Time for a Broader Use of Accelerated Failure-time Models in Cancer Clinical Trials

Tomasz Burzykowski

IDDI and Hasselt University

tomasz.burzykowski@uhasselt.be

Outline

- PH assumption
- Non-PH in cancer clinical trials
- Parametric AFT models
- Semi-parametric AFT model
- Simple simulations
- Conclusions

Proportional Hazards Model

$$\lambda(t; \boldsymbol{Z}) = \lambda_0(t) \cdot e^{\beta' Z}$$

- *Semi-parametric:* $\lambda_0(t)$ need not to be specified.
- Ubiquitous in cancer clinical trials and data analyses.

PH is a Very Strong Assumption (1)

- Consider two covariates, a binary Z_1 and Z_2.
- Assume the PH model for both, i.e.,

$$\lambda(t \mid z_1, z_2) = \lambda_0(t)\mathrm{e}^{\beta_1 z_1 + \beta_2 z_2}$$

- Assume that we use the model only for Z_1.

PH is a Very Strong Assumption (2)

- If $\beta_2 \neq 0$, omitting Z_2 will induce time-dependence of HR for Z_1.
  - Difficult to distinguish the effect from a true time-dependent coefficient

- It will cause bias in estimation of β_1.
  - Even if the distribution of Z_2 is balanced for the levels of Z_1.
  - An issue in clinical trials!

PH is a Very Strong Assumption (3)

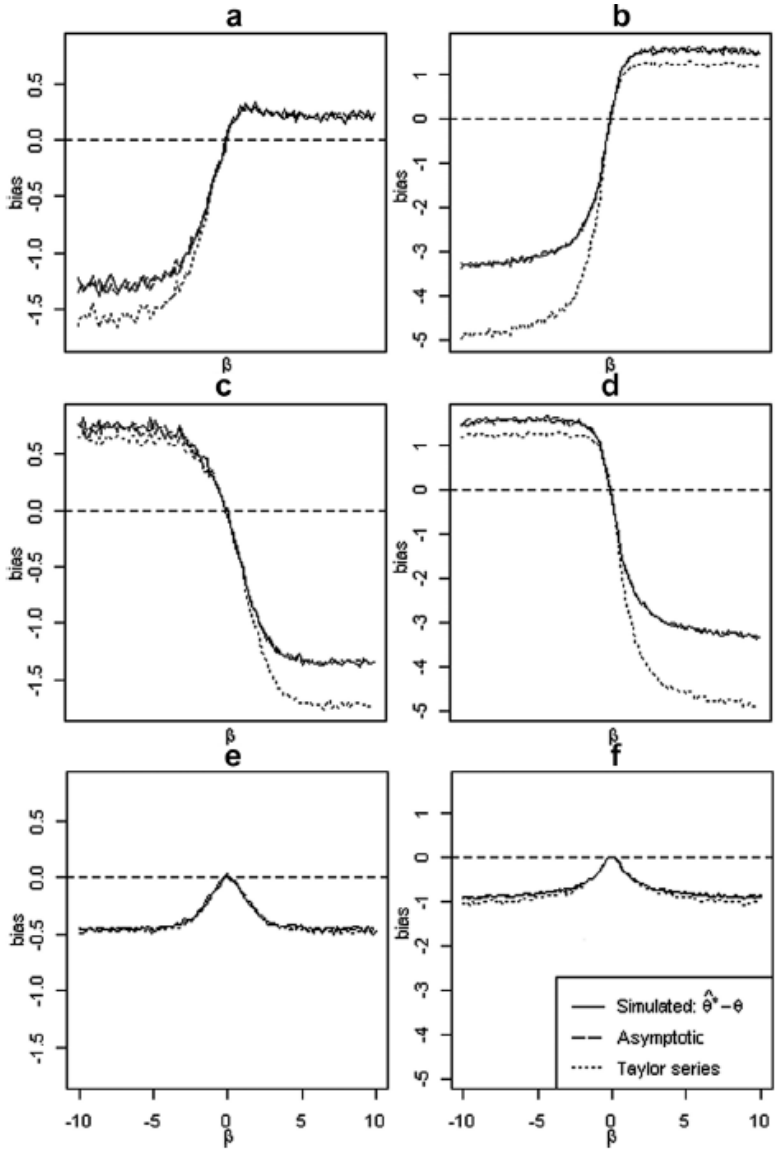


Figure 1. Comparison of simulated biases, asymptotic biases and first-order Taylor series approximations for different types of omitted covariate and censorship. Since θ^* is the asymptotic value of the MLE $\hat{\theta}^*$ and the sample size=10,000 is large, we calculated the simulated bias by $\hat{\theta}^* - \theta$. The asymptotic biases and Taylor series approximations were obtained from (9) and (11), respectively. Monte Carlo integration was used to approximate the expectations in formulae. (a) Binary confounder c: $(\rho_0 = 0.3, \rho_1 = 0.7)$, censored; (b) Normal confounder c: $(\mu_0 = -1, \mu_1 = 1)$, censored; (c) Binary confounder c: $(\rho_0 = 0.7, \rho_1 = 0.3)$, censored; (d) Normal confounder c: $(\mu_0 = 1, \mu_1 = -1)$, censored; (e) Binary balanced c: $(\rho_0 = \rho_1 = 0.5)$, uncensored; (f) Normal balanced c: $(\mu_0 = \mu_1 = 0)$, uncensored.

