UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

A studentized permutation test for the treatment effect in individual participant data meta-analysis

Thien Phuc Tran, Long-Hao Xu, Christian Röver, Tim Friede

Symposium: Recent Advances in Meta-Analysis, TU Dortmund, 01. July 2025

TABLE OF CONTENTS

- Motivation
- Proposed test
- Proposed confidence intervals
- Simulation
- Conclusion

MOTIVATION

- Meta-analysis can be performed on the basis of aggregate data or **individual participant data (IPD)**.
- For continuous outcome, assume a **linear mixed model (LMM)**.

EXAMPLE (KAMBIC ET AL., 2024)

- IPD MA on the safety and efficacy of exercise training on exercise capacity (e.g., six-minute walk test distance) and quality of life compared to standard care in left ventricular asssist device patients.
- 4 studies, each ranges from 14 to 54 patients.

→ Test and confidence intervals for the treatment effect based on the asymptotic normal distribution might not well control the Type I error rate.

…Satterthwaite's approach and Kenward-Roger's approach might be conservative.

→ Permutation tests might be useful.

MODEL (RILEY ET AL., 2013)

- Consider model:

stratified by study
random effect
treatment effect
residual variance

$$y_{ij} = \beta_{0i} + \beta_{1i} y_{0ij} + u_i z_{ij} + \varepsilon_{ij}, \; u_i \sim N(\theta, \tau^2), \qquad \varepsilon_{ij} \sim N(0, \sigma_i^2),$$

outcome
baseline value
group indicator (binary)
residual
between-study heterogeneity

- k: number of studies. study index: $i = 1, \dots, k$. study size: n_i. subject index: $j = 1, \dots, n_i$.
- β_{0i} and β_{1i} can be set random.

MODEL (MATRIX NOTATION)

$(y_{11}, \dots, y_{1n_1}, \dots, y_{k1}, \dots, y_{kn_k})^\top$ $\quad (u_1, \dots, u_k)^\top$ $\quad (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \dots, \varepsilon_{k1}, \dots, \varepsilon_{kn_k})^\top$ $\quad diag(\sigma_1^2, \dots, \sigma_1^2, \dots, \sigma_k^2, \dots, \sigma_k^2)$

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon},\ \boldsymbol{u} \sim N(\theta \mathbf{1}_k, \tau^2 \boldsymbol{I}_k), \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{R}), \tag{8}$$

$$\begin{pmatrix}
1 & 0 & \cdots & 0 & y_{011} & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
1 & 0 & & \vdots & y_{01n_1} & 0 & & \vdots \\
0 & 1 & & \vdots & 0 & y_{021} & & \vdots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\vdots & 1 & & \vdots & \vdots & y_{02n_2} & & \vdots \\
\vdots & 0 & & \vdots & \vdots & 0 & & \vdots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & & 0 & \vdots & \vdots & & 0 \\
\vdots & \vdots & & 1 & \vdots & \vdots & & y_{0k1} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & y_{0kn_k}
\end{pmatrix}$$

$$(\beta_{01}, \dots, \beta_{0k}, \beta_{11}, \dots, \beta_{1k})^\top$$

$$\begin{pmatrix}
z_{011} & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
z_{01n_1} & 0 & & \vdots \\
0 & z_{021} & & \vdots \\
\vdots & \vdots & & \vdots \\
\vdots & z_{02n_2} & & \vdots \\
\vdots & 0 & & \vdots \\
\vdots & \vdots & & \vdots \\
\vdots & \vdots & & 0 \\
\vdots & \vdots & & z_{0k1} \\
\vdots & \vdots & & \vdots \\
0 & 0 & \cdots & z_{0kn_k}
\end{pmatrix}$$

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

ESTIMATION & INFERENCE

$$\rightarrow \boldsymbol{Y} \sim N(\boldsymbol{X\beta} + \theta \boldsymbol{Z1}_k, \tau^2 \boldsymbol{ZZ}^\top + \boldsymbol{R})$$

- Approach: maximum likelihood, **restricted maximum likelihood (REML)**.
- No closed-form expressions for estimators.
- Maximize the restricted likelihood function using iterative algorithms.
- **Interest in estimation and inference on the treatment effect $\boldsymbol{\theta}$ (testing $\boldsymbol{\theta = 0}$).**

STANDARD APPROACHES

- Asymptotic normal distribution:

$$t = \frac{\theta}{\widehat{se}(\hat{\theta})} \sim N(0,1), \qquad CI_N = \hat{\theta} \pm z_{1-\alpha/2}\,\widehat{se}(\hat{\theta})$$

