



# Efficient tests to demonstrate the similarity of dose response curves with regard to small sample sizes

Kathrin Möllenhoff, Ruhr-Universität Bochum  
Holger Dette, Ruhr-Universität Bochum  
Frank Bretz, Novartis Basel



Birmingham, August 2016

FP7 HEALTH 2013 - 602552

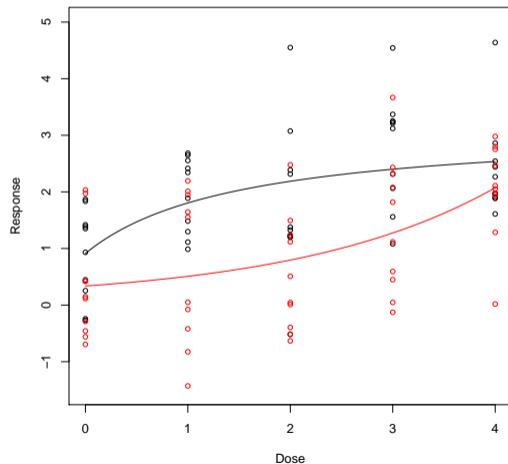
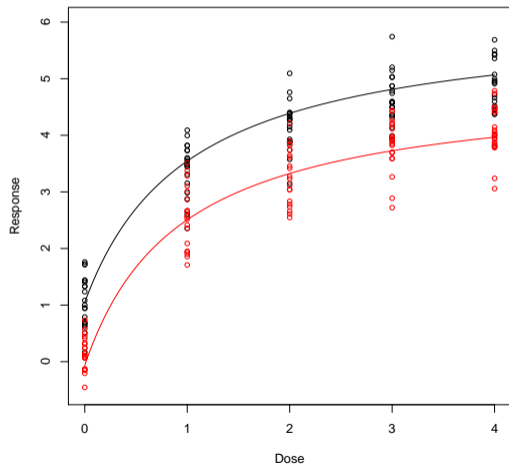
# Outline

- 1 Motivation
- 2 Similarity of curves
- 3 Efficient tests for the similarity of curves
- 4 Finite sample properties
- 5 A Case Study- the IBS data set
- 6 Conclusions & Extensions

# Motivation

- populations of different geographic regions may bear differences in dose response and/or safety issues
  - goal: ability of extrapolation of study results
  - ICH E5 provides a guideline for conducting a bridging study which claims similarity in dose response, efficacy and safety
  - demonstrating similarity of curves becomes an issue
- another application of the new tests: comparison of different drugs containing the same active substance in order to claim bioequivalence. Traditionally approaches as the observation of AUC or Cmax may be misleading
  - goal: developing a test which takes the whole curve into account

# Comparing curves - The setting



## Comparing curves - The setting II

- Two dose response curves (from two samples)

$$Y_{\ell,i,j} = m_{\ell}(x_{\ell,i}, \vartheta_{\ell}) + \varepsilon_{\ell,i,j} ; \quad \ell = 1, 2;$$
$$j = 1, \dots, n_{\ell,i} ; \quad i = 1, \dots, k_{\ell},$$

- $\varepsilon_{\ell,i,j}$  independent  $\sim \mathcal{N}(0, \sigma_{\ell}^2)$  ( $\ell = 1, 2$ )
- $x_{\ell,i} \in \mathcal{X}$  (dose levels)
- $\vartheta_{\ell} \in \mathbb{R}^{d_{\ell}}$  (parameter in model  $m_{\ell}$ )

**Problem:** Are the dose response curves  $m_1$  and  $m_2$  similar?

## Comparing curves - The setting II

- Two dose response curves (from two samples)

$$Y_{\ell,i,j} = m_{\ell}(x_{\ell,i}, \vartheta_{\ell}) + \varepsilon_{\ell,i,j} ; \quad \ell = 1, 2;$$
$$j = 1, \dots, n_{\ell,i} ; \quad i = 1, \dots, k_{\ell},$$

- $\varepsilon_{\ell,i,j}$  independent  $\sim \mathcal{N}(0, \sigma_{\ell}^2)$  ( $\ell = 1, 2$ )
- $x_{\ell,i} \in \mathcal{X}$  (dose levels)
- $\vartheta_{\ell} \in \mathbb{R}^{d_{\ell}}$  (parameter in model  $m_{\ell}$ )

**Problem:** Are the dose response curves  $m_1$  and  $m_2$  **similar**?

# Measures of similarity

- $d$ : a *metric* measuring the distance between  $m_1$  and  $m_2$ .
- Hypothesis of similarity:

$$H_0 : d(m_1, m_2) \geq \epsilon \quad \text{versus} \quad H_1 : d(m_1, m_2) < \epsilon$$

(here  $\epsilon$  is a pre-specified constant).