← Clinical trial (balanced) setting; attenuation $|\beta^*_1| < |\beta_1|$

Ref: Lin et al., *Biometrics* 2013, 69:850

PH is a Very Strong Assumption (4)

- The results imply that the “bottom-up” (“forward selection”) strategies of building PH models are potentially flawed.

- In clinical trials, treatment effect “adjusted” for prognostic covariates should always be looked at.
  - Have we always got information about all important covariates?

Violations of the proportional hazards assumption in randomized phase III oncology clinical trials.

Rifaquat Rahman, Geoffrey Fell, Lorenzo Trippa, Brian Michael Alexander

DOI:
10.1200/JCO.2018.36.15_suppl.2543
Journal of Clinical Oncology 36, no. 15_suppl (May 2018) 2543-2543.

"We performed a PubMed search for randomized phase III trials in breast cancer, lung cancer, prostate cancer and colorectal cancer published in high-impact journals between 2014 and 2016. (...)

We identified 157 publications with 115 KM curves of overall survival (OS) and 139 KM curves of a non-survival time-to-event outcome.

There was evidence of non-proportionality of hazards in a total of 62 (24%) time-to-event outcomes including 20 of 115 (18%) OS KM curves and 42 of 139 (30%) non-survival KM curves. (...)"

Innovative estimation of survival using log-normal survival modelling on ACCENT database

J W Chapman*[,1], C J O'Callaghan[1], N Hu[1], K Ding[1], G A Yothers[2,6], P J Catalano[3], Q Shi[4], R G Gray[5], M J O'Connell[6], D J Sargent[4] and for the ACCENT collaborative group

British Journal of Cancer (2013) 108, 784–790 | doi: 10.1038/bjc.2013.34

. Further, the ACCENT data exhibited classical log-normal hazard shape for both stage II and III patients (see Figure 1 of previous work, Sargent *et al*, 2007). Thus, we hypothesised, based on recent breast cancer literature, that the ACCENT data would have substantive differences in survival estimates with log-normal survival modelling, rather than with the Cox model (Royston, 2001; Chapman *et al*, 2008).

End Points for Colon Cancer Adjuvant Trials: Observations and Recommendations Based on Individual Patient Data From 20,898 Patients Enrolled Onto 18 Randomized Trials From the ACCENT Group

Daniel J. Sargent, Smitha Patiyil, Greg Yothers, Daniel G. Haller, Richard Gray, Jacqueline Benedetti, Marc Buyse, Roberto Labianca, Jean Francois Seitz, Christopher J. O'Callaghan, Guido Francini, Axel Grothey, Michael O'Connell, Paul J. Catalano, David Kerr, Erin Green, Harry S. Wieand, Richard M. Goldberg, and Aimery de Gramont

J Clin Oncol 25:4569-4574. © 2007 by American Society of Clinical Oncology

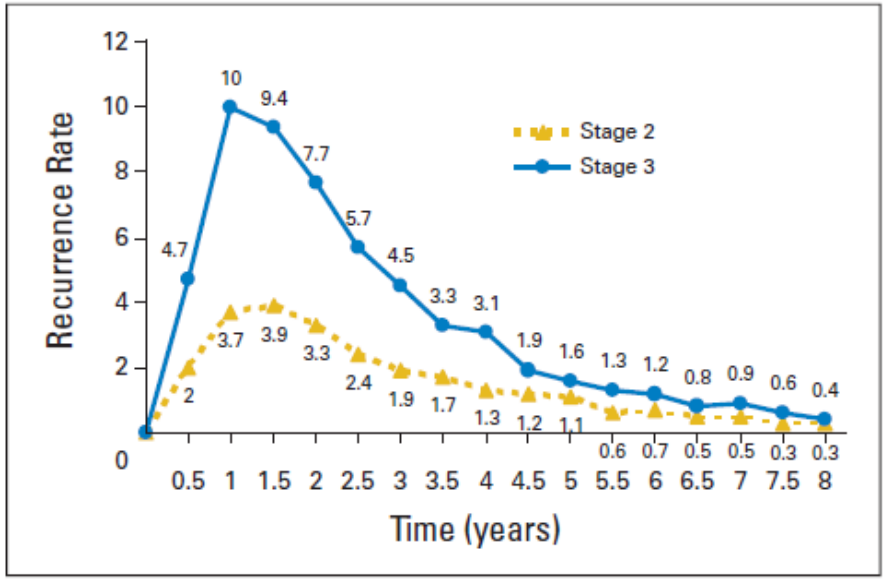


Fig 1. Recurrence rate by stage and time from random assignment. Risk of recurrence in each 6-month interval after random assignment among those remaining recurrence free at the start of each interval, by time and stage.

