Planning and analyzing clinical trials with competing risks: Recommendations for choosing appropriate statistical methodology

JC Poythress[1], Misun Yu Lee[2], and James Young[2]

[1]Department of Statistics, University of Georgia

[2]Data Science, Astellas Pharma Inc.

UNIVERSITY OF GEORGIA

Research Questions

1. How are non-parametric hypothesis tests affected when competing events are treated as censored?
2. Under what conditions is the estimated subdistribution hazard ratio (SHR) in the Fine-Gray (F-G) model [1] substantially different than the estimated cause-specific hazard ratio (csHR) for the event of interest in the cause-specific hazards (CSH) model?
   - Does the treatment effect on the competing event matter?
   - Does the proportion of competing events matter?
3. Can model diagnostics detect lack-of-fit when one of the CSH model or F-G model holds, but the other is misspecified? How does model misspecification affect inference?

Background

- **Context**: Time-to-event data analysis.
- **Problem**: How to handle more than one type of event?
  - Competing risk: An event whose occurrence precludes the occurrence of the event of interest.
- **One solution**: Treat competing events as censored and use traditional time-to-event analysis methodology.
  - This strategy can result in **biased** inference.
    - e.g. 1-KM as an estimator of the CIF is biased upwards.
  - Violates assumption of non-informative censoring.
- **Better solution**: Use methods that properly account for competing events.
- The main functions of interest, their interpretations, and appropriate models/estimators for the functions in both traditional time-to-event analysis and the competing risks setting are described in Table 1.

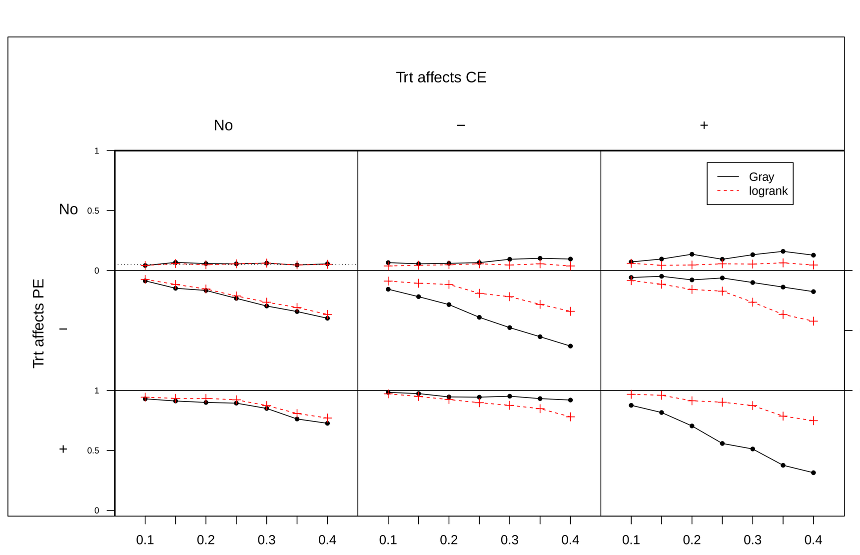
Simulation Study

- Treatment and control, with $N = 250$ per arm.
- Data simulated under CSH and F-G models.
- Treatment effects for both competing event (CE) and primary event (PE).
  - "No," "Decreases (-)," and "Increases (+)" correspond to $csHR$ or $SHR = 1, 0.67, 1.5$, respectively.
- Proportion of CEs varied from 10% to 40%.
- Censoring fixed at 30%.
- **Non-parametric hypothesis testing**: logrank test vs. Gray's test [2] for H_0: "$CIF_{1,trt}(t) = CIF_{1,ctrl}(t)\ \forall t$."
- **Semi-parametric modelling**: $\widehat{csHR}$ vs. $\widehat{SHR}$.
- **Goodness-of-fit**: simulation parameters changed to induce mild or severe lack-of-fit.
  - Overlay model-based estimator of CIF on non-parametric estimator.

