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
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)

Validation Based on Precision of Prediction

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

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 $\approx$ 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 ?

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

Prediction of Treatment Effect: Multiple Trials

$R^2$ indicates quality of regression.

Treatment effects observed in all 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)

## PFS and OS Hazard Ratios

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

\[ R = 0.87 \ (95\% CI = 0.68 - 1.00) \]

\[ \log \text{HR}_{\text{OS}} = -0.06 + 0.80 \times \log \text{HR}_{\text{EFS}} \]
Consider two covariates, a binary $Z_1$ and $Z_2$

Assume the PH model for both, i.e.,

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

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 $\beta_1$

- 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 censoring. Since $\theta^*$ 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 $|\beta^*_1| < |\beta_1|$
(Parametric) AFT Models

♦ Assume

\[ \ln T = \mu + x'\beta + \sigma \cdot \varepsilon \]

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

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

♦ A linear model on the logarithmic scale with random error \( \varepsilon \)

• Less vulnerable to the bias due to omission of a covariate

♦ Simple interpretation of \( \beta \) 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

\[ \ln T = \mu + x'\beta + \varepsilon \]

with distribution of \( \varepsilon \) left unspecified

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

- Estimation
  - Rank-based (inverting the weighted logrank test)
  - Least-squares
  - IPCW loss function
Multivariate Semi-parametric AFT Model

♦ Assume

\[ \ln T_{ij} = \mu_j + x_{ij}'\beta + \varepsilon_{ij} \]

with distributions of \( \varepsilon_{ij} \) left unspecified

• (some) \( \varepsilon_{ij} \) 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)

• \textit{aftgee} R-package (needs a revision)
## PFS and OS Mean-time Ratios

<table>
<thead>
<tr>
<th>Trial</th>
<th>Contrast</th>
<th>EFS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>median time</td>
<td>mean-time ratio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Exp/Ctrl, days)</td>
<td>(95% CI)</td>
</tr>
<tr>
<td>AMLSG 06-04</td>
<td>1</td>
<td>57/96</td>
<td>0.94 (0.61-1.43)</td>
</tr>
<tr>
<td>AMLSG 12-09</td>
<td>2</td>
<td>81/235</td>
<td>0.46 (0.23-0.91)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>75/243</td>
<td>0.53 (0.28-1.00)</td>
</tr>
<tr>
<td>AMLHD 98B</td>
<td>4</td>
<td>104/52</td>
<td>1.60 (1.10-2.33)</td>
</tr>
<tr>
<td>AMLSG 07-04</td>
<td>5</td>
<td>219/399</td>
<td>0.61 (0.27-1.37)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>249/172</td>
<td>1.14 (0.57-2.27)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>291/273</td>
<td>1.00 (0.72-1.39)</td>
</tr>
</tbody>
</table>
Treatment Effects on EFS and OS, AFT (estimation-error-adjusted) Analysis

\[ R = 0.93 \ (95\% CI = 0.61 - 1.00) \]

\[ \log \text{MR}_{OS} = 0.02 + 0.90 \times \log \text{MR}_{EFS} \]
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 interpret than HR
  - *Time to change clinical-trials practice in oncology?*

- Software available
  - *aftgee* R package (Chiou *et al.*, 2014)