Anytime-valid log-rank testing for randomised trials

Joren Brunekreef[1], Renée X. Menezes[2], Rianne de Heide[3,4]
[1] Netherlands Cancer Institute (Department of Radiotherapy and AI for Oncology), Amsterdam, NL
[2] Netherlands Cancer Institute (Biostatistics Centre and Department of Psychosocial Research and Epidemiology), Amsterdam, NL
[3] University of Twente, NL · [4] Centrum Wiskunde & Informatica, Amsterdam, NL

Joren Brunekreef
Department of Radiotherapy & AI for Oncology
Netherlands Cancer Institute
j.brunekreef@nki.nl

At a glance

Anytime-valid survival monitoring approaches classical log-rank power. Simple stopping rules can help exceed classical power under continuation.

Anytime-valid (AV) log-rank tests continuously monitor survival trials with Type-I error control. We characterise anytime-valid survival analysis by simulation and deploy the framework retrospectively on a real-world randomised trial.

1. Setup

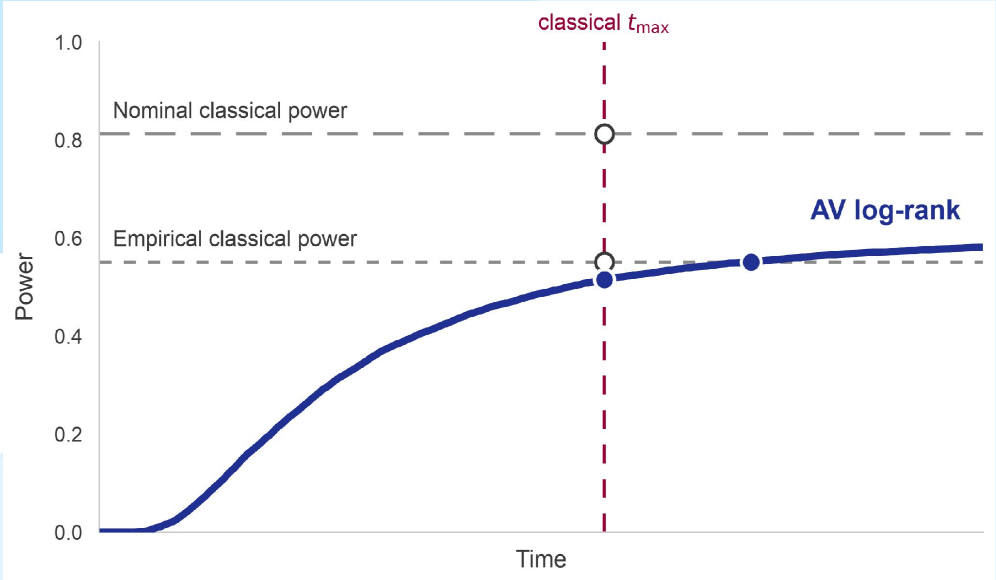
- Fixed-δ test martingale on the event sequence (Ter Schure et al.)
- Optional stopping & continuation: **reject early when E ≥ 1/α, or keep accumulating**
- Type-I error ≤ α at any stopping time (Ville's inequality)

2. Simulation framework

- AV log-rank inherits the classical design parameter: **an assumed δ**
- Classical practice fixes a nominal design power (e.g., 80%) under the assumed δ; the empirically realised power depends on how close δ is to the true HR

