Evaluation of Time-to-event Surrogate Endpoints Using Accelerated Failure-time Models

Tomasz Burzykowski

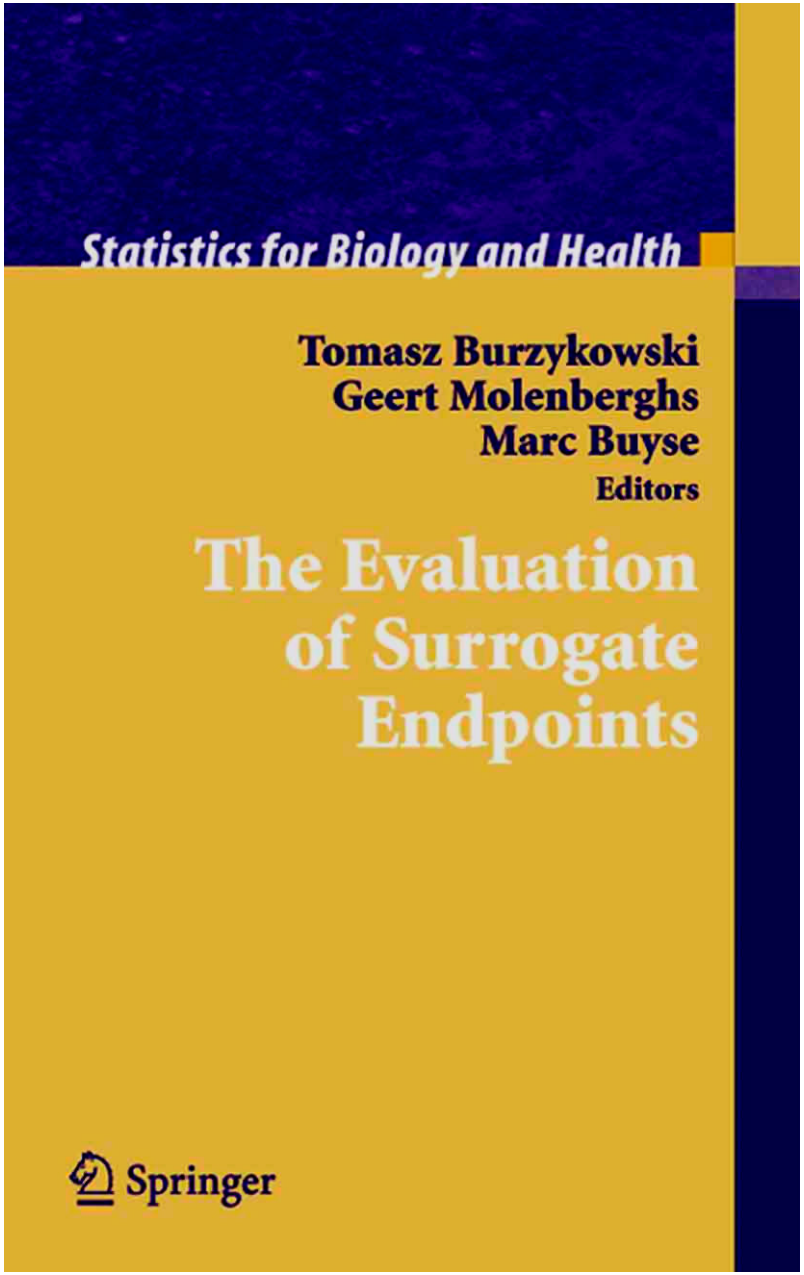
International Drug Development Institute (IDDI)

