Advancing Decision-Efficiency in (Pre)-clinical Research via Novel Sequential Frameworks

Frank Konietschke

Institut für Biometrie und Klinische Epidemiologie

Charité - Universitätsmedizin Berlin, Berlin

frank.konietschke@charite.de

Sequential Tests and Monitoring

- **Pre-Clinical** Experiments
  - Animals are treated in cohorts. Results are immediately available and analyzed
  - Go/ No-Go decisions
  - Researchers stop when they feel/see efficacy

Sequential Tests and Monitoring

- **Group-Sequential** Designs
  - Interim analysis to stop for efficacy/ futility
  - Usually one (or two) interim analysis

Sequential Tests and Monitoring

- **Sequential** Data
  - Sequential data streams occur often
  - Wearable
  - Digital real-time biomarker

Why Sequential Designs?

- Clinical trials are planned under uncertainty
- Assumptions on effect sizes and event rates may be wrong
- Ethical concerns:
  - exposing too many patients to inferior treatments
  - continuing trials unlikely to succeed
- Adaptive designs allow learning from accumulating data

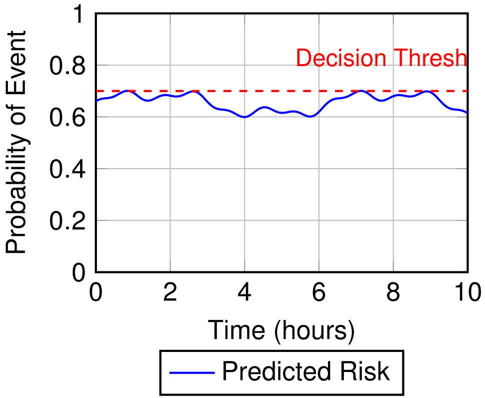
Goal: More efficient, informative, and ethical trials

Digital Biomarker

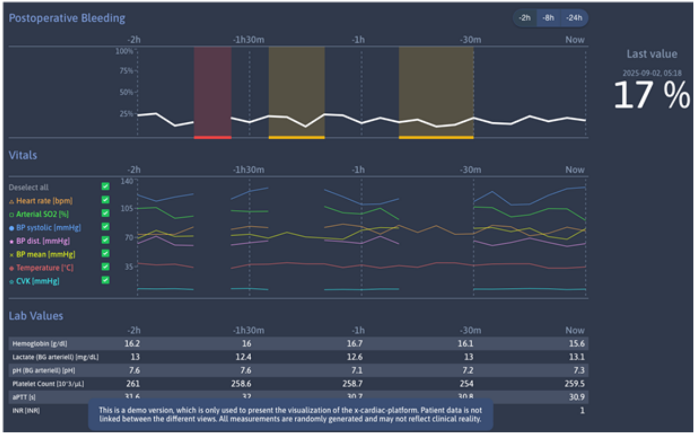
- **Real-Time Digital Biomarkers**
- Monitoring of risk in ICU patients
- Detection of arrhythmias
- Response at time point t

$$P(\text{Patient will event in the next six hours})$$

- Response measured every 5 minutes



Digital Real-Time Biomarker



www-x-cardiac.com

Digital Real-Time Biomarker

Time-Dependent Sensitivity ($Se(t)$)

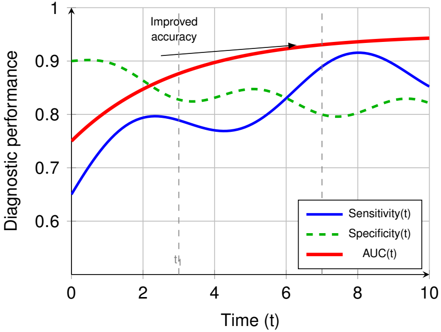
- Prob. that $X \geq c$ at time t.
- $P\left(Y(t) > c(t) | D(u) = 1;\ 0 \leqslant u \leqslant t \leqslant u + \Delta\right)$

Time-Dependent AUC ($AUC(t)$)

- AUC at a specific time horizon t.
- Summarizes discriminative power of separation

