


Testing Procedures for Group Sequential Analyses with Multiple Survival Endpoints

November 6, 2007

Rebecca Scherzer, Ph.D.

- 
- Research for this talk is related to dissertation work conducted at Western Michigan University, completed April 2006.
 - I am now at the San Francisco VA Medical Center, working on the FRAM Study:

Rebecca Scherzer, PhD
Principal Research Statistician
VAMC, Metabolism Section, 111F
4150 Clement Street
San Francisco, CA 94121
Phone: 415-221-4810 x3112
Fax: 415-750-6927
Email: rebecca.scherzer@va.gov



Outline of presentation

- Motivation for group sequential and multivariate survival methods
- Literature Review
- Research
 - Methods
 - Results of data example
 - Simulation results



Introduction

- What is Group Sequential (GS) Analysis?
- Group sequential methodology in clinical trials
 - Data monitoring committee
 - Motivation
 - Example: Long-term cardiac study, with multiple outcomes of interest



Introduction

- Why consider multiple outcomes?
 - More information than univariate => more power
=> possible to stop the trial earlier.
 - Why stop the trial earlier?

- Why consider survival data?
 - Special challenges with this type of data (not adequately addressed by existing literature).



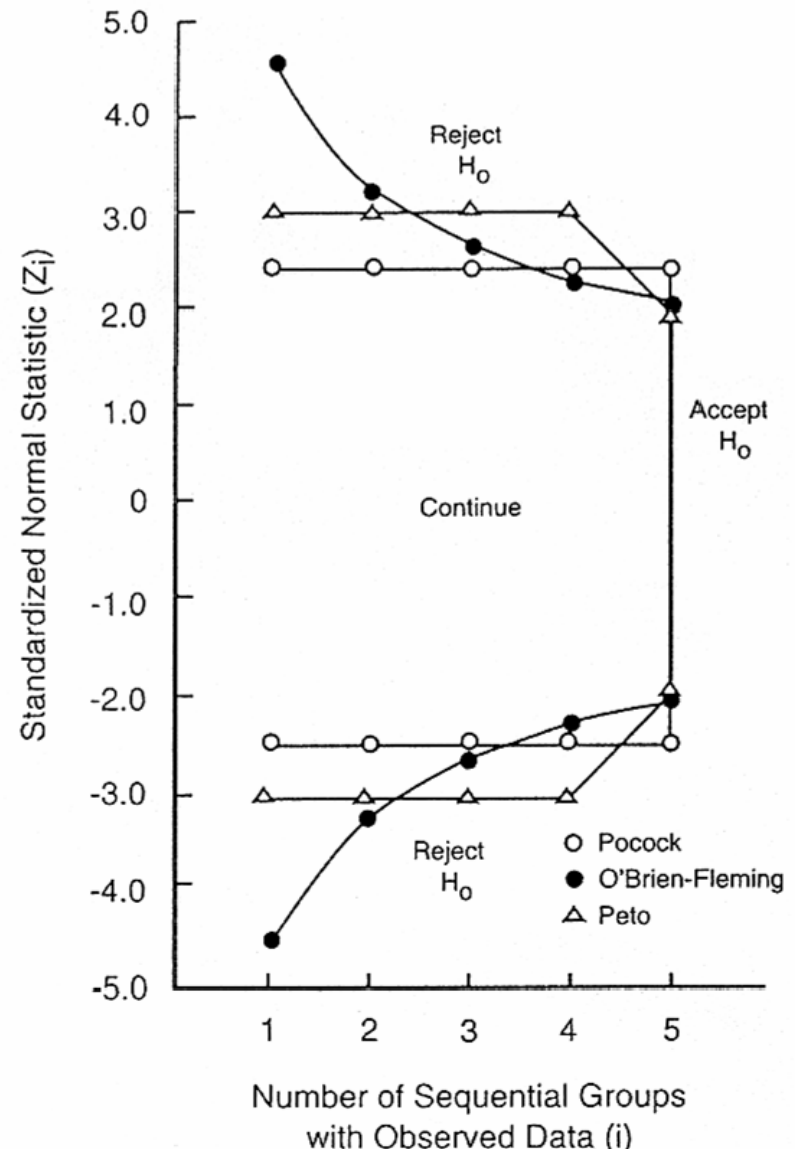
Our objective:

- ❑ Identify existing methods which were designed only individually for GS or survival to a new setting: multivariate GS survival data
- ❑ Extend these existing methods in new directions

- ❑ Compare methodologies
 - ❑ Illustrate and compare results using data examples
 - ❑ Compare error rates and power using data simulations

Literature Review

- Early Work in Group Sequential Methodology
 - Pocock (1977)
 - Uses same critical value at each interim look
 - O'Brien and Fleming (1979)
 - Alpha-level increases for later interim looks, making it easier to reject at later interim looks.





Literature Review

- Early Work in Group Sequential Methodology, cont.
 - **Haybittle (1971) and Peto et al (1976)**
 - Reject at any interim look if :
 $| Z_i | \geq 3; \quad i = 1, \dots, K$
 - **Lan-DeMets (1983)**
 - Flexible error-spending functions




Literature Review

- Early Work in Group Sequential Methodology, cont.
 - Wang and Tsatis(1987)
 - Test allows boundaries of different shapes:
Pocock and O'Brien-Fleming as special cases
 - Benefit
 - Drawback



Approaches to multivariate GS data

- ❑ Tang, Geller, Pocock (1993)
- ❑ Tang and Geller (1999)
- ❑ Jennison and Turnbull (1999)
 - (1) Bonferroni
 - (2) GS Hotelling Test
 - (3) GS O'Brien's Test
 - (4) Tests based on marginal criteria



Approaches to multivariate survival data

Therneau and Grambsch (2000) :

- Time to first event
- Frailty Model
- Marginal Models Approach
 - *Wei, Lin et al (1989 & later)*
 - *Prentice et al (1981)*
 - *Andersen-Gill (1982)*
- Model subject's correlation within Cox framework
(*Prentice & Cai, 1992*)



Approaches to multivariate GS survival data

- Much less has been published in our specific area of interest:
 - Lin (1991): nonparametric method
 - Williams (1996): repeated CIs using counting process methods
 - Kosorok et al (2004): global alpha-spending with multiple decision rule.



Research for this study

- We proposed the following:
 - Apply existing methods for survival analysis to a new setting
 - Develop new methodology by exploiting what is available, borrowing and combining properties of existing methods



Combining existing methods

Example:

- *Survival*: WLW (1989) marginal prop hazards
- *MCP*: Tang & Geller's (1999) Closed Testing Procedure (CTP)
- *GS*: Lan-DeMets (1983) error spending function

Closed Testing Schematic

System of hypotheses for M=4 endpoints

step

1:

$$H_0 = H_0^{\{1,2,3,4\}}$$

2:

$$H_0^{\{1,2,3\}}$$

$$H_0^{\{1,2,4\}}$$

$$H_0^{\{1,3,4\}}$$

$$H_0^{\{2,3,4\}}$$

3:

$$H_0^{\{1,2\}}$$

$$H_0^{\{1,3\}}$$

$$H_0^{\{1,4\}}$$

$$H_0^{\{2,3\}}$$

$$H_0^{\{2,4\}}$$

$$H_0^{\{3,4\}}$$

4:

$$H_0^{\{1\}}$$

$$H_0^{\{2\}}$$

$$H_0^{\{3\}}$$

$$H_0^{\{4\}}$$

From Lehmacher, et al, 1991



Data Example:

Primary Biliary Cirrhosis (PBC)

Lindor et al (1994)

- N = 170 subjects, 4-year study, UDCA vs. placebo
- Death, transplant, histologic progression, and development of varices
- Interim looks ...