- Student-t distribution:

degree of freedom

$$t \sim T_{df}, \qquad CI_T = \hat{\theta} \pm t_{1-\frac{\alpha}{2},df}\,\widehat{se}(\hat{\theta})$$

Pinheiro's method (Pinheiro, 2000): df = $\sum_i n_i - 2k$

Satterthwaite's method (Giesbrecht and Burns, 1985) and **Kenward & Roger's method** (Kenward and Roger, 1997): df = $2\left[\widehat{var}(\hat{\theta})\right]^2/\widehat{var}\left[\widehat{var}(\hat{\theta})\right]$

- **Kenward & Roger's method** modifies $\widehat{se}(\hat{\theta})$ to account for the biasness in $\widehat{cov}(\hat{\beta},\hat{\theta})$.

PROS & CONS

- Normal-based approach and Pinheiro's approach cannot control Type I error rate (Riley et al., 2021).
- Satterwaite's approach and Kenward & Roger's approach might be conservative in IPD MA (Legha et al., 2018).
- All the above approaches are with distributional assumption.
- **Resampling approaches and permutation approaches are well-known for controlling Type I error, even for small data. → Might be useful.**

PROPOSED METHODS - IDEA

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon},\ \boldsymbol{u} \sim N(\theta\mathbf{1}_k, \tau^2\boldsymbol{I}_k), \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{R})$$

center **u**

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z}(\theta\mathbf{1}_k + \boldsymbol{u}_0) + \boldsymbol{\varepsilon},\ \boldsymbol{u}_0 \sim N(\mathbf{0}, \tau^2\boldsymbol{I}_k)$$

separate fix effects from random effects

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z}\theta\mathbf{1}_k + (\boldsymbol{Zu}_0 + \boldsymbol{\varepsilon}\ )$$

define new residuals

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z}\theta\mathbf{1}_k + \boldsymbol{\varepsilon}_0,$$

where $\boldsymbol{\varepsilon}_0 = \boldsymbol{Zu}_0 + \boldsymbol{\varepsilon}$ and $\sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0} = \tau^2\boldsymbol{ZZ}^\top + \boldsymbol{R}\right)$.

PROPOSED METHODS – IDEA (CONT.)

Under H_0: $\theta = 0$,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}_0$$
$$\rightarrow \boldsymbol{\varepsilon}_0 = \boldsymbol{Y} - \boldsymbol{X\beta} \qquad (7)$$

Impossible to permute ε_{0ij}'s since they are not **exchangeable** as $\boldsymbol{\Sigma}_{\varepsilon_0}$.

Solution (standardization):

Let **W** is an upper triangle matrix such that $\boldsymbol{\Sigma}_{\varepsilon_0} = \boldsymbol{W}^\top \boldsymbol{W}$ (Cholesky's decomposition).

So

$$\tilde{\boldsymbol{\varepsilon}} = (\boldsymbol{W}^\top)^{-1} \boldsymbol{\varepsilon}_0 \sim N(\boldsymbol{0}, \boldsymbol{I})$$

$\rightarrow$ $\tilde{\varepsilon}_{ij}$'s (asymptotic) exchangeable $\rightarrow$ **permute** $\tilde{\varepsilon}_{ij}$'s, **not** ε_{0ij}'s (first proposed in Lee & Braun, 2012).

PROPOSED PERMUTATION TEST

Table 1: Pseudo code for the proposed studentized permutation test, generating a permutation distribution of t-statistic values.

1. Fit Model (7) to the response $\boldsymbol{Y}$ and compute t-statistic $t = \hat{\theta}/\widehat{\text{se}}(\hat{\theta})$.
2. Fit Model (8) to the response $\boldsymbol{Y}$ and compute the standardized error $\hat{\tilde{\varepsilon}} = (\hat{\boldsymbol{W}}^{\top})^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$, where $\hat{\boldsymbol{W}}$ is the Cholesky decomposition of $\hat{\boldsymbol{\Sigma}}_{\varepsilon_0} = \hat{\tau}^2\boldsymbol{Z}\boldsymbol{Z}^{\top} + \hat{\boldsymbol{R}}$.
3. For permutation $p = 1, \cdots, N$,
   - Permute $\hat{\tilde{\varepsilon}}$ as $\hat{\tilde{\varepsilon}}_{(p)}$ and compute the permuted response $\boldsymbol{Y}_{(p)} = \boldsymbol{X}\hat{\beta} + \hat{\boldsymbol{W}}^{\top}\hat{\tilde{\varepsilon}}_{(p)}$.
   - Refit Model (7) to $\boldsymbol{Y}_{(p)}$ and compute t-statistic $t_{(p)} = \hat{\theta}_{(p)}/\widehat{\text{se}}(\hat{\theta}_{(p)})$.

