CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

Group-sequential methods for generalized pairwise comparisons

Stephen Schüürhuis
Frank Konietschke

Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin Berlin

Adaptive Designs - Modern Approaches in Clinical Research

Göttingen
27th of February, 2026

Part I

Generalized Pairwise Comparisons (GPC)

RECAP: THE WILCOXON-MANN-WHITNEY TEST

- Statistical model $X_{gi} \sim F_g, g = T, C, i = 1, \ldots, n_g$
- Null hypothesis: $\mathcal{H}_0 : F_T = F_C$

RECAP: THE WILCOXON-MANN-WHITNEY TEST

- Statistical model $X_{gi} \sim F_g, g = T, C, i = 1, \ldots, n_g$
- Null hypothesis: $\mathcal{H}_0 : F_T = F_C$
- Procedure: Compare all X_{Ti} with all X_{Cj} using

$$g(i,j) = c(X_{Ti}, X_{Cj}) = \begin{cases} 0 & \text{if} \quad X_{Cj} > X_{Ti} \quad \text{(unfavorable)}, \\ 1/2 & \text{if} \quad X_{Cj} = X_{Ti} \quad \text{(tie)}, \\ 1 & \text{if} \quad X_{Cj} < X_{Ti} \quad \text{(favorable)} \end{cases}$$

resulting in $n_T \times n_C$ *pairwise comparisons*

RECAP: THE WILCOXON-MANN-WHITNEY TEST

- Statistical model $X_{gi} \sim F_g, g = T, C, i = 1, \ldots, n_g$
- Null hypothesis: $\mathcal{H}_0 : F_T = F_C$
- Procedure: Compare all X_{Ti} with all X_{Cj} using

$$g(i,j) = c(X_{Ti}, X_{Cj}) = \begin{cases} 0 & \text{if} \quad X_{Cj} > X_{Ti} \quad \text{(unfavorable)}, \\ 1/2 & \text{if} \quad X_{Cj} = X_{Ti} \quad \text{(tie)}, \\ 1 & \text{if} \quad X_{Cj} < X_{Ti} \quad \text{(favorable)} \end{cases}$$

  resulting in $n_T \times n_C$ *pairwise comparisons*
- Inference is based on Mann–Whitney statistic $U = \sum_{i=1}^{n_T} \sum_{j=1}^{n_C} g(i,j)$

LIMITATIONS

- The Wilcoxon–Mann–Whitney test is limited to a single outcome

LIMITATIONS

- The Wilcoxon–Mann–Whitney test is limited to a single outcome
- Increasing interest in clinical trials to evaluate **multiple outcomes** (clinical, safety, patient-relevant)
- Outcomes differ in **clinical severity** (e.g., death vs. hospitalization)
- **Relative importance** must be reflected in the analysis
- Outcomes may be of **different types** (metric, discrete, survival)

LIMITATIONS

- The Wilcoxon–Mann–Whitney test is limited to a single outcome
- Increasing interest in clinical trials to evaluate **multiple outcomes** (clinical, safety, patient-relevant)
- Outcomes differ in **clinical severity** (e.g., death vs. hospitalization)
- **Relative importance** must be reflected in the analysis
- Outcomes may be of **different types** (metric, discrete, survival)

Generalized pairwise comparisons provide a principled solution

KEY REFERENCE

Research Article

Statistics in Medicine

Received 27 October 2009, Accepted 2 March 2010 Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/sim.3923

Generalized pairwise comparisons of prioritized outcomes in the two-sample problem

Marc Buyse[a,b]*†

This paper extends the idea behind the U-statistic of the Wilcoxon–Mann–Whitney test to perform generalized pairwise comparisons between two groups of observations. The observations are outcomes captured by a single variable, possibly repeatedly measured, or by several variables of any type (e.g. discrete, continuous, time to event). When several outcomes are considered, they must be prioritized. We show that generalized pairwise comparisons extend well-known non-parametric tests, and illustrate their interest using data from two randomized clinical trials. We also show that they lead to a general measure of the difference between the groups called the 'proportion in favor of treatment', denoted Δ, which is related to traditional measures of treatment effect for a single variable.

