Randomization-based Inference for MCP-Mod

Lukas Pin, Alex Sverdlov, Frank Bretz, Björn Bornkamp

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

PSI Conference

London, 11th of June 2025

Outline

Motivation

Why?

MCP-Mod in **small samples** with a **binary endpoint**

Which challenges?

Asymptotic inference and **non-existence of maximum-likelihood estimators (MLEs)**

What is MCP-Mod?

MCP-Mod is used in Phase-II trials for **testing** and estimation of dose-response relationships

Solutions:

Randomization-based inference and **penalised MLEs**

Trial Example

- Randomized, double-blinded Phase-II trial
- Binary outcome: 1 = Success, 0 = Failure.
- Objective: Assess efficacy and safety across 3 dose levels (10 mg, 25 mg, 100 mg) vs. placebo (0 mg).
- Assumed response rates: Placebo: 20% Highest dose (100 mg): 80%
- Total of $n = 49$ patients randomized 1:2:2:2 ratio (placebo vs. active doses).
- 80% power to detect dose-response signal at one-sided 10% significance level.

Outline

Outline

Generalised MCP-Mod

Purpose:

- Multiple Comparison Procedures and Modeling (MCP-Mod) approach for testing and modeling the relationship between doses and responses (Bretz et al., 2005).
- Generalised MCP-Mod extends to more complex scenarios like binary outcomes, count data, or time-to-event endpoints (Pinheiro et al., 2014).

Implementation:

- Available in the `DoseFinding` package in `R`.

Used in practice and regulatory recognition:

- Recognized by **EMA** and **FDA** for use in Phase-II dose-finding studies.

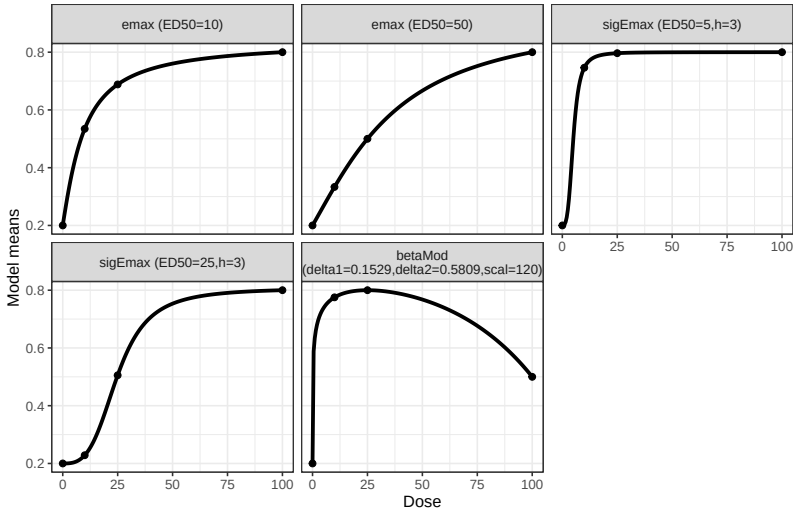
The Two Steps of MCP-Mod

1. **Testing dose-response signal:**
   - Use of **multiple contrast tests** to evaluate the null hypothesis:

$$H_0 : \mu_0 = \mu_1 = \ldots = \mu_{k-1}$$

   - Optimized contrast coefficients detect specific dose-response shapes.
   - Controls **type-I error rate** at a predefined significance level.
2. **Estimating dose-response relationship:**
   - If the null hypothesis is rejected, estimate the dose-response relationship.
   - Uses **model averaging** to ensure robust estimation.

Candidate Models



Outline

Complete Separation in Logistic Regression

Complete separation:

- Occurs when predictor variables perfectly predict the outcome.
  - e.g. when we have zero responders on placebo group
- A hyperplane separates the two classes perfectly.
- This leads to the **non-existence of MLEs**.

Solution: Firth's Method (Firth, 1993)

- Modifies the score function to reduce bias of MLEs.
- **Ensures finite estimates** even in cases of complete separation.
- Implemented in `R` using the `logistf`-package.

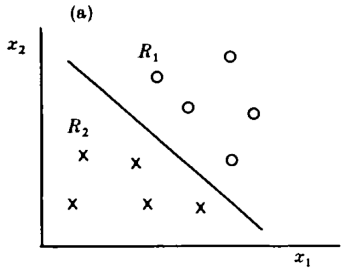


Figure 1: Complete Separation (Albert & Anderson, 1984)

Impact of Sample Size on Complete Separation

Complete separation and sample size:

- More likely to occur with **smaller sample sizes**.
- Evident in our trial example

Simulation results (10,000 simulations):

- **Sample Size = 49:** Likelihood of complete separation: **18.02%**.
- **Sample Size = 98:** Likelihood decreases to **3.36%**.
- **Sample Size = 490: No cases** of complete separation observed.

