

Equivalence of regression curves sharing common parameters

Kathrin Möllenhoff¹, Frank Bretz², Holger Dette¹

¹ Department of Mathematics, Ruhr-Universität Bochum, Germany,

²Novartis Pharma AG, CH-4002 Basel, Switzerland

February 6, 2019

Abstract

In clinical trials the comparison of two different populations is a frequently addressed problem. Non-linear (parametric) regression models are commonly used to describe the relationship between covariates as the dose and a response variable in the two groups. In some situations it is reasonable to assume some model parameters to be the same, for instance the placebo effect or the maximum treatment effect. In this paper we develop a (parametric) bootstrap test to establish the similarity of two regression curves sharing some common parameters. We show by theoretical arguments and by means of a simulation study that the new test controls its level and achieves a reasonable power. Moreover, it is demonstrated that under the assumption of common parameters a considerable more powerful test can be constructed compared to the test which does not use this assumption. Finally, we illustrate potential applications of the new methodology by a clinical trial example.

Keywords and Phrases: Similarity of regression curves, equivalence testing, parametric bootstrap, nonlinear regression, dose finding studies

1 Introduction

Regression models are commonly used to describe the relationship between multiple covariates and a response variable. In certain applications, more than one regression model is available, such as when assessing the relationship between the covariates and the response variables in more than one population (e.g. in males and females). It is then often of interest to demonstrate the equivalence of the regression curves: If equivalence can be claimed, conclusions can be drawn from the pooled sample and a single regression model is sufficient to describe the data. This can be achieved by testing a suitable null hypothesis that the distance between the regression curves (measured in an appropriate sense) is smaller than

a pre-specified equivalence margin at a controlled Type I error rate. Note that the problem of equivalence testing, as considered in this paper, is conceptually different from the more frequent problem of testing for equality of curves and is much less studied in the literature due to methodological difficulties.

The problem of testing for equality of regression models has been intensively discussed in the nonparametric context and we refer to the recent work of Feng et al. (2015), which contains a rather comprehensive list of references. In applied regression analysis, however, parametric models are usually preferred to a purely nonparametric approach as they admit a direct interpretation of the observed effects in terms of the model parameters. In addition, the available information of the observations is increased by applying more efficient estimation or test procedures, provided that the assumed model is valid. Despite its importance, the problem of establishing equivalence of two parametric regression models while controlling the Type I error rate has only recently found attention in the literature. Using the intersection-union test device from Berger (1982), Liu et al. (2009) investigated the assessment of non-superiority, non-inferiority and equivalence when comparing two regression models over a restricted covariate region. Building upon this work, Gsteiger et al. (2011) derived equivalence tests based on simultaneous confidence bands for nonlinear regression models, with application to population pharmacokinetic analyses. Likewise, Bretz et al. (2016) assessed the similarity of dose response curves in two non-overlapping subgroups of patients. Alternatively, Dette et al. (2018) suggested directly estimating the distance between the regression curves and using a non-standard bootstrap test to decide for equivalence of the two curves if the estimate is less than a certain threshold. Expanding this approach, Moellenhoff et al. (2018) assessed the comparability of drug dissolution profiles via maximum deviation, whereas Hoffelder (2018) demonstrated the equivalence of dissolution profiles using the Mahalanobis distance; see also Collignon et al. (2018).

In these papers, the authors assumed that the regression models have different parameters and can therefore be evaluated separately. In some applications, however, this assumption cannot be justified and it is more reasonable to assume that the regression models may have some common parameters. The total number of parameters to estimate is then reduced to the common and remaining parameters of each model, affecting the asymptotic behavior of the estimators. Consider, for example, the Phase II dose finding trial for a weight loss drug described in Bretz et al. (2016). This trial aimed at comparing the dose response relationship for two regimens administered to patients suffering from overweight or obesity: Three doses each for once daily (o.d.) and twice daily (b.i.d.) use of the medication, and placebo. It is reasonable to assume that the placebo response is the same under both the o.d. and the b.i.d. regimen. Since the regression models typically used for dose response modeling contain a parameter for the placebo response (Pinheiro et al. (2006)), they will

thus share this common parameter for both the o.d. and the b.i.d. regimen. In some instances, it might even be reasonable to assume that the maximum efficacy for high doses is similar in both groups. Moreover, clinical trial sponsors may even decide to use the same placebo group for logistical reasons. The response of each patient on placebo is then used twice in the estimation of the o.d. and b.i.d. dose response models, further complicating the statistical problem.

In this paper, we investigate the equivalence of two parametric regression curves that share common parameters. In Section 2 we first introduce the regression models to be estimated under the assumption of common parameters. We then develop a non-standard bootstrap test which performs the resampling under the constraints of the interval hypotheses implied by the equivalence test problem. The new tests improves the procedure proposed in Dette et al. (2018) using the additional information of common parameter in both groups. We also discuss testing the equivalence of model parameters to assess whether the assumption of common parameters is plausible. In Section 3 we investigate the finite sample properties of the proposed bootstrap test proposed in terms of power and size. In Section 4 we illustrate the methods using a multi-regional clinical trial example where it is conceivable that the placebo and maximum treatment responses are the same across geographic regions but the onset of treatment differs due to intrinsic and extrinsic factors (Malinowski et al. (2008); ICH (2017)). Technical details and proofs are deferred to an appendix.