Keywords: generalized pairwise comparisons; prioritized outcomes; measure of treatment effect; randomization test

GENERALIZED PAIRWISE COMPARISONS

- **GPC** extend the Wilcoxon–Mann–Whitney test to this multivariate setting
- For each patient i in group $g \in \{T, C\}$, we observe d **prioritized outcomes**

$$\mathbf{X}_{gi} = \left(X_{gi}^{(1)}, \ldots, X_{gi}^{(d)}\right)^\top, \qquad X_{gi}^{(k)} \sim F_g^{(k)}$$

GENERALIZED PAIRWISE COMPARISONS

- **GPC** extend the Wilcoxon–Mann–Whitney test to this multivariate setting
- For each patient i in group $g \in \{T, C\}$, we observe d **prioritized outcomes**

$$\mathbf{X}_{gi} = \left(X_{gi}^{(1)}, \ldots, X_{gi}^{(d)}\right)^\top, \qquad X_{gi}^{(k)} \sim F_g^{(k)}$$

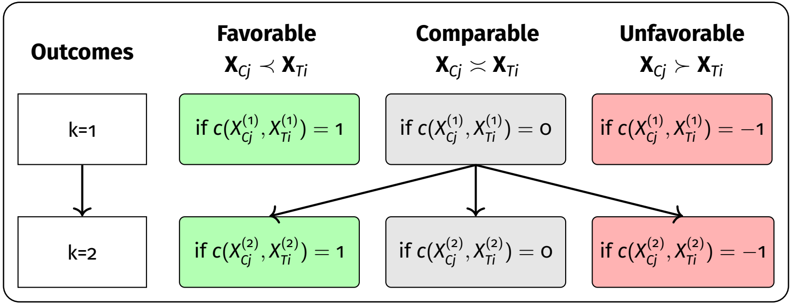
- Compare all $\mathbf{X}_{Ti}$ with all $\mathbf{X}_{Cj}$ using

$$g(i,j) = c(\mathbf{X}_{Ti}, \mathbf{X}_{Cj}) = \begin{cases} -1 & \text{if} \quad \mathbf{X}_{Cj} \succ \mathbf{X}_{Ti} \quad \text{(unfavorable)}, \\ 0 & \text{if} \quad \mathbf{X}_{Cj} \asymp \mathbf{X}_{Ti} \quad \text{(comparable)}, \\ 1 & \text{if} \quad \mathbf{X}_{Cj} \prec \mathbf{X}_{Ti} \quad \text{(favorable)} \end{cases}$$

- The operators $\{\succ, \asymp, \prec\}$ allow to generalize $\{>, =, <\}$ to multiple prioritized outcomes

GPC FOR $d = 2$

- Perform all $n_T \times n_C$ as follows



- Compare outcomes in priority order: start at $k = 1$; if tied, increment k; else stop.

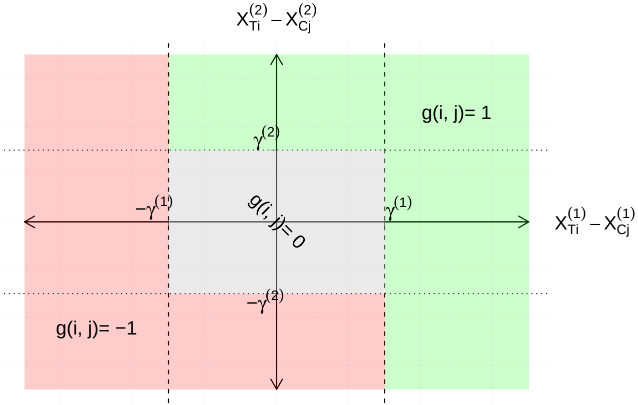
THRESHOLDS OF CLINICAL RELEVANCE

- Thresholds of clinical relevance $\gamma^{(k)}$ may be used for each outcome $k = 1, \ldots, d$

$\succ_\gamma$: $\mathbf{X}_{Ti}$ favorable to $\mathbf{X}_{Cj}$ by γ

$\prec_\gamma$: $\mathbf{X}_{Ti}$ unfavorable to $\mathbf{X}_{Cj}$ by γ

$\asymp_\gamma$: $\mathbf{X}_{Ti}$ in a range of $\mathbf{X}_{Cj} \pm \gamma$



ESTIMANDS IN THE GPC FRAMEWORK

- Various estimands exist in the context of generalized pairwise comparisons:

	Definition	Relation to θ
Mann-Whitney Effect θ	$\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})$	X
Net treatment benefit Δ	$\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj})$	$2\theta - 1$
Win ratio Ψ	$\frac{\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj})}{\mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj})}$	$\frac{\theta - 1/2\tau}{1 - \theta - 1/2\tau}$
Success Odds Λ	$\frac{\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})}{\mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})}$	$\frac{\theta}{1-\theta}$

