

Simulation study on methods for ordinal data in central statistical monitoring and their application on the German Multiple Sclerosis Register

**ISCB-GMDS 2026
Sept 27 - Oct 01**

**Firas Fneish, David Ellenberger, Niklas Frahm, Alexander Stahmann, Gerhard Fortwengel,
Frank Schaarschmidt**

msregister



30-09-2026

- The monitoring team ensures the compliance of each center to the study protocol and the requirements of good clinical practice (ICH-GCP)
- ICH E6(R2) Good Clinical Practice Guideline
- Purpose of CSM: Identifying atypical centers
- Current methods focused on outlier/inlier detection on a patient level and demonstrated the usage of principle component analysis on center level
- **Our focus is on comparing centers to identify potential deviations in clinical practice**



1 December 2016
EMA/CHMP/ICH/135/1995
Committee for Human Medicinal Products

Guideline for good clinical practice E6(R2)

CLINICAL TRIALS ARTICLE

A statistical approach to central monitoring of data quality in clinical trials





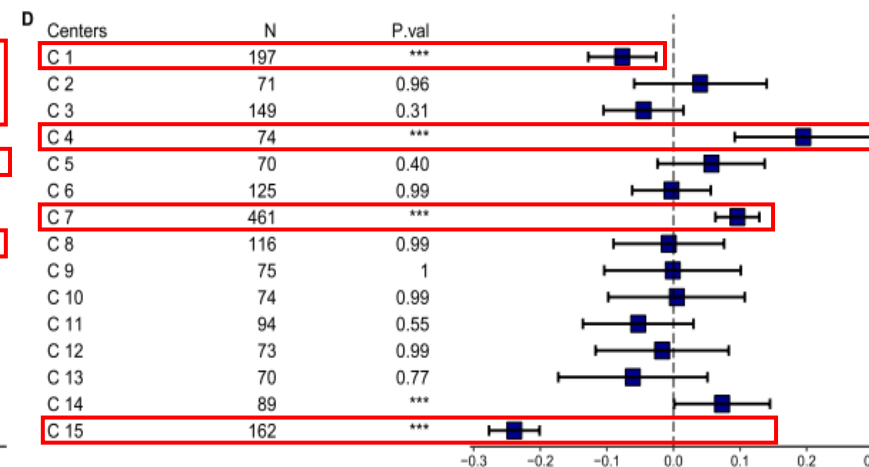
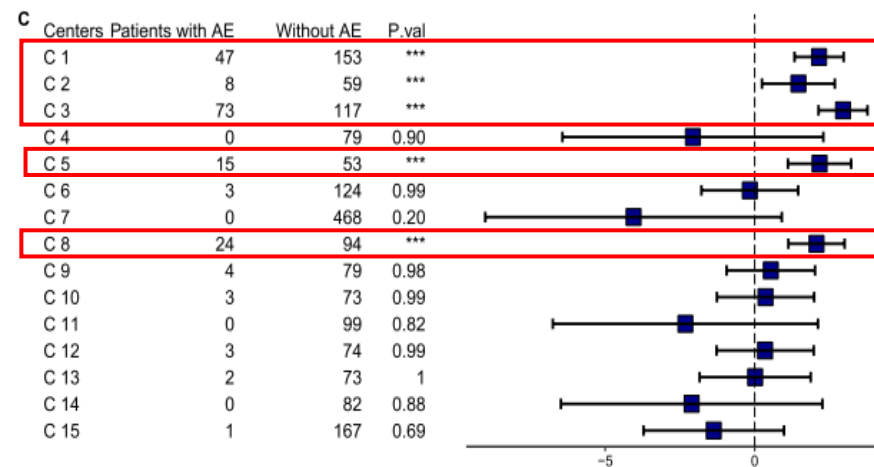
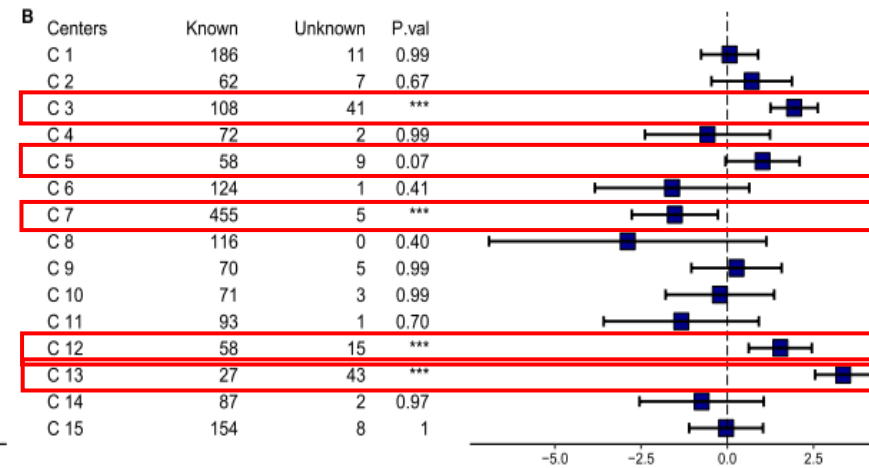
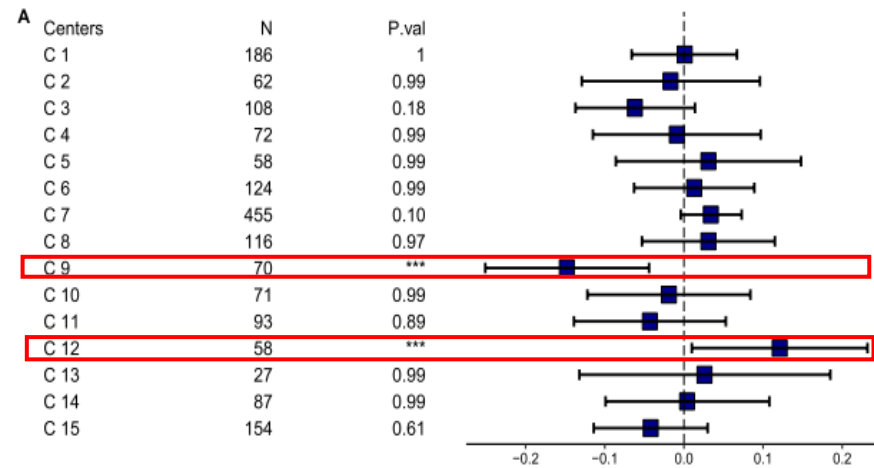
ORIGINAL RESEARCH