Statistical Challenges in the Design of Late-Stage Cancer Immunotherapy Studies

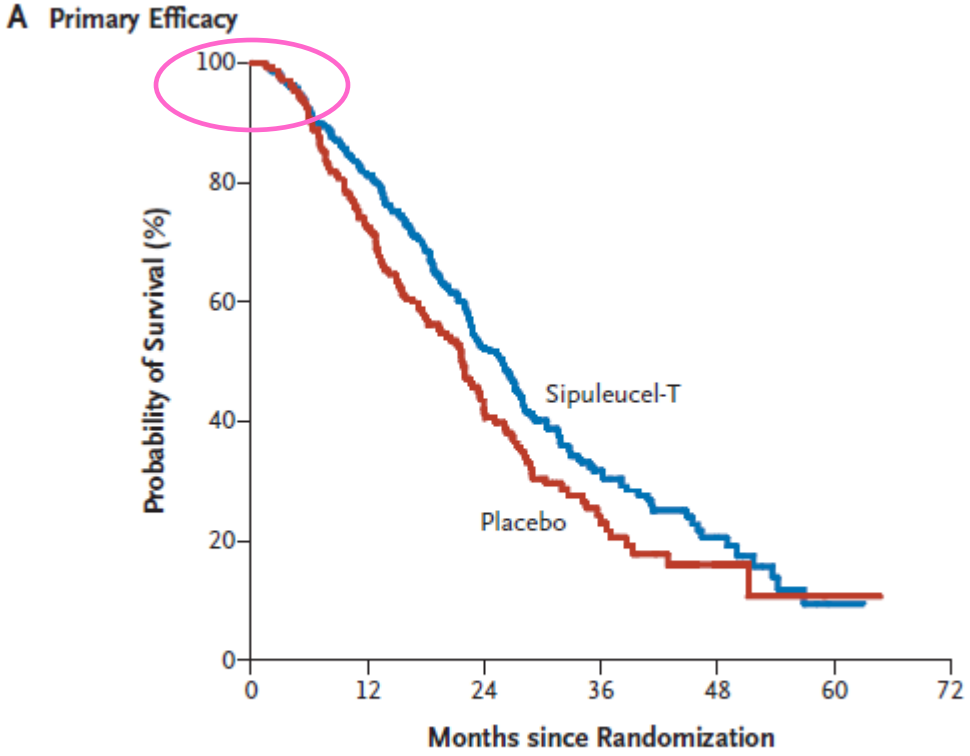
Rosemarie Mick[1] and Tai-Tsang Chen[2,3]

The past several years have witnessed a revival of interest in cancer immunology and immunotherapy owing to striking immunologic and clinical responses to immune-directed anti-cancer therapies and leading to the selection of "Cancer Immunotherapy" as the 2013 Breakthrough of the Year by *Science*. But statistical challenges exist at all phases of clinical development. In phase III trials of immunotherapies, survival curves have been shown to demonstrate delayed clinical effects, as well as long-term survival. These unique survival kinetics could lead to loss of statistical power and prolongation of study duration. Statistical assumptions that form the foundations for conventional statistical inference in the design and analysis of phase III trials, such as exponential survival and proportional hazards, require careful considerations. In this article, we describe how the unique characteristics of patient response to cancer immunotherapies will impact our strategies on statistical design and analysis in late-stage drug development. *Cancer Immunol Res; 3(12); 1292–8. ©2015 AACR.*

Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer

Philip W. Kantoff, M.D., Celestia S. Higano, M.D., Neal D. Shore, M.D., E. Roy Berger, M.D., Eric J. Small, M.D., David F. Penson, M.D., Charles H. Redfern, M.D., Anna C. Ferrari, M.D., Robert Dreicer, M.D., Robert B. Sims, M.D., Yi Xu, Ph.D., Mark W. Frohlich, M.D., and Paul F. Schellhammer, M.D., for the IMPACT Study Investigators*

N Engl J Med 2010;363:411-22.



Accelerated Failure-time Model (AFT)

- Assumption: the effect of covariate is expressed as shortening or lengthening of the time to event.