2 Methodology

2.1 Models with common parameters

Let

$$Y_{\ell,i,j} = m_{\ell}(d_{\ell,i}, \beta_{\ell}) + \eta_{\ell,i,j}, \quad j = 1, \dots, n_{\ell,i}, \quad i = 1, \dots, k_{\ell}, \quad (2.1)$$

denote the observed response of the j th subject at the i th dose level $d_{\ell,i}$ under the ℓ th dose response model m_{ℓ} , where $\ell = 1, 2$ denotes the index of the two groups under consideration. We assume that the (non-linear) regression model m_{ℓ} is parametrized through a p_{ℓ} -dimensional vector β_{ℓ} , $\ell = 1, 2$. Note that the regression models m_1 and m_2 may be different. Likewise, the parameters β_1 and β_2 may be different even if $m_1 = m_2$. We further assume that the error terms $\eta_{\ell,i,j}$ are independent and identically distributed with expectation 0 and variance σ_{ℓ}^2 . The dose levels $d_{\ell,i}$ may be different in both groups but they are attained on the same (restricted) covariate region \mathcal{D} . In this paper \mathcal{D} is assumed to be the dose range, although the results can be generalized to include other covariates. Further, $n_{\ell} = \sum_{i=1}^{k_{\ell}} n_{\ell,i}$ denotes the sample size in group ℓ where we assume $n_{\ell,i}$ observations in the

i th dose level ($i = 1, \dots, k_\ell$, $\ell = 1, 2$). The sample sizes n_ℓ can be unequal and the total number of observations is denoted by $n = n_1 + n_2$.

In this paper we consider the situation, where the regression models have some common parameters. More precisely, we assume without loss of generality that these parameters are given by the first p' model parameters of the parameter β_ℓ in model (2.1) that is

$$\beta_\ell = (\beta_0, \tilde{\beta}_\ell) \in \mathbb{R}^{p_\ell}, \quad \ell = 1, 2, \quad (2.2)$$

where $\beta_0 \in \mathbb{R}^{p'}$ denotes the vector of common parameters in both regression models and $\tilde{\beta}_1$ and $\tilde{\beta}_2$ denote the remaining parameters in the models m_1 and m_2 , respectively, which do not necessarily coincide. The case where the models m_1 and m_2 do not share any common parameters is included and corresponds to $\beta_\ell = \tilde{\beta}_\ell$ for $\ell = 1, 2$ (that is $p' = 0$). As a consequence the $p_1 + p_2 - p'$ -dimensional vector of all parameters of the regression functions in model (2.1) under the assumption (2.2) is given by $\beta = (\beta_0, \tilde{\beta}_1, \tilde{\beta}_2)$. Throughout this paper we assume that $\beta \in B$ where $B \subset \mathbb{R}^{p_1 + p_2 - p'}$ is a compact set.

These parameters are now estimated by least squares using the combined sample $\{Y_{\ell,i,j} : j = 1, \dots, n_{\ell,i}, i = 1, \dots, k_\ell, \ell = 1, 2\}$, that is

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg \min_{(b, \tilde{b}_1, \tilde{b}_2) \in B} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(d_{\ell,i}, (b_0, \tilde{b}_\ell)))^2. \quad (2.3)$$

2.2 Testing equivalence of regression curves

Following Liu et al. (2009) and Gsteiger et al. (2011) we consider the regression curves m_1 and m_2 to be equivalent if the maximum distance between the two curves is smaller than a given pre-specified constant, say $\varepsilon > 0$, that is,

$$d_\infty(\beta_1, \beta_2) = \max_{d \in \mathcal{D}} |m_1(d, \beta_1) - m_2(d, \beta_2)| < \varepsilon.$$

In clinical trial practice ε is often referred to as a relevance threshold in the sense that if $d_\infty(\beta_1, \beta_2) < \varepsilon$ the difference between the two curves is believed not to be clinically relevant. In order to establish equivalence of the two curves m_1 and m_2 at a controlled type I error, we will develop a test for the hypotheses

$$H_0 : d_\infty(\beta_1, \beta_2) \geq \varepsilon \quad \text{versus} \quad H_1 : d_\infty(\beta_1, \beta_2) < \varepsilon. \quad (2.4)$$

In the following we extend the bootstrap approach from Dette et al. (2018) to test the hypotheses (2.4) in the situation of common parameters. Note that the test procedure proposed below could also be applied to alternative measures of equivalence, such as the integrated deviation $\int_{\mathcal{D}} |m_1(t, \beta_1) - m_2(t, \beta_2)| dt$.

Algorithm 2.1. (parametric bootstrap for testing equivalence under the assumption of common parameters)

- (1) Calculate the ordinary least-square (OLS) parameter estimate (2.3) assuming a common parameter β_0 . The corresponding variance estimates are given by

$$\hat{\sigma}_\ell^2 = \frac{1}{n_\ell} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(d_{\ell,i}, \hat{\beta}_\ell))^2, \quad \ell = 1, 2, \quad (2.5)$$

where $\hat{\beta}_\ell = (\hat{\beta}_0, \hat{\beta}_\ell)$, $\ell = 1, 2$. Calculate the estimate

$$\hat{d}_\infty = d_\infty(\hat{\beta}_1, \hat{\beta}_2) = \max_{d \in \mathcal{D}} |m_1(d, \hat{\beta}_1) - m_2(d, \hat{\beta}_2)|$$

for the maximal deviation between the two regression curves.