Application of Statistical Methods for Central Statistical Monitoring and Implementations on the German Multiple Sclerosis Registry

Firas Fneish^{1,2} · David Ellenberger² · Niklas Frahm² · Alexander Stahmann² · Gerhard Fortwengel³ · Frank Schaarschmidt¹

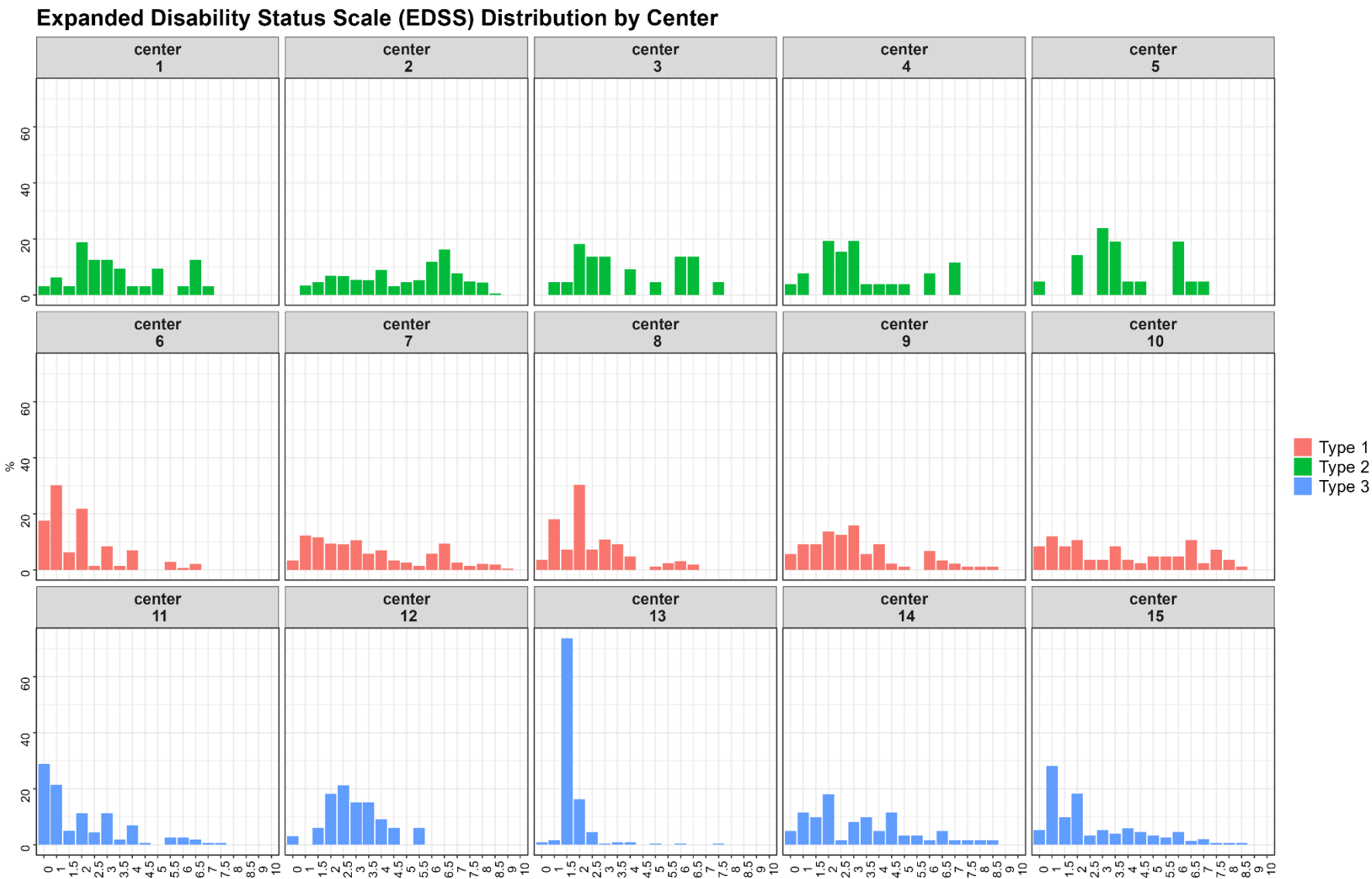
Received: 2 January 2023 / Accepted: 8 June 2023 / Published online: 14 July 2023
© The Author(s) 2023

msregister



- Multiple comparisons of single centers to the Grand Mean of all centers
- Applying statistical methods to cover ordinal data
 - Identify requirements/settings needed for each potential method
 - The appropriate sample size N per center
- Application of potential statistical methods on medical registry data to identify atypical center(s) in comparison to the Grand Mean of the data as a whole