Conclusion:

- Complete separation is **much more common** in smaller samples.
- Becomes **exceedingly rare** as sample size grows.

Outline

Population-based Inference

- Assumes trial sample is drawn from a larger super-population.
- Statistical model describes how data are generated from the population.
- Strong assumptions: independent sampling, parametric distributional assumptions.

  $\rightarrow$ Violation of assumptions can lead to biased estimates and incorrect inference.
- Reminder: In generalized MCP-Mod we evaluate the null hypothesis:

$$H_0 : \mu_0 = \mu_1 = \ldots = \mu_{k-1},$$

  where $\mu_0, ..., \mu_{k-1}$ are the dose response means.

Randomization-based Inference

- Focuses on the finite trial sample, no assumptions about a larger population.
- Tests strong null hypothesis:

$$H_0 : y_{i,d_0} = \ldots = y_{i,d_{k-1}} \quad \text{for all } i \in \{1, \ldots, n\},$$

  where $y_{i,d_0}, ..., y_{i,d_{k-1}}$ are potential outcomes of patient i.
- Does not rely on distributional assumptions for the outcome, only on randomization.
- P-value can be approximated via Monte Carlo sampling.
- Read more: Imbens & Rubin (2015) & Rosenberger & Lachin (2016)

Population-based vs. Randomization-based

$$H_0 : \mu_0 = \mu_1 = \ldots = \mu_{k-1}$$

$$\nRightarrow$$

$$H_0 : y_{i,d_0} = \ldots = y_{i,d_{k-1}}$$

Randomization-based Inference for MCP-Mod (1/2)

Proposing two randomization-based MCP-Mod test statistics:

Test statistic 1: Generalized MCP-Mod

- Fit a GLM with linear predictor (incl. covariates):

$$\eta_i = \mathbf{d}(\mathbf{Z})_i'\boldsymbol{\delta} + \mathbf{x}_i'\boldsymbol{\beta}$$

  where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{k-1})'$ contains treatment differences versus placebo and $\mathbf{d}(\mathbf{Z})_i'$ is the treatment assignment vector for patient i

- Estimate group means $\widehat{\boldsymbol{\mu}}_{\mathbf{Z}}$ and covariance matrix $\mathbf{S}_{\mathbf{Z}}$.
- Compute test statistic:

$$S_1(\mathbf{Z}) = \max_{m=1,\ldots,M} T_{1m}(\mathbf{Z}), \quad \text{where} \quad T_{1m}(\mathbf{Z}) = \frac{\mathbf{c}_m'\widehat{\boldsymbol{\mu}}_{\mathbf{Z}}}{\sqrt{\mathbf{c}_m'\mathbf{S}_{\mathbf{Z}}\mathbf{c}_m}},$$

  where $\mathbf{c}_1, \ldots, \mathbf{c}_M$ are optimal contrasts for the M candidate model shapes.

Randomization-based Inference for MCP-Mod (2/2)

Motivation: Avoid fitting the GLM for each Monte Carlo sample (Parhat et al., 2014).

Test statistic 2: Residual-based inference

1. Fit a GLM with linear predictor: $\eta_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta}$
2. Compute residuals: $r_i = y_i - g(\hat{\eta}_i) = y_i - g(\hat{\alpha} + \mathbf{x}_i'\widehat{\boldsymbol{\beta}})$
3. Residuals $r_1, \ldots, r_n$ contain information about treatment differences. Calculate residual group means and variances:

$$\bar{\boldsymbol{r}}_{\mathbf{Z}} = (\bar{r}_0, \ldots, \bar{r}_{k-1}), \quad \boldsymbol{s}^2_{\mathbf{Z}} = (s^2_0, \ldots, s^2_{k-1})$$

4. Compute:

$$S_2(\mathbf{Z}) = \max_{m=1,\ldots,M} T_{2m}(\mathbf{Z}), \quad \text{where} \quad T_{2m}(\mathbf{Z}) = \frac{\mathbf{c}_m'\bar{\boldsymbol{r}}_{\mathbf{Z}}}{\sqrt{\sum_{j=0}^{k-1} c^2_{m,j}\frac{s^2_j}{n_j}}}$$

Monte Carlo Randomization Test

Both test statistics S_1 and S_2 can be used in a Monte Carlo randomization test.

Procedure:

1. Calculate test statistic $S_i(\mathbf{Z}_{\text{obs}})$ for observed randomization sequence.
2. Repeatedly sample treatment assignments $\mathbf{Z}_l = (Z_{l,1}, \ldots, Z_{l,n})$ for $l = 1, \ldots, n_{\text{rand}}$.
3. Samples are drawn from the reference set $\mathcal{R}_Z$ using the randomization probability distribution $\mathcal{P}(\mathbf{Z})$, where $\mathcal{R}_Z$ contains all possible treatment assignments under the employed randomization procedure.
4. Calculating test statistics: For each re-randomization $\mathbf{Z}_l$, compute $S_i(\mathbf{Z}_l)$ for $i = 1, 2$.