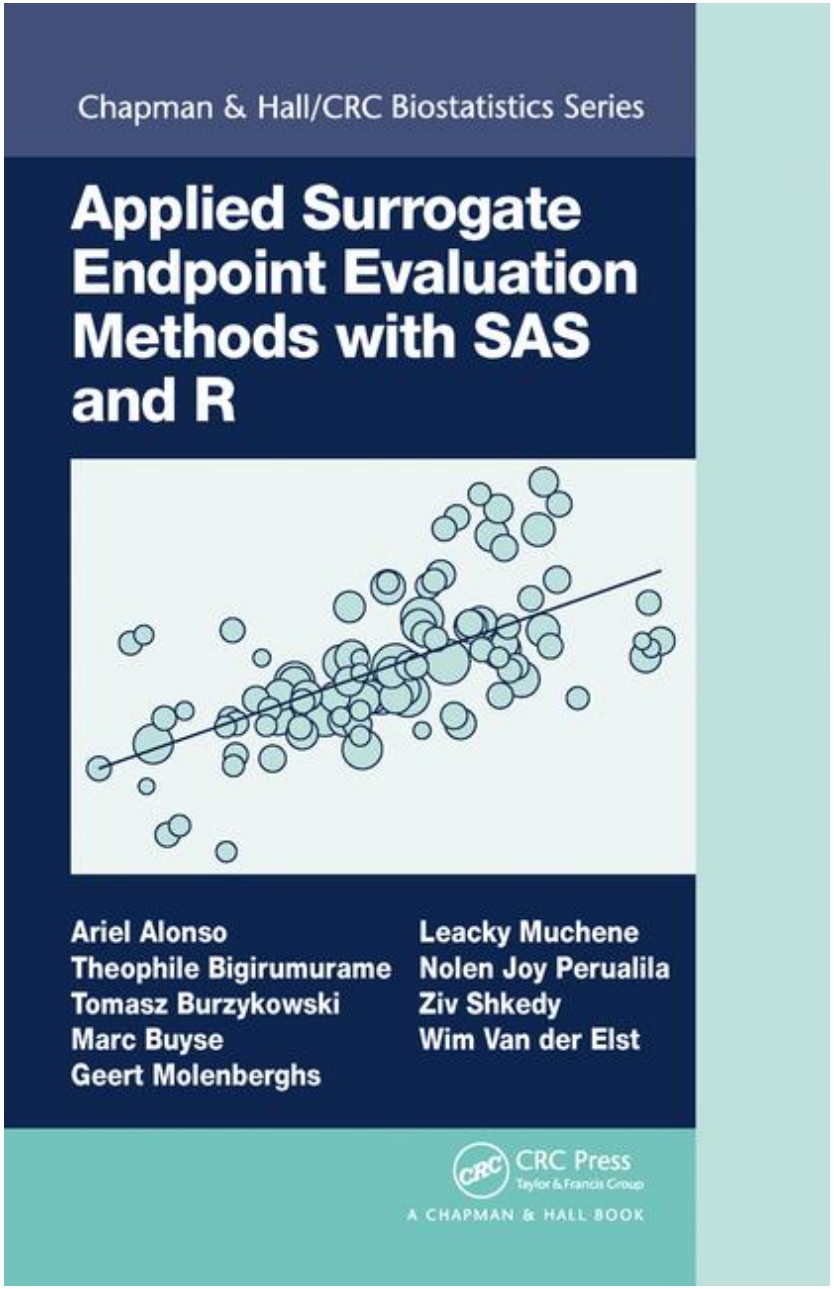
&

Hasselt University

Belgium

tomasz.burzykowski@iddi.com


Statistics for Biology and Health
Tomasz Burzykowski
Geert Molenberghs
Marc Buyse
Editors
The Evaluation of Surrogate Endpoints
Springer


Chapman & Hall/CRC Biostatistics Series
Applied Surrogate Endpoint Evaluation Methods with SAS and R
Ariel Alonso
Theophile Bigirumurame
Tomasz Burzykowski
Marc Buyse
Geert Molenberghs
Leacky Muchene
Nolen Joy Perualila
Ziv Shkedy
Wim Van der Elst
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Outline

- Surrogate validation in multiple trials
- Time-to-event surrogates
- Example: AML
- PH model issues
- AFT model
  - parametric
  - semi-parametric
  - multivariate semi-parametric
- Example: AML
- Conclusions

Terminology

- Clinical endpoint:
  a characteristic or variable that reflects how a patient feels, functions, or survives
- Biomarker:
  a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention
- Surrogate endpoint:
  a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm)

Refs: *Biomarkers Definition Working Group, Clin Pharmacol Ther 2001, 69: 89;*
De Gruttola et al, Controlled Clin Trials 2001, 22: 485.