- (2) Define the constrained estimates

$$\hat{\beta}_\ell = \begin{cases} \hat{\beta}_\ell & \text{if } \hat{d}_\infty \geq \varepsilon \\ \bar{\beta}_\ell & \text{if } \hat{d}_\infty < \varepsilon \end{cases}, \quad \ell = 1, 2, \quad (2.6)$$

where $\bar{\beta}_1, \bar{\beta}_2$ minimize the objective function in (2.3) under the additional restriction

$$d_\infty(\beta_1, \beta_2) = \max_{d \in \mathcal{D}} |m_1(d, \beta_1) - m_2(d, \beta_2)| = \varepsilon. \quad (2.7)$$

Define $\hat{d}_\infty = d_\infty(\hat{\beta}_1, \hat{\beta}_2)$ and note that $\hat{d}_\infty \geq \varepsilon$.

The next two steps describe the (parametric) bootstrap procedure.

- (3) Generate data

$$Y_{\ell,i,j}^* = m_\ell(d_{\ell,i}, (\hat{\beta}_0, \hat{\beta}_\ell)) + \eta_{\ell,i,j}^*, \quad i = 1, \dots, n_{\ell,i}, \quad \ell = 1, 2, \quad (2.8)$$

with independent and normally distributed errors $\eta_{\ell,i,j}^* \sim \mathcal{N}(0, \hat{\sigma}_\ell^2)$.

- (4) Calculate the OLS estimate $\hat{\beta}^*$ as in Step (1) and the test statistic

$$\hat{d}_\infty^* = \max_{d \in \mathcal{D}} |m_1(d, \hat{\beta}_1^*) - m_2(d, \hat{\beta}_2^*)|,$$

where $\beta_\ell^* = (\beta_0^*, \tilde{\beta}_\ell^*)$, $\ell = 1, 2$. The α -quantile of the distribution of the distribution of the statistic \hat{d}_∞^* is denoted by q_α^* and the null hypotheses in (2.4) is rejected, whenever

$$\hat{d}_\infty < q_\alpha^*. \quad (2.9)$$

In practice the \hat{q}_α^* can be calculated repeating steps (3) and (4), say B times, in order to obtain replicates $\hat{d}_{\infty,1}^*, \dots, \hat{d}_{\infty,B}^*$ of \hat{d}_∞^* . An estimate of \hat{q}_α^* then is defined by $\hat{q}_\alpha^{(B)} := \hat{d}_\infty^{*(\lfloor B\alpha \rfloor)}$, where $\hat{d}_\infty^{*(1)} \leq \dots \leq \hat{d}_\infty^{*(B)}$ denotes the corresponding order statistic, and this estimate is used in (2.9)

The following theorem states that this algorithm yields a valid test procedure. The proof is left to the Appendix 6.

Theorem 2.1. *The test defined by (2.9) is a consistent, asymptotic α -level test. That is*

$$\lim_{n_1, n_2 \rightarrow \infty} \mathbb{P}(\hat{d}_\infty < \hat{q}_\alpha^*) = 1, \quad (2.10)$$

whenever $d_\infty < \varepsilon$, and

$$\limsup_{n_1, n_2 \rightarrow \infty} \mathbb{P}(\hat{d}_\infty < \hat{q}_\alpha^*) \leq \alpha. \quad (2.11)$$

if $d_\infty \geq \varepsilon$.

Remark 2.2. The results presented in this section remain correct in trials with a common placebo group, where n_0 observations are taken at dose level $d_1 = 0$ (corresponding to placebo), which are modelled by the random variables $Y_{0,1}, \dots, Y_{0,n_0}$. For the sake of a simple presentation we consider location-scale type models, such that the common effect at the placebo can easily be modelled, but we note that more general models can be considered as well introducing additional constraints for the parameter.

To be precise, we assume that the models in (2.1) are given by

$$m_\ell(d, \beta_\ell) = \beta_{0,1} + \tilde{\beta}_{\ell,1} \cdot m_\ell^0(d, \tilde{\beta}_\ell^0), \quad \ell = 1, 2, \quad i = 1, \dots, k_\ell. \quad (2.12)$$

where $m_\ell^0(0, \tilde{\beta}_\ell^0) = 0$ ($\ell = 1, 2$), such that the condition $m_1(0, \beta_\ell) = m_2(0, \beta_\ell) = \beta_{0,1}$ reflects the fact that there is only one placebo group (and as a consequence a common placebo parameter). Models of this type cover the most frequently used functional forms used in drug development and several examples can be found in Ting (2006). Beside the location parameter $\beta_{0,1}$ there may be also other shared parameters, which we do not reflect in our notations for a better readability. The ℓ -th model is completely characterized by its parameter $\beta_\ell = (\beta_0, \tilde{\beta}_\ell) = (\beta_0, \tilde{\beta}_{\ell,1}, \tilde{\beta}_\ell^0)$, $\ell = 1, 2$, and we obtain estimates of the model parameters by minimizing the sum of squares

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg \min_{b \in B} \sum_{j=1}^{n_0} (Y_{0,j} - b_{0,1})^2 + \sum_{\ell=1}^2 \sum_{i=2}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - (b_{0,1} + \tilde{b}_{\ell,1} \cdot m_\ell^0(d_{\ell,i}, \tilde{b}_\ell^0)))^2. \quad (2.13)$$

Theorem 2.1 remains valid in this situation and a proof can be found in the Appendix (see Section 6.4).

