A General Framework For Sequential And Adaptive Methods In Survival Studies*

Zhiliang Ying

IMS @ NUS

October 27, 2011

*Joint work with Xiaolong Luo (Celgene Corp.) and Gongjun Xu (Columbia U.)

Outline

- Survival analysis and martingale theory framework
- Group sequential in survival analysis
- Sellke and Siegmund (1983)
- Gu and Lai (1991)
- Extensions to adaptive designs
- Profile likelihood
- Conclusions

Survival analysis

- Clinical trials
- Epidemiological cohort studies
- Sociology (event history analysis)
- Microeconomics (unemployment duration)
- Marketing (product growth, brand switching)
- Finance (default probability, rating)
- Ecology (capture-recapture)
- Software engineering and reliability
- Education testing (response time analysis for speed test or mixed speed-power)

Survival Data

- Usual statistical problems are about effects of X on Y.
- Modeling and analysis largely depend on Y.
- Y continuous: linear model, least squares and t test.
- Y binary: logistic, probit.

Survival Data

- Survival data: $(\tilde{T}, \Delta)$, $\tilde{T} = \min\{T, C\}$ and $\Delta = I(T \leq C)$.
- Outcome variables: $\tilde{T}$-continuous, Δ-binary.
- Two approaches:
- (1) Accelerated failure time (AFT) model: extending the linear model (continuous)
- (2) Mantel-Haenszel: conversion to binary outcomes.

Counting process/martingale

A revolutionary development is the counting process approach to censored data, implicitly used by N. Mantel (1958) and D.R. Cox (1972), and formalized by O. Aalen (1978).

Counting process

- Let T_i be the failure time for individual i.
- Observing random variable T_i is tantamount to observing entire counting process $N_i(t) = I(T_i \leq t)$, $t \geq 0$.
- Difference # 1: T_i is continuous but N_i binary.
- Difference # 2: 0-1 information accumulation vs. continuous accumulation.

Compensator and martingale

- Let $\{\mathcal{F}_t = \sigma(N_i(s), s \leq t),\ t \geq 0,\ i = 1, \ldots, n\}$ be the σ filtration, representing information accumulation.
- We have $P(dN(t) = 1|\mathcal{F}_{t-}) = I(T \geq t)\lambda(t)dt$, where λ is the hazard function of T.
- $dN(t)$-observed vs. $\lambda(t)$-expected.
- Doob-Meyer: $M(t) = N(t) - \int_0^t I(T \geq s)\lambda(s)ds$ is martingale wrt to $\mathcal{F}$.
- Modeling and inference about population reduce to those about λ.

Censoring

- T-survival time, C-censoring time (omitting i)
- Observe $\tilde{T} = T \wedge C$ and $\Delta = I(T \leq C)$
- Conversion to counting processes: $N(t) = \Delta I(\tilde{T} \leq t)$ and $Y(t) = I(\tilde{T} \geq t)$, $t \geq 0$.
- Note that $N(t) = I(\tilde{T} \leq t \wedge C)$, which is $I(T \leq t)$ being stopped at C.
- $M(t) = N(t) - \int_0^t Y(s)\lambda(s)ds$ is now the original martingale (for uncensored data) stopped at C.
- Noninformative censoring: C is a stopping time.

Cox Model

For regression analysis (with covariates Z), the Cox model specifies the Z-specific hazard function

$$\lambda(t|Z) = e^{\beta Z}\lambda_0(t)$$

- Two-sample comparison: $Z = 0$ for control and $Z = 1$ for treatment.
- Control: $\lambda_0(t)$
- Treatment: $\lambda_1(t) = e^{\beta}\lambda_0(t)$, where e^{β} is known as the relative risk.

Partial Likelihood

Analysis of the Cox regression can be made through the partial likelihood approach. Specifically, by maximizing the partial likelihood function, one can estimate the regression parameter β. In addition, one can differentiate the log-partial likelihood function to get a score test statistic for $H_0 : \beta = \beta_0$. Suppose we have simultaneous entry and the current time is t.