PROPOSED CONFIDENCE INTERVALS

- Quantile-based confidence interval:

$$CI_{p1} = \left[\hat{\theta} - t^*_{1-\frac{\alpha}{2}} se(\hat{\theta}); \hat{\theta} - t^*_{\frac{\alpha}{2}} se(\hat{\theta})\right]$$

- Simple but may not well control Type I error when $\theta \neq 0$.
- Search-based confidence interval:
  - Idea: confidence bounds are the closest values to $\hat{\theta}$ such that they are just significantly different from θ (Follmann & Proschan, 1999).

→ Search for values θ_0 closest to $\hat{\theta}$ and perform permutation test $H_0: \theta = \theta_0$ until the test significant.

  - Let $\tilde{\theta} = \theta - \theta_0$,

$$H_0: \theta = \theta_0 \Leftrightarrow H_0: \tilde{\theta} = 0,$$

PROPOSED CONFIDENCE INTERVALS

- Search-based confidence intervals:

$$CI_{p2} = [\theta_{lower}, \theta_{upper}]$$

- Idea: θ_{lower} and θ_{upper} are the closest values to $\hat{\theta}$ such that they are just significantly different from θ (Follmann & Proschan, 1999).

→ Search for values θ_0 closest to $\hat{\theta}$ and perform permutation test $H_0: \theta = \theta_0$ until the test significant.

- Let $\tilde{\theta} = \theta - \theta_0$,

$$H_0: \theta = \theta_0 \Leftrightarrow H_0: \tilde{\theta} = 0,$$

→ Analogous to Table (1).

- Better control Type I error rate, but computationally expensive!!!

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

PROPOSED CONFIDENCE INTERVALS

Table 2: Pseudo code for constructing the upper bound of the search- and permutation-based $100\%(1-\alpha)$ confidence interval. Construction of the lower bound works analogously.

For $\theta_{0i} = i \times (b-a)/(M-1), i = 1, \cdots, M$, where M denotes the number of grid points in $[a;b], b \geq a \geq \hat{\theta}$,

1. Compute $\boldsymbol{Y}_{\theta_{0i}} = \boldsymbol{Y} - \theta_{0i}\boldsymbol{Z}\mathbf{1}_k$.
2. Replace $\boldsymbol{Y}$ with $\boldsymbol{Y}_{\theta_{0i}}$ and perform steps 1-3 in Table 1.
3. Compute p-value for the test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta < \theta_0$,:
   p-value $= 1 - \sum_{p=1}^{N} \mathbb{1}(t_{(p)} \geq t)/N$.
4. If p-value $\leq \alpha/2$, stop and return $\theta_{upper} := \theta_{0i}$.

SIMULATION SETTING

1. Simulate data from Model (8)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon},\ \boldsymbol{u} \sim N(\theta \mathbf{1}_k, \tau^2 \boldsymbol{I}_k), \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{R}),$$

2. Fit Model (8) assuming σ_i are the equal.

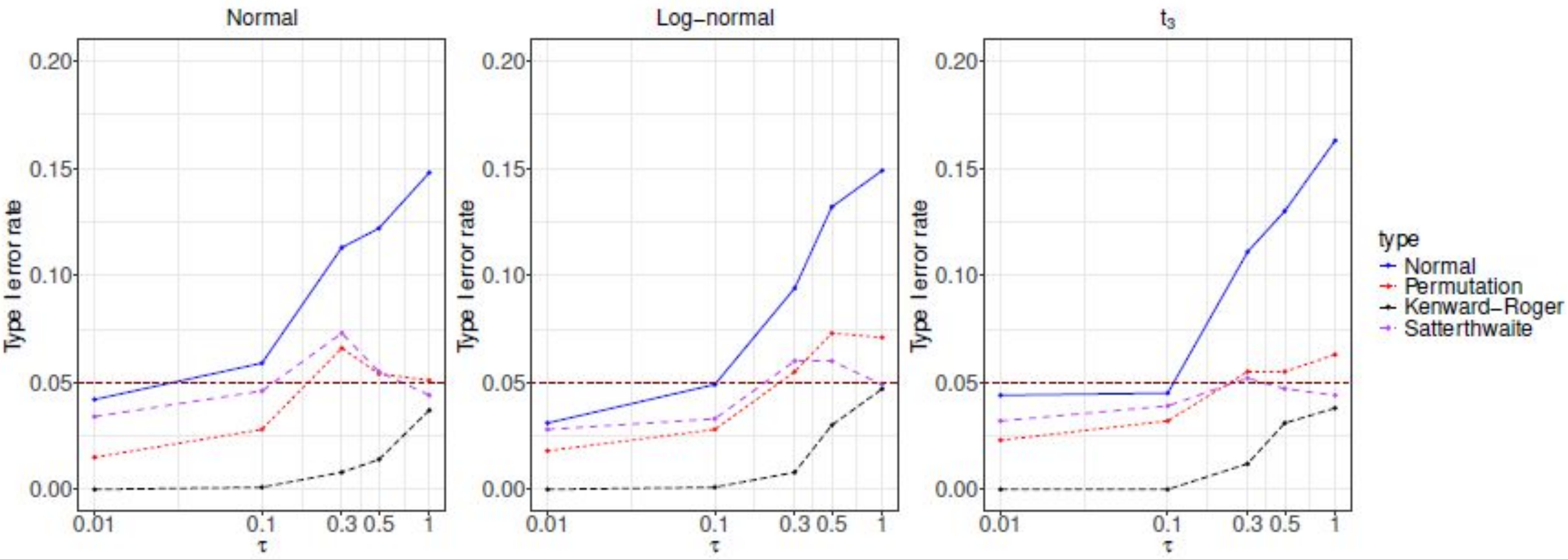
3. Compare tests and CIs based on: the proposed methods, Normal distribution, Satterthwaite‘s approach, and Kenward & Roger‘s approach.