- Examples

- ▶ maximum absolute difference

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)|$$

- ▶ squared  $L_2$ -distance

$$d_2(m_1, m_2) = \int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx$$

# Measures of similarity

- $d$ : a *metric* measuring the distance between  $m_1$  and  $m_2$ .
- Hypothesis of similarity:

$$H_0 : d(m_1, m_2) \geq \epsilon \quad \text{versus} \quad H_1 : d(m_1, m_2) < \epsilon$$

(here  $\epsilon$  is a pre-specified constant).

- Examples

- ▶ maximum absolute difference

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)|$$

- ▶ squared  $L_2$ -distance

$$d_2(m_1, m_2) = \int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx$$



## Efficient tests for similarity

## Tests based on distances

- calculate (parametric) estimates and therefore the distance of  $m_1$  and  $m_2$ :

$$\hat{m}_1 = m_1(\cdot, \hat{\vartheta}_1), \quad \hat{m}_2 = m_2(\cdot, \hat{\vartheta}_2), \quad \hat{d} = d(\hat{m}_1, \hat{m}_2)$$

- Decide for similarity, i.e. reject the hypothesis

$$H_0 : d(m_1, m_2) \geq \epsilon$$

whenever

$$\hat{d} = d(\hat{m}_1, \hat{m}_2) < q$$

- **Problem:** how do we find the critical value  $q$ ?

# Estimating the distance between two curves- the asymptotic approach

- Empirical process theory :  $d_2(\hat{m}_1, \hat{m}_2) - d_2(m_1, m_2) \stackrel{a}{\sim} \mathcal{N}\left(0, \frac{\tau^2}{n_1+n_2}\right)$ , where

$$\tau^2 = \int_{\mathcal{X} \times \mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2)) \cdot (m_1(y, \vartheta_1) - m_2(y, \vartheta_2)) k(x, y) dx dy$$

- If the true absolute difference curve has only one extremal point, we have

$$d_\infty(\hat{m}_1, \hat{m}_2) - d_\infty(m_1, m_2) \stackrel{a}{\sim} \mathcal{N}\left(0, \frac{\sigma_{d_\infty}^2}{n_1 + n_2}\right),$$

where  $\sigma_{d_\infty}^2$  depends on the location of the extremal point

- **It is very difficult to estimate  $\tau^2$  and  $\sigma_{d_\infty}$ , especially for small sample sizes**

# Bootstrap test for similarity

Generate data under the **null hypothesis**

- Estimate  $\vartheta_1, \vartheta_2$  under the restriction of the null hypothesis, that is the estimates fulfill  $d(m_1, m_2) = \epsilon$   
 $\rightsquigarrow$  **constrained** estimates  $\hat{\vartheta}_1, \hat{\vartheta}_2$   
e.g. for  $d = d_\infty$ ,  $\hat{\vartheta}_1, \hat{\vartheta}_2$  satisfy  $\sup_{x \in \mathcal{X}} |m_1(x, \hat{\vartheta}_1) - m_2(x, \hat{\vartheta}_2)| = \epsilon_\infty$
- Generate bootstrap data (parametric bootstrap)

$$Y_{\ell, i, j}^* = m_\ell(x_{\ell, i}, \hat{\vartheta}_\ell) + \hat{\sigma}_\ell \varepsilon_{\ell, i, j}^*$$