Partial Likelihood Score

- Score function for data up to time t:

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_j Z_j e^{\beta Z_j} Y_j(s)}{\sum_j e^{\beta Z_j} Y_j(s)} \right) dN_i(s)$$

- It can be written as a martingale integral (wrt a predictable process):

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_j Z_j e^{\beta Z_j} Y_j(s)}{\sum_j e^{\beta Z_j} Y_j(s)} \right) dM_i(s)$$

where $dM_i(s) = dN_i(s) - Y_i(s)e^{\beta Z_i}\lambda_0(s)ds$- "martingale difference".

Two-sample Log-rank Test

- For testing $\beta = 0$ or $\lambda_1 = \lambda_0$,

$$U(t) = U(0,t) = \sum_i^n \int_0^t \left(Z_i - \frac{\sum_j Z_j Y_j(s)}{\sum_j Y_j(s)} \right) dN_i(s)$$

$$= \sum_{i:\tilde{T}_i \le t} \Delta_i \left(Z_i - \frac{\#\text{ at risk in treatment group at } T_i}{\text{total } \#\text{ at risk at } T_i} \right)$$

- Information at $t \approx$ variance $\approx$ predictable variation at t

$$\langle U \rangle_t = \sum_{i=1}^n \int_0^t \left(Z_i - \frac{\sum_j Z_j Y_j(s)}{\sum_j Y_j(s)} \right)^2 Y_i(s) \lambda_0(s) ds$$

$$\approx \sum_{i:\tilde{T}_i \le t} \Delta_i \left(Z_i - \frac{\#\text{ at risk in treatment group at } T_i}{\text{total } \#\text{ at risk at } T_i} \right)^2$$

Two-sample Log-rank Test

- Information at t:

$$\langle U \rangle_t \approx \sum_{i:\tilde{T}_i \le t} \Delta_i \left(Z_i - \frac{\#\text{ at risk in treatment group at } T_i}{\text{total } \#\text{ at risk at } T_i} \right)^2$$

- For randomized balanced design, under the null of no treatment difference, the information at t becomes

$$I(t) \approx \frac{1}{4} \times \text{ number of events up to } t$$

Simultaneous entry with interim analyses

- Basic fact: a continuous sample path (locally square integrable) martingale W can be expressed as time re-scaled Brownian motion:

$$W(t) = B(\langle W \rangle_t)$$

- Therefore, for $U(t)$, we have

$$U(t) \approx B(I(t))$$

  and sequential boundaries for the Brownian motion can be applied with information (number of events) as the new clock.
- The test is valid (α-level protected) without the Cox model assumption, but power (sample size) calculation does require the assumption.

Covariate Adjustment

- Suppose, in addition to treatment Z, we also have baseline covariates X.
- Assume the Cox model

$$\lambda(t|Z,X) = e^{\beta Z+\gamma X}\lambda_0(t).$$

- Score test for $\beta = 0$ is the score for β with γ replaced by its estimator $\hat{\gamma}(0)$.

Covariate Adjustment

- Score for testing $\beta = 0$ at t:

$$\sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_j Z_j e^{\gamma X_j} Y_j(s)}{\sum_j e^{\gamma X_j} Y_j(s)} \right) dN_i(s)$$

Covariate Adjustment

- Information (for testing $\beta = 0$) at t:

$$I_{\beta\beta}(t) - I_{\beta,\gamma}(t) I_{\gamma,\gamma}^{-1}(t) I_{\gamma,\beta}(t)$$

- Here $I_{\beta,\gamma}(t) =$

$$\sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_j Z_j e^{\gamma X_j} Y_j(s)}{\sum_j e^{\gamma X_j} Y_j(s)} \right) \left(X_i - \frac{\sum_j X_j e^{\gamma X_j} Y_j(s)}{\sum_j e^{\gamma X_j} Y_j(s)} \right) dN_i(s)$$

which is approximation 0 (divided by n), since Z is independent of X, N and Y.