SIMULATION SETTING

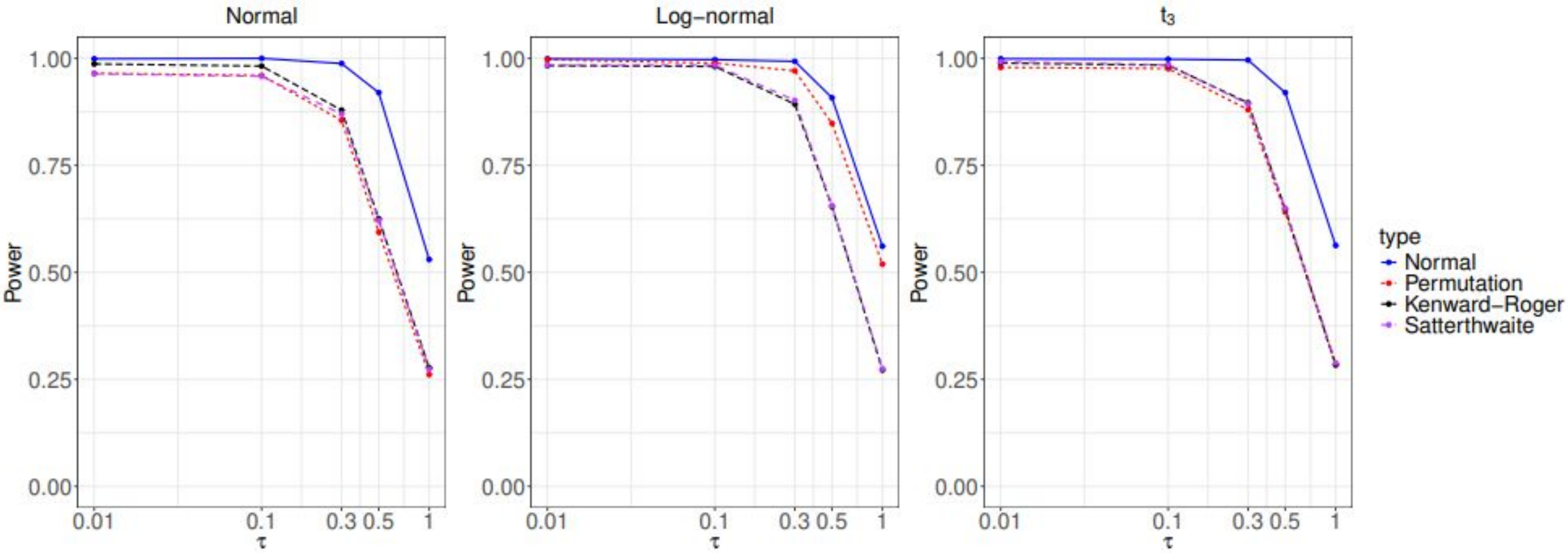
Table 3: Parameter settings for the simulations discussed in Section 4.