- Symbolically: $T = T_0 \cdot e^{\boldsymbol{X}' \cdot \boldsymbol{\beta}}$

- Equivalently: $\ln T = \ln T_0 + \boldsymbol{X}' \cdot \boldsymbol{\beta}$

Interpretation of an AFT Model

$$T = T_0 \cdot e^{x' \cdot \beta}$$

- $\beta > 0 \rightarrow e^{\beta} > 1 \rightarrow T > T_0 \rightarrow$ longer time to an event
- $\beta < 0 \rightarrow$ shorter time
- $E(T \mid X) = e^{X \cdot \beta} E(T_0) \rightarrow E(T \mid X=x+1) / E(T \mid X=x) = e^{\beta}$
  - "mean ratio"

General Form of (Parametric) AFT Models

- Let $\varepsilon \sim f_{\varepsilon}(w)$, $E(\varepsilon)=0$, and $Var(\varepsilon)=1$.
- Assume

$$\ln T = \mu + \boldsymbol{X}'\boldsymbol{\beta} + \sigma \cdot \varepsilon$$

- A linear model on the logarithmic scale with random error ε.
  - $\ln T_0 = \mu + \sigma \cdot \varepsilon$

Most Frequently Used Parametric Models

- *T ~ Weibull (PH, AFT)* $\rightarrow \varepsilon \sim$ *Gumbel,* $S_\varepsilon(w) = \exp(-e^w)$
- *T ~ log-normal (AFT)* $\rightarrow \varepsilon \sim N(0,1)$, $S_\varepsilon(w) = 1-\Phi(w)$
- *T ~ log-logistic (AFT)* $\rightarrow \varepsilon \sim$ *logistic*, $S_\varepsilon(w) = 1/(1+e^w)$

- *T ~ (Generalized) Gamma*
- *T ~ (Generalized) F*

The Choice of a Parametric Model (1)

- *Generalized gamma* GΓ(κ ,σ):

$$S(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(z), & \text{if } \kappa = 0 \\ I(\gamma, u), & \text{if } \kappa < 0 \end{cases}$$

$$f(t) = \begin{cases} \frac{\gamma^{\gamma}}{\sigma t \sqrt{\gamma}\Gamma(\gamma)} \exp\left(z\sqrt{\gamma} - u\right), & \text{if } \kappa \neq 0 \\ \frac{1}{\sigma t \sqrt{2\pi}} \exp(-z^2/2), & \text{if } \kappa = 0 \end{cases}$$

where $\gamma = |\kappa|^{-2}$, $z = \text{sign}(\kappa)\{\log(t) - \mu\}/\sigma$, $u = \gamma \exp(|\kappa| z)$, $\Phi(z)$ is the standard normal cumulative distribution function, and $I(a, x)$ is the incomplete gamma function. See the `gammap(a,x)`

- A family of models
  - κ = 1 → Weibull
  - κ = σ = 1 → exponential
  - κ = 0 → log-normal

The Choice of a Parametric Model (2)

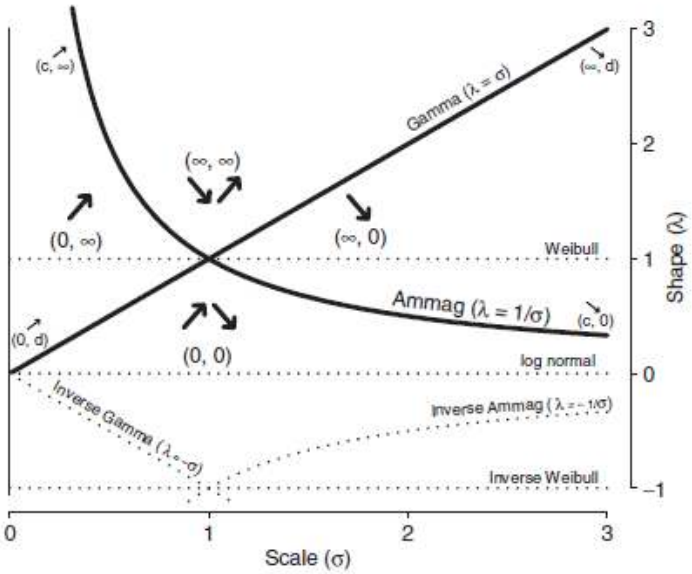


Figure 1. A schematic representation in the (σ, λ) half-plane of the generalized gamma distribution family. The four regions defined by the two curves include the four common types of hazard function: increasing and decreasing failure rate, bathtub and arc-shaped.