German MS Register (GMSR) data



Approach: Comparisons to Grand Mean

- For a given model with parameters m_i and possibly unbalanced samples sizes n_i , the grand mean $m_{\cdot}$ can be computed by $m_{\cdot} = \sum_{i=1}^I \frac{n_i}{N} m_i$
- Comparisons of each center parameter m_i to the grand mean $m_{\cdot}$ can then be written as a set of $k=1, \dots, K$ linear contrasts, with contrast coefficients $c_k = (c_{k1}, c_{k2}, c_{k3}, \dots, c_{kI})$

$$c_1 = (1 - \frac{n_1}{N}, -\frac{n_2}{N}, \frac{n_3}{N}, \dots, -\frac{n_I}{N})$$

$$c_2 = (-\frac{n_1}{N}, 1 - \frac{n_2}{N}, -\frac{n_3}{N}, \dots, -\frac{n_I}{N})$$

- The deviations d_k of each center from the grand mean can then be written as:

$$d_k = m_{i=k} - m_{\cdot} = \sum_{i=1}^I c_{ki} m_i$$

Approach: Comparisons to Grand Mean

- Testing the null hypothesis: no center deviates from the overall mean

$$H_0 = d_k = (m_{i=k} - m_{.}) = 0$$

versus the alternative hypothesis that at least one center deviates from the overall mean

$$H_A = d_k = (m_{i=k} - m_{.}) \neq 0$$

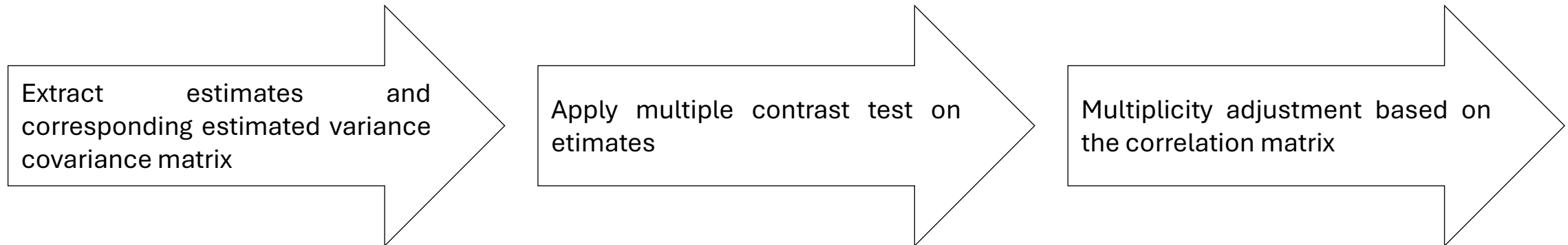
- Family-wise error rate (FWER=0.05): the probability of obtaining at least one falsely significant result among all centers.

Method 1: Cumulative link model

- Fit a cumulative link model (CLM)
- Logit function:

$$\log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \theta_j - \mathbf{x}'\boldsymbol{\beta}$$

- Differences between centers as only predictor
- Allows including additional covariates (ex: center types)



Implemented in R:

- Model fitting: package `ordinal`, function `clm()`, (Christensen, 2025, version 2025.12-29)
- Multiple contrast tests: package `emmeans`, function `emmeans()`, (Lenth & Piaskowski, 2025, version 2.0.0)

Method 2: Non-parametric approach (Nparcomp)

- No specific distributional assumption for data outcome
- Center effects are defined as relative effects $p_i = (Y_i \leq Y)$
- Relative effects are estimated by samples comparing observations

$$Y: p_{ii'} = P(Y_i < Y_{i'}) + \frac{1}{2}P(Y_i = Y_{i'})$$

- Explicitly allows for discrete, tied data
- Performs multiple contrast tests between centers' relative effects
- Multiplicity adjustment is based on the correlation obtained from the standardized variance covariance matrix (of relative effects)