Validation Based on Precision of Prediction

"The effect of treatment on a surrogate endpoint must be reasonably likely to predict clinical benefit"

*Ref: Biomarkers Definitions Working Group, Clin Pharmacol Ther 2001;**69**:89.*

Important implications

- A prediction model is needed
  - not in the approaches by Prentice (1989), Freedman et al. (1992), ...
  - ... present in the one by Buyse and Molenberghs (1998)
- Validity of a surrogate ≈ quality of prediction
- Model extrapolated to a new treatment (mechanism)
  - validation across a range of classes of treatments
  - a “leap of faith”; biological argumentation in addition to the statistical

Analysis Based on Multiple Trials

- Context:
  - Multicenter trials
  - Meta-analysis

- Surrogacy levels:
  - **Trial-level (predictive)**

    How close is the relationship between the treatment effects on the surrogate and true endpoints, based on the different trials (units) ?
  - **Individual-patient-level (prognostic)**

    How close is the relationship between the surrogate and true outcomes, after accounting for trial and treatment effects ?

*Ref: Buyse et al, Biostatistics 2000;**1**:49*.

Formal Statistical Definition of Surrogate Endpoints: Multiple Trials

Based on a two-stage model

First stage: a (joint) model for individual observations on surrogate and true endpoints

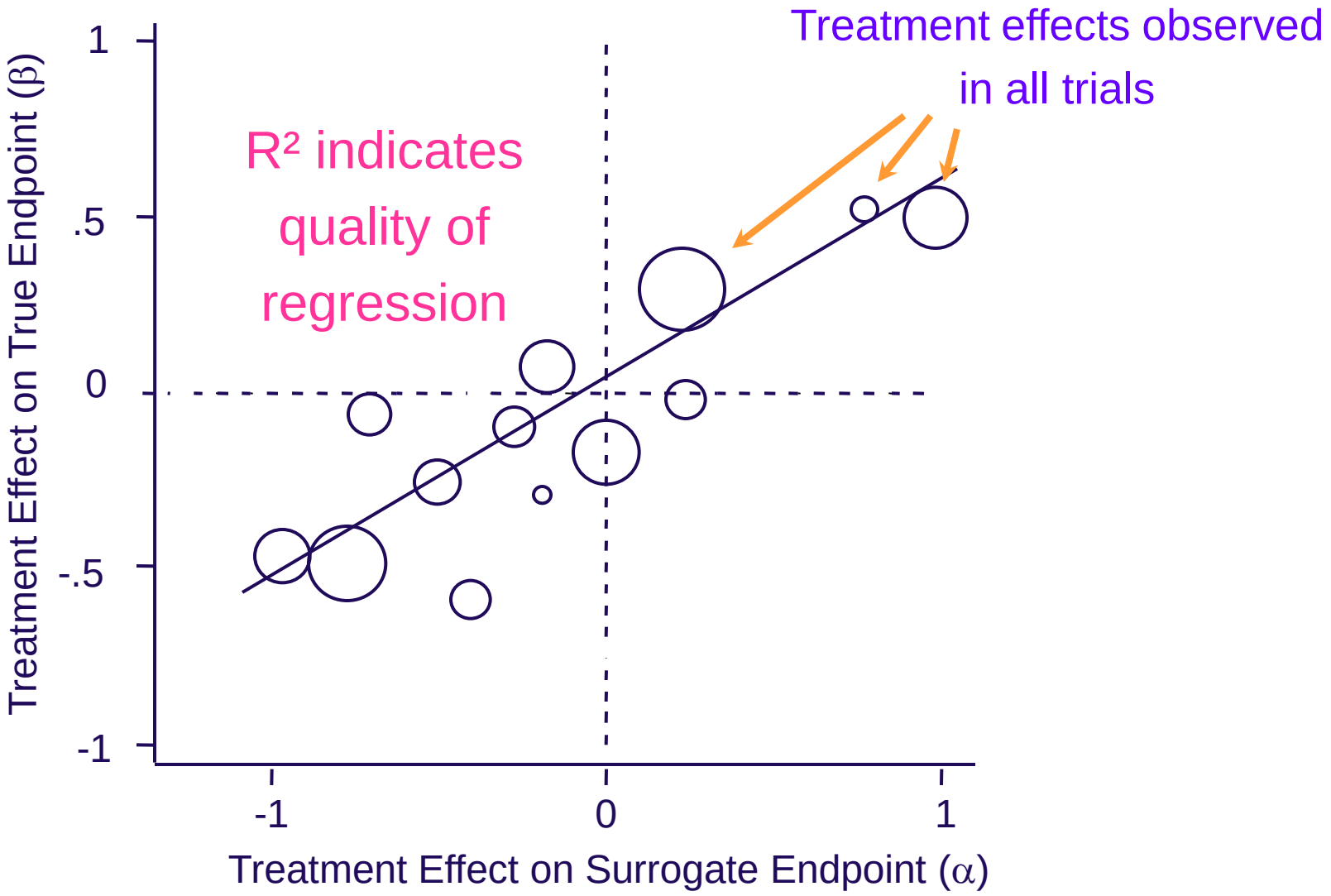
- (individual-level) association between endpoints
- (trial-specific) effects of treatment on surrogate/true endpoint