Monitoring Framework

- Alerts
- Clinical Decision support
- Changepoint estimation



Monitoring Framework

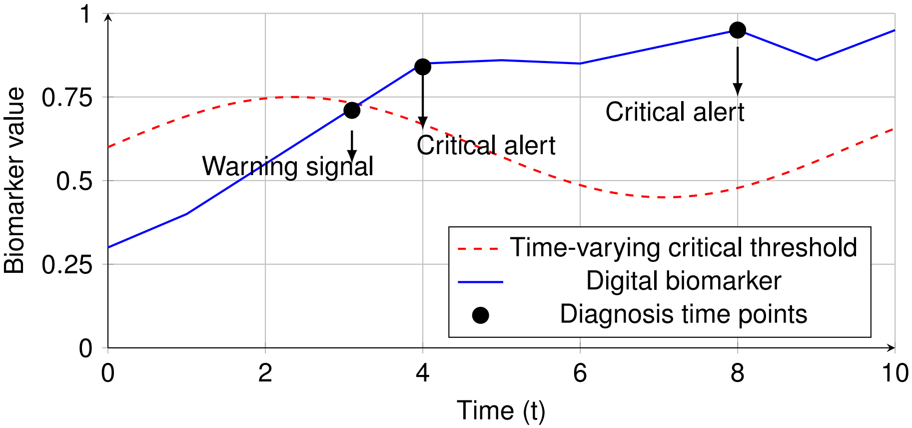
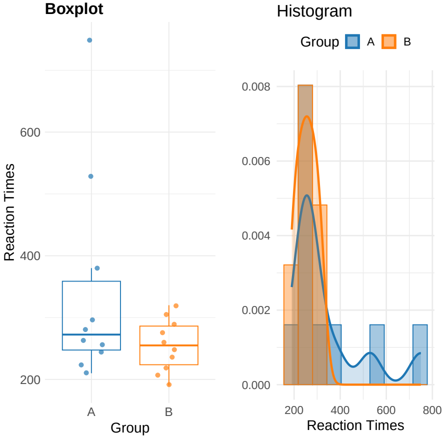


Abb.: Monitoring framework to be established for digital real-time biomarkers.

Example: Reaction Times

Pre-Clinical trial

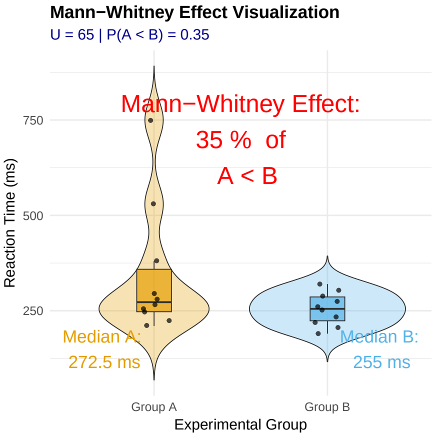
Measuring reaction times [ms] in two groups (n = 10)



Group	N	Mean	SD	Med.	(Min,Max)
A	10	343.5	272.5	171.0	(210, 750)
B	10	255.0	255.0	43.1	(190 , 320)

- **Impact of the outliers:**
  - Mean (1:7): 253.57
  - Mean (1:8): 269.38
  - Mean (1:9): 298.33
  - Mean (1:10): 343.50

Descriptive Overview



- **Effects:**
  - MW = 35%
  - Odds = $65\%/35\% = 1.9$
  - The chance of observing a larger reaction time in A, is 1.9 times as large

Statistical Model

- $X_{ik} \sim F_i, \quad i = 1, 2; k = 1, \ldots, n_i$
- Effect of the populations
  - Relative effect $\theta = P(X_{11} < X_{21}) + 0.5P(X_{11} = X_{21})$
  - Interpretation
    - θ is the true proportion of the values in population 1 being smaller than those in population 2
    - If $\theta < \frac{1}{2}$: X1 tends to be larger than $X2$
    - If $\theta = 1/2$: No tendency to smaller or larger values

Nonparametric Null Hypotheses

- $X_i : X_{i1}, \ldots, X_{in_i} \sim F_i,\ i = 1, 2;\ k = 1, \ldots, n_i$
- Two possible ways to formulate the null hypothesis

- $\mathbf{H_0^F : F_1 = F_2}$
- **Wilcoxon-Mann-Whitney Test**
- Test whether distributions are identical
- Homoscedasticity (equal variances)
- Asymptotic and Exact

- $\mathbf{H_0} : \theta = \frac{1}{2}$
- **Nonparametric Behrens-Fisher Problem**
- Heteroscedasticity
- Unequal variances and shapes

- Remember: $H_0^F \Rightarrow H_0^\theta$: If $F_1 = F_2 \Rightarrow \theta = \frac{1}{2}$
- Does the reverse also hold?

