

ONGOING METHODOLOGICAL RESEARCH  
AND  
COLLABORATIVE PROJECTS

JONG-HYEON JEONG, PH.D.

Assistant Professor

Department of Biostatistics

University of Pittsburgh

October 8, 2004

## INTRODUCTION

- Joined the department in Aug. 1, 1998
- Ph.D. in Statistics from University of Rochester in 1996 (Thesis topic: New applications of frailty model in survival analysis)
- Post-graduate researcher in ADCS at UCSD (1996-1998)
- Statistician for UPCI; Statistician for NSABP
- Teaching: Applied Regression Analysis (BST 2049); Introduction to Probability and Mathematical Statistics (BST 2043, BST 2044)
- Research interest: Parametric and nonparametric inference in survival analysis; Phase III clinical trials

## OUTLINE

- Recent research papers and work in progress-Methodology
- Recent collaborative work in NSABP

Paper 1: On the asymptotic relative efficiency of cumulative hazard estimates from Cox's model by J. JEONG AND D. OAKES (*Sankhya*, 2003)

- “Semiparametric” inference
- Proportional hazards model; for the  $j$ th subject with a covariate  $z$

$$h_j(t; z) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} Pr(T \leq t + \Delta | T \geq t; z_j) = \exp(\beta z_j) h_0(t)$$

- Two estimators are considered under Cox's model;  $\beta$  and  $S(t; z) = \exp\{\int H(t; z) dt\}$
- Test for  $H_0 : \beta = 0$  from Cox's model is fully efficient (Efron, 1977; Oakes, 1977)
- No consideration on the efficiency of survival function estimates under Cox's model
- Variance formula for those estimators from Cox's model are given in Tsiatis (1981)
- Frailty model

$$h(t|z_{ij}, w) = w_i e^{\beta z_{ij}} h_0(t)$$

- Using gamma frailty theory under Cox's model, simple closed forms of the variance formula can be derived parametrically and nonparametrically
- Survival function estimates under Cox's model could lose a substantial amount of efficiency relative to the exponential parametric model

Paper 2: Rank tests for clustered survival data when dependent subunits are randomized by J. JEONG AND S. JUNG (in revision for *Statistics in Medicine*)

- “Nonparametric” inference
- Rank test statistics for comparing two marginal survival distributions;

$$W = \sqrt{n} \int_0^\infty G(t) \{d\hat{\Lambda}_1(t) - d\hat{\Lambda}_2(t)\}$$

- Correlated subunits from the same cluster are often randomized to different treatment groups
- Example
  - A litter-matched tumorigenesis experiment
  - Three rats (subunits) in each of 50 female litters (clusters) were randomly allocated to a drug treatment group or the control group.
  - Major interest; assess whether the time to tumor appearance in the drug-treated group tends to be longer than that in the control group.
- In this case, the variance formulas of the standard rank tests need to be adjusted for correlations both within and between treatment groups.
- Derived a general form of simple variance formulas of the rank tests
- Extension of Jung and Jeong (2003, *LIDA*)
- Compared with the optimal frailty model approach when the inter- and intra-cluster correlations are the same

Paper 3: Parametric inference for cumulative incidence function: application to breast cancer data by J. JEONG (in review for *Journal of the Royal Statistical Society-Series A*)

- “Parametric” inference
- Cause-specific hazard function

$$h_j(t) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta, J = j | T \geq t)}{\Delta}$$

- Cumulative incidence function

$$F_j(t) = \int_0^t S(u) d\Lambda_j(u)$$

- Nonparametric (Korn and Dorey, 1992; Gaynor et al., 1993; Pepe and Mori, 1003...) and semiparametric (Bryant and Dignam, 2004) estimators exist
- Proposed a flexible parametric estimator to improve efficiency
- Applied to NSABP B-14 and B-20 data sets

Paper 4: Median regression on residual lifetimes by J. JEONG AND S. JUNG  
(in preparation)

- “Semiparametric” inference
- Interpretation of the hazard function is not straightforward to the laymen
- Would be useful to have a semiparametric regression model to relate the median remaining lifetimes and some prognostic factors
- Median residual life time
  - At a given time  $t$ ,  $P(T-t \geq \theta_t | T \geq t) = 1/2$  implies  $P(T \geq t + \theta_t) = (1/2)P(T \geq t)$
  - With an independent censoring

$$S_n(\theta_t) = \frac{\sum_{i=1}^n I(Y_i \geq t + \theta_t)}{\hat{G}(t + \theta_t)} - \frac{\sum_{i=1}^n I(Y_i \geq t)}{2\hat{G}(t)} \approx 0,$$

where  $I(Y \geq y) = 1$  if  $Y \geq y$  and 0 if  $Y < y$  and  $\hat{G}$  is Kaplan-Meier estimates of survival function of the censoring distribution.

- Can consider regression model

$$\text{med}(T_i - t | T_i \geq t, Z_i) = \beta'_t \mathbf{Z}_i$$

- Score function

$$\mathbf{S}_{tn}(\beta_t) = \sum_{i=1}^n \mathbf{Z}_i \left\{ \frac{I(Y_i \geq t + \beta'_t \mathbf{Z}_i)}{\hat{G}(t + \beta'_t \mathbf{Z}_i)} - \frac{I(Y_i \geq t)}{2\hat{G}(t)} \right\} \approx 0$$

- Developing a semiparametric testing procedure for the regression parameters

## COLLABORATIVE WORK IN NSABP

1. Statistician for ongoing protocols B-30 and B-31
2. Paper 1: Fisher B, Jeong J-H, Bryant J, Anderson S, Dignam J, Fisher ER, and Wolmark N. Treatment of Lymph Node-Negative, Estrogen Receptor-Positive Breast Cancer: Long-Term Findings from National Surgical Adjuvant Breast and Bowel Project Randomized Clinical Trials. *Lancet*, 364, 858-868, 2004.
3. Paper 2: Fisher B, Jeong J-H, Anderson S, Wolmark N. Treatment of lymph-node-negative, ER-negative breast cancer: Updated findings from NSABP clinical trials. *Journal of the National Cancer Institute*, in press.
4. Paper 3: Taghian A, Jeong J-H, Mamounas E, Anderson S, Bryant J, Deutsch M, Wolmark N. Patterns of locoregional failure in patients with operable breast cancer treated by mastectomy and adjuvant chemotherapy with or without tamoxifen and without radiotherapy: Results from five National Surgical Adjuvant Breast and Bowel Project randomized clinical trials. *Journal of Clinical Oncology*, in press.