MATH590 Proposal Presentation

Topic: Survival Analysis - Comparative Evaluation of Max-Combo and RMST Approaches under Non-Proportional Hazards in Phase 3 Clinical Trials

Chang Rong, Loh

July 2025

Background: Standard Log-Rank and Cox PH

- Cox Proportional Hazards (PH) model assumes constant hazard ratio.
- Log-rank test: most powerful test under PH.
- **Problem:** Noticeably in immuno-oncology trials, PH is often violated (e.g. KEYNOTE-024).
  - Delayed treatment effects
  - Crossing hazards
- Raises need for alternatives

Non-PH example: KEYNOTE-024 [4]

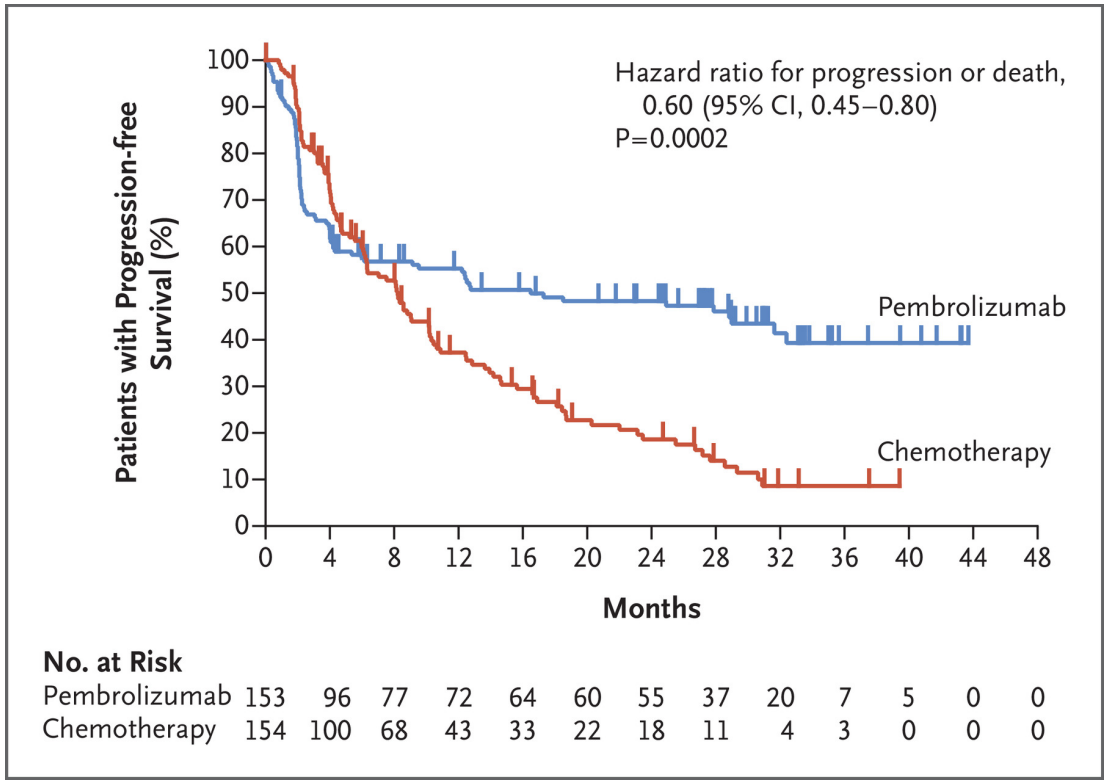

Figure: Reconstructed Kaplan-Meier curve for progression-free survival in KEYNOTE-024

- Alternative treatment benefit appears after ˜6 months.
- Indicates possible delayed treatment effect, violating PH

No Gold Standard: Handling Non-PH

- Current literature lacks consensus [1]
- Main approaches discussed in Freidlin and Korn (2019) [3]:
  - **Max-Combo Test**: combines Fleming-Harrington weights to test for differences in survival
  - **Restricted Mean Survival Time (RMST)**: difference in area under survival curve, which quantifies on average, how much more time one treatment offers within that fixed time frame.

Formulations: Max-Combo

1. **Max-Combo Test (Max of FH-weighted log-rank tests)** [2]
   - Construct weighted log-rank statistics using Fleming–Harrington weights:

$$w^{(\rho,\gamma)}(t) = \hat{S}(t)^{\rho}(1-\hat{S}(t))^{\gamma}$$

   - Let $Z_1, Z_2, \ldots, Z_K$ be the standardised test statistics under different (ρ, γ) pairs.
   - The Max-Combo test statistic is:

$$Z_{\max} = \max\left(|Z_1|, |Z_2|, \ldots, |Z_K|\right)$$

   - Adjusted p-value computed from joint multivariate normal distribution of $(Z_1, \ldots, Z_K)$ [7].