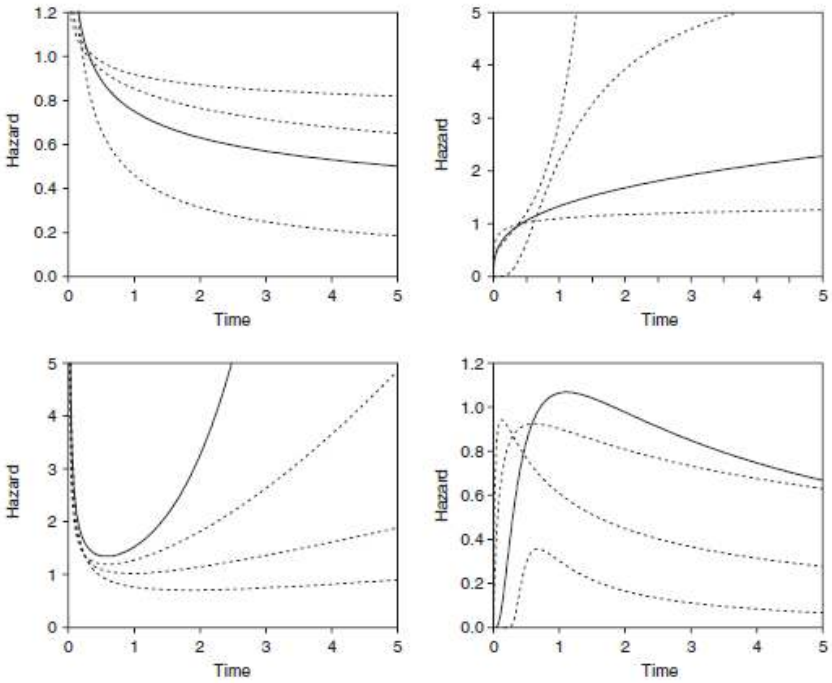


Figure 2. Examples of the four types of hazard function for the generalized gamma distribution, corresponding to the four regions in Figure 1.

Ref: Cox et al., *Statist in Med* 2007, 4352

The Choice of a Parametric Model (3)

- *Generalized F:* GF(β, σ>0, m_1, m_2)

$$f_{GF}(t) = \frac{e^{-\beta m_1/\sigma_t (m_1/\sigma)-1}(m_1/m_2)^{m_1}}{\sigma B(m_1, m_2)[1+(m_1/m_2)(e^{-\beta t})^{1/\sigma}]^{(m_1+m_2)}}$$

- For $q=(m_1^{-1}-m_2^{-1})\ (m_1^{-1}+m_2^{-1})^{-1/2}$ i $p= 2(m_1^{-1}+m_2^{-1})^{-1}$ we get
  - q=0, p=0 → log-normal
  - q=0, p=1 → log-logistic
  - q=1, p=0 → Weibull
  - q>0, p=0 → generalized gamma

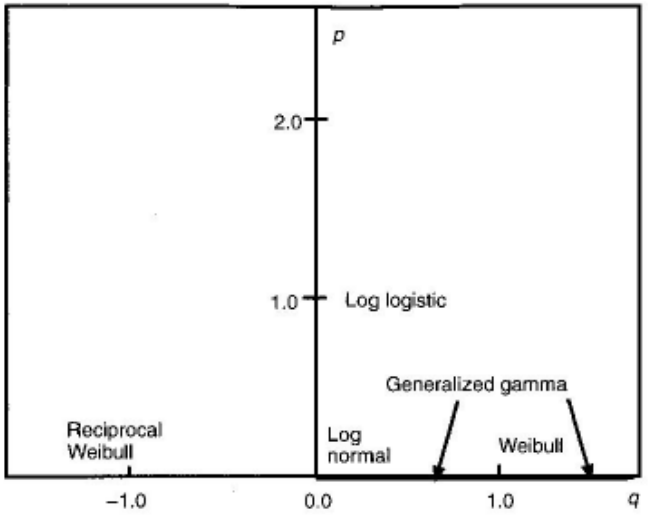


Figure 3.2 Special cases of the log F model. Note that $\{(q, 0) : q \geq 0\}$ gives the generalized gamma model.

Parametric AFT Models

- Assume a particular distribution of the failure time.
- Advantages: allow more precise inferences; more robust to omission of covariates.
- Disadvantages: the parametric assumption.

Estimation of a Parametric AFT Model

- Likelihood function (right-censored data):

$$L(\beta,\mu,\sigma)=\prod_{j=1}^{n} f\left(t_j\right)^{\delta_j} S\left(t_j\right)^{1-\delta_j} = \prod_{j=1}^{n} \left\{\left(\sigma t_j\right)^{-1} f_{\varepsilon}\left(w_j\right)\right\}^{\delta_j} S_{\varepsilon}\left(w_j\right)^{1-\delta_j}$$

where $w_j = (\ln t_j - \mu - \boldsymbol{X}_j'\boldsymbol{\beta})/\sigma$

- Estimation of the variance of the estimates of μ, σ, and β based on the inverse of the observed information matrix.

Diagnostics

- Parametric models allow for various kinds of model diagnostics.
  - Compare observed and predicted survival curves.
  - Residuals $\epsilon = \{\ln T - (\mu + \beta \cdot X)\}/\sigma$ should behave as a (censored) sample from a known distribution - Kaplan-Meier estimate should corrrespond to the parametric $S_\varepsilon(\epsilon)$.
  - Cox-Snell residuals: $r_{CS,i} = \Lambda_i(t_i) = -\ln S_i(t_i) = -\ln S_\varepsilon(\epsilon_i)$; should behave as a (censored) sample from the exponential distribution with $\lambda=1$.

Semiparametric AFT Model

- Assume

$$\ln T = \boldsymbol{X}'\boldsymbol{\beta} + \varepsilon$$

  without specifying the distribution of ε.