2.3 Testing equivalence of model parameters

So far we assumed that the two regression models m_1 and m_2 share the common parameter β_0 . In practice it may be necessary to assess whether this assumption is plausible using an appropriate equivalence test for the shared model parameters. To be more precise, we recall the definition the parameters β_ℓ in model (2.1), i.e.

$$\beta_\ell = (\beta_{\ell,1}, \dots, \beta_{\ell,p'}, \dots, \beta_{\ell,p_\ell}) , \quad \ell = 1, 2.$$

and note that assumption (2.2) of p' common parameters in the models m_1 and m_2 can be represented as $(\beta_{1,1}, \dots, \beta_{1,p'}) = (\beta_{2,1}, \dots, \beta_{2,p'})$ for $\ell = 1, 2$. In order to investigate if this assumption holds at least approximately we construct a test for the hypotheses

$$K_0 : \max_{i=1, \dots, p'} |\beta_{1,i} - \beta_{2,i}| \geq \delta \quad \text{versus} \quad K_1 : \max_{i=1, \dots, p'} |\beta_{1,i} - \beta_{2,i}| < \delta, \quad (2.14)$$

where δ denotes the equivalence margin. To be precise let $\hat{\beta}^{(\ell)}$ denote the least squares estimates in model m_ℓ for the sample $\{Y_{\ell,i,j} : j = 1, \dots, n_{\ell,i}, i = 1, \dots, k_\ell\}$ ($\ell = 1, 2$), and assume that for large sample considerations the sample sizes n_ℓ and $n_{\ell,i}$ converge to infinity such that

$$\lim_{n_\ell \rightarrow \infty} \frac{n_{\ell,i}}{n_\ell} = \zeta_{\ell,i} > 0 , \quad i = 1, \dots, k_\ell , \quad \ell = 1, 2, \quad (2.15)$$

$$\lim_{n_1, n_2 \rightarrow \infty} \frac{n}{n_1} = \lambda \in (1, \infty). \quad (2.16)$$

Under standard assumptions, which are listed in Section 6 it can be shown that the least squares estimate $\hat{\beta}^{(\ell)}$ of the parameter β_ℓ in model m_ℓ is approximately normal distributed, that is

$$\sqrt{n_\ell}(\hat{\beta}^{(\ell)} - \beta_\ell) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\ell^{-1}) , \quad \ell = 1, 2, \quad (2.17)$$

where the symbol $\xrightarrow{\mathcal{D}}$ means convergence in distribution and the matrix Σ_ℓ is defined by

$$\Sigma_\ell = \frac{1}{\sigma_\ell^2} \sum_{i=1}^{k_\ell} \zeta_{\ell,i} \left. \frac{\partial}{\partial b_\ell} m_\ell(d_{\ell,i}, b_\ell) \right|_{b_\ell = \beta_\ell} \left(\left. \frac{\partial}{\partial b_\ell} m_\ell(d_{\ell,i}, b_\ell) \right|_{b_\ell = \beta_\ell} \right)^T , \quad \ell = 1, 2. \quad (2.18)$$

Here and throughout this paper we assume that the matrices Σ_1 and Σ_2 are non-singular. Consequently the difference $\sqrt{n}(\hat{\beta}^{(1)} - \hat{\beta}^{(2)})$ is also asymptotically normal distributed, and in particular it follows for the first p' components of the difference that

$$\sqrt{n}((\hat{\beta}_{1,1}, \dots, \hat{\beta}_{1,p'}) - (\hat{\beta}_{2,1}, \dots, \hat{\beta}_{2,p'}) - ((\beta_{1,1}, \dots, \beta_{1,p'}) - (\beta_{2,1}, \dots, \beta_{2,p'}))) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega), \quad (2.19)$$

where the matrix Ω is defined by

$$\Omega := \lambda \Lambda_1^{-1} + \frac{\lambda}{\lambda-1} \Lambda_2^{-1}, \quad (2.20)$$

$\Lambda_\ell^{-1} = ((\Sigma_\ell^{-1})_{ij})_{i,j=1}^{p'}$ denotes the upper-left $p' \times p'$ -block of the matrix Σ_ℓ^{-1} ($\ell = 1, 2$) and λ is defined in (2.16). Therefore we obtain the approximation

$$(\hat{\beta}_{1,1}, \dots, \hat{\beta}_{1,p'}) - (\hat{\beta}_{2,1}, \dots, \hat{\beta}_{2,p'}) \stackrel{\mathcal{D}}{\approx} \mathcal{N}((\beta_{1,1}, \dots, \beta_{1,p'}) - (\beta_{2,1}, \dots, \beta_{2,p'}), \frac{1}{n} \Omega),$$

where Ω is defined in (2.20). We can now apply the test (2.2) proposed in Wang et al. (1999) by rejecting the null hypothesis K_0 in (2.14), whenever

$$|\hat{\beta}_{1,i} - \hat{\beta}_{2,i}| < \delta - t_{1-\alpha, n-2} \left(\frac{\hat{\Omega}_{ii}}{n(n-2)} \right)^{1/2} \text{ for all } i = 1, \dots, p', \quad (2.21)$$

where $t_{1-\alpha, n-2}$ denotes the $1 - \alpha$ quantile of the t -distribution with $n - 2$ degrees of freedom and $\hat{\Omega}_{ii}$ the i th diagonal element of the matrix $\hat{\Omega}$ which is an estimate for the (unknown) covariance matrix Ω (this is obtained by replacing the unknown parameters β_ℓ , σ_ℓ^2 and weights $\zeta_{\ell,i}$ in (2.18) by their corresponding estimates and $n_{\ell,i}/n_\ell$, respectively).