The Wilcoxon-Mann-Whitney Test

- Two independent samples: $X_{ik} \sim F_i,\ i = 1, 2;\ k = 1, \dots, n_i;\ N = n_1 + n_2$
- Under $H_0^F : F_1 = F_2 = F$,

$$W_N = \frac{\widehat{\theta} - \frac{1}{2}}{\widehat{\sigma}_0 \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \ \dot{\sim}\ N(0, 1)$$

where

$$\widehat{\sigma}_0^2 = \frac{1}{N^2(N-1)} \sum_{i=1}^{2} \sum_{k=1}^{n_i} \left(R_{ik} - \frac{N+1}{2}\right)^2$$

(Asymptotic WMW Test)

- Reject H_0 at level α, if $|W_N| \geq z(1 - \alpha/2)$: $(1 - \alpha/2)$-quantile from $N(0, 1)$
- Assume: $\sigma_0^2 = Var\left(F(X_{11})\right) > 0$ and $min\{n_1, n_2\} \to \infty$ such that $N/n_i \leq N_0 < \infty$

Variance Estimator

- No unbiased estimator of its variance (in finite n) existed
- We recently derived an unbiased estimator

$$\widehat{\sigma}_N^2 = \frac{1}{n_1 n_2}\left[(n_2 - 1)\widehat{\sigma}_1^2 + (n_1 - 1)\widehat{\sigma}_2^2 + \widehat{\theta}(1 - \widehat{\theta}) - \frac{1}{4}\widehat{\tau}\right]$$

  - $\widehat{\sigma}_i^2 =$ Estimated variance of $\widehat{F}_j(X_{ik})$
  - $\widehat{\tau}$: Estimated probability of ties

REGULAR ARTICLE

An unbiased rank-based estimator of the Mann–Whitney variance including the case of ties

Edgar Brunner[1] · Frank Konietschke[2]

Received: 10 December 2023 / Revised: 21 September 2024

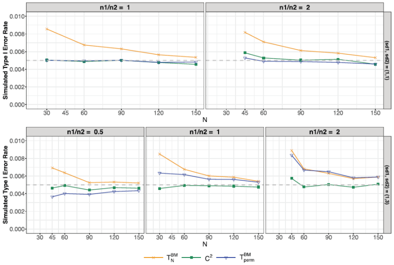
- $\widehat{\sigma}_N^2 \geq 0$, unbiased and strongly consistent

The C^2-Test

- The Brunner-Munzel test shows a liberal behavior when samples are (very) small
- Needed: A better approximation
- The novel C^2-Test
  - Relate to the maximum variance of $\widehat{\theta}$
  - $Var(\widehat{\theta}) \approx \frac{\widehat{\sigma}_N^2}{\widehat{\theta}(1-\widehat{\theta})}\theta(1-\theta)$

$$C^2 = \frac{4}{\widehat{q}}\left(\widehat{\theta} - \frac{1}{2}\right)^2$$

- $\widehat{q} = \frac{\widehat{\sigma}^2}{\widehat{\theta}(1-\widehat{\theta})}$
- Reject H_0, if $C^2 \geq \chi^2_{1,1-\alpha}$



A New Approach to the Nonparametric Behrens–Fisher Problem With Compatible Confidence Intervals

Stephen Schüürhuis[1] | Frank Konietschke[1] | Edgar Brunner[2]

[1]Institute of Biometry and Clinical Epidemiology, Charit-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany | [2]Department of Medical Statistics, Universitätsmedizin Göttingen, Göttingen, Germany

Correspondence: Stephen Schüürhuis (stephen.schueuerhuis@charite.de)

Received: 31 March 2025 | **Revised:** 24 September 2025 | **Accepted:** 18 October 2025

Keywords: Birnbaum–Klose inequality | Brunner–Munzel test | Mann–Whitney effect θ | studentized permutation test | C^2-test

Group Sequential Designs: Canonical Joint Distribution

Theorem (Canonical joint law of sequential test statistics)

