Advances in Time-to-Event Analyses in Clinical Trials
Non-Proportional Hazards and Composite End-Points

Douglas E. Schaubel[1]

[1]**Division of Biostatistics, University of Pennsylvania**

Conference on Statistical Issues in Clinical Trials: 4/17/2023

Non-Proportional Hazards

- Proportional hazards assumption frequently fails in practice.
  - treatment effect may depend on follow-up time
  - mortality risk factors may be excluded from the model
- e.g., true model: $\lambda_0(t)\exp\{\beta_A A_i + \boldsymbol{\beta}_1' \boldsymbol{Z}_i\}$
  - fitted model: $\lambda_0(t)\exp\{\beta_A A_i\}$
  - risk factors omitted from a non-linear model
  - $\widehat{\beta}_A$ will be biased
  - but, a model that may fit well: $\lambda_0(t)\exp\{\beta_A(t)\}$

Remedying Non-Proportional Hazards

- Two general themes:
  - adjust model (or log rank test)
    - address time-dependent effect OR
    - address heterogeneity
  - abandon HR; choose an alternative metric
- Restricted Mean Survival Time (RMST):

$$E[T \wedge L] = \int_0^L S(t)dt$$

  - can estimate through $S(t)$
  - can directly estimate via $T \wedge C \wedge L$
  - if not for censoring, we would probably model T directly

Devan Mehrotra: Non-PH, Composite End-Points

- Addressed risk heterogeneity through stratification (prior to unblinding):
  - covariate selection via elastic net
  - break subjects into risk strata
  - seeks to reduce/elminate risk heterogeneity
- Potential issues:
  - heterogeneity may persist (treatment omitted from 1st stage)
  - interpretation of $\sum f_k \beta_k$ (e.g., versus stratified Cox model)

Lu Tian: Survival Distribution of DOR

- Nonparametric methods for contrasting *duration of response*: $T_D - T_R$
- Standard Kaplan-Meier estimator of $(T_D - T_R)|T_R < T_D$ is problematic:
  - biased: induced dependent censoring
  - non-identifiability
- Two main choices presented:
  (1) $\text{RMDOR}(t) = \int_0^\tau \text{PBIR}(t)dt$
  (2) IPCW-based DOR survival function estimators

Lu Tian: Duration of Response (cont'd)

Potential issues:

- Range constraints, in response to identifiability considerations:
  - pre-specify range for response: $T_R \leq \tau_R$
  - implies constraint on duration: $(T_D - T_R) \leq \tau_{\max} - \tau_R$
  - events and patient-time are re-censored
- What are the considerations in choosing between the DOR survival function and the integrated PBIR?

Zhenzhen Xu: Biomarker-Defined Subgroups

- Methods address two primary causes of non-PH:
  - delayed treatment effect: APPLE/SEPPLE
  - risk heterogeneity: PRIME+
- Piece-wise weighted log rank tests
  - APPLE: analytic-based procedures
  - SEPPLE: simulation-based
  - version that allows treatment effect lag t_{ind} to follow a distribution

Zhenzhen Xu: Biomarker-Defined Subgroups (cont'd)

- Addressing heterogeneity:
  - PRIME+
  - latent responder class membership

  Questions:
- Huge catalog of methods!
  - Where should a practitioner start?
  - The causes of non-PH lead to very different solutions. Any way to use the data to suggest which direction to go?

Fan Li: Multiple Robustness with Noncompliance

- Methods for estimating the causal effect on failure time by principal stratum
- Proposal does not involve estimating treatment HRs
- Cox models are applied, but are then transformed and integrated to estimate the survival function
- "Multiple robust" in the sense that models for T, C, Z and G are fitted but not assumed correct

Fan Li: Multiple Robustness (cont'd)

Potential issues:

- Positivity would seem to be a bigger concern than usual, given the $e_g(\boldsymbol{X})$ models? (e.g., moderately large ASD values in Table 3?)
- Cox models are proposed for the failure time, which is a composite.
  - risk factors for death and CVD-hospitalization could be very different
  - estimate *total* hazard via cause-specific hazard models(?)

800 pound gorilla . . .

Causal Inference Interpretation of HR

- Hazard ratio lacks a causal interpretation
  - non-collapsibility
  - additive hazard model has been suggested
- Hazard *contrasts* have been billed as (causally) flawed:
  - taking the HR as an example,

$$\text{HR}(t) = \frac{P(T_1^* \in [t, t+dt) | T_1^* \geq t)}{P(T_0^* \in [t, t+dt) | T_0^* \geq t)}$$

- Causal interpretation,

$$\begin{aligned} \theta(t) &= \frac{\log S_1^*(t)}{\log S_0^*(t)} \\ &= \theta \quad \text{under PH} \end{aligned}$$

Shameless Promotion

- Wei and Schaubel (2008):

BIOMETRICS 64, 724–732
September 2008
DOI: 10.1111/j.1541-0420.2007.00947.x

724 **Estimating Cumulative Treatment Effects in the Presence of Nonproportional Hazards**

Guanghui Wei* and Douglas E. Schaubel
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.
**email:* ghwei@umich.edu

- Proposed cumulative hazard ratio:

To compare each treatment group to the reference group, we propose the following measure,

$$\theta_j(t) = \frac{\Lambda_{0j}(t)}{\Lambda_{00}(t)}, \quad \text{for } j = 1, \ldots, m, \tag{2}$$

Thank You!

douglas.schaubel@pennmedicine.upenn.edu