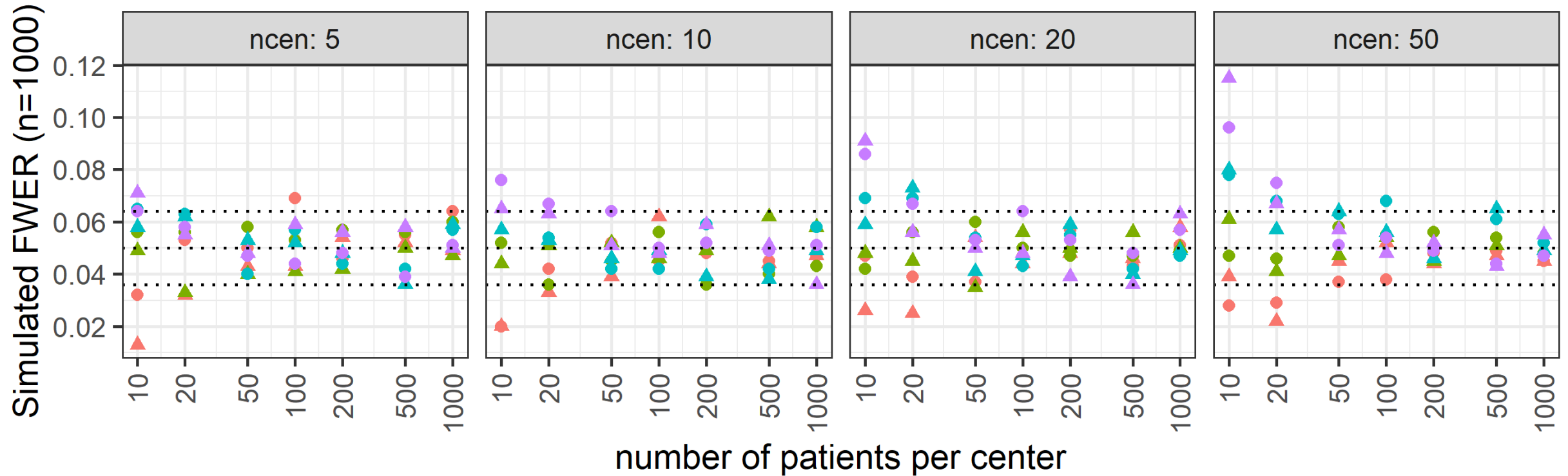
Simulation set up

- One-way layout with I centers and n_i patients per center
- Latent variable sampled from logistic distribution
- Thresholds to obtain ordinal data from latent variable: predefined quantiles of logistic distribution
- Global H_0 :
 - location (0) and scale (1) parameters equal across all centers
 - thresholds equal across all centers

Parameter	Values / Settings
Number of centers	5, 10, 20, 50
Number of patients per center	10, 20, 50, 100, 200, 500, 1000
Number of categories in ordinal response	3, 5, 10, 20
Distribution pattern 1 (relatively even categories)	$p = (0.5, 0.3, 0.2)$
Distribution pattern 2 (rare high scores)	$p = (0.5, 0.45, 0.05)$
Number of simulations per setting	$S = 1000$

