

Designing clinical trials with a delayed treatment effect

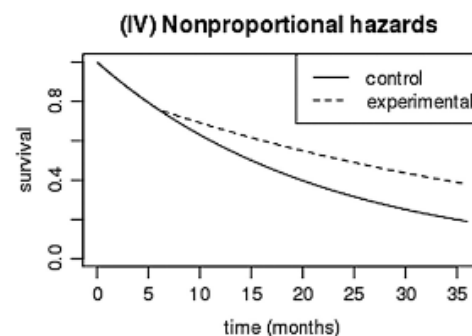
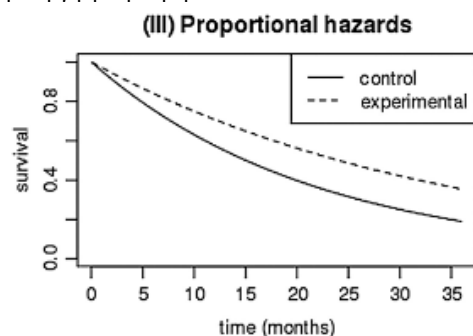
EMA Methodology Working Party

Virtual meeting, June 7th 2024

*Dominic Magirr (Novartis), Carl Fredrik Burman
(AstraZeneca)*

Significance and urgency of the topic

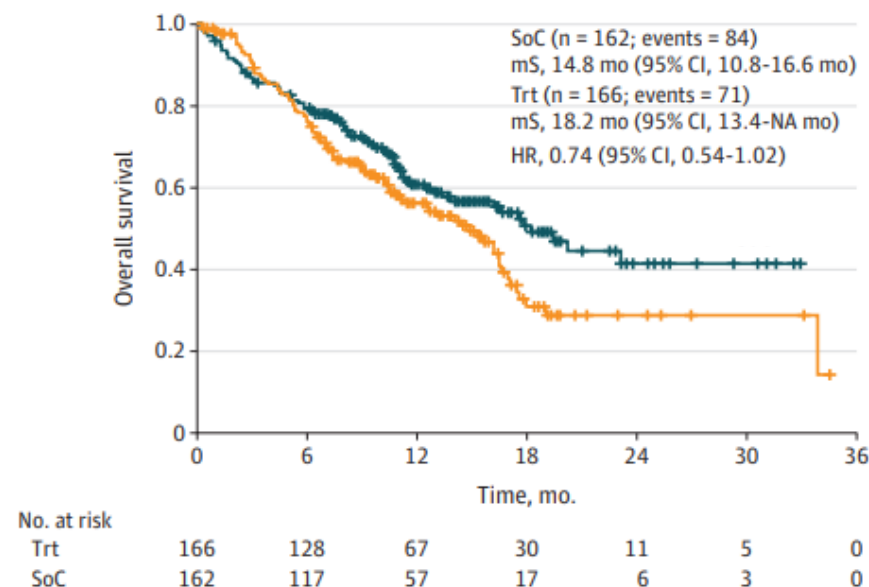
- **High risk of type 2 errors** when using a proportional hazards model by default, when it's clear from mechanism of the drug that effects will be delayed.
- Often in serious disease, hard endpoint.
- Efficiency gains of using tailored methods can be substantial



Relative Efficiency
of MWLRT compared to
LRT

95%

126%



Challenges

- There is a tension between picking a test that corresponds to a “simple” estimand, and picking a test that has high power to reject a null hypothesis of no treatment benefit.
 - E.g. Restricted Mean Survival Time (RMST) tends to emphasize early differences rather than late differences.
 - E.g. Milestone survival probabilities (at a late follow-up time) might capture long term benefit **but** also involve a dichotomization, and it is challenging to pre-specify the timepoint – with high vulnerability to large power loss if wrong choice.
 - More creative choices (window mean survival time, average hazard ratio, etc.) no longer simple.
- Care is needed with weighted log-rank tests
 - Substantial increase in power **but**
 - If we give too small weights to early event times (many commonly used weighting schemes do this), then we could be rewarding early harm → one-sided p-value gets smaller as early survival on experimental treatment gets worse.
 - Modestly-weighted log-rank test (Magirr & Burman, 2019) designed to address this issue.

Questions

- Current practice when anticipating a delayed treatment effect is to decouple testing and estimation,
 - Will this approach remain acceptable?
- Current practice is to describe the treatment effect via multiple summary measures (medians, milestones, etc.) without formal corrections for multiplicity,
 - What is the agency’s position on simultaneous inference for multiple measures of treatment effect, as proposed in Ristl et al. (2024) doi: [10.1177/09622802241231497](https://doi.org/10.1177/09622802241231497) ?
- Acceptability of weighted log-rank:
 - Is a distinction made between tests that give very low weights to early events (thus increasing risk of false conclusions) and tests that provide strong control of type 1 error rate