Example: Primary Biliary Cirrhosis

Table 1. *Cumulative No. of Subjects with Events in PBC Study*

Event	Timepoint	Cumulative Counts		
		UDCA (N = 86) n (%)	Placebo (N = 84) n (%)	Overall (N = 170) n (%)
Death	Month 12	0 (0.0)	3 (3.6)	3 (1.8)
	Month 36	3 (3.5)	7 (8.3)	10 (5.9)
	Final	6 (7.0)	10 (11.9)	16 (9.4)
Transplant	Month 12	1 (1.2)	1 (1.2)	2 (1.2)
	Month 36	3 (3.5)	5 (6.0)	8 (4.7)
	Final	6 (7.0)	6 (7.1)	12 (7.1)
Histologic Progression	Month 12	0 (0.0)	0 (0.0)	0 (0.0)
	Month 36	7 (8.1)	10 (11.9)	17 (10.0)
	Final	8 (9.3)	12 (14.3)	20 (11.8)
Development of Varices	Month 12	0 (0.0)	0 (0.0)	0 (0.0)
	Month 36	7 (8.1)	13 (15.5)	20 (11.8)
	Final	8 (9.3)	17 (20.2)	25 (14.7)
At Least One Event	Month 12	1 (1.2)	4 (4.8)	5 (2.9)
	Month 36	17 (19.8)	31 (36.9)	48 (28.2)
	Final	23 (26.7)	37 (44.0)	60 (35.3)



Data Example: Primary Biliary Cirrhosis, continued

- Lan-DeMets error spending function:

$$f(t) = 2 - 2\Phi(z_{\alpha/2}/\sqrt{t})$$

Analysis	Timepoint (Months)	Nominal α -level	Cumulative Alpha	Z-critical Value
1	12	<0.0001	<0.0001	3.750
2	36	0.0236	0.0236	1.985
Final	48	0.0431	0.0500	1.715

Methods for this data example

- We apply the WLW model to our data:

$$\lambda_{mi}(t) = \lambda_{m0}(t) \exp\{\beta_0' Z_{mi}(t)\}, t \geq 0$$

- Let $\hat{\beta}$ be the vector of maximum partial likelihood estimates $(\hat{\beta}_1 \dots \hat{\beta}_p)$

- Estimate β as:

$$L(\beta) = \prod_{i=1}^n \prod_{m=1}^M \left[\frac{\exp(\beta' Z_{im}(X_{im}))}{\sum_{l \in \mathcal{R}_m(X_{mi})} \exp(\beta' Z_{lm}(X_{l'm}))} \right]^{\Delta im}$$

Methods, continued

- Then $\hat{\beta}$ is asymptotically multivariate $N(\beta, Q)$, where Q is the covariance matrix (per Wei et al 1989).
- where Q is estimated using a robust sandwich covariance matrix estimate:

$$\hat{Q} = n^{-1} \begin{bmatrix} \hat{D}_{11}(\hat{\beta}_1, \hat{\beta}_1) & \cdots & \hat{D}_{1P}(\hat{\beta}_1, \hat{\beta}_P) \\ \vdots & & \vdots \\ \hat{D}_{P1}(\hat{\beta}_P, \hat{\beta}_1) & \cdots & \hat{D}_{PP}(\hat{\beta}_P, \hat{\beta}_P) \end{bmatrix}$$

Methods, continued

- Now we can form a single test statistic over all P covariates:

$$W = \left(\hat{\beta}_1, \dots, \hat{\beta}_P \right) \cdot \hat{Q}^{-1} \cdot \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_P \end{pmatrix} \sim \chi^2_{(P)}$$

- For a single covariate, this reduces to:

$$W_p = \left(\frac{\hat{\beta}}{SE_{robust}} \right)^2 \sim \chi^2_{(1)}$$

- Since W_p is a 1 d.f. test ...

$$z = \frac{\hat{\beta}}{SE_{robust}} \sim N(\beta, 1)$$



Methods, continued

$\beta^{(1)}, \dots, \beta^{(K)}$ is asymptotically multivariate normal, using a canonical joint distribution:

$$\hat{\beta}^{(k)} \sim N(\beta, I_k^{-1}) \text{ for each } k,$$

$$\text{and } \text{Cov}(\hat{\beta}^{(k_1)}, \hat{\beta}^{(k_2)}) = (I_{k_2})^{-1} \text{ for } 1 \leq k_1 \leq k_2 \leq K,$$

$$\text{where } I_k = \text{var}(\beta^{(k)})^{-1}.$$

Let $Z_k = (I_k)^{1/2} \hat{\beta}^{(k)}$. Then:

$$Z_k \sim N((I_k)^{1/2} \beta, 1) \text{ for each } k,$$

$$\text{and } \text{Cov}(Z_{k_1}, Z_{k_2}) = \sqrt{(I_{k_1} / I_{k_2})} \text{ for } 1 \leq k_1 \leq k_2 \leq K.$$