Let $(Z_1, \ldots, Z_K)$ *denote the sequence of standardized test statistics computed at interim analyses with corresponding information levels* $0 < I_1 \leq \cdots \leq I_K$*.*
Then the vector $(Z_1, \ldots, Z_K)^\top$ *follows a multivariate normal distribution*

$$(Z_1, \ldots, Z_K)^\top \sim \mathcal{N}(0, \Sigma),$$

with covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^{K}$ *given by*

$$\sigma_{ij} = \mathrm{Cov}(Z_i, Z_j) = \sqrt{\frac{\min\{I_i, I_j\}}{\max\{I_i, I_j\}}}.$$

Group sequential methods for the Mann-Whitney parameter

Claus P Nowak[1,2], Tobias Mütze[3], and Frank Konietschke[1]

Statistical Methods in Medical Research
1–17

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09622802221107103
journals.sagepub.com/home/smm
SAGE

Group Sequential Design: Example Boundaries

Stopping boundaries:

Look (N)	O'Brien-Fleming		Pocock	
	Efficacy (Z)	p (nom.)	Efficacy (Z)	p (nom.)
Interim 1 (25%)	4.05	$<$0.0001	2.29	0.0221
Interim 2 (50%)	2.86	0.0042	2.29	0.0221
Final (100%)	2.02	0.0430	2.29	0.0221

Interpretation:

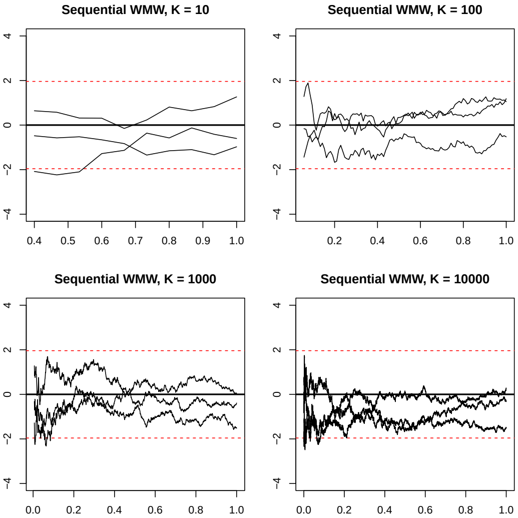
- **Efficacy boundary:** if the test statistic crosses this line, stop early for success
- Otherwise, continue until the next interim or final analysis

Group Sequential Design: Example Boundaries

Note

The total sample size increases of about 2 - 10 % usually

Sequential WMW Tests



Sequential Rank Tests: Continuous-Time Limit

Theorem (Canonical Process)

Let Z_k denote the standardized rank-based test statistics at look k, with sample sizes (n_{1k}, n_{2k}). Define the information fraction

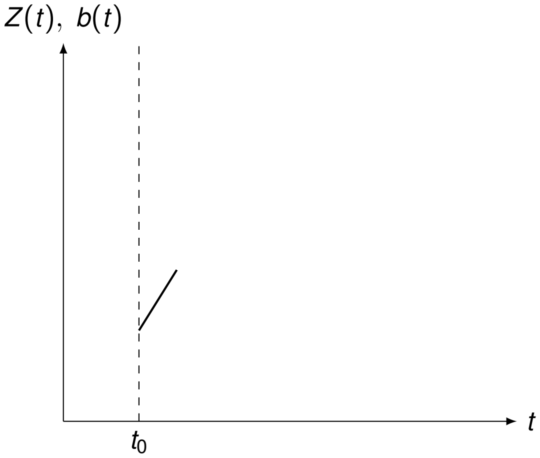
$$t_k = \frac{I_k}{I_{\max}}, \qquad I_k = \begin{cases} \frac{12 n_{1k} n_{2k}}{n_{1k}+n_{2k}+1}, & \text{WMW}, \\ c\,\frac{n_{1k} n_{2k}}{n_{1k}+n_{2k}}, & \text{BF}. \end{cases}$$

Assume $N_k = n_{1k} + n_{2k} \to \infty$ and $\sup_k N_k / \min(n_{1k}, n_{2k}) < \infty$.
Then, as $N_k \to \infty$,

$$\{Z(t_k) : t_k \in [t_0, 1]\} \;\Rightarrow\; \left\{ \frac{W(t)}{\sqrt{t}} : t \in [t_0, 1] \right\} \quad \text{in } D[t_0, 1],$$