where  $\varepsilon_{\ell, i, j}^*$  are i.i.d.  $\sim \mathcal{N}(0, 1)$

# Bootstrap test for similarity

Generate data under the **null hypothesis**

- Estimate  $\vartheta_1, \vartheta_2$  under the restriction of the null hypothesis, that is the estimates fulfill  $d(m_1, m_2) = \epsilon$   
 $\rightsquigarrow$  **constrained** estimates  $\hat{\vartheta}_1, \hat{\vartheta}_2$   
e.g. for  $d = d_\infty$ ,  $\hat{\vartheta}_1, \hat{\vartheta}_2$  satisfy  $\sup_{x \in \mathcal{X}} |m_1(x, \hat{\vartheta}_1) - m_2(x, \hat{\vartheta}_2)| = \epsilon_\infty$
- Generate bootstrap data (parametric bootstrap)

$$Y_{\ell, i, j}^* = m_\ell(x_{\ell, i}, \hat{\vartheta}_\ell) + \hat{\sigma}_\ell \varepsilon_{\ell, i, j}^*$$

where  $\varepsilon_{\ell, i, j}^*$  are i.i.d.  $\sim \mathcal{N}(0, 1)$

## Bootstrap test for similarity II

- Calculate  $\hat{\vartheta}_1^*, \hat{\vartheta}_2^*$  from the bootstrap data  $Y_{\ell, i, j}^*$
- Calculate  $\hat{d}^* = d(\hat{m}_1^*, \hat{m}_2^*)$ , for  $d = d_\infty$  that is  $\hat{\mathbf{d}}_\infty^* = \sup_{x \in \mathcal{X}} |m_1(x, \hat{\vartheta}_1^*) - m_2(x, \hat{\vartheta}_2^*)|$
- Repeat this procedure B times
- If  $\hat{\mathbf{d}}^{*(1)}, \dots, \hat{\mathbf{d}}^{*(B)}$  denote the ordered bootstrap replicates the hypothesis

$$H_0 : d(m_1, m_2) \geq \epsilon$$

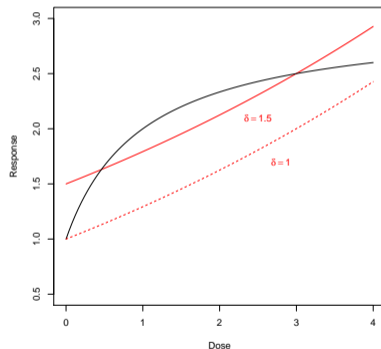
is rejected, whenever

$$\hat{\mathbf{d}} < \hat{\mathbf{d}}^{*(\lfloor B\alpha \rfloor)}$$

- Theoretical properties:
  - ▶ the bootstrap test has asymptotic level  $\alpha$
  - ▶ the bootstrap test is consistent

## Finite sample properties

- $m_1(x, \beta_1) = \beta_{11} + \frac{\beta_{12}x}{\beta_{13}+x}$ ,  $m_2(x, \beta_2) = \beta_{21} + \beta_{22} \cdot (\exp(\frac{x}{\beta_{23}}) - 1)$ ,  $\mathcal{X} = [0, 4]$ ,  
where  $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13}) = (1, 2, 1)$  and  $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23}) = (\delta, 2.2, 8)$



- $H_0 : d_\infty(m_1, m_2) \geq 1$  versus  $H_1 : d_\infty(m_1, m_2) < 1$

## The $d_\infty$ bootstrap test- Simulation of the nominal level

|              |          |            |              | $\alpha = 0.05$            |              | $\alpha = 0.1$             |  |
|--------------|----------|------------|--------------|----------------------------|--------------|----------------------------|--|
|              |          |            |              | $(\sigma_1^2, \sigma_2^2)$ |              | $(\sigma_1^2, \sigma_2^2)$ |  |
| $(n_1, n_2)$ | $\delta$ | $d_\infty$ | (0.25, 0.25) | (0.5, 0.5)                 | (0.25, 0.25) | (0.5, 0.5)                 |  |
| (10, 10)     | 0.25     | 1.5        | 0.001        | 0.001                      | 0.000        | 0.004                      |  |
| (10, 10)     | 0.5      | 1.25       | 0.005        | 0.011                      | 0.013        | 0.030                      |  |
| (10, 10)     | 0.75     | 1          | 0.045        | 0.037                      | 0.102        | 0.086                      |  |
| (20, 20)     | 0.25     | 1.5        | 0.000        | 0.000                      | 0.000        | 0.000                      |  |
| (20, 20)     | 0.5      | 1.25       | 0.001        | 0.004                      | 0.006        | 0.018                      |  |
| (20, 20)     | 0.75     | 1          | 0.034        | 0.038                      | 0.091        | 0.100                      |  |
| (50, 50)     | 0.25     | 1.5        | 0.000        | 0.000                      | 0.000        | 0.000                      |  |
| (50, 50)     | 0.5      | 1.25       | 0.000        | 0.000                      | 0.002        | 0.001                      |  |
| (50, 50)     | 0.75     | 1          | 0.051        | 0.059                      | 0.096        | 0.108                      |  |

