{i \in D} G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2\right\}^{\frac{1}{2}}}. \end{aligned}$$

$\sum_{i \in D} (X_i(S_i) - e_i\{X_k(S_i)\})$ is the score function of the partial likelihood, and thus, the numerator of the first term is asymptotically normal with mean 0 and variance

$N^{-1}\sum_{i\in D} e_i\{(X_k(S_i))^2\} - (e_i\{X_k(S_i)\})^2$. As in Schoenfeld [1] and Ewell & Ibrahim [2], consider alternatives, which are location shifts of known distribution functions, such that β is $O(n^{-\frac{1}{2}})$. As $e_i\{(X_k(S_i))^q\} \to G_i\{(X_k(S_i))^q\}$ when $\beta \to 0$, the first term $\to N(0,1)$ when $\beta \to 0$.

Expanding the numerator of the 2nd term in a Taylor's series about $\beta = 0$ shows that

$$\begin{aligned} &e_i\{X_k(S_i)\} - G_i\{X_k(S_i)\} \approx \\ &\beta\left\{G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2\right\}. \end{aligned}$$

The 2nd term approaches

$$\beta\left\{\sum_{i=1}^{D}\{G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2\right\}^{\frac{1}{2}}.$$

Since Z_k is a fixed treatment indicator and assuming that each treatment group is large,

$$\begin{aligned} G_i\{(X_k(S_i))^q\} &= \frac{\frac{1}{N}\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\hat{\alpha}Z_k\}(X_k(S_i))^q}{\frac{1}{N}\sum_{k=1}^{N} I(S_k \geq S_i)\exp\{\hat{\alpha}Z_k\}} \\ &\to \frac{E\left\{I(S_k \geq S_i)(X_k(S_i))^q\right\}}{E\left\{I(S_k \geq S_i)\right\}}. \end{aligned} \tag{1}$$

When $\beta \to 0$, S_k is independent of the θ_k's and $I(S_k \geq S_i)$ is independent of $X_k(S_i)$ conditional on S_i, thus (1) $\to E\left\{(X_k(S_i))^q\right\}$. Then

$$\begin{aligned} &G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2 \to \\ &E\left\{(X_k(S_i))^2\right\} - \left\{E(X_k(S_i))\right\}^2 = \mathrm{Var}\{X_k(S_i)\} \end{aligned}$$

as $\beta \to 0$. It follows that

$$\begin{aligned}
&\beta D^{\frac{1}{2}} \left\{ \frac{1}{D} \sum_{i \in D} \{G_i\{(X_k(S_i))^2\} - (G_i\{X_k(S_i)\})^2 \right\}^{\frac{1}{2}} \\
&\to \beta D^{\frac{1}{2}} \left\{ \frac{1}{D} \sum_{i \in D} \mathrm{Var}(X_k(S_i)) \right\}^{\frac{1}{2}} \\
&= \beta D^{\frac{1}{2}} \left\{ \frac{1}{D} \sum_{i \in D} (1 \ S_i \ \ldots \ S_i^p) \, \mathbf{\Sigma}_\theta \, (1 \ S_i \ \ldots \ S_i^p)^T \right\}^{\frac{1}{2}} \\
&= \beta D^{\frac{1}{2}} \left\{ \frac{1}{D} \sum_{i \in D} \mathbf{S_i} \ \mathbf{\Sigma}_\theta \ \mathbf{S_i^T} \right\}^{\frac{1}{2}}, \qquad (2)
\end{aligned}$$

where $\mathbf{\Sigma}_\theta$ is the covariance matrix of $(\theta_{0k} \ \theta_{1k} \ \ldots \ \theta_{pk})$. Note that

$$\frac{1}{D} \sum_{i \in D} S_i^q = \frac{N}{D} \frac{1}{N} \sum_{i \in D} T_i^q \to \mathrm{E}\left\{ I(T \le \bar{t}_f) T^q \right\} / \tau,$$