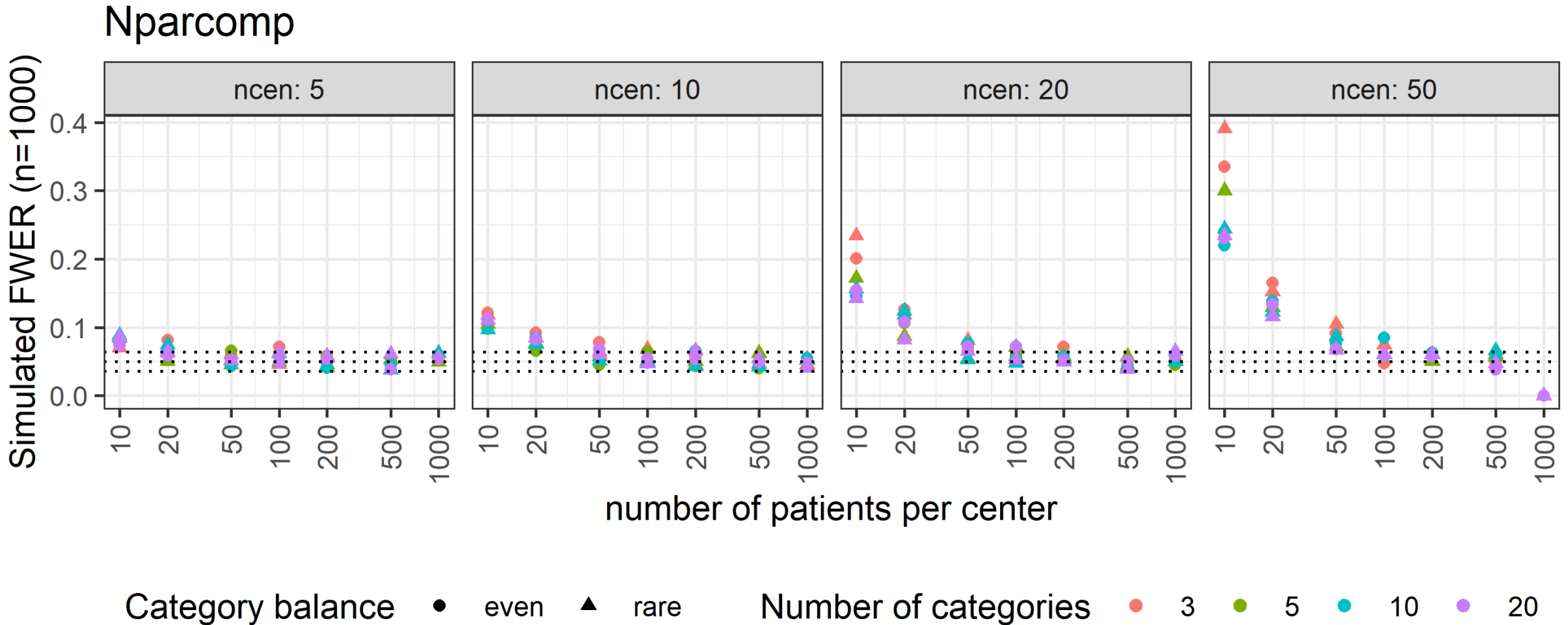
Family-Wise Error Rate

CLM

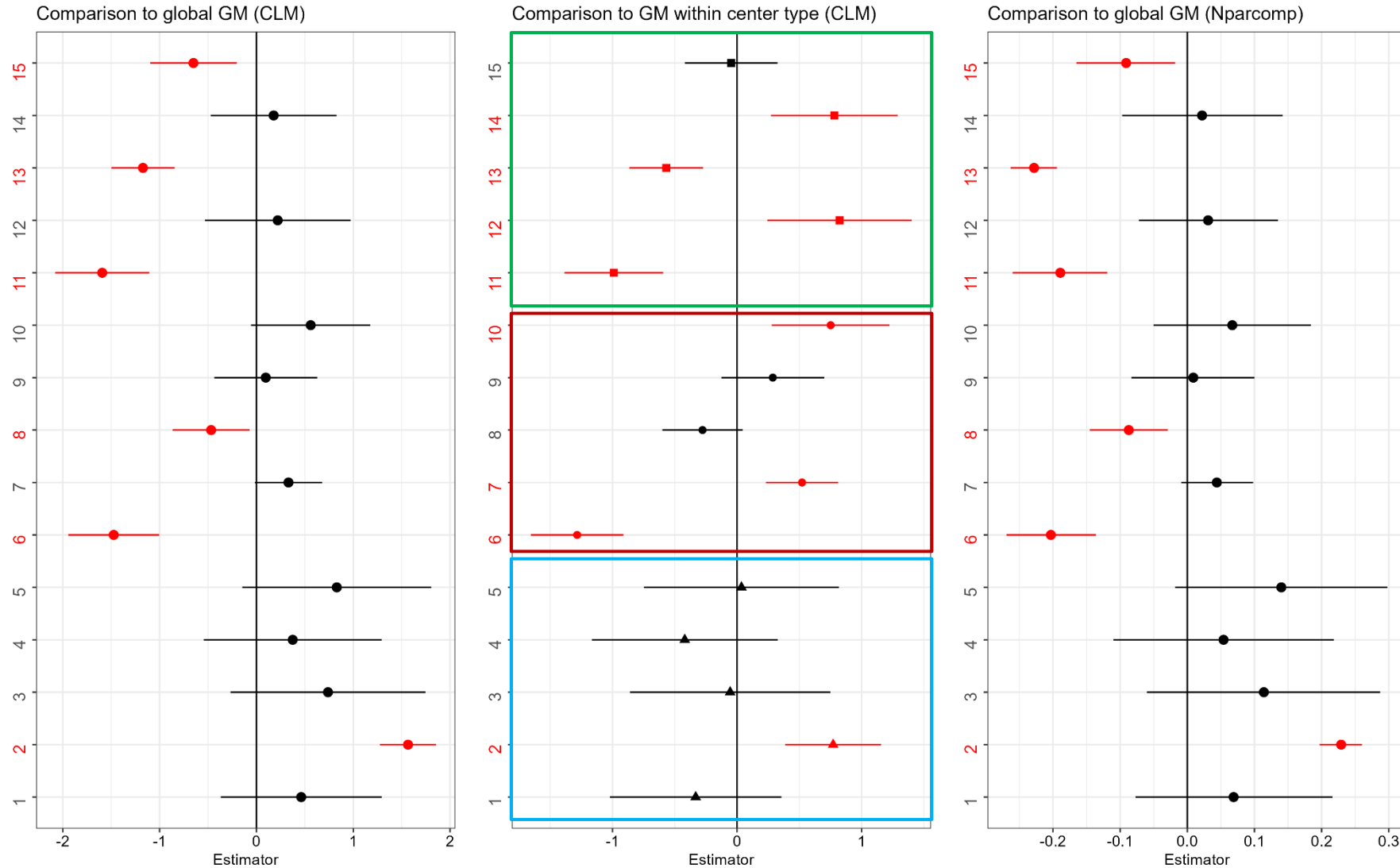


Category balance ● even ▲ rare Number of categories ● 3 ● 5 ● 10 ● 20

Family-Wise Error Rate



Contrasts of center means with Grand Mean for GMSR data set



Take home message

- Cumulative link models are more suitable in controlling FWER for GM comparisons when $n_i < 20$
- Nparcomp inflates FWER up to 40% when $i = 50$ and $n_i < 20$
- For both CLM and Nparcomp FWER control improves with increasing categorical levels
- Further simulations needed for unbalanced designs
- Power simulations are required to identify sample sizes that achieve adequate power while maintaining FWER control

- Christensen R (2025). ordinal-Regression Models for Ordinal Data. R package version 2025.12-29, <https://CRAN.R-project.org/package=ordinal>
- Lenth R, Piaskowski J (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 2.0.0, <https://CRAN.R-project.org/package=emmeans>
- Konietzschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. A. (2015). nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. Journal of Statistical Software, 64(9), 1–17. <https://doi.org/10.18637/jss.v064.i09>
- Fneish, F., Ellenberger, D., Frahm, N. et al. Application of Statistical Methods for Central Statistical Monitoring and Implementations on the German Multiple Sclerosis Registry. Ther Innov Regul Sci 57, 1217–1228 (2023). <https://doi.org/10.1007/s43441-023-00550-0>

Für weitere Informationen besuchen Sie gerne
unsere Website unter
www.msregister.de

Kontakt

MS Forschungs- und Projektentwicklungs-gGmbH
Krausenstr. 50
30171 Hannover
Deutschland



Geschäftsführer:

Alexander Stahmann
Tel.: +49(0)511 44 45 99 55
kontakt@msregister.de

msregister