Computing the p-value:

- The one-sided p-value is estimated as:

$$\hat{p} = \frac{\sum_{l=1}^{n_{\text{rand}}} \mathbf{1}\left(S_i(\mathbf{Z}_l) \geq S_i(\mathbf{Z}_{\text{obs}})\right)}{n_{\text{rand}}},$$

  where $S_i(\mathbf{Z}_{\text{obs}})$ is the test statistic for the observed assignment.

Outline

Reference Set $\mathcal{R}_Z$

The reference set $\mathcal{R}_Z$ is determined by **randomization procedure**. We analysed:

1. Complete Randomization (CR)
   - Each patient is randomized individually according to pre-specified probability distribution.
   - No predictability of treatment assignments but imbalance between groups likely.
2. Random Allocation Rule (RA)
   - Urn design: Drawing treatment allocation without replacement.
   - Ensuring exact allocation to each treatment but some predictability.
3. Permuted Block Design (PBD)
   - Repeated RA in blocks to be more robust to time trends.
   - Ensuring exact allocation to each treatment but even more predictability.

Consider our clinical trial example with $n = 49$. For PBD we can have 7 blocks of length 7, each with $1:2:2:2$ allocation ratio. The total number of randomization sequences:

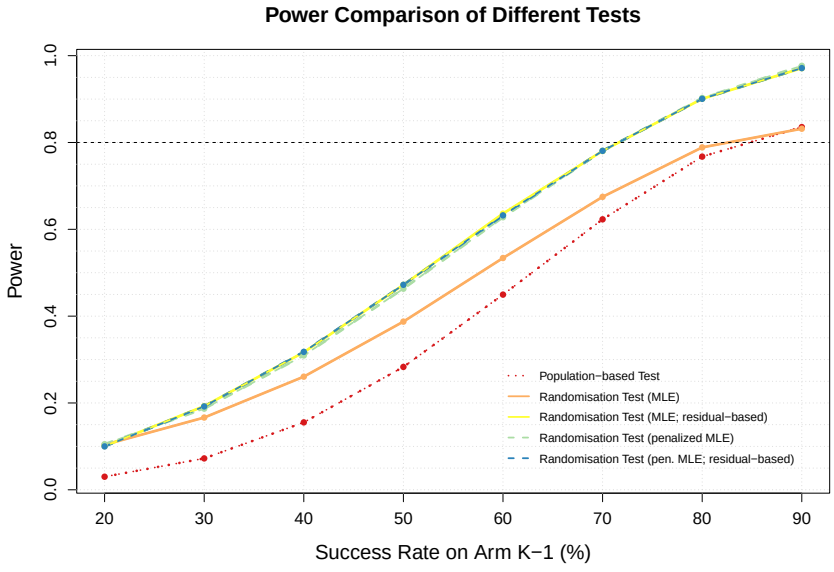
- for PBD is $\binom{7}{1\,2\,2\,2}^7 = 630^7 \approx 3.94 \cdot 10^{19}$
- then for RA is $\binom{49}{7\,14\,14\,14} = \frac{49!}{7!14!14!14!} \approx 1.82 \cdot 10^{26}$ and
- for CR it is $7^{49} \approx 2.57 \cdot 10^{41}$