*$\tau = \mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})$

ESTIMANDS IN THE GPC FRAMEWORK

- Various estimands exist in the context of generalized pairwise comparisons:

	Definition	Relation to θ
Mann-Whitney Effect θ	$\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})$	X
Net treatment benefit Δ	$\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj})$	$2\theta - 1$
Win ratio Ψ	$\frac{\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj})}{\mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj})}$	$\frac{\theta - 1/2\tau}{1-\theta-1/2\tau}$
Success Odds Λ	$\frac{\mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})}{\mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj}) + 1/2\mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})}$	$\frac{\theta}{1-\theta}$

$^*\tau = \mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj})$

- NTB Δ is the most common effect measure in GPC

PROPERTIES OF THE WIN RATIO

- Consider a trial with $n_T = n_C = 100$, yielding $n_T \times n_C = 10{,}000$ pairwise comparisons

Setting	Favorable	Neutral	Unfavorable	θ	Δ	Ψ	Λ
1	8 (0.08%)	9991 (99.91%)	1 (0.01%)	0.5004	0.0007	**8**	1.001
2	80 (0.8%)	9910 (99.1%)	10 (0.1%)	0.504	0.007	**8**	1.014
3	800 (8.0%)	9100 (91.0%)	100 (1.0%)	0.535	0.07	**8**	1.15
4	8000 (80.0%)	1000 (10.0%)	1000 (10.0%)	0.85	0.7	**8**	5.667

- Ψ cannot be interpreted in the presence of ties
- If $\tau = \mathbb{P}(\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj}) = 0$, then $\Psi = \Lambda$.

DECOMPOSITION OF THE NTB

- Consider a trial with $n_T = n_C = 100$, yielding $n_T \times n_C = 10{,}000$ pairwise comparisons

Outcome	Pairs	Favorable	Neutral	Unfavorable	θ	Δ	Ψ	Λ
1	10,000	4000 (40%)	4000 (40%)	2000 (20%)	0.6	0.2	2	1.5
2	4000	2000 (50%)	1000 (25%)	1000 (25%)	0.25	0.1	2	1.67
3	1000	500 (50%)	200 (20%)	300 (30%)	0.15	0.02	1.67	1.15
Overall	10,000	6500 (65%)	200 (2%)	3300 (33%)	0.66	0.32	1.97	1.94

- Additivity: $\Delta = 0.20 + 0.10 + 0.02 = 0.32$
- Relative contributions of outcomes 1–3: 62.5%, 31.25%, 6.25%

ESTIMATION AND HYPOTHESIS TESTING

- **Data:** $X_{gi}^{(k)} \overset{iid}{\sim} F_g^{(k)}, g = T, C, i = 1, \ldots, n_g, k = 1, \ldots, d$
- **Net treatment benefit** $\Delta = \mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Ci})$
- **One-sided hypothesis:** $\mathcal{H}_0 : \Delta \leq 0$

ESTIMATION AND HYPOTHESIS TESTING

- **Data:** $X_{gi}^{(k)} \overset{iid}{\sim} F_g^{(k)}, g = T, C, i = 1, \ldots, n_g, k = 1, \ldots, d$
- **Net treatment benefit** $\Delta = \mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Ci})$
- **One-sided hypothesis:** $\mathcal{H}_0 : \Delta \leq 0$

$\mathbf{X}_{Ti}$ *vs.* $\mathbf{X}_{Cj}$

- $\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}$ — $g(i,j) = 1$
- $\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj}$ — $g(i,j) = 0$
- $\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj}$ — $g(i,j) = -1$

ESTIMATION AND HYPOTHESIS TESTING

- **Data:** $X_{gi}^{(k)} \overset{iid}{\sim} F_g^{(k)}, g = T, C, i = 1, \ldots, n_g, k = 1, \ldots, d$
- **Net treatment benefit** $\Delta = \mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Ci})$
- **One-sided hypothesis:** $\mathcal{H}_0 : \Delta \leq 0$

$\mathbf{X}_{Ti}$ *vs.* $\mathbf{X}_{Cj}$

- $\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}$ —— $g(i,j) = 1$
- $\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj}$ —— $g(i,j) = 0$
- $\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj}$ —— $g(i,j) = -1$

$$\Longrightarrow \mathbf{G} = \begin{pmatrix} g(1,1) & \cdots & \cdots & g(1,n_C) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ g(n_T,1) & \cdots & \cdots & g(n_T,n_C) \end{pmatrix}$$