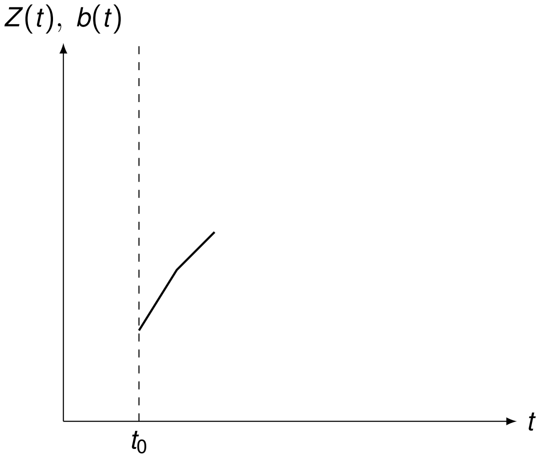
where $W(t)$ is standard Brownian motion.

Animated Supremum and Boundaries $b(t)$



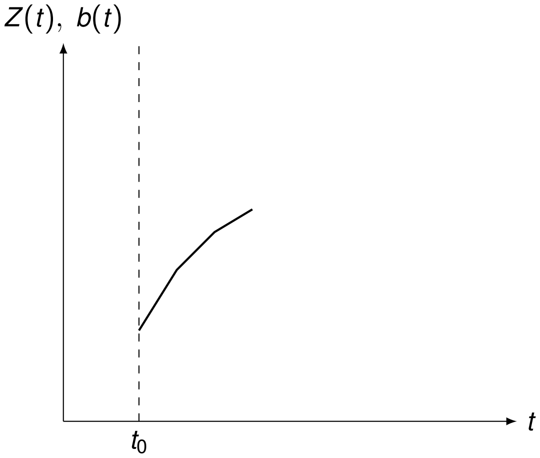
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



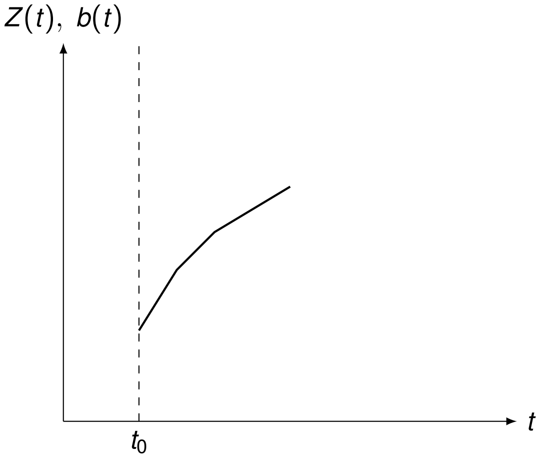
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



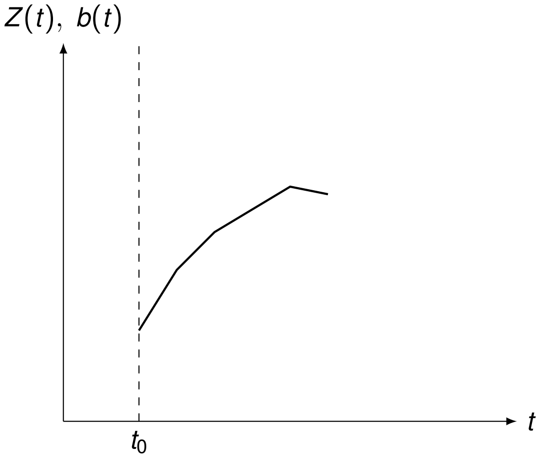
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



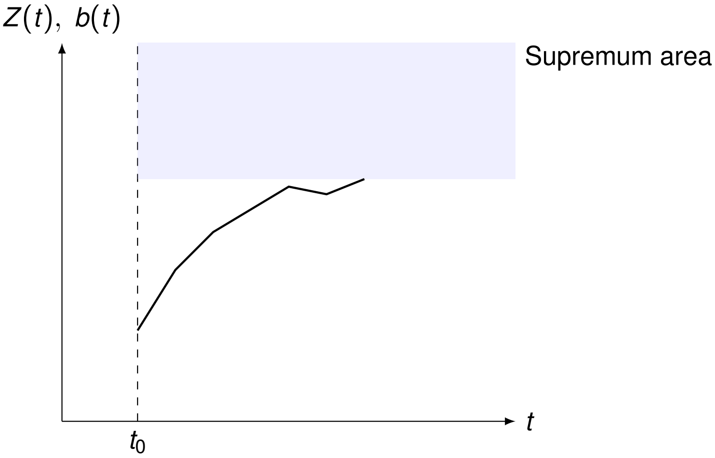
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