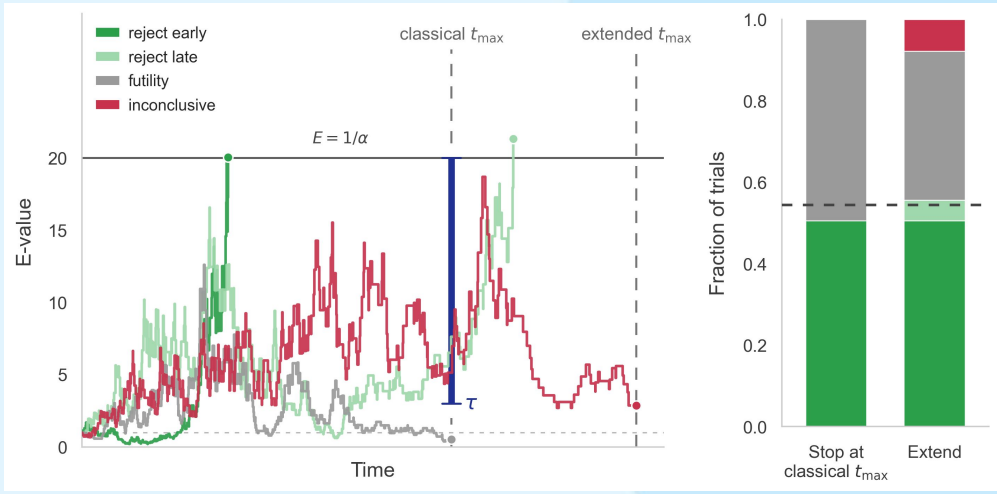
Power-over-time

Rejection probability accumulates by each follow-up moment. Generalises classical fixed-n power once anytime-validity admits unrestricted stopping.



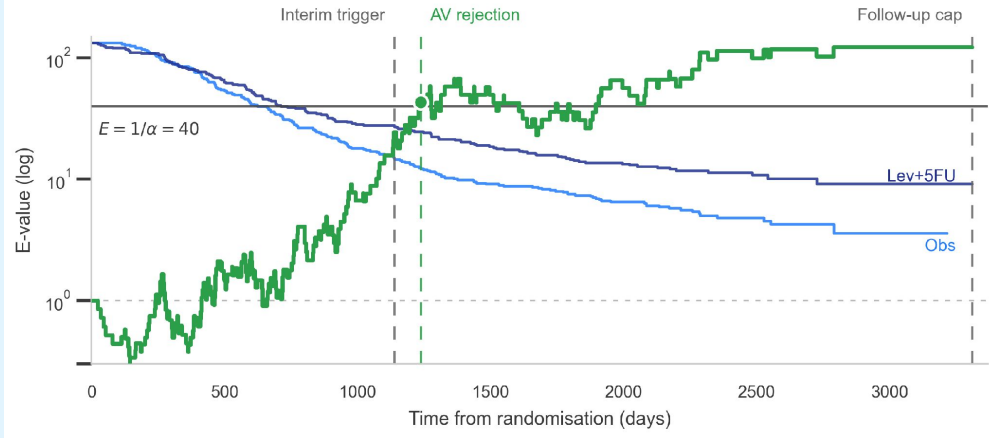
- At the classical trial end, AV power sits just below classical's empirical power: the (small) price we pay for anytime-validity
- Past the trial end, AV power keeps climbing. Eventually, it closes the gap and overtakes classical power under continuation

3. Continuation rules



- Anytime-validity admits continuation past the classical trial end time: if the evidence is **suggestive** but **not conclusive**, we continue monitoring
- Continuation rule: if E > τ at trial end, extend for fixed duration
- ”Late” rejections allow us to exceed classical power, at the cost of further inconclusive continuations. Continuation rule parameters let us pick a balance

4. Real-world deployment



- Colon Intergroup adjuvant trial (Moertel 1990, n=619): AV rejects at 207 events, 84 events earlier than the trial’s 291-event follow-up cap (but 15 events later than interim trigger at 192 events)
- With enrolment dates, AV converts event-savings into calendar-date savings

Funding
R. de Heide supported by NWO Veni grant VI.Veni.222.018. · J. Brunekreef and R. X. Menezes supported by institutional grants of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport.

References

[1] Ter Schure J, Pérez-Ortiz MF, Ly A, Grünwald PD. The safe logrank test: error control under continuous monitoring with unlimited horizon. (in press)
[2] Grünwald P, de Heide R, Koolen W. Safe testing. J R Stat Soc B. 2024.
[3] Ramdas A, Grünwald P, Vovk V, Shafer G. Game-theoretic statistics and safe anytime-valid inference. Stat Sci. 2023.
[4] Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med. 1990;322:352-358.

Where science meets inspired minds.

NETHERLANDS CANCER INSTITUTE
ANTONI VAN LEEUWENHOEK