Second stage: a linear model for the trial-specific treatment effects

- $R^2_{trial} \approx 1$: surrogate "valid at the trial-level"
- *adjustment for estimation-error necessary*

*Ref: Buyse et al, Biostatistics 2000;**1**:49*; *Burzykowski et al., Applied Statistics 2001;**50**:405*

Prediction of Treatment Effect: Multiple Trials



Validation of Time-to-event Surrogates for Time-to-event Clinical Endpoints

- Burzykowski *et al.* (2001)
  - First-stage model: a copula, *Weibull margins*
    - likelihood-based estimation of the copula parameter and treat effects
  - SAS macros

- Burzykowski (2017)
  - First-stage model: a copula, *non-parametric, treat-specific margins*
    - likelihood-based estimation of the copula parameter, Shih & Louis (1995)
    - (GEE) estimation of treatment effects from marginal PH models
      - v-cov of estimated treatment effects adjusted for individual-level correlation
  - "Regular" SAS code

Ref: Alonso et al., Applied Suurogate Endpoint Evaluation Methods with SAS and R, 2017

Acute Myeloid Leukemia: EFS as Surrogate for Survival

- 4 trials, the German-Austrian AML Study Group (AMLSG):
  - AMLHD 98B (n=254)
  - AMLSG 06-04 (NCT00151255, n=189)
  - AMLSG 07-04 (NCT00151242, n=1,100)
  - AMLSG 12-09 (NCT01180322, n=268)

- 7 treatment-contrasts, 1,811 pts
  - standard induction vs. standard + valproic acid (VA)
  - standard induction vs. standard + azacitidine
  - standard induction vs. standard + all-*trans* retinoic acid (ATRA)

Ref: Schlenk et al., Leukemia 2004:1798; Tassara et al., Blood 2014:4027; Schlenk et al., Annals of Hematology 2016:1931; Schlenk et al., Blood 2012:#412 (abstract).

PFS and OS Hazard Ratios

Trial	Contrast	Control/experimental	N	EFS HR (95% CI)	OS HR (95% CI)
AMLSG 06-04	1	SI/SI + VA	96/93	0.96 (0.72-1.29)	1.10 (0.81-1.49)
AMLSG 12-09	2	SI/SI + azacitidine (before/concurrently/after)	36/103	1.47 (0.96-2.26)	1.41 (0.89-2.23)
	3	SI/SI + azacitidine (after)	64/65	1.67 (1.13-2.46)	1.44 (0.92-2.27)
AMLHD 98B	4	SI/SI + ATRA	128/126	0.75 (0.58-0.97)	0.75 (0.58-0.97)
AMLSG 07-04	5	SI/SI + ATRA	92/94	1.19 (0.84-1.67)	0.95 (0.63-1.44)
	6	SI + VA/SI + ATRA + VA	95/91	0.93 (0.67-1.28)	1.01 (0.70-1.45)
	7	SI/SI + ATRA	369/359	0.97 (0.82-1.15)	0.86 (0.70-1.06)