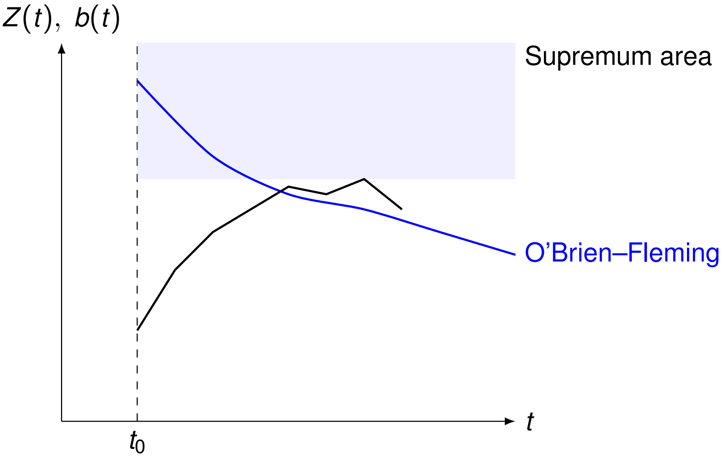
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



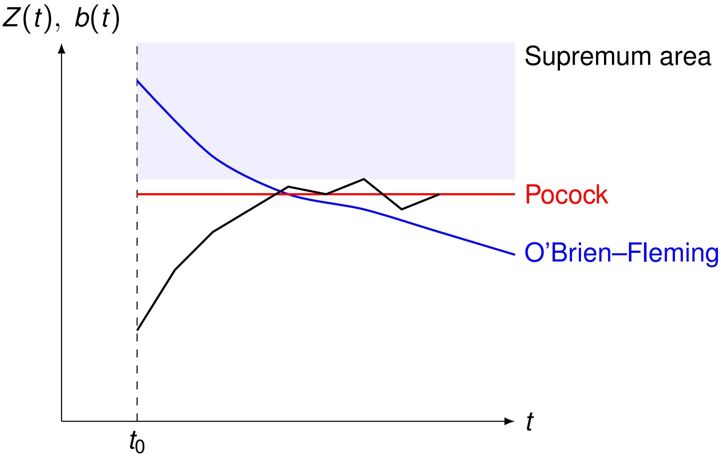
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



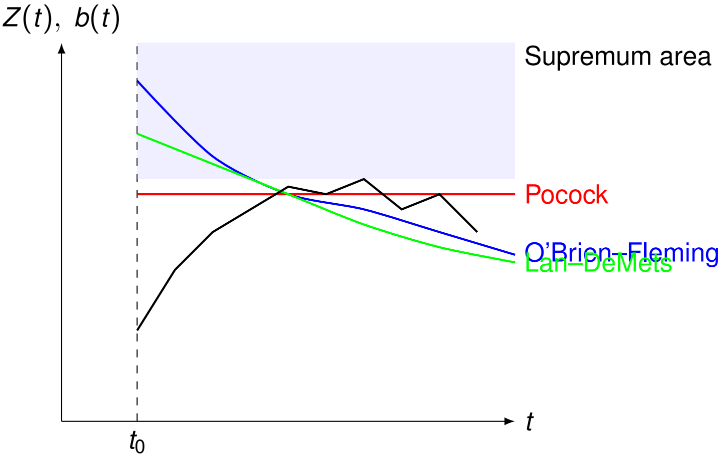
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



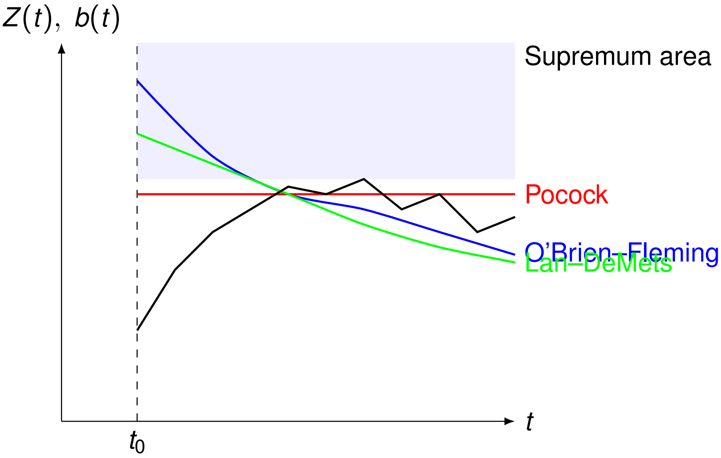
Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Animated Supremum and Boundaries $b(t)$



Note: The black path represents $Z(t)$; colored curves are boundaries. The shaded area indicates where the supremum over $[t_0, 1]$ is relevant.