ESTIMATION AND HYPOTHESIS TESTING

- **Data:** $X_{gi}^{(k)} \overset{iid}{\sim} F_g^{(k)}, g = T, C, i = 1, \ldots, n_g, k = 1, \ldots, d$
- **Net treatment benefit** $\Delta = \mathbb{P}(\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}) - \mathbb{P}(\mathbf{X}_{Ti} \prec \mathbf{X}_{Ci})$
- **One-sided hypothesis:** $\mathcal{H}_0 : \Delta \leq 0$

$\mathbf{X}_{Ti}$ *vs.* $\mathbf{X}_{Cj}$:

- $\mathbf{X}_{Ti} \succ \mathbf{X}_{Cj}$ — $g(i,j) = 1$
- $\mathbf{X}_{Ti} \asymp \mathbf{X}_{Cj}$ — $g(i,j) = 0$
- $\mathbf{X}_{Ti} \prec \mathbf{X}_{Cj}$ — $g(i,j) = -1$

$$\Longrightarrow \mathbf{G} = \begin{pmatrix} g(1,1) & \cdots & \cdots & g(1,n_C) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ g(n_T,1) & \cdots & \cdots & g(n_T,n_C) \end{pmatrix}$$

$$g(i,+) := \sum_{j=1}^{n_C} g(i,j), \quad g(+,j) := \sum_{i=1}^{n_T} g(i,j), \quad g(+,+) := \sum_{i=1}^{n_T} \sum_{j=1}^{n_C} g(i,j)$$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$
- **Information/variance estimator:** $\widehat{\mathcal{I}} := \widehat{\mathbb{V}}(\widehat{\Delta})^{-1} = \frac{1}{4N^2}\left(\frac{Q_C^2}{(n_C-1)n_C n_T^2} + \frac{Q_T^2}{(n_T-1)n_T n_C^2}\right)^{-1}$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$
- **Information/variance estimator:** $\widehat{\mathcal{I}} := \widehat{\mathbb{V}}(\widehat{\Delta})^{-1} = \frac{1}{4N^2}\left(\frac{Q_C^2}{(n_C-1)n_C n_T^2} + \frac{Q_T^2}{(n_T-1)n_T n_C^2}\right)^{-1}$
  - Q_g^2 can be computed based on $g(i,+), g(+,j)$ and $g(+,+)$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$
- **Information/variance estimator:** $\widehat{\mathcal{I}} := \widehat{\mathbb{V}}(\widehat{\Delta})^{-1} = \frac{1}{4N^2}\left(\frac{Q_C^2}{(n_C-1)n_C n_T^2} + \frac{Q_T^2}{(n_T-1)n_T n_C^2}\right)^{-1}$
  - Q_g^2 can be computed based on $g(i,+), g(+,j)$ and $g(+,+)$
- **Test statistic:** $T^{GPC} = \widehat{\Delta}\sqrt{\widehat{\mathcal{I}}}$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$
- **Information/variance estimator:** $\widehat{\mathcal{I}} := \widehat{\mathbb{V}}(\widehat{\Delta})^{-1} = \frac{1}{4N^2}\left(\frac{Q_C^2}{(n_C-1)n_C n_T^2} + \frac{Q_T^2}{(n_T-1)n_T n_C^2}\right)^{-1}$
  - Q_g^2 can be computed based on $g(i,+), g(+,j)$ and $g(+,+)$
- **Test statistic:** $T^{GPC} = \widehat{\Delta}\sqrt{\widehat{\mathcal{I}}}$
- **Under** $\mathcal{H}_0 : \Delta = 0$: $T^{GPC} \xrightarrow{d} \mathcal{N}(0,1)$

ESTIMATION AND HYPOTHESIS TESTING

- **Effect estimator:** $\widehat{\Delta} = \frac{g(+,+)}{n_T n_C}$
- **Information/variance estimator:** $\widehat{\mathcal{I}} := \widehat{\mathbb{V}}(\widehat{\Delta})^{-1} = \frac{1}{4N^2}\left(\frac{Q_C^2}{(n_C-1)n_C n_T^2} + \frac{Q_T^2}{(n_T-1)n_T n_C^2}\right)^{-1}$
  - Q_g^2 can be computed based on $g(i,+), g(+,j)$ and $g(+,+)$
- **Test statistic:** $T^{GPC} = \widehat{\Delta}\sqrt{\widehat{\mathcal{I}}}$
- **Under** $\mathcal{H}_0 : \Delta = 0$: $T^{GPC} \xrightarrow{d} \mathcal{N}(0,1)$