Formulations: RMST

2. Restricted Mean Survival Time (RMST) [5]

- The RMST up to time τ is defined as:

$$\text{RMST}(\tau) = \int_0^\tau S(t)\, dt$$

- Compare arms using difference in RMST:

$$H_0 : \mu_1(\tau) = \mu_0(\tau) \quad \text{vs} \quad H_1 : \mu_1(\tau) \neq \mu_0(\tau)$$

- Typically tested via a Wald-type statistic [6]:

$$Z_{\text{RMST}} = \frac{\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)}{\sqrt{\hat{\text{Var}}(\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau))}}$$

Limitations of proposed methods

- **Max-Combo:**
  - **Less power** than log-rank if PH holds
  - Composite test statistic, hard to interpret
- **RMST:**
  - Requires fixed truncation time τ, which is arbitrarily determined.
  - Clinical interpretation less familiar

Simulation Design Overview I

- **Aim**: Compare the statistical power of Max-Combo and RMST-based tests under various non-PH scenarios.
- **Scenarios Considered**:
  - Delayed treatment effect
  - Diminishing treatment effect
  - Crossing hazards
- **Sample Size and Simulation**:
  - Total sample size: $n = 300$ (1:1 allocation)
  - Number of simulation replicates: $R = 10{,}000$
  - Follow-up duration: $\tau_{\max} = 48$ months
- **Censoring Mechanism**:
  - Event time: $T_i \sim S(\cdot)$
  - Administrative censoring: $C_i^{(A)} \sim \mathcal{U}(a, b)$
  - Dropout time: $C_i^{(D)} \sim \text{Exp}(\lambda)$
  - Total censoring: $C_i = \min(C_i^{(A)}, C_i^{(D)})$
  - Observed time: $X_i = \min(T_i, C_i)$, with indicator $\delta_i = \mathbb{I}(T_i < C_i)$

Simulation Design Overview II

- **Test Specifications**:
  - RMST truncation time: τ to be chosen slightly below the maximum expected follow-up time per Royston and Parmar (2011)
  - Max-Combo test includes the full set of weights: $G^{0,0}$, $G^{1,0}$, $G^{0,1}$, $G^{1,1}$
  - A secondary scenario is evaluated to reflect prior information from earlier-phase studies (e.g., phase II). For example where $G^{1,0}$ is excluded when delayed treatment effect is expected.
- **Packages**:
  - SimNPH
  - simtrial

Sample Size I

Under Proportional Hazards (PH)

- Required number of events for equal treatment groups:

$$d_0 = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{\log(\mathrm{HR})} \right)^2$$

$Z_{1-\alpha/2}$ is the quantile of the standard normal distribution corresponding to a two-sided Type I error rate of α.
$Z_{1-\beta}$ is the quantile corresponding to the desired power $1-\beta$.
E.g., when $\alpha = 0.05$, $Z_{1-\alpha/2} \approx 1.96$, and 90% power implies $\beta = 0.10$, so $Z_{1-\beta} \approx 1.28$.
HR denotes the assumed hazard ratio under the alternative hypothesis HR $\neq 1$.

Where e is the expected proportion of subjects who will experience the event during the study, derived using the accrual and follow-up period,

Sample Size II

total sample size:

$$n_0 = \frac{d_0}{e}$$

Working estimate under non-PH

Apply adjustment:

$$n_{\text{adjusted}} = \frac{n_0}{\pi}$$

Where π is the empirically estimated power under non-PH using the PH-derived sample size n_0.

Possible Future Work

- Incorporate ICH E9(R1) estimand framework guidelines (likely suitable for RMST)
- Compare sample size derivation methods
  - Power correction factor
  - Simulation based

Summary

- Log-rank/Cox PH is optimal only under PH
- Proposed alternatives exist: Max-Combo, RMST
- Simulation helps quantify power loss and guide Phase 3 design
- Aim: Improve trial design through informed pre-specification

References I

[1] Bardo, M., Huber, C., and Friede, T. (2024). Methods for non-proportional hazards in clinical trials: A systematic review. *Statistical Methods in Medical Research*, 33(6):1189–1206.

[2] Duke-Margolis Center for Health Policy (2018). Public workshop: Oncology clinical trials in the presence of non-proportional hazards — meeting summary. White Paper, Duke-Margolis Center for Health Policy. Accessed July 2025.

[3] Freidlin, B. and Korn, E. L. (2019). Methods for accommodating nonproportional hazards in clinical trials: Ready for the primary analysis? *Journal of Clinical Oncology*, 37(35):3455–3459.

[4] Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., and Gottfried, M. (2016). Pembrolizumab versus chemotherapy for pd-l1–positive non–small-cell lung cancer. *New England Journal of Medicine*, 375(19):1823–1833.

[5] Royston, P. and Parmar, M. K. B. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30(19):2409–2421.

References II

[6] Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L. J., Skali, H., Solomon, S. D., Jacobus, S., Hughes, M., Packer, M., and Wei, L. J. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology*, 32(22):2380–2385.

[7] Zhao, Y., Han, Y., Zhang, Y., Liu, H., Qian, J., Pan, X., Wang, Y., and Zhang, Y. (2024). lrstat: Power and sample size calculation for non-proportional hazards and beyond. R package version 1.1.3.