Methods, continued

- Test: $H_{o,k}: \beta_{m,k} = 0$ against the alternative hypothesis $H_{a,k}: \beta_{m,k} > 0$, with at least one inequality
- $HR = \exp(\beta) = \lambda_P / \lambda_A$
- We can apply the closed testing procedure while preserving strong control of the type 1 error



Data Example: Primary Biliary Cirrhosis, continued

□ Results:

- Interim Analysis One:

$$Z_{\{1,2,3,4\}} = 1.22 < 3.75 \Rightarrow \text{continue the trial}$$

- Interim Analysis Two:

$$Z_{\{1,2,3,4\}} = 2.51 > 1.985 \Rightarrow \text{stop the trial, examine sub-hypotheses using CTP}$$



Data Example: Primary Biliary Cirrhosis, continued

- Interim Analysis Two, continued:
 - All three-way subhypotheses exceed critical value 1.985:
 $Z_{\{1,2,3\}}=2.0$, $Z_{\{1,2,4\}}=2.39$,
 $Z_{\{1,3,4\}}=2.45$, $Z_{\{2,3,4\}}=2.09$
 - Proceed to test all two-way subhypotheses:
Result: all are non-significant ($Z < 1.985$),
so *we cannot test any individual endpoints*.

What is our conclusion?



Data Example: Primary Biliary Cirrhosis, continued

- Are there better ways to look at this data?
- We also tried the following:
 - Approach 1: different alpha-spending functions
 - Approach 2: different survival methods
 - Approach 3: different multiple comparison procedures



Proposed new combinations

- Approach 2: Keep the closed testing method, but try different survival methods (alternatives to WLW method):
 - Andersen-Gill (AG)
 - Prentice-Williams-Peterson (PWP)
 - Frailty
 - Accelerated Failure Time (AFT)

We applied each of these methods to our data example to compare performance, usefulness, flexibility, ease of use, etc.



Proposed new combinations

- Approach 3: Alternatives to Closed Testing Procedure:
 - Other MCP
 - Global test procedures
 - Various MCPs will control the familywise error (FWE) rate while allowing us to obtain more specific information about the multiple outcomes.
 - Westfall and Young (1989, 1993); Troendle (1995, 1996)



Data simulations

- Goals of Simulation

- Want to test:

- $H_{o,k}: \beta_{m,k} = 0$ vs. $H_{a,k}: \beta_{m,k} > 0$, with at least one inequality (as in Tang and Geller, 1999)