Table: Simulated level of the bootstrap test for the equivalence of an EMAX and an exponential model.



## The $d_\infty$ bootstrap test- Simulation of power

|              |          |            |              | $\alpha = 0.05$            |              | $\alpha = 0.1$             |  |
|--------------|----------|------------|--------------|----------------------------|--------------|----------------------------|--|
|              |          |            |              | $(\sigma_1^2, \sigma_2^2)$ |              | $(\sigma_1^2, \sigma_2^2)$ |  |
| $(n_1, n_2)$ | $\delta$ | $d_\infty$ | (0.25, 0.25) | (0.5, 0.5)                 | (0.25, 0.25) | (0.5, 0.5)                 |  |
| (10, 10)     | 1        | 0.75       | 0.160        | 0.093                      | 0.297        | 0.225                      |  |
| (10, 10)     | 1.5      | 0.5        | 0.237        | 0.133                      | 0.383        | 0.231                      |  |
| (20, 20)     | 1        | 0.75       | 0.214        | 0.138                      | 0.393        | 0.271                      |  |
| (20, 20)     | 1.5      | 0.5        | 0.401        | 0.229                      | 0.604        | 0.398                      |  |
| (50, 50)     | 1        | 0.75       | 0.504        | 0.274                      | 0.662        | 0.416                      |  |
| (50, 50)     | 1.5      | 0.5        | 0.777        | 0.491                      | 0.877        | 0.648                      |  |
| (100, 100)   | 1        | 0.75       | 0.798        | 0.522                      | 0.878        | 0.667                      |  |
| (100, 100)   | 1.5      | 0.5        | 0.951        | 0.767                      | 0.988        | 0.877                      |  |

**Table:** Simulated power of the bootstrap test for the equivalence of an EMAX and an exponential model

## A Case Study- the IBS data set

- Biesheuvel, E. and Hothorn, L. A. : Female and male patients with Irritable Bowel Syndrome (IBS),  $n = 369$
- randomized to one of the five doses 0 (placebo), 1, 2, 3, and 4
- larger values corresponding to a better treatment effect
- fitted models
  - ▶ male:  $m_1(x, \beta_1) = 0.398 + 0.043x$
  - ▶ female:  $m_2(x, \beta_2) = 0.220 + 0.517 \frac{x}{1.396+x}$
- maximum distance: 0.1784

## A Case Study- the IBS data set II

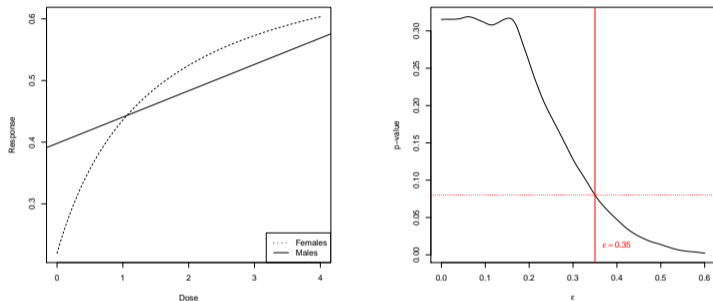


Figure: Left panel: Fitted dose response curves for male (linear model) and female (Emax model) patients. Right panel: p-values of the  $d_{\infty}$ -bootstrap test for different values of the threshold  $\epsilon_{\infty}$ .