Covariate Adjustment

- Also

$$I_{\beta\beta}(t) = \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_j Z_j e^{\gamma X_j} Y_j(s)}{\sum_j e^{\gamma X_j} Y_j(s)} \right)^2 dN_i(s)$$

  which is approximately $4^{-1} \times \#\{\text{events up to } t\}$.
- Therefore, information at t is again $4^{-1} \times \#\{\text{events up to } t\}$, same as without covariate adjustment (no efficiency improvement!).
- TO ADJUST OR NOT TO ADJUST?

Covariate Adjustment

- If non-trivial Cox model is true, then the marginal hazard ratio is NOT proportional.
- Both adjusted and unadjusted log-rank test statistics are unbiased (0-mean) due to randomization.
- If the independence assumption between T and C (non-informative censorship) is violated, but the conditional independence given X holds, then the adjusted log-rank is still ok.

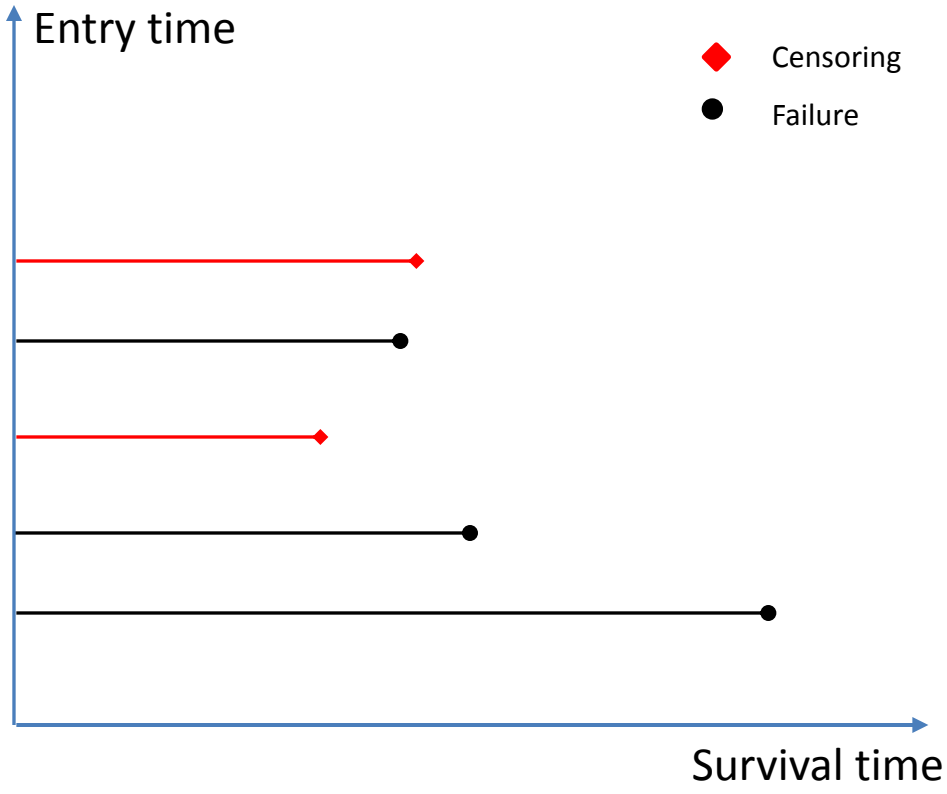
Covariate Adjustment

- Adjusted log-rank is at least as powerful as the unadjusted log-rank.
- If Cox model specification on X is not correct, then standard variance estimation may not be correct for the adjusted, but ok for the unadjusted.
- More complicated if censoring depends on treatment.