- Simulations under null and under alternative

- Alter the correlation between endpoints

- Alter the censoring patterns

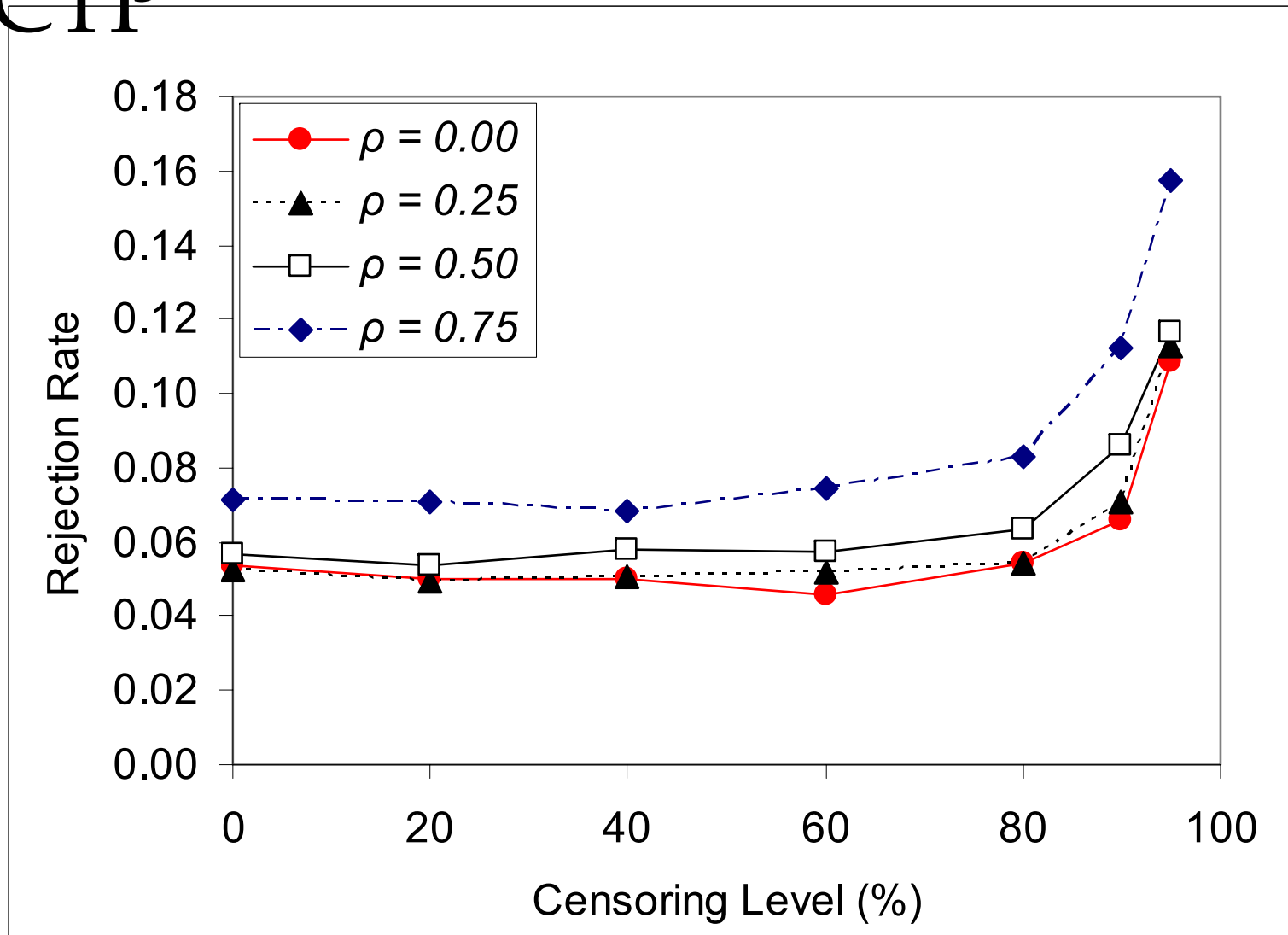
- Focus is on WLW for CTP, but we also tried BHM vs. Hochberg and others



Data simulations

- Simulation Construction
 - mimic the structure of the PBC dataset
 - lognormally-distributed survival times
 - censoring times are also simulated independently
 - we also preset the desired level of correlation between events
 - 5000 iterations
 - for each procedure, count rejections at each interim analysis using appropriate critical value

Observed rejection rate under null for CTP



Observed power in a closed testing procedure

Event Time Spacing	Censoring Level (%)	Dimension of Marginal Hypothesis	Correlation Level (ρ) Between Endpoints			
			0.00	0.25	0.50	0.75
Equal*	0	4	100	100	99.8	99.5
		3	97.3	97.5	97.9	97.9
		2	88.9	90.6	92.2	93.8
		1	65.2	69.3	74.3	79.9
	30	4	99.4	96.7	92.5	85.9
		3	94.8	91.2	87.2	81.0
		2	81.7	78.3	75.8	72.2
		1	49.4	50.0	52.2	53.7
	60	4	97.2	85.9	70.7	59.9
		3	88.9	75.8	61.6	52.5
		2	70.3	58.7	48.2	42.8
		1	36.2	31.1	28.0	27.8
	90	4	18.8	16.9	16.3	18.0
		3	12.3	11.9	12.2	15.3
		2	6.8	7.6	8.6	12.3
		1	3.1	4.1	5.2	8.7