parameter		settings
number of studies	k	4
study size	n_i	medium: Unif(100, 200), small: Unif(30, 100), very small: 0.8 Unif(15, 30) + 0.2 Unif(30, 100)
treatment allocation	p_i	Unif(0.5, 0.7)
intercepts	β_{0i}	$(0.9, 2.3, 0.3, 0.1)^\top$
slopes	β_{1i}	$(0.8, 0.7, 0.9, 0.9)^\top$
treatment effect	θ	$\{0, 1\}$
heterogeneity	τ	$\{0.01, 0.1, 0.3, 0.5, 0.7, 1.0\}$
residual s.d.	σ_i	$1, (0.9, 0.9, 0.9, 1.4)^\top$
baseline value	y_{0ij}	Normal(4, 1)
residuals	ϵ_{ij}	Normal, Student-t_3, log-Normal
permutations	N	10 000 (permutation test, CI_{p1}), 2 000 (2nd permutation, CI_{p2})

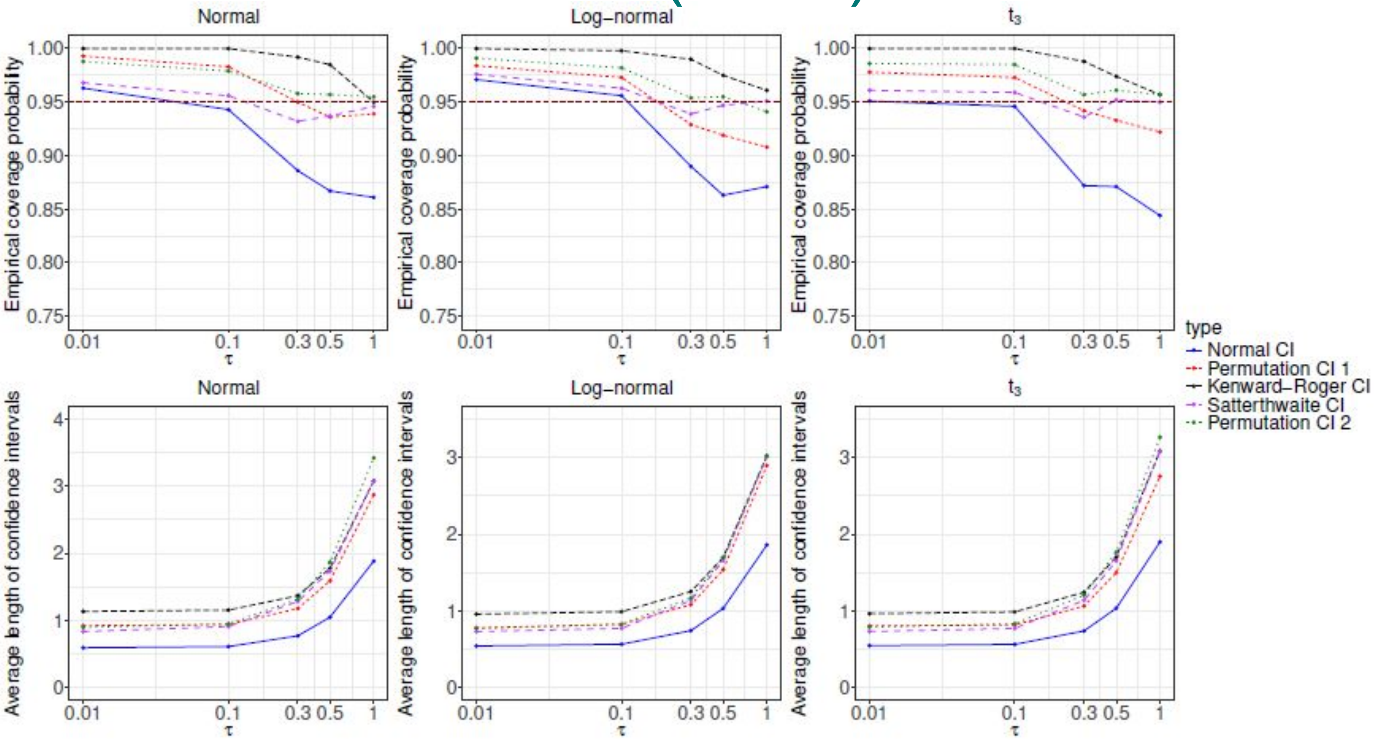
SIMULATION RESULTS



SIMULATION RESULTS (cont.)



SIMULATION RESULTS (cont.)



SUMMARY

- Proposed a studentized permutation test and two permutation-based confidence intervals for the treatment effect in IPD MA of continuous outcomes, assuming the exchangeability of the standardized residuals.
- Future work:
  - ❖ Analytical proof due the lack of a closed-form expression for the effect estimator when variance components are unknown.
  - ❖ Reduce the computational expense of the search-based confidence interval method.
  - ❖ Straightforward to extend the proposed method to cases when there is more than one random effect. About extension of the proposed method to other types of outcomes.