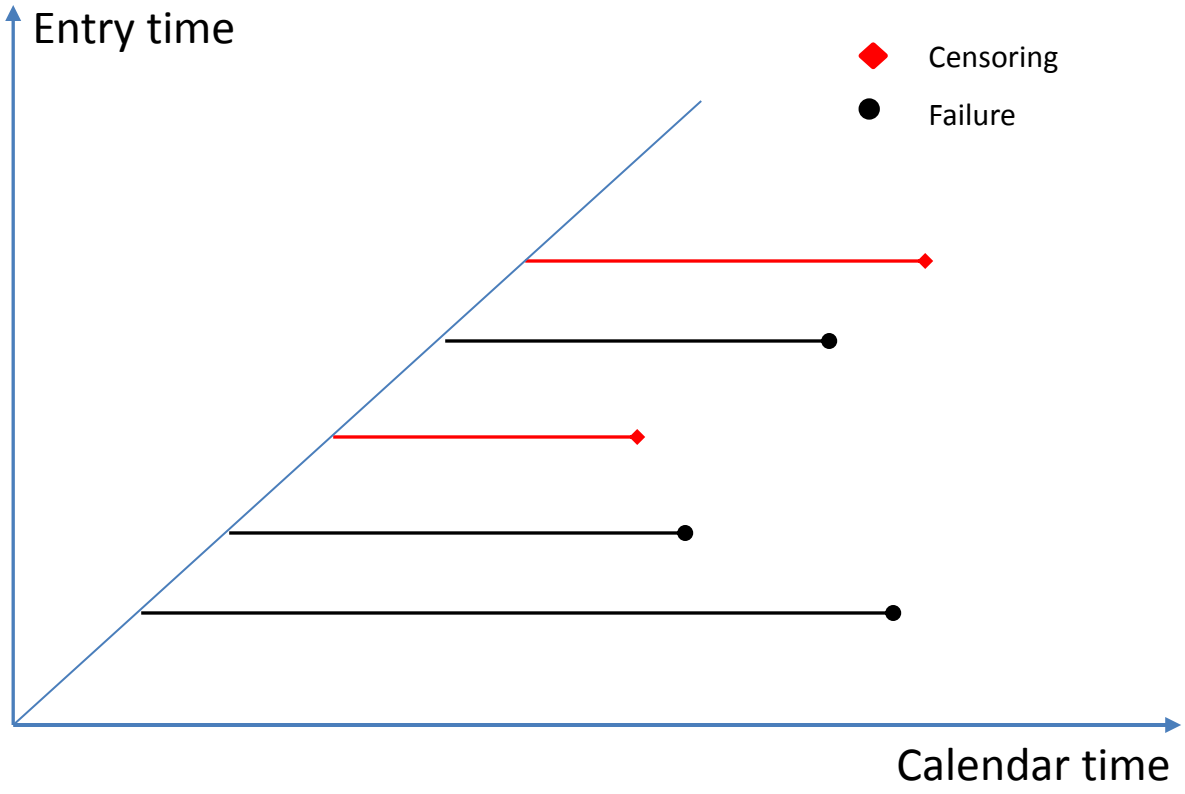
Staggered Entry Survival Data

In reality, the U_i, patients entry times, are different. At the (calendar) time of analysis, one can write down the partial likelihood score (log-rank statistic) for the survival experience.

- Analysis at (calendar) time t
- T_i censored by C_i and $(t - U_i)^+$
- In other words, $C_i(t) = C_i \wedge (t - U_i)^+$ are the effective censoring times.
- Define $N(t, s) = I(T \leq s \wedge C_i(t))$
- $M(t, ds) \equiv N(t, ds) - I(T \wedge C_i(t) \geq s)\lambda(s)ds$ is martingale difference in $s \in [0, t]$ wrt (t-specific) σ-filtration $\mathcal{F}_{t,s}$, $0 \leq s \leq t$.



Observations under survival time



Observations under calendar time and entry time

Sellke and Siegmund (1983, *Biometrika*)

- The partial likelihood score is NOT a martingale, even though the full likelihood is (has to be).
- In technical terms,

$$\sum_{i=1}^{n} \int_0^t [Z_i - \bar{Z}(t,s)] N_i(t, ds), \quad t \geq 0$$

is not a martingale. Here

$$\bar{Z}(t,s) \equiv \frac{\sum_j Z_j Y_j(t,s)}{\sum_j Y_j(t,s)}$$

is the # at risk at survival time s from available data at calendar time t.