3 Finite sample properties

We now investigate the finite sample properties of the bootstrap test proposed in Section 2.2 in terms of power and size using numerical simulations. The data is generated as follows:

- (a) We choose the functional form of the models m_1, m_2 and specify their parameters β_1, β_2 (including a common parameter β_0), which determine the true underlying models. Further we choose variances σ_ℓ^2 and the actual dose levels $d_{\ell,i}$, $\ell = 1, 2$.
- (b) For each dose $d_{\ell,i}$ we calculate $n_{\ell,i}$ values for the response given by $m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))$. By generating residual errors $\eta_{\ell,i,j} \sim N(0, \sigma_\ell^2)$ we obtain the final response data

$$Y_{\ell,i,j} = m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell)) + \eta_{\ell,i,j}, \quad j = 1, \dots, n_{\ell,i}, \quad i = 1, \dots, k_\ell, \quad \ell = 1, 2. \quad (3.1)$$

The simulation results below were obtained using 1'000 simulation runs, where $B = 500$ bootstrap replications were used to calculate quantiles of the bootstrap test.

In the following, we report the simulations results for power and size under three different scenarios. We consider the four-parameter sigmoid Emax model

$$m(d, \beta) = \beta_1 + \frac{\beta_2 d^{\beta_3}}{\beta_4^{\beta_3} + d^{\beta_3}}, \quad (3.2)$$

which is frequently used in practice when modeling dose response relationships (see for example Gabrielsson and Weiner (2007) or Thomas et al. (2014)). In model (3.2) the parameter $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ corresponds (in this order) to the placebo effect E_0 , the maximum effect E_{max} , the Hill parameter h determining the steepness of the dose-response curve and the dose ED_{50} producing half of the maximum effect (Macdougall (2006)). In what follows we add an index $\ell = 0$ for a shared parameter or $\ell = 1, 2$ for the group under consideration.

Scenario 1: We assume the dose range $\mathcal{D} = [0, 4]$ with identical dose levels $d_{\ell,i} = i - 1$, $i = 1, 2, 3, 4, 5$ for both regression models $\ell = 1, 2$. For each configuration of $\sigma_\ell^2 = 1, 2, 3$ we use (3.1) to simulate $n_{\ell,i} = 6, 18, 30$ observations at each dose level $d_{\ell,i}$, resulting in total sample sizes of $n_\ell = 30, 90, 150$, respectively. We first compare the two sigmoid Emax models

$$m_1(d, \beta_1) = \beta_{0,1} + \frac{\beta_{0,2}d^{\beta_{0,3}}}{\tilde{\beta}_{1,4}^{\beta_{0,3}} + d^{\beta_{0,3}}} \quad \text{and} \quad m_2(d, \beta_2) = \beta_{0,1} + \frac{\beta_{0,2}d^{\beta_{0,3}}}{\tilde{\beta}_{2,4}^{\beta_{0,3}} + d^{\beta_{0,3}}}. \quad (3.3)$$

assuming the shared parameters $(\beta_{0,1}, \beta_{0,2}, \beta_{0,3})$. The only difference between the two models is in the ED_{50} parameters $\tilde{\beta}_{1,4}$ and $\tilde{\beta}_{2,4}$, which results in the need to estimate five parameters in total. We consider the reference sigmoid Emax model m_1 with the parameters $(\beta_{0,1}, \beta_{0,2}, \beta_{0,3}) = (1, 5, 4)$ and $\tilde{\beta}_{1,4} = 1$. This reference model is compared to various specifications of the second model m_2 determined by $\tilde{\beta}_{2,4} = 1.99, 1.77, 1.59, 1.43, 1.37, 1$ and common shared parameters $(\beta_{0,1}, \beta_{0,2}, \beta_{0,3})$. The values for $\tilde{\beta}_{2,4}$ were chosen such that the maximum absolute distances $d_\infty = \max_{d \in \mathcal{D}} |m_2(\beta_2, d) - m_1(\beta_1, d)|$ are given by 2, 1.5, 1, 0.5, 0.25, 0 respectively. For $d_\infty > 0$ these are attained at the dose levels 1.61, 1.52, 1.44, 1.37 and 1.33; see Figure 1a. For $d_\infty = 0$, that is $\tilde{\beta}_{1,4} = \tilde{\beta}_{2,4}$, the maximum distance is attained at every point in \mathcal{D} .