Outline

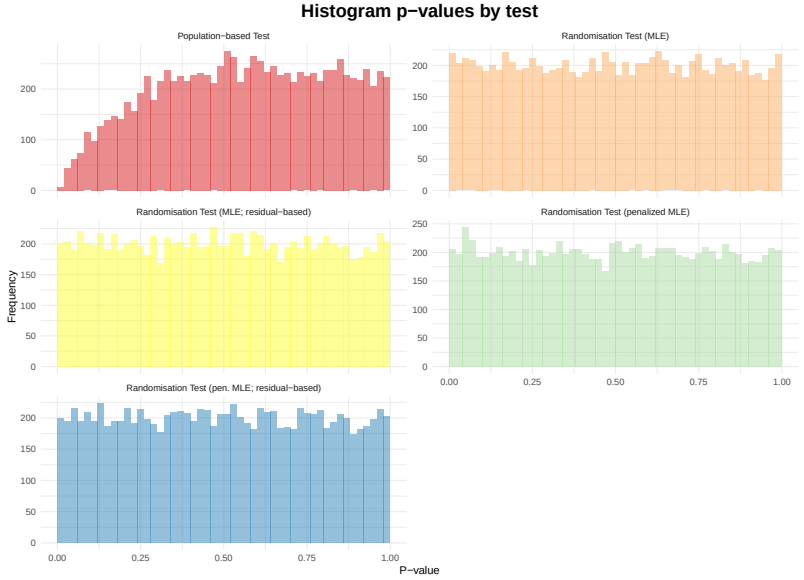
Reminder: Trial Example

- Randomized, double-blind trial Phase-II trial
- Objective: Assess efficacy and safety across 3 dose levels (10 mg, 25 mg, 100 mg) vs. placebo (0 mg).
- Assumed response rates: Placebo: 20% Highest dose (100 mg): 80%
- Total of $n = 49$ patients randomized 1:2:2:2 ratio (placebo vs. active doses).
- Binary outcome: 1 = Success, 0 = Failure.
- 80% power to detect dose-response signal at one-sided 10% significance level.
- PBD with block size 7 is used and a covariate is included

Power Comparison



Distribution of p-values under the Null



Findings Original Scenario

- **Randomization tests (Residual-based and MCP-Mod with penalized MLE):**
  - Achieve higher power than population-based test.
  - Exhausts significance level.
- **Population-based test:**
  - Underperforms due to issues with complete separation.
  - Does not exhaust significance level. (Mainly because complete seperation occurs in 18% of cases with $n = 49$ and null hypothesis cant be rejected in those cases.)

13 Other Scenarios

Sample Size	Randomization Method	Time Trend
49	{RA, **PBD**}	{**No**, Yes}
98	{RA, PBD}	{No, Yes}
490	{RA, PBD, CR}	{No, Yes}

Table 1: The 14 data generating scenarios.

- Outcomes are simulate under Emax model from the candidate set.
- Linear time trend $\gamma_i^t = \max(0, \min(\gamma_i + t_i, 1))$ where $t_i = 0.4 \cdot i/n - 0.2$ for $i = 1, .., n$.
- One covariate always included in analysis.

Findings Additional Scenarios

Impact of time trends:

- **Type-I error rate:**
  - Population-based test: deflated for all sample sizes under PBD when time trends are present.
  - Randomization tests always maintain control per definition.

Impact of randomization procedures:

- **Power comparison:**
  - Randomization test: higher power with PBD compared to Random Allocation (RA) and Complete Randomization (CR). CR shows slightly lower power than RA.
- **Sample size effects:**
  - With larger n ($n = 98, 490$), randomization tests exhaust significance level even more exact.
  - Power differences between all tests diminish with increasing n.

Computational efficiency of randomization tests:

- Residual-based approach over 14 times faster.
- **Conclusion:** residual-based tests preferred due to efficiency and robustness.

Findings Confirmed: in simulation study with continuous endpoint in pharmacometric setting

Outline

Conclusions

Penalized MLE:

- **Complete separation issue** in logistic regression, especially with small sample sizes.
  $\rightarrow$ Leads to non-existent MLEs.
- Using penalized MLE effectively addresses complete separation.
- Particularly significant in **small to medium-sized samples**.

Randomization test:

- Randomization-based tests showed **improved power** in certain scenarios and control & **exhaust type-I error** per definition.
- Enhanced power consistent across various scenarios, including with and without covariates / time trends.

Residual-based vs. standard MCP-Mod

- **Residual-based randomization tests** using test performed as well as or better than the one using the standard MCP-Mod test statistic.
- Demonstrated **computational advantages** (more efficient).

Future Research

- **Explore additional randomization procedures:**
  - Investigate methods beyond complete randomization, random allocation, and permuted block design.
  - Consider complex designs like **covariate-adaptive** and **response-adaptive randomization**.
- **Examine different test statistics:**
  - Evaluate alternative test statistics beyond those motivated by MCP-Mod.
  - Aim to improve performance under specific conditions.
- **Broaden applications in complex trial designs:**
  - Address challenges in complex designs, high-dimensional data, unbalanced groups.

Paper avaliable (Statistics in Medicine):

References I

Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10.

Bretz, F., Pinheiro, J. C. and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61** 738–748.

Buatois, S., Ueckert, S., Frey, N., Retout, S. and Mentré, F. (2021). cLRT-Mod: An efficient methodology for pharmacometric model-based analysis of longitudinal phase II dose finding studies under model uncertainty. *Statistics in Medicine* **40** 2435–2451.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine* **21** 2409–2419.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Parhat, P., Rosenberger, W. F. and Diao, G. (2014). Conditional Monte Carlo randomization tests for regression models. *Statistics in Medicine* **33** 3078–3088.

Pinheiro, J. C., Bornkamp, B., Glimm, E. and Bretz, F. (2014). Model-based dose finding under model uncertainty using general parametric models. *Statistics in Medicine* **33** 1646–1661.

Rosenberger, W. F. and Lachin, J. M. (2016). *Randomization in clinical trials: theory and practice*. John Wiley Sons.

Questions?

Thank you to everyone for lisiting!

lukas.pin@mrc-bsu.cam.ac.uk

Questions?