Sellke and Siegmund (1983, *Biometrika*)

- However, at the diagonal $s = t$,

$$M(t,t) \equiv N(t,t) - A(t,t))$$

  is a martingale.
- Sellke and Siegmund (1983) showed that if $\bar{Z}(t,s)$ converges to a limit $z(s)$ (large t) that does not depend on t, then the partial likelihood score $U(t,t)$ is approximated by a time re-scaled Brownian motion.
- Assumption of $\bar{Z}(t,s) \to z(s)$: (1) large t vs. large n; (2) dependence between entry time of treatment allocation.

Sellke and Siegmund (1983)

- Technical reason for the partial likelihood score not being a martingale is the term $\bar{Z}(t,s)$, which is not predictable.
- Their approach is to approximate $\bar{Z}(t,s)$ by its limit $\bar{z}(s)$ and to show that the partial likelihood score process is asymptotically equivalent to, uniformly in t,

$$\sum_{i=1}^{n} \int_0^t [Z_i - \bar{z}(s)] M_i(t, ds).$$

Gu and Lai (1983, *Annals*)

- Consider both survival and calendar times (s, t), i.e.,

$$U(t, s) = \sum_{i=1}^{n} \int_0^s [Z_i - \bar{Z}(t, u)] N_i(t, du)$$

- Show weak convergence to a Gaussian random field.
- Useful for the weighted log-rank statistics (such as the Wilcoxon):

$$U_W(t, t) = \int_0^t W(s) U(t, ds)$$

where W is a weight function.

Gu and Lai (1983, *Annals*)

- Exponential inequalities to obtain justification.
- Empirical process theory can be used to provide a general theory (Bilias, Gu and Y., 1997)

Gu and Lai (1983), Bialis et al. (1997)

- The empirical process theory requires i.i.d. sampling assumption.
- Such an assumption may be violated in adaptive designs.
- Not as elegant as the martingale approach, which is natural in terms of information accumulation and likelihood paradigm.

Adaptive Designs

- Sample size re-estimation: sample size may be adjusted through conditional power analysis, or other considerations.
- Covariate-adjusted allocation schemes: to balance treatment allocations in subgroups.
- Outcome-dependent treatment allocation schemes: for ethical and other considerations.

Sample Size Re-estimation

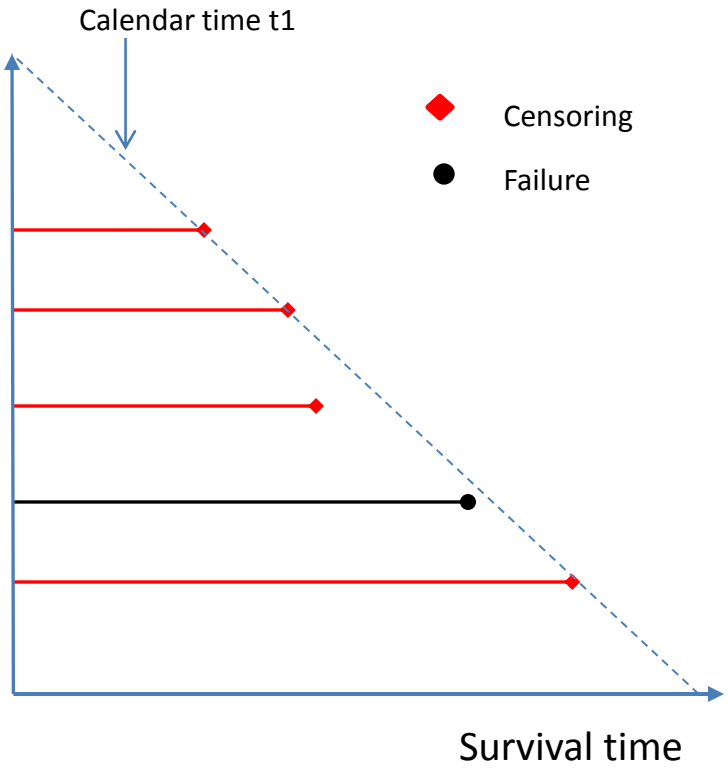
If a re-estimation scheme is used, then log-rank statistic needs to be modified accordingly.