In Table 1 we summarize the simulated rejection probabilities of the bootstrap test (2.9) under the null hypothesis (2.4) with $d_\infty = 2, 1.5, 1$ and $\varepsilon = 1$. We conclude that the bootstrap test controls its level in all cases under consideration. At the margin of the null hypothesis (i.e. $d_\infty = 1$) the approximation of the level is very precise, even for sample sizes as small as $n_{\ell,i} = 6$.

We also investigated the relative residual mean squared errors (RRMSE) of the parameters estimates. Table 2 summarizes the simulation results only for $d_\infty = 1$ (i.e. at the margin of the null hypothesis), as the results are similar for other choices of d_∞ . We conclude that the RRMSE for estimating the Hill parameter $\beta_{0,3}$ is (by far) the largest. This phenomenon has also been observed by Mielke (2016). We also observe that all estimation errors decrease with larger sample sizes and smaller variances. Table 2 also summarizes the RRMSE when fixing the Hill parameter at $\beta_{0,3} = 4$ (see the numbers in brackets). In this case,

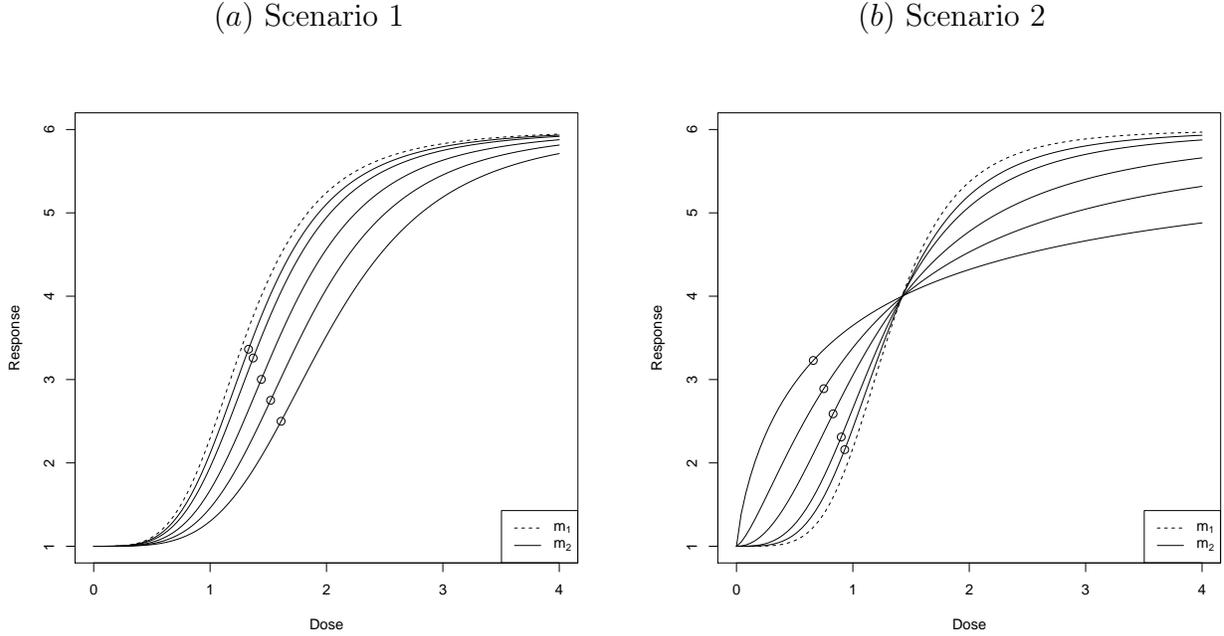


Figure 1: Graphical illustration of Scenarios 1 and 2. Open dots indicate the doses and corresponding responses where the maximum distance to the reference curve m_1 (dashed line) is attained.

| n_ℓ | d_∞ | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
|----------|------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| | | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ |
| 30 | 2 | 0.002 (0.000) | 0.010 (0.005) | 0.012 (0.009) | 0.002 (0.004) | 0.018 (0.013) | 0.031 (0.019) |
| 30 | 1.5 | 0.014 (0.013) | 0.035 (0.028) | 0.027 (0.043) | 0.026 (0.020) | 0.055 (0.049) | 0.052 (0.070) |
| 30 | 1 | 0.068 (0.058) | 0.054 (0.065) | 0.051 (0.060) | 0.106 (0.099) | 0.109 (0.114) | 0.115 (0.121) |
| 90 | 2 | 0.000 (0.000) | 0.000 (0.000) | 0.001 (0.002) | 0.000 (0.000) | 0.000 (0.000) | 0.002 (0.002) |
| 90 | 1.5 | 0.004 (0.001) | 0.010 (0.006) | 0.013 (0.015) | 0.005 (0.005) | 0.021 (0.014) | 0.026 (0.020) |
| 90 | 1 | 0.048 (0.071) | 0.053 (0.043) | 0.062 (0.059) | 0.104 (0.122) | 0.101 (0.097) | 0.129 (0.117) |
| 150 | 2 | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.004) |
| 150 | 1.5 | 0.000 (0.000) | 0.003 (0.000) | 0.002 (0.000) | 0.002 (0.000) | 0.011 (0.002) | 0.012 (0.009) |
| 150 | 1 | 0.056 (0.061) | 0.040 (0.061) | 0.042 (0.063) | 0.103 (0.102) | 0.090 (0.102) | 0.096 (0.109) |

