Assessing Covariate-Adjusted Risk Differences in Small-Sample Trials

A Comparative Evaluation of Statistical Methods

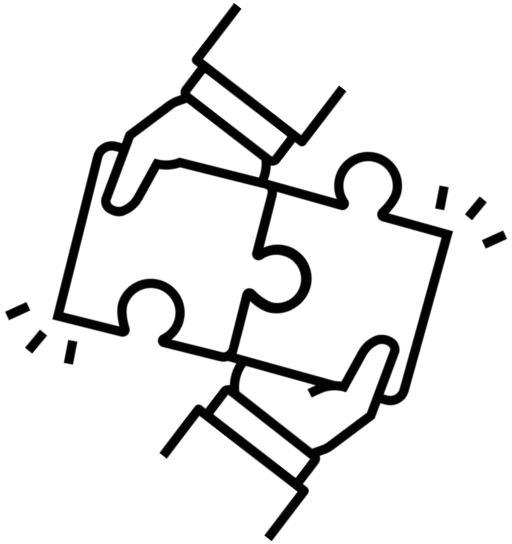
Martin Schnuerch | Global Biostatistics and Data Sciences, Boehringer Ingelheim

Life forward

Acknowledgements

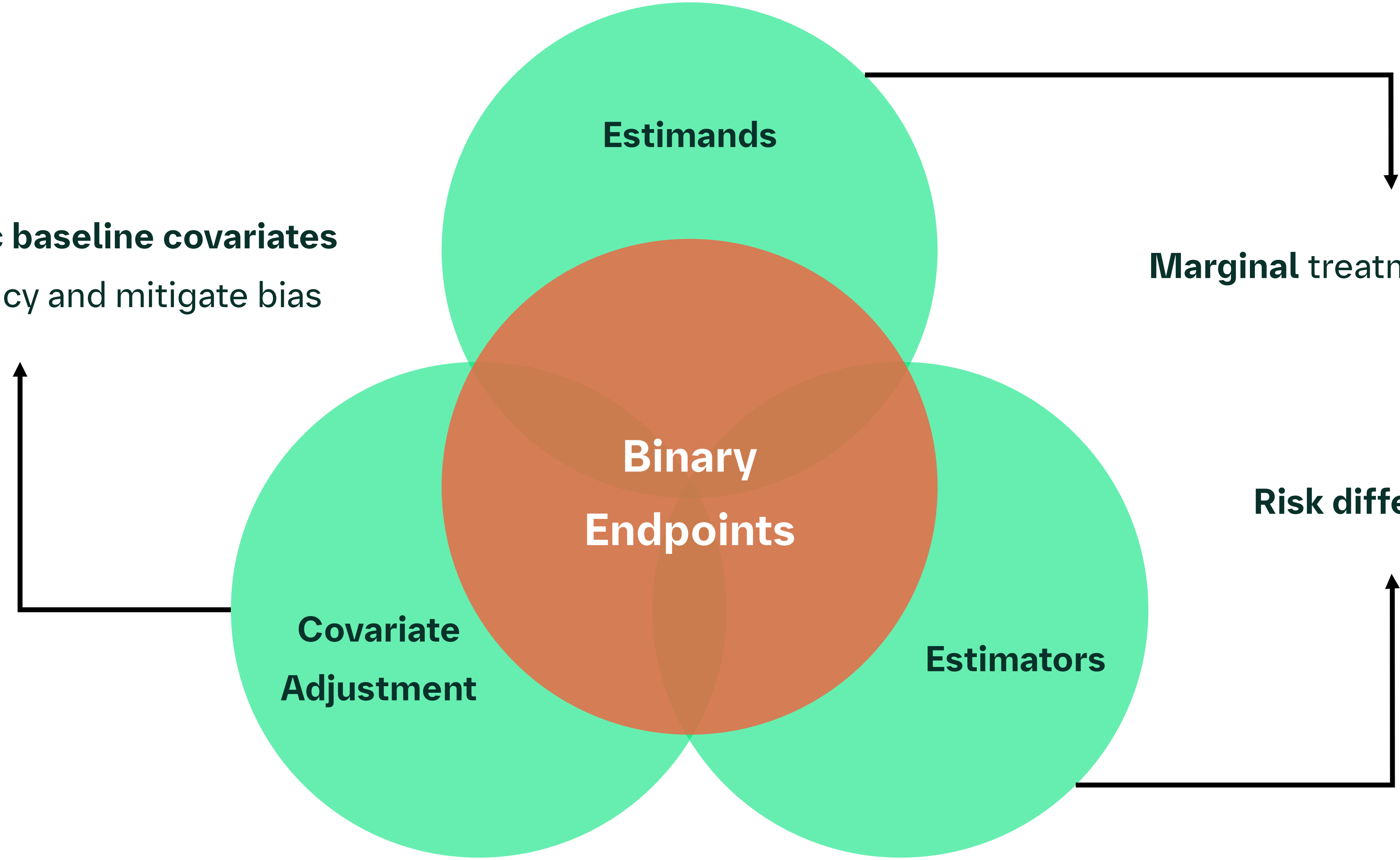
Joint work with:

- **Christian Stock**, Boehringer Ingelheim Pharma GmbH und Co. KG, Germany
- **Alex Ocampo**, F. Hoffmann-La Roche Ltd, Switzerland
- **Klaus Kähler Holst**, Novo Nordisk, Denmark



Background

Include **prognostic baseline covariates** to increase efficiency and mitigate bias



Marginal treatment effects

Risk differences

Boehringer Ingelheim

Popular Method: *g*-Computation

- Fit **working model** (treatment assignment $A = 1, 0$, baseline covariates $\boldsymbol{X}$)

$$\text{logit}\{P(Y = 1 \mid A, \boldsymbol{X})\} = \beta_0 + \beta_A A + \beta_X^T \boldsymbol{X}$$

- **Predict conditional response probabilities** for each individual

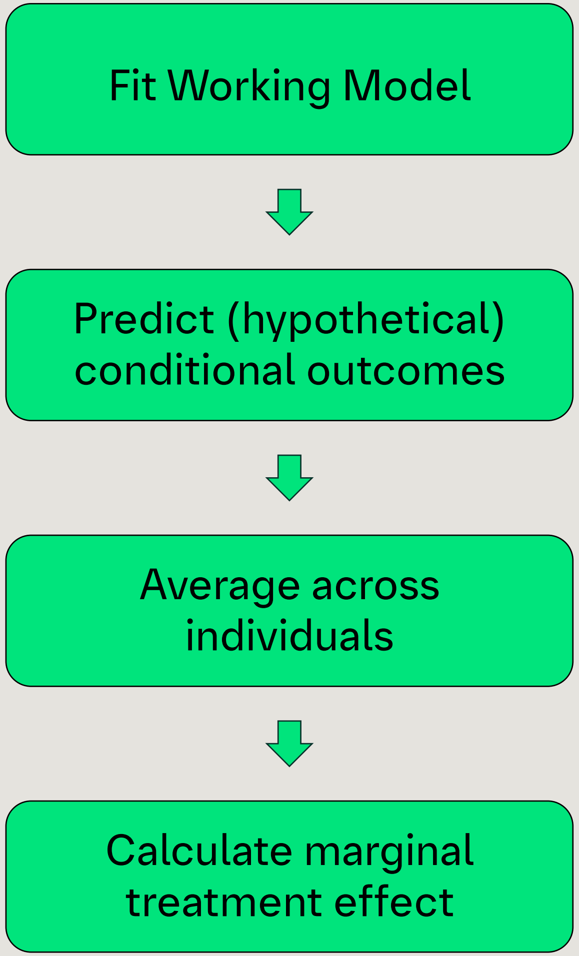
$$\hat{\pi}_i^{(1)} = \text{logit}^{-1}\{\hat{\beta}_0 + \hat{\beta}_A + \hat{\boldsymbol{\beta}}_X^T \boldsymbol{X}_i\}, \qquad \hat{\pi}_i^{(0)} = \text{logit}^{-1}\{\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_X^T \boldsymbol{X}_i\}$$