Stopping Times

- **Stops with error control**
  - Let $z_{1-\alpha}$ denote the $(1-\alpha)$-quantile of

$$\sup_{t\in[t_0,1]} \frac{W(t)}{\sqrt{t}},$$

    where $W(t)$ is a standard Wiener process.
  - Define the stopping time

$$\tau := \inf\left\{t \in [t_0, 1] : Z(t) \geq z_{1-\alpha}\right\},$$

    with the convention $\tau = \infty$ if the boundary is never crossed.
  - Reject H_0 if $\tau \leq 1$.
  - Then the type-I error probability is strongly controlled at level α.

Stopping Times with Continuous O'Brien-Fleming Boundaries

- **Continuous O'Brien-Fleming boundary:**

$$b(t) = \frac{C}{\sqrt{t}}, \quad t \in [t_0, 1]$$

  where C is the constant adjusted for sequential testing (e.g., $C \approx 1.1\, z_{1-\alpha/2}$ for $\alpha = 0.05$).
- Define the stopping time

$$\tau_{\mathrm{OF}} := \inf\left\{t \in [t_0, 1] : Z(t) \geq b(t)\right\},$$

  with $\tau_{\mathrm{OF}} = \infty$ if the boundary is never crossed.
- Reject H_0 if $\tau_{\mathrm{OF}} \leq 1$.
- Then the type-I error probability is strongly controlled at level α, and early stopping is more conservative at small t due to the O'Brien-Fleming shape.

Simulation Results: Sequential Tests

Tab.: Sequential WMW test simulation results under H_0 ($\mu_2 = 0$, $\alpha = 0.05$, $n_{01} = n_{02} = 10$)

incrs	critBM	critMax	BM	Maxstat	OBF	FixedWMW	N_{max}	ESS (cond/uncond)
10	2.371	2.389	0.0491	0.0474	0.0417	0.0474	40	26.16 / 39.34
30	2.587	2.579	0.0463	0.0467	0.0500	0.0494	80	39.44 / 78.11
50	2.675	2.661	0.0443	0.0454	0.0507	0.0498	120	53.26 / 116.97
100	2.755	2.764	0.0506	0.0496	0.0554	0.0493	220	80.61 / 213.09

▷ Type I error is well controlled at the nominal level.

Simulation Results: Continuous-Time Sequential Tests

Tab.: Sequential WMW test simulation results under varying μ_2 ($\alpha = 0.05$, $n_{01} = n_{02} = 10$, incrs = 30)

μ_2	critBM	critMax	BM	Maxstat	OBF	FixedWMW	N_{max}	ESS (cond / uncond)
0.1	2.599	2.581	0.0560	0.0601	0.0704	0.0725	80	42.26 / 77.73
0.2	2.565	2.581	0.1087	0.1036	0.1280	0.1359	80	44.80 / 76.35
0.3	2.577	2.579	0.1821	0.1808	0.2365	0.2557	80	46.27 / 73.90
0.4	2.583	2.586	0.3031	0.3023	0.3886	0.4093	80	46.57 / 69.90
0.5	2.597	2.582	0.4401	0.4462	0.5509	0.5794	80	46.35 / 64.99
0.6	2.586	2.584	0.6026	0.6026	0.7059	0.7358	80	45.12 / 58.98
0.7	2.585	2.582	0.7553	0.7558	0.8331	0.8514	80	43.20 / 52.19
0.8	2.560	2.577	0.8604	0.8565	0.9163	0.9312	80	39.91 / 45.66
0.9	2.590	2.583	0.9291	0.9302	0.9665	0.9740	80	37.54 / 40.50
1.0	2.592	2.580	0.9694	0.9702	0.9879	0.9906	80	33.91 / 35.28


Thank You!

References

- Jennison C., Turnbull B. (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press.
- Pocock S. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64, 191–199.
- OBrien P.C., Fleming T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35, 549–556.
- Wald A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.*, 16, 117–186.