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

THANK YOU

Preprint:
http://arxiv.org/abs/2505.24774

References

T. P. A. Debray, K. G. M. Moons, G. van Valkenhoef, O. Efthimiou,N. Hummel, R. H. H. Groenwold, J. B. Reitsma, and on behalf of the GetReal methods review group. Get real in individual participant data(IPD) meta-analysis: a review of the methodology. Research Synthesis Methods, 6(4):293–309, 2015.

D. A. Follmann and M. A. Proschan. Valid inference in random effects meta-analysis. Biometrics, 55(3):732–737, 05 2004.

U. Halekoh and S. Højsgaard. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The Rpackage pbkrtest. Journal of Statistical Software, 59(9):1–32, 2014.

J. Higgins and S. Green. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Collaboration, 2011.

References

R. N. Kackar and D. A. Harville. Approximations for standard errors of estimators of fixed and random effect in mixed linear models. Journal of the American Statistical Association, 79(388):853–862, 1984.

T. Kambiˇc, A. Feuerstein, P. T. Tran, T. Friede, F. Edelmann, M. Lainscak, and ExTraLVAD Collaborators. Exercise training in left ventricular assist device patients: Protocol of an individual participant data meta-analysis. ESC Heart Failure, 2024.

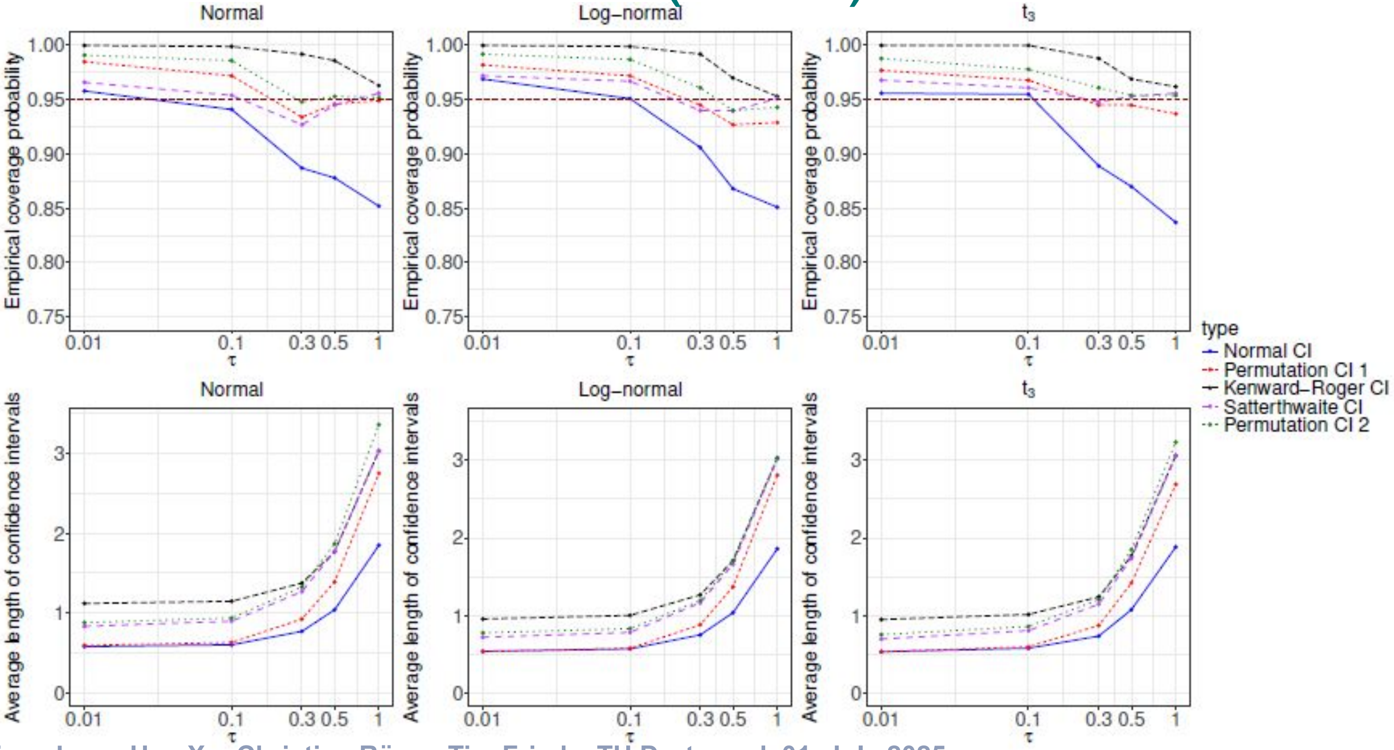
O. E. Lee and T. M. Braun. Permutation tests for random effects in linear mixed models. Biometrics, 68(2):486–493, 2011

J. C. Pinheiro. Theory and computational methods for linear mixed-effects models. In Mixed-Effects Models in S and S-PLUS, pages 57–96.Springer New York, New York, NY, 2000.