- **Average** to estimate marginal response probabilities

$$\hat{\mu}_1 = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_i^{(1)}, \qquad \hat{\mu}_0 = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_i^{(0)}$$

- Calculate contrast to estimate the **marginal treatment effect**

$$\hat{\delta} = \hat{\mu}_1 - \hat{\mu}_0.$$



Practical and Inferential Challenges

1. (Robust) **variance estimation** for *g*-computation point estimator
2. Statistical **assumptions** and alignment with targeted **estimands**
3. Further complication: **Small sample sizes** (e.g., rare diseases, pediatric diseases)
   - Violation of asymptotic assumptions?
   - Convergence issues?

Current research:

- Simulation-based evaluation of **range of statistical methods** in small-sample trial scenario
- Overview and **practical guidance** for selection of statistical methods

Boehringer Ingelheim

Hypothetical Trial & Potential Estimands of Interest

- 2-arm RCT comparing efficacy of investigational treatment vs. control
- Binary endpoint, two binary baseline covariates

- **Marginal treatment effect:**

$$\text{MTE} = \mathbb{E}(Y^{(1)} - Y^{(0)})$$

- **Conditional treatment effect:**

$$\text{CTE}(\boldsymbol{x}) = \mathbb{E}(Y^{(1)} - Y^{(0)} \mid \boldsymbol{X} = \boldsymbol{x})$$

- **Conditional population-average treatment effect:**

$$\text{CPATE} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(Y^{(1)} - Y^{(0)} \mid \boldsymbol{X} = \boldsymbol{x}_i)$$

Boehringer Ingelheim

Statistical Methods: Overview

Method	Covariate Adjustment	Statistical Test	Estimand
Suissa-Shuster Test	**none**	Exact unconditional test	MTE
Cochran-Mantel-Haenszel Test	**stratification**	Stratified χ^2 test	CTE
Mantel-Haenszel RD: Sato	**stratification**	Wald *z* test	CPATE
Mantel-Haenszel RD: mGR	**stratification**	Wald *z* test	MTE
g-computation: Ge	**regression model**	Wald *z* test	CPATE
g-computation: Liu	**regression model**	Wald *z* test	MTE
g-computation: Ye	**regression model**	Wald *z* test	MTE
g-computation: Steingrimsson	**regression model**	Wald *z* test	MTE
g-computation: Zhang	**regression model**	Score test	MTE
g-computation: Firth	**regression model**	Wald z test	CPATE*

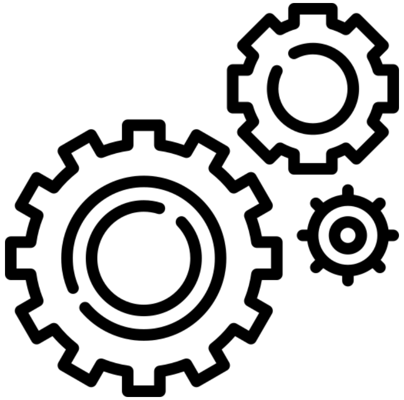
*Firth is a working model decision rather than an inferential one, so it could be used to target any estimand. RD = Risk Difference; mGR = modified Greenland-Robins

Boehringer Ingelheim

Simulation Study

Data Generation:

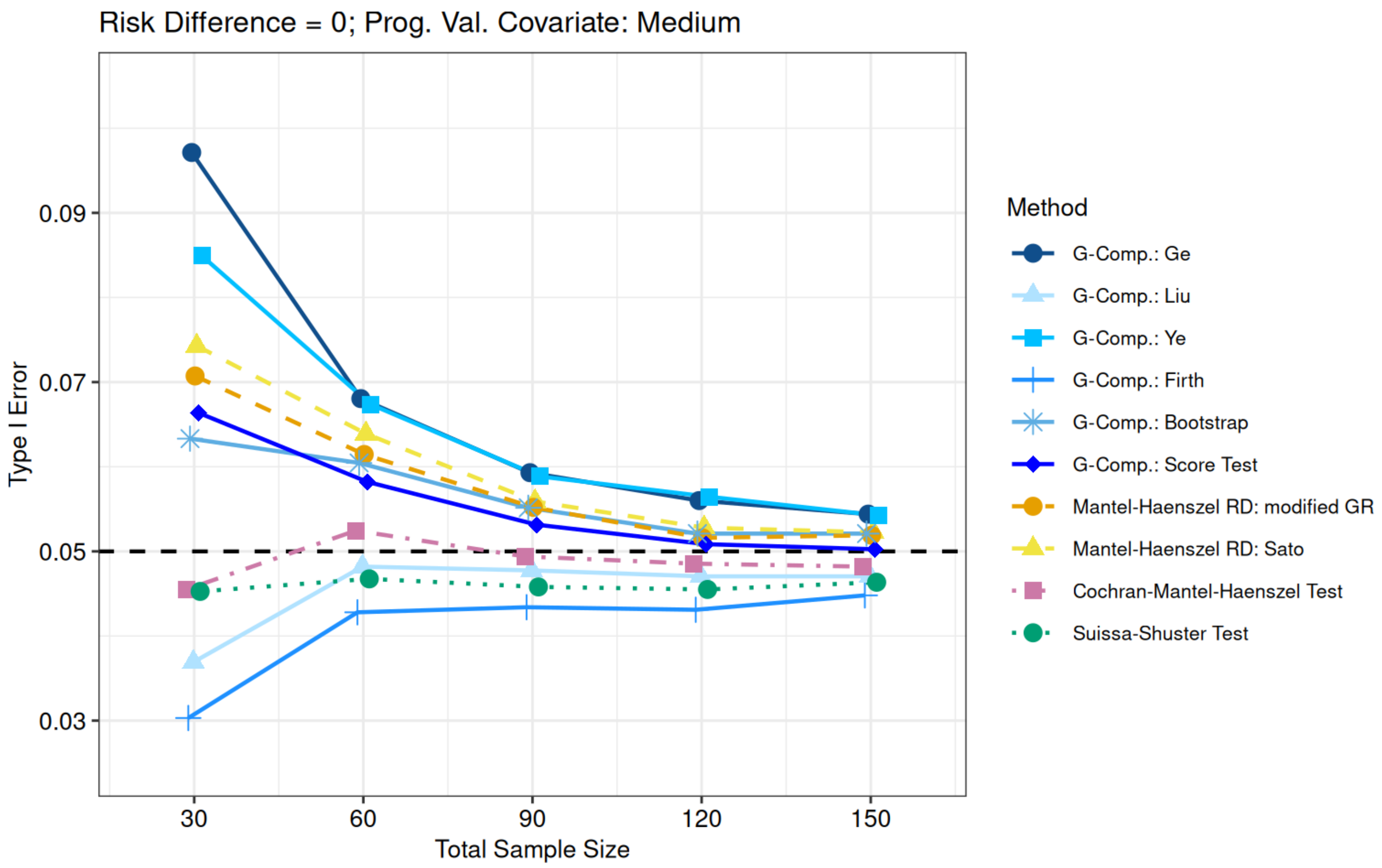
- Multivariable logistic model with parameters determined by an iterative bisection search procedure to achieve target marginal proportions
- 2 treatment arms (active vs. control) with 1:1 randomization
- 2 binary covariates with $\mathrm{P}(\mathrm{X}_1 = 1) = \mathrm{P}(\mathrm{X}_2 = 1) = 0.50$
- Marginal response rate under control: 0.20