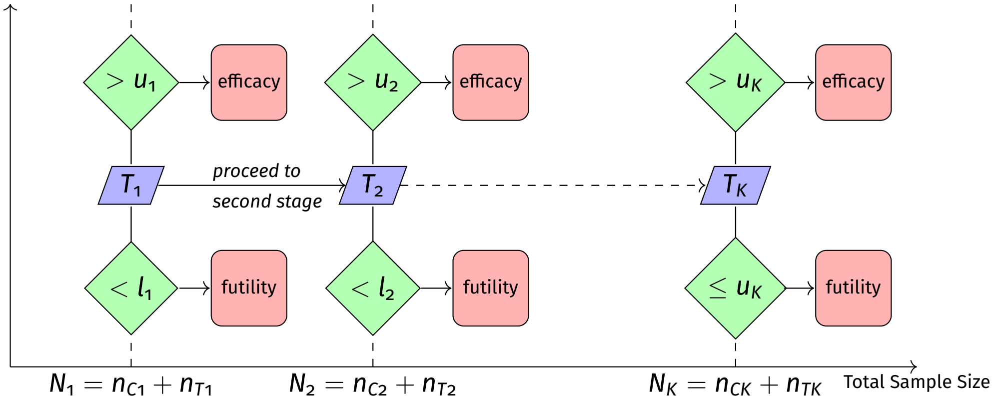
- For small sample sizes, we approximate $T^{GPC} \dot{\sim} t_f$, where f is estimated as $\widehat{f}$ from the data
- We reject $\mathcal{H}_0$ if $T^{GPC} > t_{1-\alpha,\widehat{f}}$

- GPC analogue of the Brunner–Munzel test for $\mathcal{H}_0 : \theta = 1/2$ (equivalently, $\mathcal{H}_0 : \Delta = 0$)

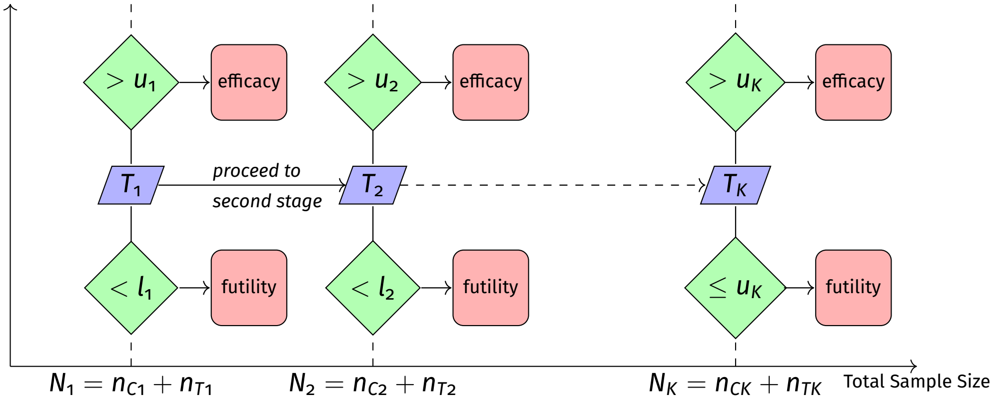
Part II

Group-Sequential Designs

GROUP-SEQUENTIAL DESIGNS



GROUP-SEQUENTIAL DESIGNS



Type-1 error rate control: $\mathbb{P}_{\mathcal{H}_0}(T_1 > u_1) + \sum_{i=2}^{K} \mathbb{P}_{\mathcal{H}_0}(\cap_{j=1}^{i-1}\{T_j \in [l_j, u_j]\}, T_i > u_i) \stackrel{!}{=} \alpha$

CANONICAL JOINT DISTRIBUTION

- A vector of test statistics $\mathbf{T} = (T_1, \ldots, T_K)^\top$ with $T_k = \widehat{\theta}_k \sqrt{\mathcal{I}_k}$ is said to follow the canonical joint distribution (asymptotically) if

CANONICAL JOINT DISTRIBUTION

- A vector of test statistics $\mathbf{T} = (T_1, \ldots, T_K)^\top$ with $T_k = \widehat{\theta}_k \sqrt{\mathcal{I}_k}$ is said to follow the canonical joint distribution (asymptotically) if

1. $\mathbf{T}$ is (asymptotically) multivariate normal,
2. $\mathbb{E}[T_k] = \theta\sqrt{\mathcal{I}_k}$ (asymptotically),
3. $\mathbb{C}ov(T_{k_1}, T_{k_2}) = \sqrt{\mathcal{I}_{k_1}/\mathcal{I}_{k_2}}$ for $1 \leq k_1 \leq k_2 \leq K$ (asymptotically) .