- Usually, no intercept (difficult to estimate from censored data; Wei, *Statist in Med* 1992; Jin, *Comm for Statist Appl & Meth* 2016).
  - E(ε) not constrained to 0

Estimation of the Semiparametric AFT Model (1)

- Estimating equations: $$U_{n,\varphi}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\varphi_i(\beta)\delta_i\left[X_i-\frac{\frac{1}{n}\sum_{j=1}^{n}X_j I\{e_j(\beta)\geq e_i(\beta)\}}{\frac{1}{n}\sum_{j=1}^{n}I\{e_j(\beta)\geq e_i(\beta)\}}\right]$$

  where $e_i(\beta)=\ln T_i - X_i'\beta$ and weight φ_i

  - $\varphi_i = 1$ (logrank)
  - $\varphi_i = \Sigma_j\, I\{e_j(\beta)\geq e_i(\beta)\}/n$ (Gehan)
  - $\varphi_i = \{1\text{-}KM_{ei(\beta)}(t)\}^p$ ($p\geq 0$, Harrington-Fleming)

- For Gehan's weight: $$U_{n,G}(\beta)=\sum_{i=1}^{n}\delta_i\left[\sum_{j=1}^{n}(X_i-X_j)I\{e_j(\beta)\geq e_i(\beta)\}\right]$$

Estimating Equation of the Semiparametric AFT Model

- For Gehan's weight: $U_{n,G}(\beta) = \sum_{i=1}^{n} \delta_i \left[\sum_{j=1}^{n} \left(X_i - X_j \right) I\left\{ e_j(\beta) \geq e_i(\beta) \right\} \right]$

- No censoring

$U_{n,G}(\beta) = \sum_{trt} \sum_{ctr} I(\ln T_{trt} - \ln T_{ctr} < \beta) - \sum_{ctr} \sum_{trt} I(\ln T_{trt} - \ln T_{ctr} > \beta)$

- $U_{n,G}(\beta) = 0 \rightarrow P(\ln T_{trt} - \ln T_{ctr} < \beta) = P(\ln T_{trt} - \ln T_{ctr} > \beta)$

Estimation of the Semiparametric AFT Model (2)

- Estimating equations: $$U_{n,\varphi}(\beta) = \frac{1}{n}\sum_{i=1}^{n} \varphi_i(\beta)\delta_i \left[X_i - \frac{\frac{1}{n}\sum_{j=1}^{n} X_j I\{e_j(\beta) \geq e_i(\beta)\}}{\frac{1}{n}\sum_{j=1}^{n} I\{e_j(\beta) \geq e_i(\beta)\}} \right]$$

- In general, it is difficult to solve the equation $U_{n,\varphi}(\beta) = 0$:

  U is neither continuous nor componentwise monotone in β.

Estimation of the Semiparametric AFT Model (3)

- Least-squares (Buckley & James, *Biometrika* 1979; Jin, *Biometrika* 2006)
- Linear programming (Jin et al., *Biometrika* 2003)
- Induced smoothing (Brown and Wang, *Statist in Med* 2006; Chiou et al. *Statist in Med* 2015)

Induced Smoothing

$$U_{n,G}(\beta) = \sum_{i=1}^{n} \delta_i \left[\sum_{j=1}^{n} \left(X_i - X_j \right) I\left\{ e_j(\beta) \geq e_i(\beta) \right\} \right]$$

- Consider $E_Z\{U_{n,G}(\beta+\Gamma Z/\sqrt{n})\}$, with $Z \sim N(0,1)$ and Γ a pxp matrix (usually, an identity).

- Then one gets the smoothed estimating equation:

$$\tilde{U}_{n,G}(\beta) = \sum_{i=1}^{n} \delta_i \left[\sum_{j=1}^{n} \left(X_i - X_j \right) \Phi\left\{ \sqrt{n} \frac{e_j(\beta) - e_i(\beta)}{\sqrt{\left(X_i - X_j \right)' \Gamma^2 \left(X_i - X_j \right)}} \right\} \right]$$

- Easy to solve, derivatives exist, etc.

Variance-covariance Estimation

- Sandwich estimator:

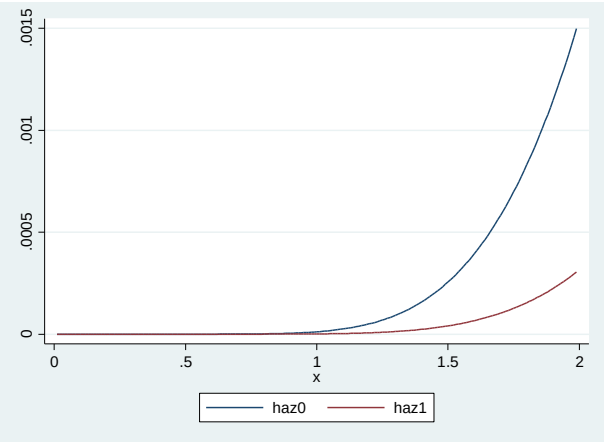
$$\hat{V}(\beta_{n,G}) = \left\{ \frac{1}{n} \frac{\partial \tilde{U}_{n,G}(\beta_{n,G})}{\partial \beta} \right\}^{-1} \hat{V}\left\{ U_{n,G}(\beta_{n,G}) \right\} \left\{ \frac{1}{n} \frac{\partial \tilde{U}_{n,G}(\beta_{n,G})}{\partial \beta} \right\}^{-1}$$