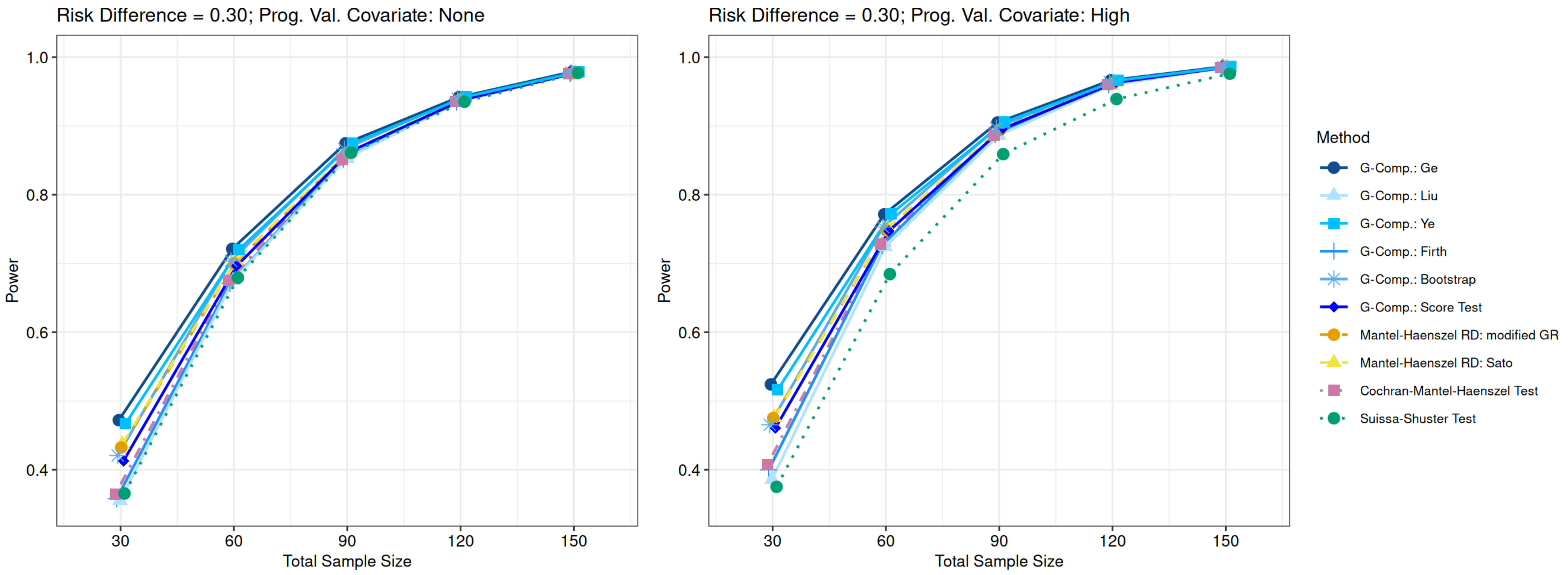
Manipulated Parameters:

- True marginal risk difference: $\delta = 0, 0.15, 0.30$
- Prognostic value of baseline covariates: $\beta = \log(1), \log(1.5), \log(3)$
- Total sample size: $\mathrm{N} = 30, 60, 90, 120, 150$

Results: Type I Error



Results: Statistical Power



Boehringer Ingelheim

Summary

- **Inflated Type I error rate** for Ge- and Ye-approach and for MH RD estimator when sample sizes are (too) small
  ⟹ Keep in mind: **estimand mismatch**

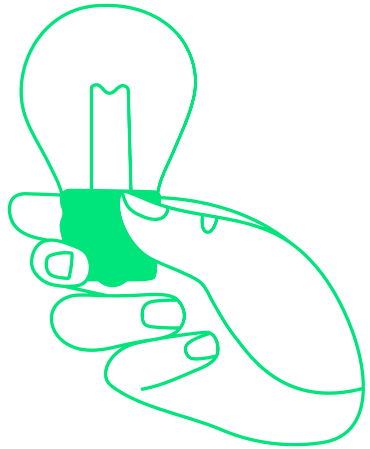
- Conservative Type I error control and reduced power for Liu- and Firth-approach

- Moderate Type I error rate inflation and moderate power loss for **bootstrap** and **score-test** method

- **Excellent Type I error rate control** but reduced power for **CMH** and **Suissa-Shuster test**
  ⟹ Safe approach when sample sizes are extremely small

No method to rule them all.