CANONICAL JOINT DISTRIBUTION

- A vector of test statistics $\mathbf{T} = (T_1, \ldots, T_K)^\top$ with $T_k = \widehat{\theta}_k \sqrt{\mathcal{I}_k}$ is said to follow the canonical joint distribution (asymptotically) if

1. $\mathbf{T}$ is (asymptotically) multivariate normal,
2. $\mathbb{E}[T_k] = \theta\sqrt{\mathcal{I}_k}$ (asymptotically),
3. $\mathbb{C}ov(T_{k_1}, T_{k_2}) = \sqrt{\mathcal{I}_{k_1}/\mathcal{I}_{k_2}}$ for $1 \leq k_1 \leq k_2 \leq K$ (asymptotically) .

- If 1) - 3) hold, standard group-sequential routines may be used

  $\Longrightarrow$ Famous examples include Pocock- and O'Brien-Fleming boundaries

  $\Longrightarrow$ Software packages are readily available (such as `rpact`)

Part III

Group-Sequential Designs for Generalized Pairwise Comparisons

KEY REFERENCE

Statistical Methods in Medical Research
2022, Vol. 31(10) 2004–2020

DOI: 10.1177/09622802221107103
journals.sagepub.com/home/smm
SAGE

Group sequential methods for the Mann-Whitney parameter

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

Abstract

Late phase clinical trials are occasionally planned with one or more interim analyses to allow for early termination or adaptation of the study. While extensive theory has been developed for the analysis of ordered categorical data in terms of the Wilcoxon-Mann-Whitney test, there has been comparatively little discussion in the group sequential literature on how to provide repeated confidence intervals and simple power formulas to ease sample size determination. Dealing more broadly with the nonparametric Behrens-Fisher problem, we focus on the comparison of two parallel treatment arms and show that the Wilcoxon-Mann-Whitney test, the Brunner-Munzel test, as well as a test procedure based on the log win odds, a modification of the win ratio, asymptotically follow the canonical joint distribution. In addition to developing power formulas based on these results, simulations confirm the adequacy of the proposed methods for a range of scenarios. Lastly, we apply our methodology to the FREEDOMS clinical trial (ClinicalTrials.gov Identifier: NCT00289978) in patients with relapse-remitting multiple sclerosis.

- Group-sequential methods are developed for:
  - Wilcoxon-Mann-Whitney test
  - Brunner-Munzel test
  - Test based on success odds
- As GPC, they are expressible via pairwise comparisons

A SEQUENTIAL TEST FOR THE NET TREATMENT BENEFIT

- To apply GPC sequentially
  1. Compute GPC matrix $\mathbf{G}_k \in \{1, 0, -1\}^{n_{Tk} \times n_{Ck}}$ at each stage $k = 1, \dots, K$

A SEQUENTIAL TEST FOR THE NET TREATMENT BENEFIT

- To apply GPC sequentially
  1. Compute GPC matrix $\mathbf{G}_k \in \{1, 0, -1\}^{n_{Tk} \times n_{Ck}}$ at each stage $k = 1, \ldots, K$
  2. Compute (dependent) test statistics $\mathbf{T}^{GPC} = (T_1^{GPC}, \ldots, T_K^{GPC})^\top$ with $T_k^{GPC} = \widehat{\Delta}_k \sqrt{\widehat{\mathcal{I}}_k}$