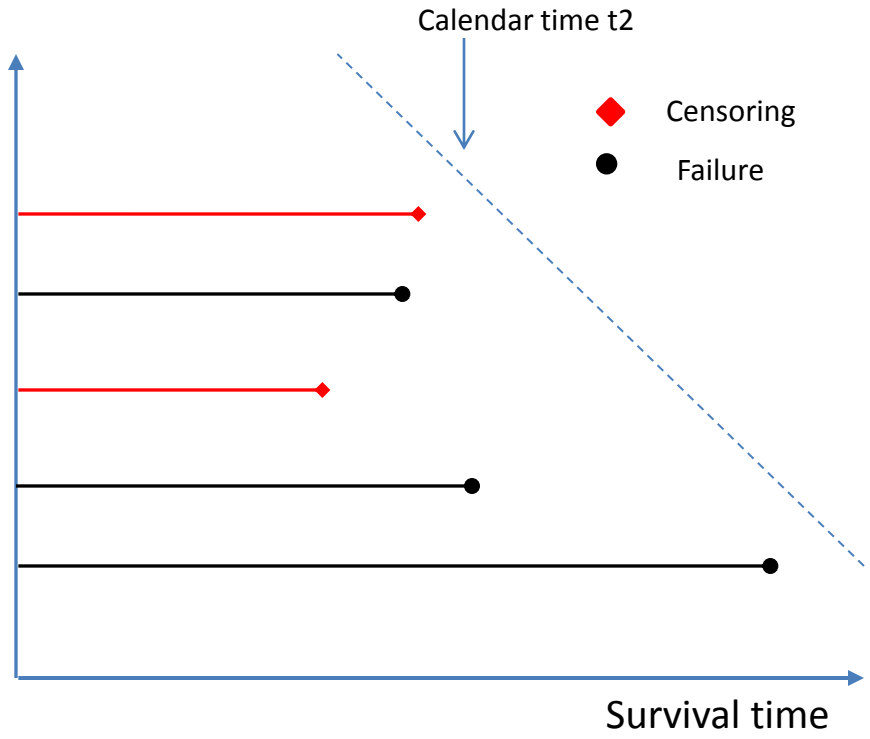
Covariate-adjusted Allocations ("individualized medicine")

- Covariate-adjusted allocation scheme to balance treatment allocations in subgroups.
- Adjusted analysis vs. unadjusted analysis.
- Perspectives of (1) FDA regulation, (2) statistical validity, (3) robustness when assumptions are violated.