Choice of method must align with targeted **estimands**, trial **objectives/priorities**, and modeling **assumptions**.



Thank you for your attention!

Contact: martin.schnuerch@boehringer-Ingelheim.com

https://arxiv.org/abs/2605.19772

Preprint

Boehringer Ingelheim

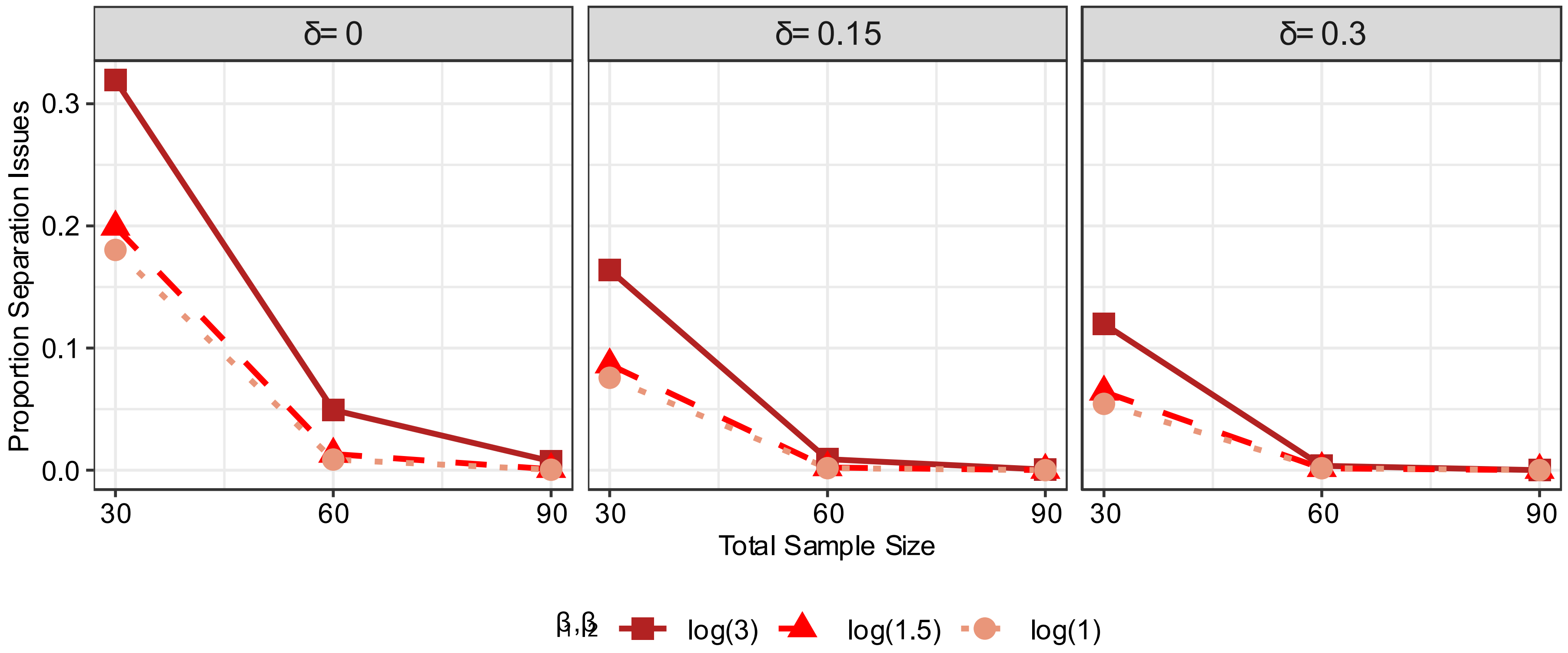
References

- Firth (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80(1): 27–38.Ge et al. (2011). Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Information Journal*, 45(4): 481–493.
- Liu and Xi (2024). Covariate adjustment and estimation of difference in proportions in randomized clinical trials. *Pharmaceutical Statistics*, 23(6): 884–905.
- Magirr et al. (2025). Estimating the variance of covariate-adjusted estimators of average treatment effects in clinical trials with binary endpoints. *Pharmaceutical Statistics*, 24(4): e70021.
- Mantel and Haenszel (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4): 719–748.
- Qiu et al. (2025). Clarifying the role of the Mantel–Haenszel risk difference estimator in randomized clinical trials. *Biometrics*, 81(4): ujaf142.
- Steingrimsson et al. (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary Clinical Trials*, 54: 18–24.
- Suissa and Shuster (1985). Exact unconditional sample sizes for the 2×2 binomial trial. *Journal of the Royal Statistical Society: Series A*, 148(4): 317–327.
- Ye et al. (2023). Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Statistical Theory and Related Fields*, 7(2): 159–163.
- Zhang et al. (2025). A robust score test in g-computation for covariate adjustment in randomized clinical trials leveraging different variance estimators via influence functions. *Statistics in Medicine*, 44(7): e70080.

Appendix

Boehringer Ingelheim

Results: Separation Issues in Logistic Regression Working Model



Statistical Methods I: (Non-Parametric) Tests

Suissa-Shuster Test

- Non-parametric exact unconditional test of two independent binomial proportions
- Null hypothesis on marginal RD: $\mu_1 = \mu_0 \Leftrightarrow \delta = 0$
- No covariate adjustment; targets **MTE**

Cochran-Mantel-Haenszel Test

- Stratified χ^2 test assessing conditional independence across a set of K 2x2 tables
- Strict null hypothesis on *common* RD: $\delta_1 = \delta_2 = \cdots = \delta_K \coloneqq \delta = 0$