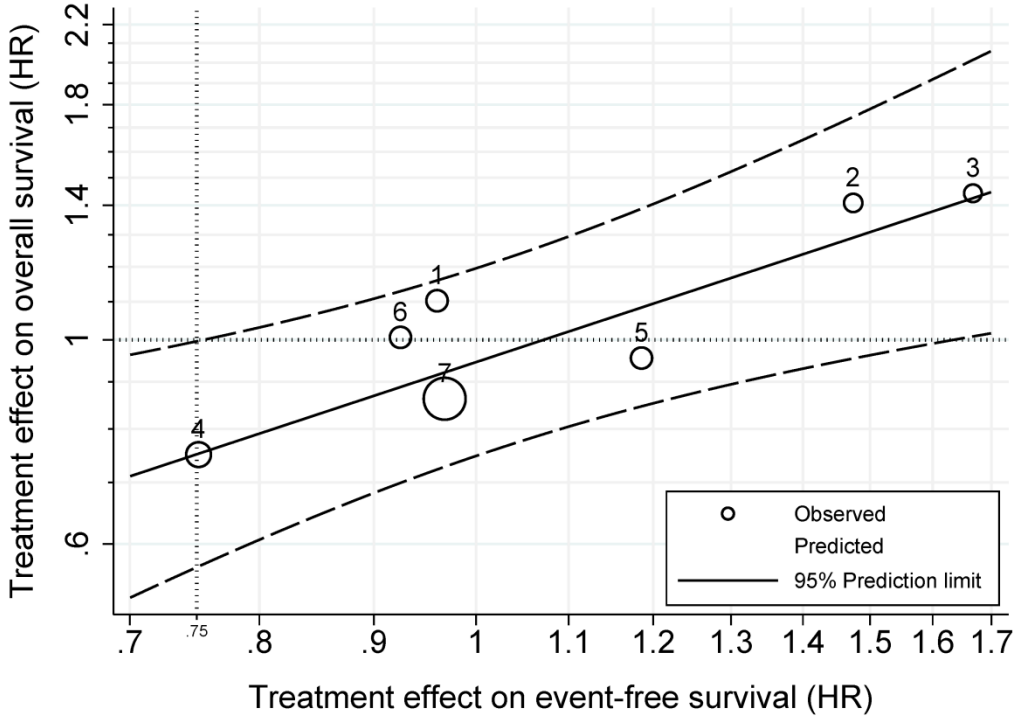
Prediction of Indivdual Outcomes

- Spearman's rank correlation, 0.70 (95% CI, 0.67-0.72)
  - Clayton copula, non-parametric margins
    - issues with the Weibull-margin assumption
- A moderate correlation between EFS and OS for an individual patient

Treatment Effects on EFS and OS, PH (sample-size-weighted) Analysis

R = 0.87 (95%CI = 0.68 - 1.00)

$\log HR_{OS} = -0.06 + 0.80 \times \log HR_{EFS}$



PH Model: Complications (1)

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

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

- If $\beta_2 \neq 0$, omitting Z_2 induces time-dependence of HR for Z_1
  - Difficult to distinguish the effect from a true time-dependent coefficient
- Hence, omitting Z_2 can 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 Model: Complications (2)