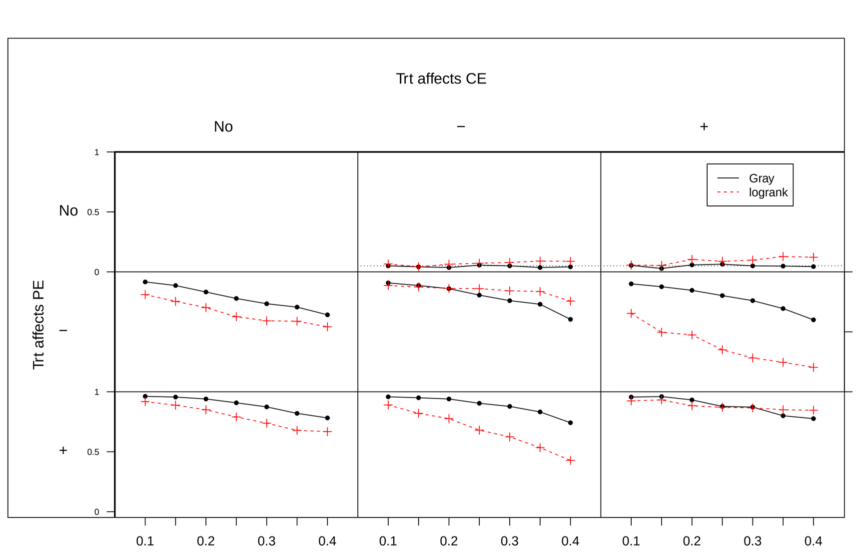
Table 1: Main functions of interest in traditional time-to-event analysis and the competing risks setting.

Name	Definition	Interpretation	Model/Estimator
Traditional			
Survivor function	$S(t) = P(T > t)$	probability that the event occurs after time t	Kaplan-Meier (KM) estimator
Hazard function	$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t \mid T \ge t)}{\Delta t}$	instantaneous event rate at time t, given that the event has not occurred before time t	Cox model: $h_i(t) = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta}\right] h_0(t)$
Competing Risks			
Cumulative Incidence Function (CIF)	$CIF_j(t) = P(T \le t,\ \delta = j), \quad j = 1, \dots, J$	probability of experiencing event j before time t	"KM-like" estimator (But not 1-KM)
Cause-specific Hazard (CSH)	$h_j(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t,\ \delta = j \mid T \ge t)}{\Delta t}, \quad j = 1, \dots, J$	instantaneous rate of event j at time t, among individuals who are event-free up to time t	CSH model: $h_{j,i}(t) = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\beta}_j\right] h_{j,0}(t), \quad j = 1, \dots, J$
Subdistribution Hazard (SH)	$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t,\ \delta = j \mid \{\{T \ge t\} \cup \{T \le t \text{ and } \delta \ne j\}\})}{\Delta t}, \quad j = 1, \dots, J$	instantaneous rate of event j at time t, among individuals who are event-free up to time t or experienced a competing event before time t	F-G model: $\lambda_{1,i}(t) = \exp\left[\boldsymbol{x}_i^T \boldsymbol{\theta}_1\right] \lambda_{1,0}(t)$ (other SHs left unspecified)