Table 1: Simulated Type I error of the bootstrap test (2.9) for the equivalence of two sigmoid Emax models defined in Scenario 1 with $\varepsilon = 1$. The numbers in brackets show the simulated Type I error when fixing the Hill parameter at $\beta_{0,3} = 4$.

four parameters need to be estimated in total and the estimation errors become slightly smaller. We also repeated the Type I error rate simulations when fixing the Hill parameter at $\beta_{0,3} = 4$. The the results are reported in Table 1 (numbers in brackets) and we conclude that the size is well controlled within the simulation error.

| n_ℓ | σ^2 | $\beta_{0,1}$ | $\beta_{0,2}$ | $\beta_{0,3}$ | $\tilde{\beta}_{1,4}$ | $\tilde{\beta}_{2,4}$ |
|----------|------------|---------------|---------------|---------------|-----------------------|-----------------------|
| 30 | 1 | 0.288 (0.263) | 0.091 (0.062) | 0.493 | 0.125 (0.108) | 0.114 (0.106) |
| 30 | 2 | 0.389 (0.358) | 0.118 (0.090) | 0.739 | 0.174 (0.157) | 0.172 (0.148) |
| 30 | 3 | 0.460 (0.427) | 0.134 (0.105) | 0.841 | 0.228 (0.198) | 0.211 (0.185) |
| 90 | 1 | 0.166 (0.152) | 0.054 (0.036) | 0.184 | 0.067 (0.064) | 0.063 (0.061) |
| 90 | 2 | 0.237 (0.219) | 0.076 (0.054) | 0.355 | 0.107 (0.086) | 0.091 (0.086) |
| 90 | 3 | 0.280 (0.261) | 0.091 (0.062) | 0.450 | 0.130 (0.112) | 0.113 (0.105) |
| 150 | 1 | 0.123 (0.115) | 0.040 (0.029) | 0.129 | 0.050 (0.057) | 0.049 (0.046) |
| 150 | 2 | 0.171 (0.170) | 0.057 (0.041) | 0.241 | 0.073 (0.072) | 0.068 (0.063) |
| 150 | 3 | 0.219 (0.204) | 0.070 (0.050) | 0.322 | 0.095 (0.086) | 0.091 (0.084) |

Table 2: *RRMSE of the parameters obtained in the model estimation step of the bootstrap test (2.9) for the equivalence of two sigmoid Emax models defined in Scenario 1 with $d_\infty = 1$. The numbers in brackets show the values for the RRMSE when fixing the Hill parameter at $\beta_{0,3} = 4$.*

In Table 3 we summarize the power of the bootstrap when generating the data under the alternative $d_\infty = 0.5, 0.25, 0$ and $\varepsilon = 1$. As expected, the power increases with larger sample sizes and smaller variances and is reasonably high across all configurations. Fixing the Hill parameter significantly improves the power which can be explained by the difficulty of estimating this parameter precisely, as discussed above.

| n_ℓ | d_∞ | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
|----------|------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| | | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ |
| 30 | 0.5 | 0.137 (0.154) | 0.075 (0.078) | 0.070 (0.073) | 0.238 (0.266) | 0.172 (0.155) | 0.137 (0.145) |
| 30 | 0.25 | 0.208 (0.190) | 0.102 (0.101) | 0.081 (0.088) | 0.344 (0.349) | 0.196 (0.188) | 0.152 (0.170) |
| 30 | 0 | 0.181 (0.203) | 0.105 (0.105) | 0.086 (0.092) | 0.333 (0.361) | 0.196 (0.213) | 0.154 (0.154) |
| 90 | 0.5 | 0.341 (0.424) | 0.180 (0.230) | 0.132 (0.153) | 0.505 (0.581) | 0.311 (0.357) | 0.246 (0.279) |
| 90 | 0.25 | 0.550 (0.675) | 0.249 (0.315) | 0.166 (0.190) | 0.733 (0.802) | 0.428 (0.484) | 0.305 (0.348) |
| 90 | 0 | 0.664 (0.783) | 0.286 (0.353) | 0.191 (0.188) | 0.822 (0.884) | 0.463 (0.562) | 0.338 (0.367) |
| 150 | 0.5 | 0.481 (0.593) | 0.297 (0.359) | 0.207 (0.273) | 0.635 (0.729) | 0.460 (0.502) | 0.355 (0.406) |
| 150 | 0.25 | 0.826 (0.868) | 0.448 (0.569) | 0.280 (0.357) | 0.902 (0.933) | 0.635 (0.719) | 0.477 (0.545) |
| 150 | 0 | 0.917 (0.961) | 0.559 (0.665) | 0.342 (0.415) | 0.966 (0.989) | 0.740 (0.812) | 0.520 (0.596) |

Table 3: *Simulated power of the bootstrap test (2.9) for the equivalence of two sigmoid Emax models defined in Scenario 1 with $\varepsilon = 1$. The numbers in brackets show the simulated power when fixing the Hill parameter at $\beta_{0,3} = 4$.*