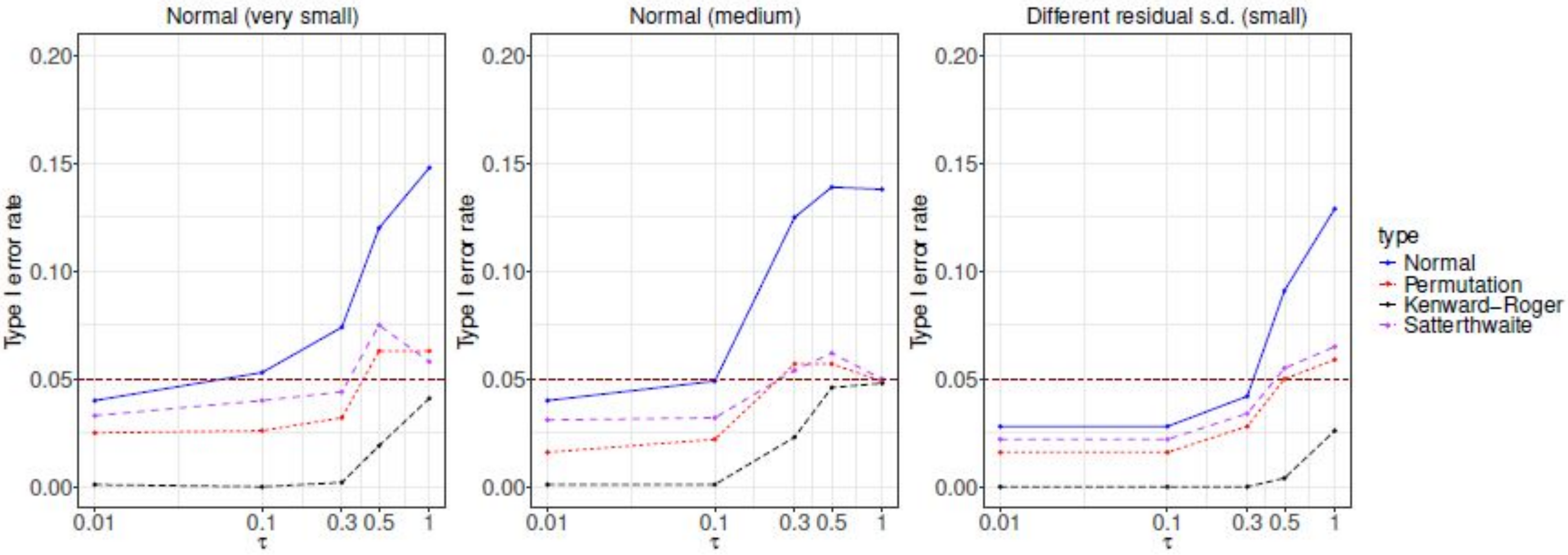
References

F. E. Satterthwaite. Approximate distribution of estimates of variance components. Biometrics Bulletin, 2(6):110–114, 1946.

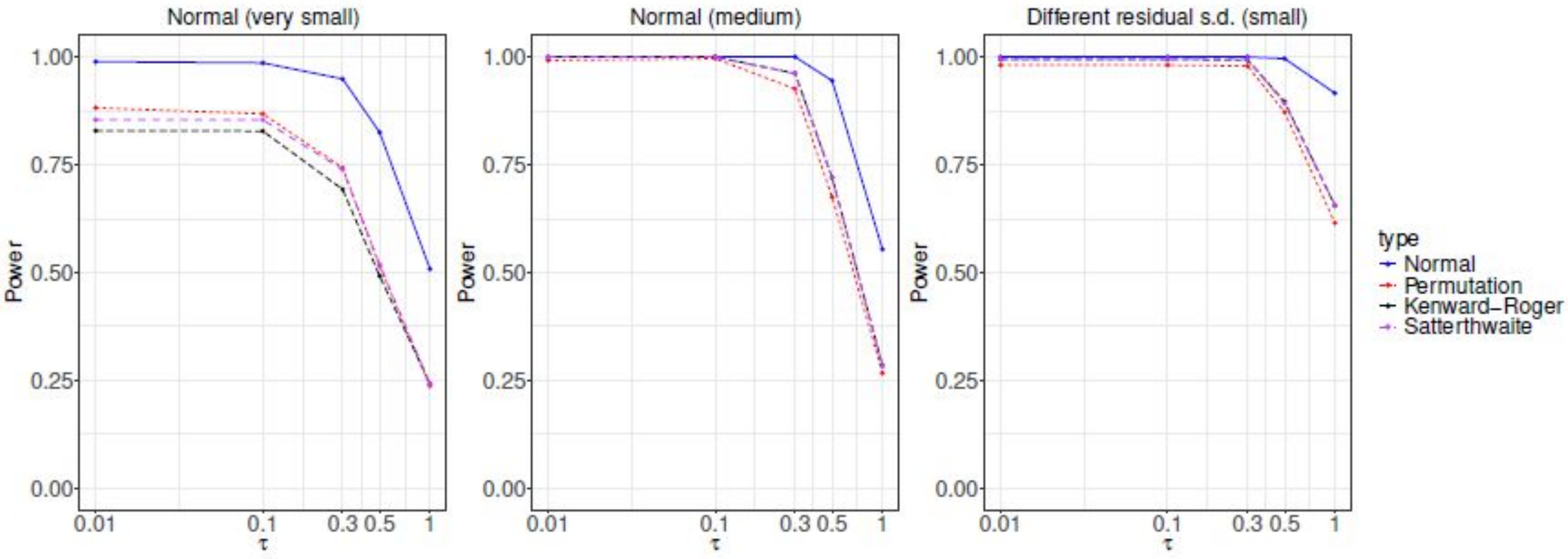
SIMULATION RESULTS (cont.)