Calendar time vs. survival time: 2-dim random field

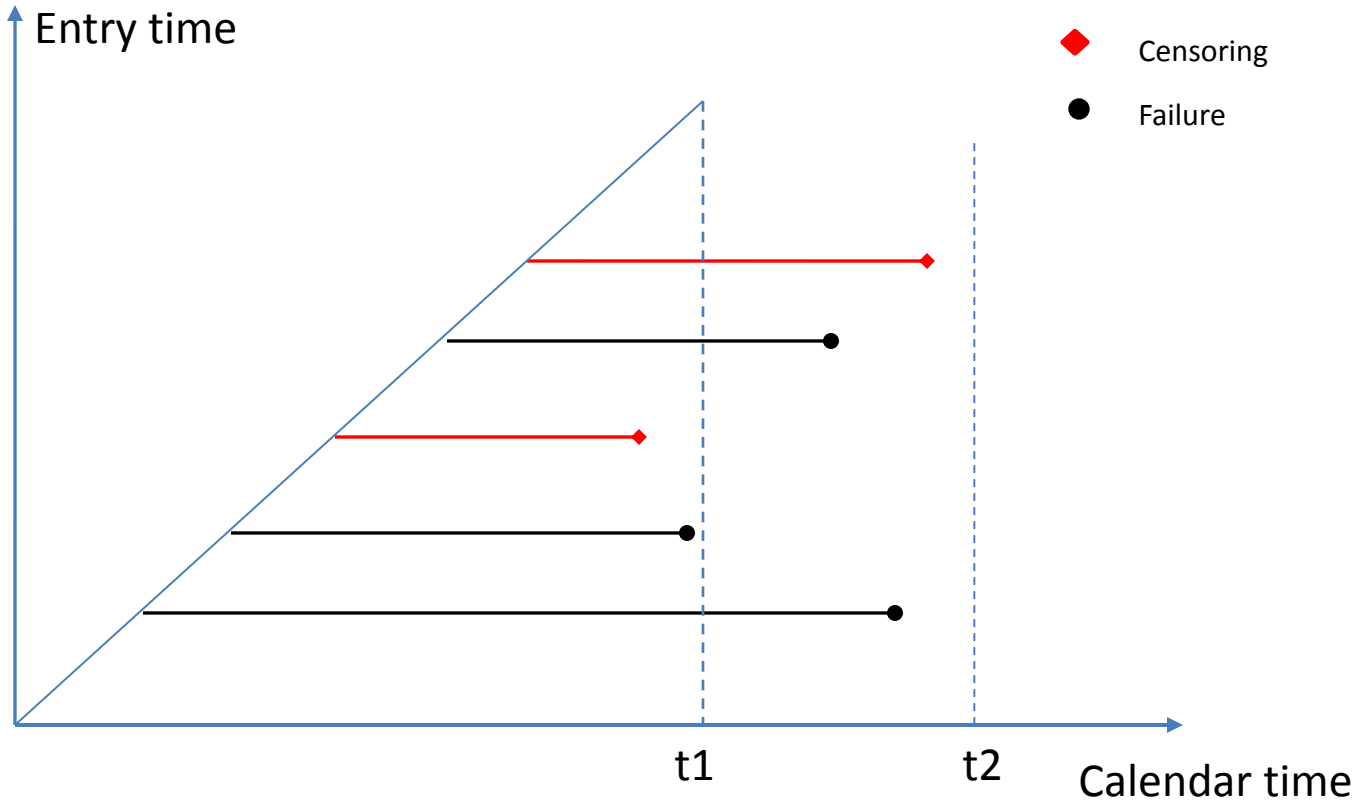
- Joint analysis of calendar and survival times
- Normally, we would consider primarily along the survival time.
- Under an outcome dependent adaptive design, there may NOT be a martingale structure.
- Empirical process theory may not be applicable either, at least directly, since observations are not independent.