where $\tau = \frac{D}{N}$ is the event rate, and $\bar{t}_f$ is the mean follow-up time in all subjects. It is a truncated moment of T^q, as we do not observe all the T_i's. Therefore (2) above converges to

$$\beta \left\{ D \sigma_s^2 \right\}^{\frac{1}{2}},$$

where

$$\begin{aligned}
\sigma_s^2 &= \mathrm{Var}(\theta_{0k}) + \sum_{j=1}^{p} \mathrm{Var}(\theta_{jk}) \mathrm{E}\{ I(T \le \bar{t}_f) T^{2j} \} / \tau \\
&+ 2 \sum_{j=0}^{p} \sum_{l>j}^{p} \mathrm{Cov}(\theta_{jk}, \theta_{lk}) \mathrm{E}\{ (I \le \bar{t}_f) T^{j+l} \} / \tau, \qquad (3)
\end{aligned}$$

and p is the degree of polynomial in the trajectory. For example, when $p = 1$ (linear trajectory),

$$\begin{aligned}
\sigma_s^2 &= \mathrm{Var}(\theta_{0k}) + \mathrm{Var}(\theta_{1k}) \mathrm{E}\{ I(T \le \bar{t}_f) T^2 \} / \tau \\
&+ 2\mathrm{Cov}(\theta_{0k}, \theta_{1k}) \mathrm{E}\{ I(T \le \bar{t}_f) T \} / \tau.
\end{aligned}$$

Thus, the score statistic, S_{score}, is asymptotically normal with unit variance and mean equal to $\beta \left\{ D \sigma_s^2 \right\}^{\frac{1}{2}}$ as $D \to \infty$. It follows that the number of events required for a one-sided level $\tilde{\alpha}$ test with power $\tilde{\beta}$ is given by

$$D = \frac{(z_{\tilde{\beta}} + z_{1-\tilde{\alpha}})^2}{\sigma_s^2 \beta^2},$$

where σ_s^2 is defined in (3).

Supplement II. The Full Joint Modeling Approach Versus the Two-Step Inferential Approach

When the true trajectory is unknown, we examined two joint modeling approaches. The first one was a two-step inferential approach proposed by Tsiatis et. al. [3], which has been described in detail in previous sections. The second approach was based on the full joint likelihood as specified in (2.3) of the main paper. Wulfsohn and Tsiatis [4] developed an EM algorithm of the model to obtain the parameter estimates. Guo and Carlin [5] develop a fully Bayesian version and implemented it via Markov chain Monte Carlo (MCMC) methods using the WinBUGS software. We used a standard SAS procedure, NLMIXED, which fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects using a dual quasi-Newton algorithm (SAS Online Documentation for Version 9.1.3). Standard deviations for the estimates are based on the 2^{nd} derivatives of the log-likelihood function. Data was simulated based on a fully parametric exponential model with constant baseline hazard. The two-step inferential approach is based on Cox's partial likelihood. It is expected that the full joint modeling approach using exactly the same exponential model will have more efficiency over the partial likelihood model. However, in practice, we rarely use a fully parametric model with constant hazard to analyze the time-to-event data. To have a fair comparison of efficiency, we simulated survival data based on a piecewise exponential model with two time intervals in which the baseline hazard changed from λ_{01} to λ_{02} at time t_q. We used the same parameters in both periods, and therefore, β is the same. Five repeated measurements of the longitudinal data were simulated based on the θ's. The measurements were set to be missing after an event or censoring.

We show in Table 1 that the full joint modeling approach based on a parametric exponential model is more efficient than the two-step inferential approach based on Cox's partial likelihood. However, the full joint exponential model is sensitive to whether the baseline hazard is constant over time. When this is true, it yields an unbiased estimate of β but yields biased estimates of β when the constant baseline hazard assumption is violated. In this case, it overestimates the trajectory effect when the baseline hazard increases after time t_q. Furthermore, it underestimates the trajectory effect when the baseline hazard decreases after time t_q. The larger the difference between the two baseline hazards, the

larger the bias. The two-step inferential approach may be more robust, although less efficient. The impact is smaller when testing the overall treatment effect. Both approaches have similar efficiency, but the misspecified exponential joint model yields a slightly biased estimate of the treatment effect. This finding is not surprising, as it corresponds to known theory between parametric and semi-parametric modeling. A retrospective power analysis from a real study data is provided in Section 7 of the main paper.