Observed power in a closed testing procedure

Event Time Spacing	Censoring Level (%)	Dimension of Marginal Hypothesis	Correlation Level (ρ) Between Endpoints			
			0.00	0.25	0.50	0.75
Unequal**	0	4	100.0	100.0	99.8	99.5
		3	98.7	99.0	99.2	99.2
		2	93.1	94.1	95.8	97.1
		1	71.6	75.0	79.7	84.0
	30	4	99.3	96.6	93.2	86.3
		3	95.8	92.1	88.3	82.2
		2	83.7	80.5	77.9	75.2
		1	51.6	52.7	54.7	58.0
	60	4	97.7	86.2	70.5	59.2
		3	89.4	76.6	62.5	53.2
		2	70.7	60.2	50.0	44.9
		1	36.8	32.9	28.9	29.4
90	4	19.1	17.6	15.7	17.2	
	3	12.2	13.0	12.1	14.7	
	2	7.3	8.5	8.7	11.7	
	1	3.4	4.6	5.4	7.9	



Discussion and summary

- ❑ Review
- ❑ Benefits
- ❑ Drawbacks
- ❑ Recommendations depend on goals, or idiosyncrasies of the data
- ❑ Directions for further research



References

- Follmann, D.A., Proschan, MA, Geller, NL. (1994), "Monitoring pairwise comparisons in multiarmed clinical trials," *Biometrics*, 50, 325-336.
- Gangnon, R.E., Roecker, E, Cook, T. (1999), "Sequential Monitoring of Multiple Endpoints in Clinical Trials," *Society for Clinical Trials Meeting*, Anaheim, CA.
- Jennison, C, Turnbull, B.W. (2000), *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC.
- Kosorok, MR, Shi, Y, DeMets, DL (2004), "Design and analysis of group sequential clinical trials with multiple primary endpoints," *Biometrics*, 60, 134-145.
- Lan, K.K.G., DeMets, D.L. (1983), "Discrete sequential boundaries for clinical trials," *Biometrika*, 70, 659-663.
- Lehman, W., Wassmer, G., and Reitmeir, P. (1991), "Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate," *Biometrics*, 47, 511-521.
- Lindor, K.D., Dickson, E.R., Baldus, W.P., et al. (1994), "Ursodeoxycholic acid in the treatment of primary biliary cirrhosis," *Gastroenterology*, 106, 1284-1290.
- Logan, B.R., Tamhane, A.C. (2001), "Combining global and marginal tests to compare two treatments on multiple endpoints," *Biometrical Journal*, 5, 591-604.
- Marcus, R, Peritz, E, Gabriel, K.R. (1976), "On closed testing procedures with special reference to ordered analysis of variance," *Biometrika*, 63, 655-660.
- O'Brien, P.C., Fleming, T.R. (1979), "A multiple testing procedure for clinical trials," *Biometrics*, 35, 549-556.



References

- Pocock, S.J. (1977), “Group sequential methods in the design and analysis of clinical trials,” *Biometrika*, 64, 191-199.
- Spiekerman, C.F., Lin, D.Y. (1998), “Marginal regression models for multivariate failure time data,” *Journal of the American Statistical Association*, 93, 1164 –1174.
- Tang, D.I., Geller, N.L. (1999), “Closed testing procedures for group sequential clinical trials with multiple endpoints,” *Biometrics*, 55, 1188-1192.
- Therneau, T.M., Grambsch, P.M. (2000), *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer-Verlag.
- Wei, L.J., Lin, D.Y., Weissfeld, L. (1989), “Regression analysis of multivariate incomplete failure time data by modeling marginal distributions,” *Journal of the American Statistical Association*, 84, 1065-1073.
- Wei, L.J., Glidden, D.V. (1997), “An overview of statistical methods for multiple failure time data in clinical trials,” *Statistics in Medicine*, 16, 833-839.
- Williams, P.L. (1996), “Sequential monitoring of clinical trials with multiple survival endpoints,” *Statistics in Medicine*, 15, 2341-2357.