Beyond Log-Rank: Advanced Survival Analysis for Non-Proportional Hazards in Oncology Trials

Jagdish Nikam

Background

- Proportional hazards assumes that the treatment effect on the event risk is constant over time.
- The log-rank test and Cox proportional hazards model are standard tools for time-to-event analyses in oncology but rely on the proportional hazard's assumption.
- Reliance on hazard ratios alone may limit interpretability for clinicians and patients, as hazard ratios represent an instantaneous relative effect rather than an absolute or time-integrated treatment benefit.
- In many oncology trials, non-proportional hazards arise due to delayed treatment effects, long-term survivors, or treatment switching, leading to potential loss of power or misleading results.
- Advanced methods such as restricted mean survival time (RMST), weighted log-rank tests, and time-varying effect models offer more robust and interpretable alternatives when proportional hazards do not hold.

Challenges

Challenge: The log-rank test and Cox model assume proportional hazards, an assumption frequently violated in modern oncology trials (e.g., immuno-oncology, delayed treatment effects, treatment switching).

Impact: Non-proportional hazards can reduce power and lead to misleading estimates of treatment effect if standard methods are applied without adjustment.

Regulatory Perspective: FDA and EMA guidance encourage alternative and supportive estimands and analyses when proportional hazards assumptions are not met, emphasizing clinical interpretability and robustness.

Focus: We review advanced methods—including weighted log-rank tests, RMST, and time-varying models—to support valid inference under non-proportional hazards.

FDA Perspective

FDA-Initiated Collaboration

Based on the Public Workshop on Oncology Clinical Trials in the Presence of Non-Proportional Hazards, the FDA recognized the need for broader collaboration to address methodological challenges associated with non-proportional hazards in oncology trials. To this end, the FDA initiated dialogue with industry statisticians and other stakeholders.

- Meetings were held in 2016 and subsequently in 2017, during which participants concluded that a systematic and methodical evaluation of available statistical methods is needed.
- The goal of this collaborative effort is to identify appropriate analysis methods for different patterns of non-proportional hazards, recognizing that no single approach is universally optimal.
- This effort is intended to be non-product-specific, with participation from multiple industry partners working collaboratively, and with FDA actively participating in the evaluation and discussion.

Protocol and ICH E9

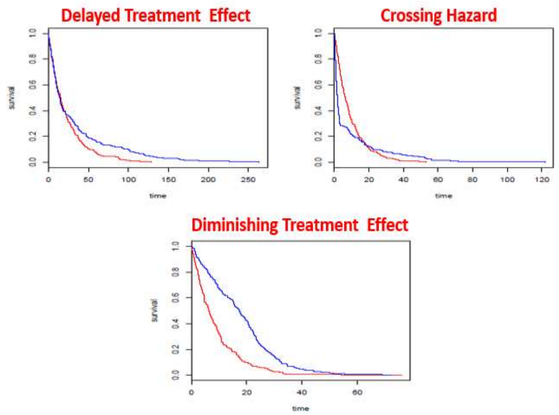
Regarding primary analysis, ICH E9 states:

For each clinical trial contributing to a marketing application, all important details of its design and conduct and the principal features of its proposed statistical analysis should be clearly specified in a protocol written before the trial begins. The extent to which the procedures in the protocol are followed and the primary analysis is planned a priori will contribute to the degree of confidence in the results and conclusions of the trial.

Key points:

- All key details of trial design, conduct, and statistical analysis must be specified in a protocol prior to trial start.
- Adherence to the protocol and pre-specification of the primary analysis increases confidence in trial results.
- Planning for the appropriate statistical approach is challenging when the NPH type is unpredictable.
- The timing and magnitude of the treatment effect may not be known before the trial begins.

Types of NPH



In the figure, the X-axis represents time in months, and the Y-axis represents survival probability.

Type	Mini Survival Curve	Label / Key Point
1. Crossing Hazards	(simulate two curves crossing)	Curves cross → treatment effect reverses over time
2. Delayed Effect / Late Separation	(curves separate later)	Early overlap, benefit appears later
3. Diminishing / Early Effect	(curves separate early then converge)	Strong early effect, reduces over time
4. Changing Risk Over Time	(curves diverge/merge continuously)	Hazard ratio changes continuously