A SEQUENTIAL TEST FOR THE NET TREATMENT BENEFIT

- To apply GPC sequentially

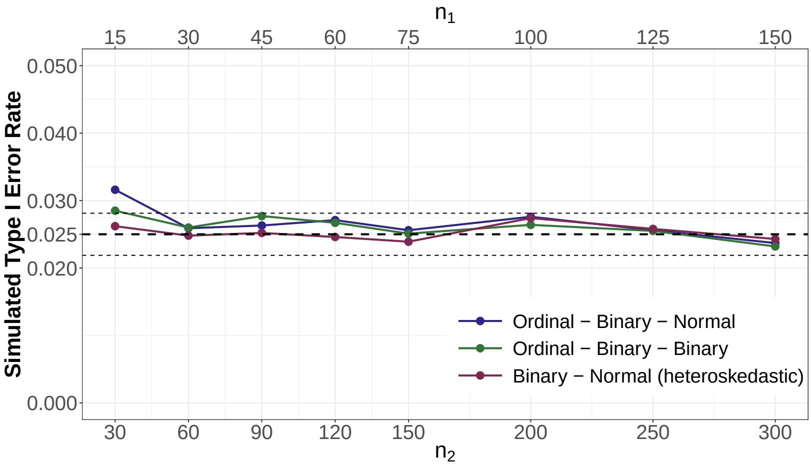
  1. Compute GPC matrix $\mathbf{G}_k \in \{1, 0, -1\}^{n_{Tk} \times n_{Ck}}$ at each stage $k = 1, \dots, K$
  2. Compute (dependent) test statistics $\mathbf{T}^{GPC} = (T_1^{GPC}, \dots, T_K^{GPC})^\top$ with $T_k^{GPC} = \widehat{\Delta}_k \sqrt{\widehat{\mathcal{I}}_k}$

Main theoretical result:

The vector of test statistics $\mathbf{T}^{GPC} = (T_1^{GPC}, \dots, T_K^{GPC})^\top$ asymptotically adheres to the canonical joint distribution.

TYPE I ERROR RATE

- Simulated type-1 error rate for 2-stage group-sequential Pocock design ($u_1 = u_2 \approx 2.147$)

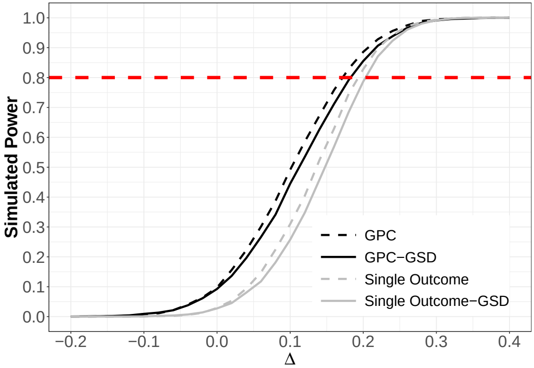


POWER AND EXPECTED SAMPLE SIZE

- 2-stage Pocock design with $n_2 = 100$/arm, $n_1 = n_2/2$
- Endpoint 1: $X_{Ci} \sim B(0.5), X_{Ti} \sim B(p_T)$; Endpoint 2: $X_{Ci} \sim N(0, 1), X_{Ti} \sim N(0.2, 1)$

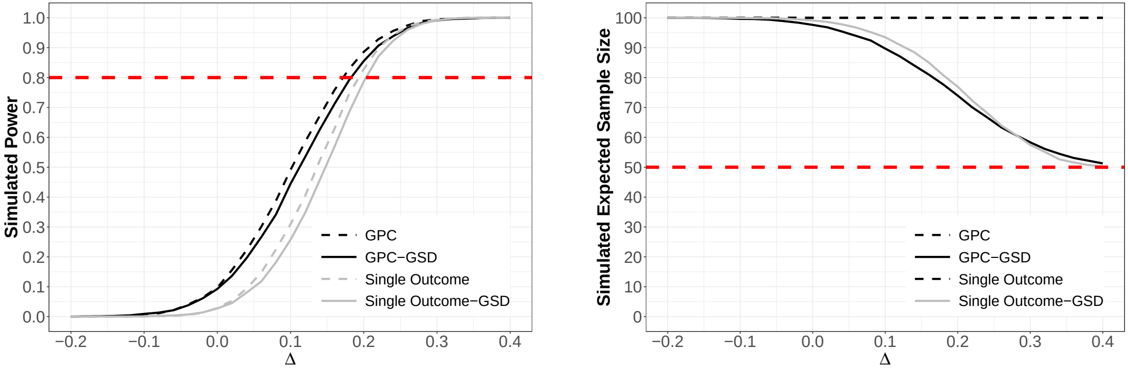
POWER AND EXPECTED SAMPLE SIZE

- 2-stage Pocock design with $n_2 = 100$/arm, $n_1 = n_2/2$
- Endpoint 1: $X_{Ci} \sim B(0.5), X_{Ti} \sim B(p_T)$; Endpoint 2: $X_{Ci} \sim N(0,1), X_{Ti} \sim N(0.2,1)$