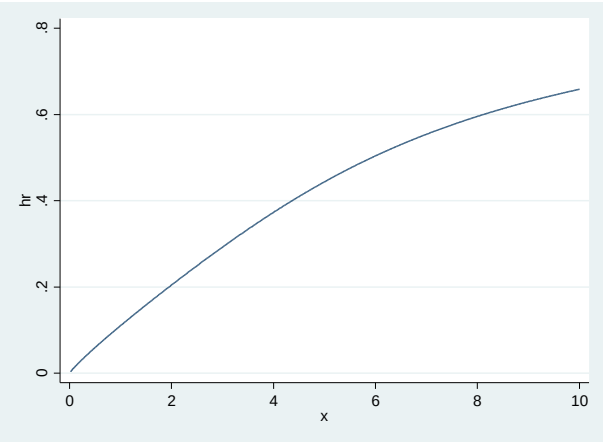
- Estimation of $V\{U_{n,G}(\beta_{n,G})\}$:
  - explicit formula: Johnson and Strawderman, *Biometrika* 2009
  - multiplier bootstrap: Chiou et al., *Statist in Med* 2015

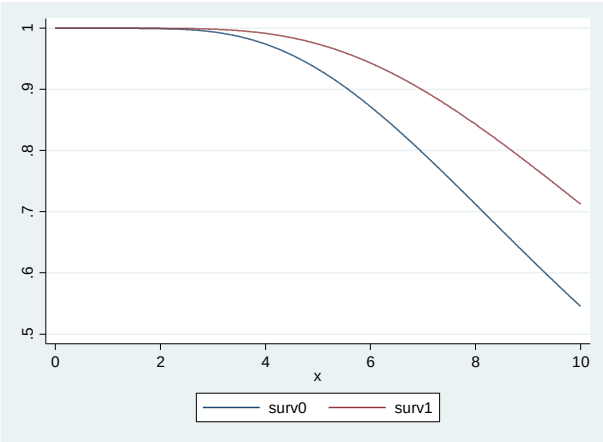
Log-normal AFT Data (1)

$$\ln T = \mu + \beta \cdot trt + \sigma \cdot \varepsilon$$

- $\varepsilon \sim N(0,1)$, $\sigma = 0.5$
- Mean T: control = 12 mths, experimental = 15 mth
- Mean ln(T): control=2.36, experimental = 2.58
  - μ=2.36, β = 0.223, e^{β} = 15/12=1.25







Semi-parametric AFT Model Fitted to Log-normal Data (1)

- Accrual 24 mths, follow-up 12 mths, *n*=2 x 240 pts
- 100 simulations

Follow-up (censoring)	mean diff	emp. SD	model SD	parametric	emp. SD
(no cens)	0.221	0.045	0.047	0.221	0.044
6 (25%/36%)	0.224	0.052	0.050	0.224	0.051
9 (16%/26%)	0.222	0.048	0.048	0.224	0.046
12 (10%/18%)	0.216	0.039	0.047	0.217	0.039

Log-normal AFT Data (2)

$$\ln T = \mu + \beta \cdot trt + \theta \cdot X + \sigma \cdot \varepsilon$$

- $\varepsilon \sim N(0,1)$, $\sigma = 0.5$
- Mean $\ln(T) \mid X$: control=2.36, experimental = 2.58
  - μ=2.36, $\beta = 0.223$, $e^{\beta} = 15/12 = 1.25$
- $X \sim N(0,0.25)$, $\theta = 0.5$
- Mean $T \mid X$: control = $12e^{\theta \cdot X}$, experimental = $15e^{\theta \cdot X}$

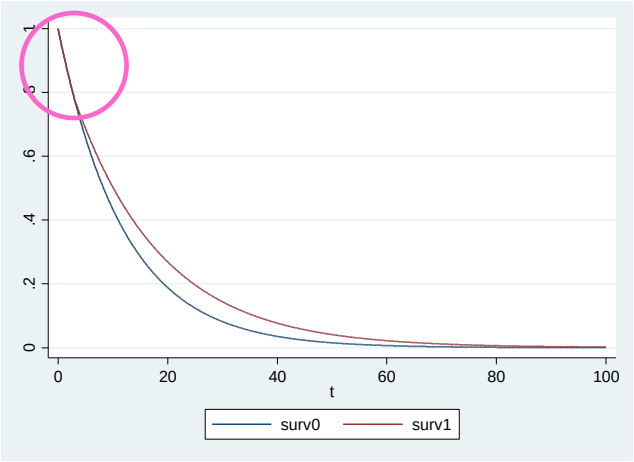
Semi-parametric AFT Model Fitted to Log-normal Data (2)