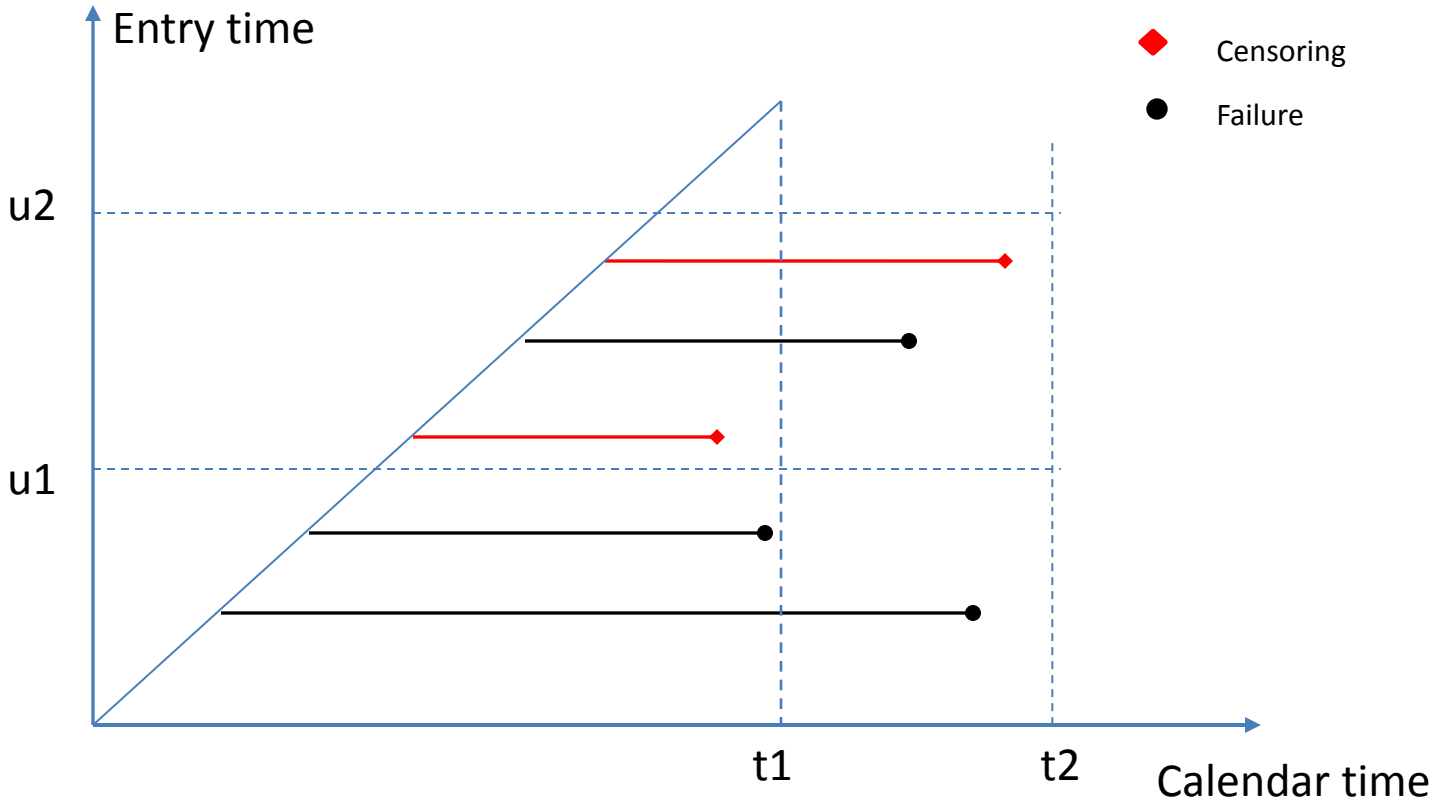


Left: observations up to calendar time t1.

Right: observations up to calendar time t2.



Observations under calendar time and entry time



Observations under calendar time and entry time

Calendar time vs. entry time: 2-dim random field

- Simultaneous consideration of calendar and entry times
- Asymptotic independent increments along both calendar time and entry time.
- Such independent increments structure continue to hold under adaptive designs.
- Martingale inequalities may be applied to obtain a large sample theory.

Extensions to Transformation Models

- Class of transformation models

$$H(T) = \beta Z + \gamma X + \epsilon$$

  H-unknown monotone function, ϵ-completely specified distribution.

- $\epsilon \sim$ extreme value: Cox model.
- $\epsilon \sim$ logistic: proportional odds model.
- $\epsilon \sim$ normal: extension of Box-Cox.

Extensions to Transformation Models

- Profile likelihood approach (Zeng and Lin, 2007)
- Sequentially calculated score from the time-sequential profile likelihood function
- Still has (asymptotically) independent increments.
- Brownian approximation.
- Theory ?

Conclusions

- Survival endpoints are common in long-term clinical trials in which sequential methods are very relevant.
- Counting process formulation is natural for sequential analysis.
- Martingale structure implies time-sequentially calculated score has (asymptotically) independent increments, leading towards Brownian approximation and standard group sequential boundaries.
- Adaptive designs (covariate adjusted, outcome dependent etc.) pose some technical challenges for theoretical developments.
- Consider calendar and entry times simultaneously.
- Violation of independent censoring needs to be studiesd.
- Extensions to alternatives to the Cox model.

Thank You!