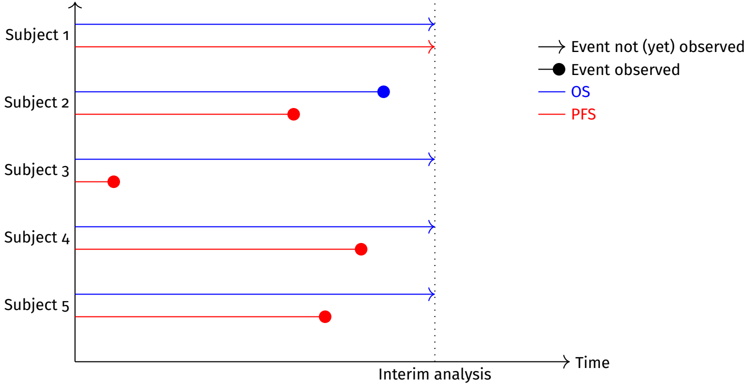


POWER AND EXPECTED SAMPLE SIZE

- 2-stage Pocock design with $n_2 = 100/\text{arm}$, $n_1 = n_2/2$
- Endpoint 1: $X_{Ci} \sim B(0.5), X_{Ti} \sim B(p_T)$; Endpoint 2: $X_{Ci} \sim N(0,1), X_{Ti} \sim N(0.2,1)$



SURVIVAL OUTCOMES

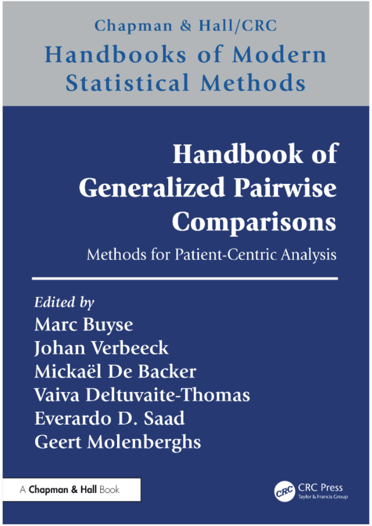


- Contribution of OF and PFS depends on timepoint of interim

KEY REFERENCE

Details

$\Longrightarrow$



CONCLUSION AND OUTLOOK

Conclusion

Group-sequential designs

- Improve efficiency of clinical trials
- Are advantageous from an ethical point of view

Generalized pairwise comparisons

- Allow to use multiple prioritized outcomes
- Allow to provide a more patient-centric view on drug efficacy

CONCLUSION AND OUTLOOK

Conclusion

Group-sequential designs

- Improve efficiency of clinical trials
- Are advantageous from an ethical point of view

Generalized pairwise comparisons

- Allow to use multiple prioritized outcomes
- Allow to provide a more patient-centric view on drug efficacy

Outlook

- Improved small-sample properties (alternative test procedures)
- Trial planning and sample size determination
- Incorporation of marginal inference
- Extension to further adaptive elements

REFERENCES I

Edgar Brunner, Arne C. Bathke, and Frank Konietschke.
Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs, Using R and SAS.
Springer, Berlin Heidelberg, 2019.

Marc Buyse.
Generalized pairwise comparisons of prioritized outcomes in the two-sample problem.
Statistics in medicine, 29(30):3245–3257, 2010.

Marc Buyse, Johan Verbeeck, Everardo D Saad, Mickaël De Backer, Vaiva Deltuvaite-Thomas, and Geert Molenberghs.
Handbook of generalized pairwise comparisons: methods for patient-centric analysis.
CRC Press, 2025.

REFERENCES II

Gernot Wassmer and Friedrich Pahlke.
rpact: Confirmatory Adaptive Clinical Trial Design and Analysis, 2022.
R package version 3.2.1.

HB Mann and DR Whitney.
On a test of whether one of two random variables is stochastically larger then the other.
Annals of Mathematical Statistics, 18:50–60, 1947.

Claus P Nowak, Tobias Mütze, and Frank Konietschke.
Group sequential methods for the mann-whitney parameter.
Statistical Methods in Medical Research, 31(10):2004–2020, 2022.

Stuart J Pocock.
Group sequential methods in the design and analysis of clinical trials.
Biometrika, 64(2):191–199, 1977.

Thanks for the attention!