- the p-value corresponding to the choice  $\epsilon_{\infty} = 0.35$  is given by 0.078

# Conclusions & Extensions I

- Two powerful tests have been proposed by
  - ▶ estimating the distance and the variance of the test statistic directly (asymptotic approach)
  - ▶ generating quantiles by a parametric bootstrap
- the bootstrap approach can be applied to any metric without deriving the asymptotic distribution
- the bootstrap  $d_\infty$ -test was implemented in the R package `TestingSimilarity`
- we investigated the performance of all tests under the assumption of model misspecification and observed a robust performance
- the proposed tests were adapted to
  - ▶ the case of multiple curve comparison
  - ▶ dependent data
  - ▶ models with common parameters
- especially the adaption to dependent data offers a variety of new applications, e.g. the demonstration of bioequivalence

# Conclusions & Extensions I

- Two powerful tests have been proposed by
  - ▶ estimating the distance and the variance of the test statistic directly (asymptotic approach)
  - ▶ generating quantiles by a parametric bootstrap
- the bootstrap approach can be applied to any metric without deriving the asymptotic distribution
- the bootstrap  $d_\infty$ -test was implemented in the R package `TestingSimilarity`
- we investigated the performance of all tests under the assumption of model misspecification and observed a robust performance
- the proposed tests were adapted to
  - ▶ the case of multiple curve comparison
  - ▶ dependent data
  - ▶ models with common parameters
- especially the adaption to dependent data offers a variety of new applications, e.g. the demonstration of bioequivalence

# Conclusions & Extensions I

- Two powerful tests have been proposed by
  - ▶ estimating the distance and the variance of the test statistic directly (asymptotic approach)
  - ▶ generating quantiles by a parametric bootstrap
- the bootstrap approach can be applied to any metric without deriving the asymptotic distribution
- the bootstrap  $d_\infty$ -test was implemented in the R package `TestingSimilarity`
- we investigated the performance of all tests under the assumption of model misspecification and observed a robust performance
- the proposed tests were adapted to
  - ▶ the case of multiple curve comparison
  - ▶ dependent data
  - ▶ models with common parameters
- especially the adaption to dependent data offers a variety of new applications, e.g. the demonstration of bioequivalence

# Conclusions & Extensions I

- Two powerful tests have been proposed by
  - ▶ estimating the distance and the variance of the test statistic directly (asymptotic approach)
  - ▶ generating quantiles by a parametric bootstrap
- the bootstrap approach can be applied to any metric without deriving the asymptotic distribution
- the bootstrap  $d_\infty$ -test was implemented in the R package `TestingSimilarity`
- we investigated the performance of all tests under the assumption of model misspecification and observed a robust performance
- the proposed tests were adapted to
  - ▶ the case of multiple curve comparison
  - ▶ dependent data
  - ▶ models with common parameters
- especially the adaption to dependent data offers a variety of new applications, e.g. the demonstration of bioequivalence

## Conclusions & Extensions II

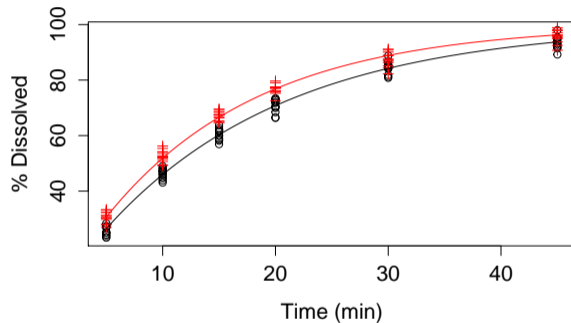


Figure: Comparison of concentration profiles.



## Acknowledgments

# Thank you very much!

The research is embedded in the IDeAI EU-FP7 project, Grant-Agreement No. 602 552

### References:

- Biesheuvel, E. and Hothorn, L. A. (2002). Many-to-one comparisons in stratified designs, *Biometrical Journal*, 44, 101–116
- Bjoern Bornkamp, Jose Pinheiro and Frank Bretz (2010) *DoseFinding: Planning and Analyzing Dose Finding experiments*. R package version 0.1-1. Available at <http://cran.r-project.org/web/packages/DoseFinding/index.html>
- Moellenhoff, K. (2015) *TestingSimilarity: Bootstrap Test for Similarity of Dose Response Curves Concerning the Maximum Absolute Deviation*. R package version 1.0, available at <http://cran.r-project.org/web/packages/TestingSimilarity/index.html>