Results: Non-parametric Hypothesis Testing



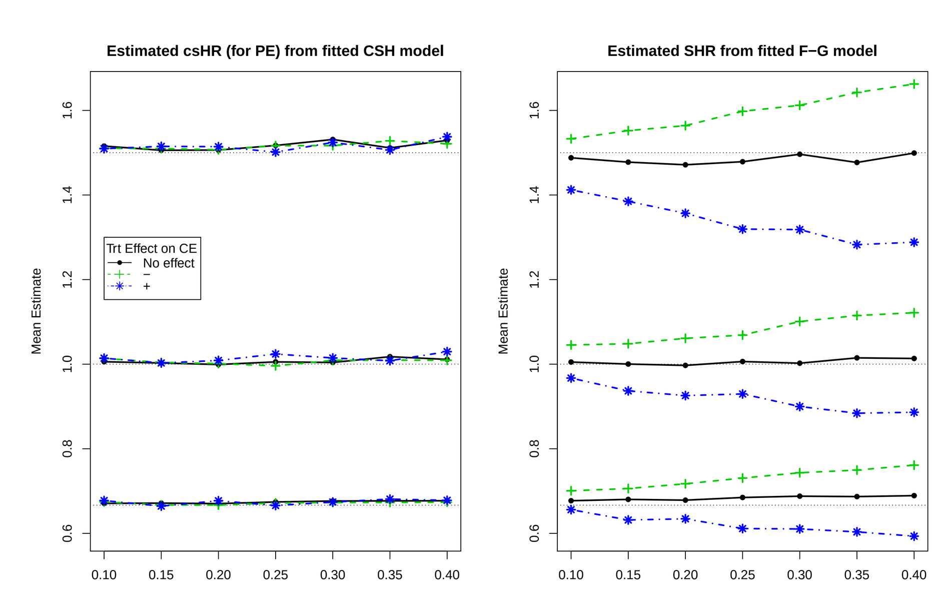
(i) True CSH model



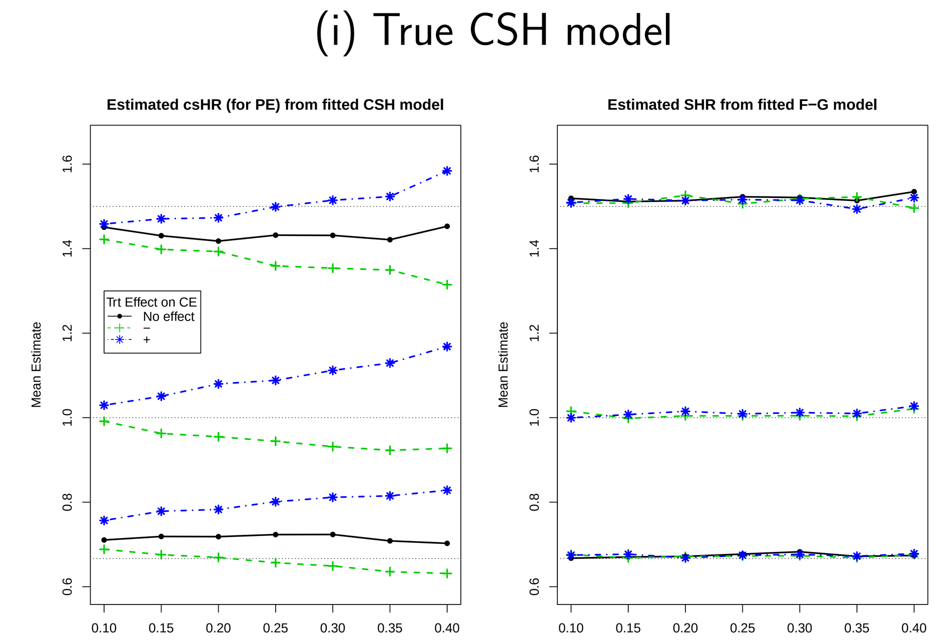
(ii) True F-G model

Figure 1: Proportion of simulations in which $P < 0.05$.

Results: Semi-parametric Modelling



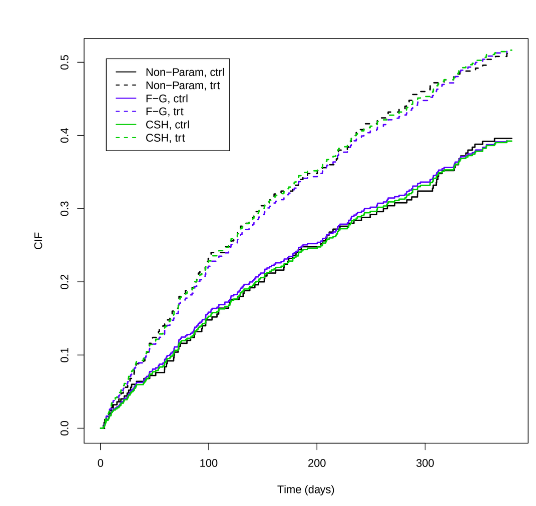
(i) True CSH model



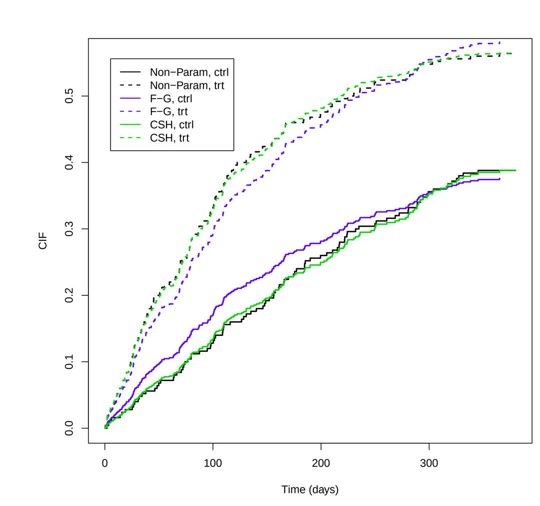
(ii) True F-G model

Figure 2: Means of estimated hazard ratios.