- Accrual 24 mths, follow-up 12 mths, *n*=2 x 240 pts
- 100 simulations

Follow-up (censoring)	**mean diff (no X)**	**emp. SD**	**model SD**	**mean diff (+X)**	**emp. SD**	**model SD**
(no cens)	0.225	0.050	0.052	0.223	0.043	0.047
6 (26%/37%)	0.229	0.054	0.057	0.225	0.046	0.051
9 (18%/28%)	0.221	0.056	0.055	0.221	0.052	0.049
12 (11%/20%)	0.223	0.055	0.054	0.222	0.050	0.048

AFT Model and Delayed Treatment Effect

- $S_{ctrl}(t)=S_{trt}(t)$ for $t \le C$, and
  $S_{ctrl}(t)<S_{trt}(t)$ for $t>C$



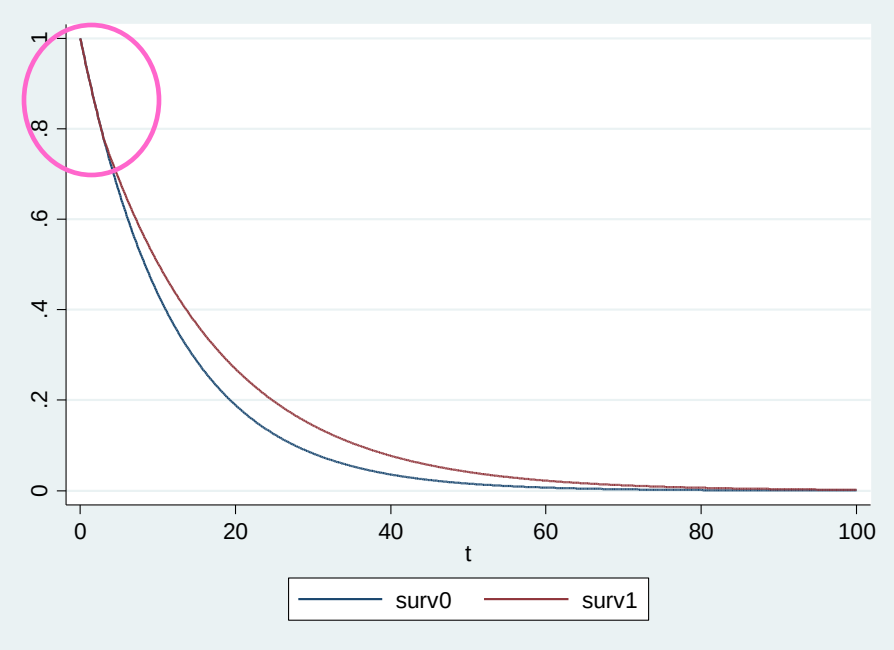
- Under AFT model, $S_{trt}(t) = S_{ctrl}(t / e^{\beta})$

- Hence,
  for $t \le C$, $S_{ctrl}(t)=S_{trt}(t) \rightarrow \beta = 0$
  for $t>C$, $S_{ctrl}(t)<S_{trt}(t) \rightarrow \beta > 0$

Delayed Treatment Effect Data (1)

- Control: hazard $\lambda = 1/12$
- Experimental:

$\lambda = 1/12$ up to C=3 mths, $\lambda e^{\beta} = e^{-0.288}/12 = 0.75/12$ after 3 mths



Delayed Treatment Effect Data (2)

- Control:

  mean $T = 1/\lambda = 12$ mths

  mean ln $T = -\ln \lambda - \gamma = -\ln(1/12) - 0.5772 = 1.908$

- Experimental:

  mean $T = \{1+e^{-\lambda C}(e^{-\beta}-1)\}/\lambda = 15.1$ mths

  mean ln $T = \exp(-\lambda C(1-e^{\beta})\Gamma(0,\lambda C e^{\beta}) - \Gamma(0,\lambda e^{\beta}) - \ln\lambda - \gamma = 2.045$

- Mean difference: $T = 3.1$ mths, ln $T = 0.137$

Semiparametric AFT Model Fitted to the Delayed Treatment Effect Data

- Accrual 24 mths, follow-up 12 mths, n=2 x 240 pts

  → 80% power for HR=0.75 if λ_{ctrl}=1/12

- 100 simulations

Follow-up (censoring)	mean diff	emp. SD	model SD	PH	Exp
(no cens)	0.185	0.107	0.110	-0.220	-0.227
6 (26%/34%)	0.179	0.120	0.127	-0.217	-0.223
9 (21%/28%)	0.151	0.101	0.122	-0.192	-0.197
12 (16%/24%)	0.169	0.131	0.117	-0.214	-0.219

- for λ_{trt} = 1/15.1, mean ln T = -ln(1/15.1) – 0.5772 = 2.137
- 2.137-1.908 = 0.229

Conclusions

- Empirical evidence against the PH assumption
- Complications due to the strong nature of the assumption
  - Bias due to omitted covariates
- Semi-parametric AFT modelling practically feasible
- A serious alternative to the semi-parametric PH model
  - Robust to omitted covariates
  - Interpretation in terms of the “mean ratio”