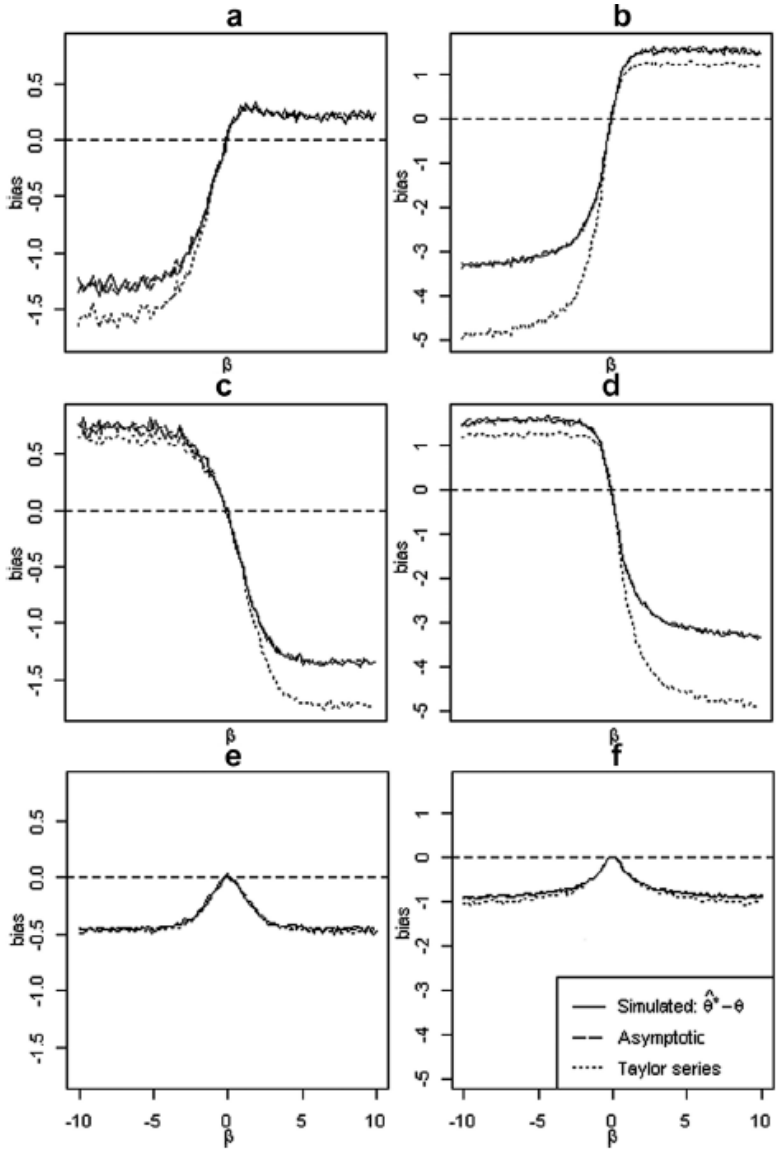


Figure 1. Comparison of simulated biases, asymptotic biases and first-order Taylor series approximations for different types of omitted covariate and censorship. Since θ^* is the asymptotic value of the MLE $\hat{\theta}^*$ and the sample size=10,000 is large, we calculated the simulated bias by $\hat{\theta}^* - \theta$. 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

(Parametric) AFT Models

- Assume

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

with $\varepsilon \sim f_{\varepsilon}(w)$, $E(\varepsilon)=0$, and $Var(\varepsilon)=1$

  - Most popular: Weibull, log-normal , log-logistic models

- A linear model on the logarithmic scale with random error ε
  - *Less vulnerable to the bias due to omission of a covariate*

- Simple interpretation of β in terms of shortening/extending the mean time to event
  - *Not used in clinical trials due to the parametric assumption?*

Semi-parametric AFT Model

- Assume

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

with distribution of ε *left unspecified*

- A serious alternative to the (semi-parametric) PH model

- Estimation
  - Rank-based (inverting the weighted logrank test)
    - Prentice (1978), Tsiatis (1990), Ying (1993), Huang (2002), Jin et al. (2003), Strawderman (2005), Brown & Wang (2007)
  - Least-squares
    - Buckley-James + EM (1979), Ritov (1990), Lai & Ying (1991), Jin et al. (2006)
  - IPCW loss function
    - Robins & Rotnitzky (1992), Zhou (1992), Stute (1993, 1996)

Multivariate Semi-parametric AFT Model

- Assume

$$\ln T_{ij} = \mu_j + \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij}$$

  with distributions of ε_{ij} left unspecified
  - (some) ε_{ij} can have the same distribution

- Estimation (independent working model)
  - Rank-based
    - Jin et al. (2006), Johnson & Strawderman (2009), Li & Yin (2009) , Wang & Fu (2011)
  - Least-squares
    - Jin et al. (2006)

- GEE: LS-based approach (Chiou *et al.*, 2014)
  - *aftgee* R-package (needs a revision)