- Adjustment for categorical covariates; targets **CTE**

Mantel and Haenszel (1959); Suissa and Shuster (1985)

Statistical Methods II: Risk-Difference Estimator

Mantel-Haenszel Risk Difference Estimator

- Weighted average of estimated stratum-specific risk differences
- Inference based on normal approximation/z test
- Adjustment for categorical covariates
- Targets **different estimands** (Qiu et al., 2025):
    - **CPATE**: Weighted average of stratum-specific risk differences
      ⇒ Conditional variance estimator, e.g. as proposed by **Sato** et al. (1989)
    - **MTE**: Marginal risk difference across random sample of strata
      ⇒ Unconditional variance estimator, **modified Greenland-Robins** (mGR; Qiu et al., 2025)

Mantel and Haenszel (1959)

Statistical Methods III: G-Computation

Ge et al. (2011)

- Conditional variance estimator using delta method and logistic regression working model-based covariance
- Inference based on normal approximation/z test
- Targeted estimand: **CPATE**

Liu and Xi (2024)

- Robust unconditional variance estimator using delta method with sandwich covariance estimator and additional covariate-variability component
- Inference based on normal approximation/z test
- Targeted estimand: **MTE**

Statistical Methods IV: G-Computation

Ye et al. (2023)

- Doubly robust, semi-parametric unconditional variance estimator
- Inference based on normal approximation/z test
- Targeted estimand: **MTE**

Steingrimsson et al. (2017)

- Non-parametric bootstrap to empirically approximate unconditional variance of g-computation estimator
- Inference based on normal approximation/z test
- Targeted estimand: **MTE**

Statistical Methods V: G-Computation

Zhang et al. (2025)

- Robust score test for g-computation estimator using generalized score statistic with finite-sample penalty
- Admits any valid estimator of unconditional variance for risk-difference estimator
- Targeted estimand: **MTE**

Firth (1993)

- Penalized-likelihood logistic regression working model to address separation and convergence issues in small samples
- Inference based on normal approximation/z test using model-based variance estimator (Ge et al., 2011)
- Targeted estimand: **CPATE**