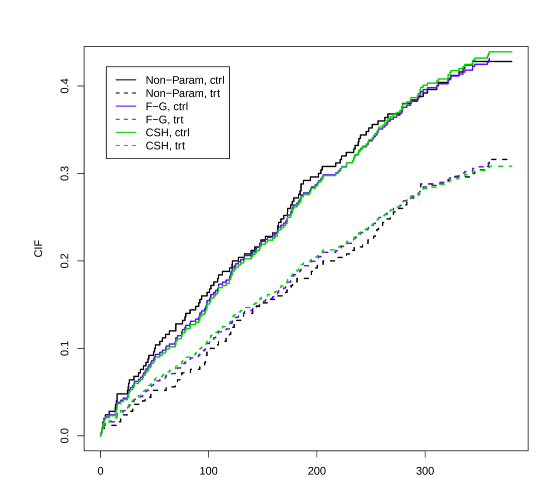
Results: Goodness-of-fit



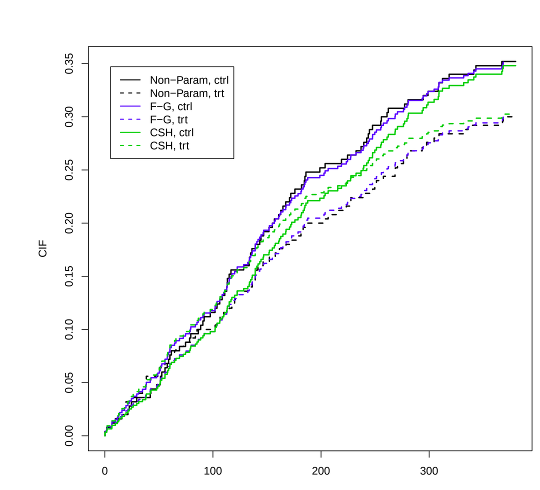
(i) True CSH model, mild lack-of-fit



(ii) True CSH model, severe lack-of-fit



(iii) True F-G model, mild lack-of-fit



(iv) True F-G model, severe lack-of-fit

Figure 3: Plots of non-parametric and model-based estimators of the CIF for the event of interest under true CSH model and true F-G model.

Summary of Results

- **Non-parametric Hypothesis Testing**
  - Use Gray's test instead of the logrank test for testing equality of CIFs.
  - The logrank test can have inflated Type I error rate when H_0 is true and poor power when H_0 is false.
- **Semi-parametric modelling**
  - $\widehat{csHR}$ and $\widehat{SHR}$ differ most when the treatment affects the competing event and the proportion of competing events is large.
- **Goodness-of-fit**
  - The CSH model and F-G model both properly account for competing risks, but are **not interchangeable**.
    - If one model fits the data adequately, it does not imply the other will also!
    - Misspecification of the F-G or CSH model can result in poor inference if the other model is the true model.
  - Traditional GOF methods only useful for detecting lack-of-fit if the proportionality assumption is severely violated.

Recommendations

- Do not ignore competing risks!
- Fit and report the results from both the CSH model and F-G model.
- Prespecify a preferred model and base decisions regarding the trial outcome on that model.
  - e.g. one might choose a preferred model based on convenience of model interpretation.
- Provide for a contingency plan: If there is evidence for significant lack-of-fit in the preferred model, but the other model appears to fit the data adequately, base decisions regarding the trial outcome on the model that fits the data adequately.

References

[1] Jason P. Fine and Robert J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

[2] Robert J. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16(3):1141–1154, 1988.

Contact Information

- Email: jpoythre@uga.edu
- Phone: (706) 542 5232