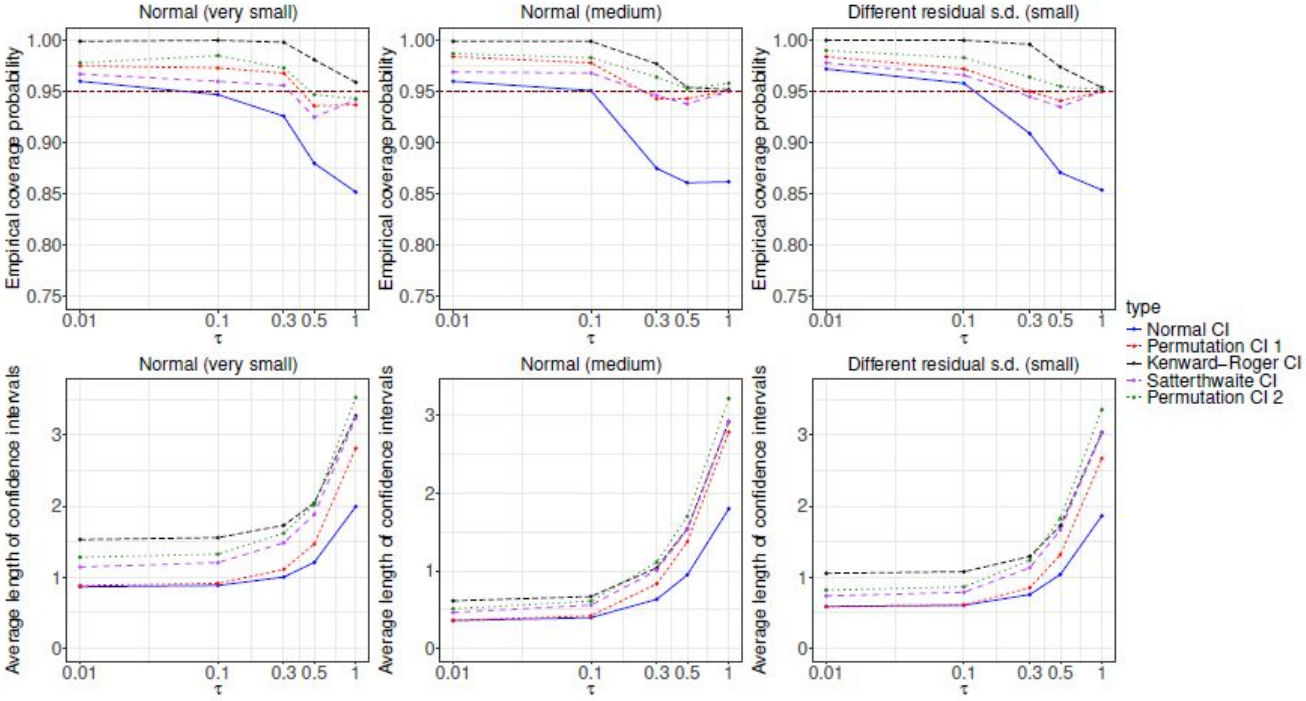
SIMULATION RESULTS (cont.)



SIMULATION RESULTS (cont.)



SIMULATION RESULTS (cont.)



SIMULATION RESULTS (cont.)