Primary Analysis Methods

Method Category	Method	Description / When to Use
Rank-Based Tests	Log-Rank Test	Standard test comparing survival distributions; optimal under proportional hazards but may lose power when hazards are non-proportional.
	Weighted Log-Rank Test	Applies time-dependent weights to emphasize early or late survival differences; useful when delayed or early treatment effects are expected.
Combination Tests	Combination of Weighted Log-Rank Tests (Max-Combo)	Combines multiple weighted log-rank tests targeting different hazard patterns (early, proportional, delayed); robust under NPH and recommended in regulatory guidance.
	Breslow (Generalized Wilcoxon) Test	Places greater weight on early events; useful when treatment effects are expected early in follow-up.
KM-Based Methods	Weighted Kaplan–Meier Test	Compares survival curves using weighted KM estimates; useful for descriptive and supportive analyses under NPH.
	Restricted Mean Survival Time	Compares average survival time up to a fixed time point; provides an absolute and clinically interpretable treatment effect without PH assumptions.

Beyond Log-Rank

Introduce advanced methods suitable for NPH

1. Restricted Mean Survival Time
- Measures average survival time up to a pre-defined time point.
- Provides an absolute, clinically meaningful difference between groups (e.g., "+2.3 months survival").
- No PH assumption required.

2. Weighted Log-Rank Test
- Applies time-dependent weights to emphasize survival differences at specific periods (e.g., early or late).
- Useful when delayed treatment effects are expected.

3. Max-Combo test (Combination of Fleming-Harrington Log-rank)
- It combines multiple weighted log-rank tests, each targeting different hazard patterns (early, proportional, or delayed effects).
- The test statistic is the maximum of these standardized tests, with multiplicity adjustment to control type I error.
- Recognized in regulatory guidance (FDA/EMA) as a robust primary or supportive analysis under non-proportional hazards.
- Can be pre-specified easily at protocol stage satisfies ICH E9 condition

4. Milestone Survival Analysis
- Compares survival probabilities at fixed time points (e.g., 12 or 24 months).
- Highly interpretable for clinicians and regulators.
- Useful for communicating long-term benefits in immunotherapy trials.

5. Time-Varying Cox Models (Piecewise Cox Model)
- The follow-up time is divided into predefined intervals (e.g., 0–5, 5–10, >10 months).
- Can incorporate interactions between treatment and time.
- Provides a flexible and detailed view of evolving treatment effects.
- A separate hazard ratio is estimated for each interval, resulting in piecewise hazard ratios.

Simulation Study

Study Design
This analysis uses a simulation-based example designed to illustrate non-proportional hazards with crossing survival curves.

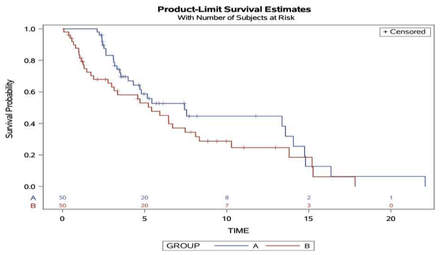
Survival times were generated by two treatment groups:

Group A: Survival times were simulated from an exponential distribution with an added delay to induce a delayed treatment effect, resulting in poorer early survival but improved later outcomes.

Group B: Survival times were simulated from an exponential distribution without delay, leading to earlier events and better early survival.

Sample Size
50 subjects per group were simulated to generate adequate events for demonstrating non-proportional hazards.

This data-generating mechanism was intentionally constructed to produce crossing Kaplan–Meier curves, thereby violating the proportional hazards assumption and motivating the use of alternative analytical approaches such as piecewise hazard ratios.



Piecewise HR

Interval (Months)	HR (Group B vs A)	95% CI	Interpretation
Overall	1.434	0.857 – 2.416	The difference is not statistically significant at the conventional α = 0.05 level.
0–5	1.532	0.815 – 2.910	Slightly higher hazard in B early on; not significant
5–10	0.537	0.077 – 2.509	Lower hazard in B mid-period; not significant
10–15	1.993	0.648 – 7.362	Higher hazard in B late; not significant

This pattern indicates non-proportional hazards: the effect of treatment differs across time. Early and mid-intervals do not show statistically significant differences, but the overall hazard shows a significant effect.

Discussion/Conclusion

This simulation demonstrates non-proportional hazards with crossing survival curves, a common scenario in oncology trials. A single overall hazard ratio may not capture time-dependent treatment effects, as seen in the varying piecewise hazard ratios across intervals.

Advanced methods like piecewise Cox models and Restricted Mean Survival Time provide complementary insights:
- Piecewise HRs show how treatment effects change over time.
- RMST gives an absolute, clinically interpretable survival difference without assuming proportional hazards.

Key takeaway: Combining robust testing methods with estimation-based approaches improves interpretation and ensures clinically meaningful conclusions under non-proportional hazards.