[Table 1 about here.]

Wulfsohn and Tsiatis [4] found that the asymptotic standard error of $\hat{\beta}$ when using the joint estimation procedure is slightly larger than that from the two-step model. It was suggested that it might be because the random effects were assumed to be influenced by the uncertainty in the estimated trajectory parameters, and more variability is incorporated. Therefore, although the full joint estimation approach should be more efficient as compared to the two-step model, since it uses information more efficiently. This may not turn out to be the case in real data settings if the real data violate the modeling assumptions. Wulfsohn and Tsiatis [4] cited earlier work concerning biased estimates of the trajectory effect when using the two-step model (slightly towards the null) and suggested that the estimate from the joint model is further away from the null, and therefore more likely to reduce the bias. We found in the simulation studies that the trajectory effect can be over-estimated or under-estimated in the fully parametric joint model if the model assumptions, such as a constant baseline hazard in the case of the exponential model, is violated. The two-step model in this case may be more robust. Further studies are needed to compare the two joint modeling approaches and other parametric or semi-parametric models.

References

[1] Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983; **39**: 499-503.

[2] Ewell M, Ibrahim JG. The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis* 1997; **3**: 5-12.

[3] Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of American Statistical Association* 1995; **90**: 27-37.

[4] Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330-339.

[5] Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* 2004; **58**: 1-9.

Table 1: Comparison of the two-step inferential approach with the full joint modeling approach in testing β and the overall treatment effect

λ_{01}	λ_{02}		Two-Step Approach		Full Joint Modeling	
$[0,\ 0.75]$	$(0.75,\ \infty)$	Parameter	Estimates (StdErr)	Power	Estimates (StdErr)	Power
0.85	0.85	$\hat{\beta}$	0.203 (0.074)	79.9	0.206 (0.054)	96.4
		$\hat{\alpha}+\hat{\beta}\hat{\gamma}$	0.319 (0.162)	50.0	0.321 (0.163)	49.6
0.85	0.65	$\hat{\beta}$	0.204 (0.075)	78.4	**0.164** (0.053)	85.5
		$\hat{\alpha}+\hat{\beta}\hat{\gamma}$	0.322 (0.164)	50.8	0.328 (0.164)	51.5
0.85	0.45	$\hat{\beta}$	0.208 (0.077)	76.2	**0.108** (0.053)	52.0
		$\hat{\alpha}+\hat{\beta}\hat{\gamma}$	0.326 (0.167)	49.7	0.339 (0.167)	51.6
0.65	0.85	$\hat{\beta}$	0.208 (0.073)	80.10	**0.248** (0.054)	99.4
		$\hat{\alpha}+\hat{\beta}\hat{\gamma}$	0.323 (0.168)	49.1	0.318 (0.170)	45.7

Note: Estimates were based on 1000 simulations, each with 100 subjects per arm. Survival time was simulated with a piecewise exponential model, minimum follow-up time is 0.75 years (9 months), and maximum follow-up time is 2 years. $\alpha = 0.3$, $\gamma = 0.1$, $\beta = 0.2$, $\mathrm{E}(\theta_{0i}) = 0$, $\mathrm{E}(\theta_{1i}) = 3$, $\mathrm{Var}(\theta_{0i}) = 0.7$, $\mathrm{Var}(\theta_{1i}) = 1.2$, $\mathrm{Cov}(\theta_{0i}, \theta_{1i}) = 0.2$, and $\sigma_e^2 = 0.16$ A maximum of 5 repeated measurements were simulated with missing data after an event or censoring. Both analyses assumed an unknown Σ_θ.