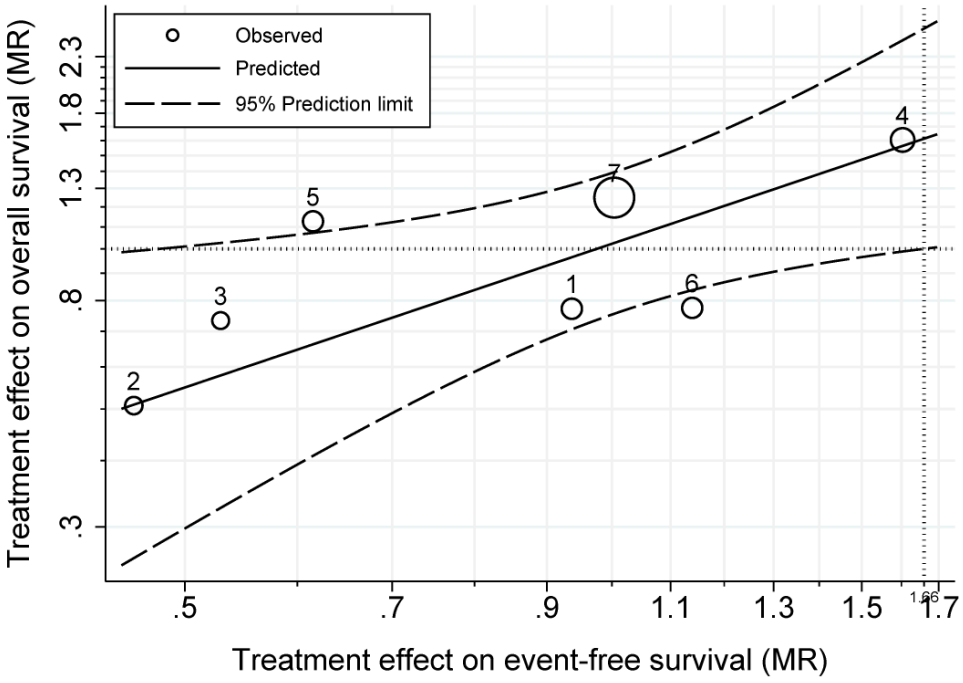
PFS and OS Mean-time Ratios

Trial	Contrast	EFS		OS	
		median time (Exp/Ctrl, days)	mean-time ratio (95% CI)	median time (Exp/Ctrl, days)	mean-time ratio (95% CI)
AMLSG 06-04	1	57/96	0.94 (0.61-1.43)	220/312	0.77 (0.48-1.25)
AMLSG 12-09	2	81/235	0.46 (0.23-0.91)	366/801	0.51 (0.31-0.83)
	3	75/243	0.53 (0.28-1.00)	471/945	0.73 (0.39-1.36)
AMLHD 98B	4	104/52	1.60 (1.10-2.33)	371/210	1.60 (1.10-2.34)
AMLSG 07-04	5	219/399	0.61 (0.27-1.37)	2892/1571	1.13 (0.49-2.58)
	6	249/172	1.14 (0.57-2.27)	789/806	0.77 (0.40-1.45)
	7	291/273	1.00 (0.72-1.39)	2013/1384	1.25 (0.89-1.75)

Treatment Effects on EFS and OS, AFT (estimation-error-adjusted) Analysis

R = 0.93 (95%CI = 0.61 - 1.00)

$$\log MR_{OS} = 0.02 + 0.90 \times \log MR_{EFS}$$



EFS as Surrogate for OS in Acute Myeloid Leukemia: Conclusions

- EFS correlates moderately well with OS
- Treatment effects on EFS correlate very well with treatment effects on OS ($R_{trial} \approx 0.9$)
  - 95% CIs still relatively wide, though
- EFS may be used as a surrogate for OS

AFT-based Approach to Validation of Time-to-event Surrogates: Conclusions

- Avoids the (stringent) PH assumption
- Treatment-effect measure (MR) easier to intepret than HR
  - *Time to change clinical-trials practice in oncology ?*
- Software available
  - *aftgee* R package (Chiou *et al.*, 2014)