Scenario 2: We maintain the basic settings from Scenario 1. We consider again two sigmoid Emax models

$$m_1(d, \beta_1) = \beta_{0,1} + \frac{\beta_{0,2}d^{\tilde{\beta}_{1,3}}}{\tilde{\beta}_{1,4}^{\tilde{\beta}_{1,3}} + d^{\tilde{\beta}_{1,3}}} \quad \text{and} \quad m_2(d, \beta_2) = \beta_{0,1} + \frac{\beta_{0,2}d^{\tilde{\beta}_{2,3}}}{\tilde{\beta}_{2,4}^{\tilde{\beta}_{2,3}} + d^{\tilde{\beta}_{2,3}}}. \quad (3.4)$$

but assume now that both, the placebo response $\beta_{0,1}$ and the maximum treatment effect $\beta_{0,2}$, are the same. For the reference model we chose $(\beta_{0,1}, \beta_{0,2}) = (1, 5)$, $\tilde{\beta}_{1,3} = 4.5$ and $\tilde{\beta}_{1,4} = 1.3$. We investigated the maximum distances $d_\infty = 2, 1.5, 1, 0.5, 0.25, 0$, which resulted in the following parameter configurations for the second model:

$$\begin{aligned} &(\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (0.81, 0.86), (\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (1.4, 1.07), (\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (2.15, 1.18), \\ &(\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (3.15, 1.25), (\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (3.75, 1.28), (\tilde{\beta}_{2,3}, \tilde{\beta}_{2,4}) = (\tilde{\beta}_{1,3}, \tilde{\beta}_{1,4}) = (4.5, 1.3). \end{aligned} \quad (3.5)$$

The maximum distances between the two curves are now attained at the dose levels 0.66, 0.75, 0.83, 0.9 and 0.93; see Figure 1b.

In Table 4 we summarize the simulated rejection probabilities under the null hypothesis (2.4) for $d_\infty = 2, 1.5, 1$ and $\varepsilon = 1$. We conclude again that the bootstrap test controls the designated significance level in all cases under consideration. Especially at the margin $d_\infty = 1$ the simulated Type I error rates are close to the nominal level α . These observations apply regardless of whether the Hill parameter is estimated or fixed at the true underlying values given in (3.5).

| n_ℓ | d_∞ | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
|----------|------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| | | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ |
| 30 | 2 | 0.002 (0.000) | 0.001 (0.000) | 0.001 (0.000) | 0.000 (0.005) | 0.003 (0.007) | 0.006 (0.011) |
| 30 | 1.5 | 0.000 (0.001) | 0.003 (0.000) | 0.008 (0.001) | 0.007 (0.033) | 0.002 (0.030) | 0.016 (0.054) |
| 30 | 1 | 0.036 (0.035) | 0.044 (0.040) | 0.053 (0.039) | 0.083 (0.099) | 0.110 (0.116) | 0.113 (0.112) |
| 90 | 2 | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.002) | 0.004 (0.002) |
| 90 | 1.5 | 0.000 (0.000) | 0.000 (0.000) | 0.004 (0.000) | 0.000 (0.007) | 0.004 (0.021) | 0.012 (0.022) |
| 90 | 1 | 0.052 (0.057) | 0.028 (0.040) | 0.016 (0.036) | 0.104 (0.113) | 0.068 (0.117) | 0.056 (0.105) |
| 150 | 2 | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| 150 | 1.5 | 0.000 (0.000) | 0.000 (0.000) | 0.004 (0.000) | 0.000 (0.000) | 0.000 (0.004) | 0.004 (0.004) |
| 150 | 1 | 0.056 (0.032) | 0.036 (0.036) | 0.040 (0.036) | 0.100 (0.088) | 0.080 (0.100) | 0.076 (0.088) |

Table 4: *Simulated Type I error of the bootstrap test (2.9) for the equivalence of two sigmoid Emax models defined in Scenario 2 with $\varepsilon = 1$. The numbers in brackets show the simulated Type I error when fixing the Hill parameters at their true underlying values.*

In Table 5 we summarize the simulated power of the bootstrap test under the alternative $d_\infty = 0.5, 0.25, 0$ and $\varepsilon = 1$. As expected, the power decreases for increasing values of d_∞ and for higher variances or smaller sample sizes. One noticeable exception occurs at $d_\infty = 0$, where in some cases the power is smaller than for $d_\infty = 0.25$. This effect can be explained theoretically when considering the proofs for the bootstrap test. In case of $d_\infty = 0$ the set $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ containing all points where the maximum distance between the two curves is attained (see Appendix 6.3) consists of the entire dose range \mathcal{D} . Therefore, the asymptotic distribution of the test statistic is not Gaussian but a maximum of Gaussian

processes. This complex structure of the asymptotic distribution has an impact on the bootstrap procedure and explains the decrease in power for $d_\infty = 0$. This phenomenon can also be observed, although to a lesser degree, in Scenario 1. Finally, we observe higher power values when fixing the Hill parameter compared to the situation where it has to be estimated.

| n_ℓ | d_∞ | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
|----------|------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| | | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ | $\sigma^2 = 1$ | $\sigma^2 = 2$ | $\sigma^2 = 3$ |
| 30 | 0.5 | 0.130 (0.147) | 0.096 (0.075) | 0.094 (0.072) | 0.231 (0.245) | 0.191 (0.133) | 0.174 (0.125) |
| 30 | 0.25 | 0.156 (0.194) | 0.102 (0.089) | 0.087 (0.085) | 0.275 (0.319) | 0.197 (0.178) | 0.157 (0.147) |
| 30 | 0 | 0.155 (0.164) | 0.108 (0.076) | 0.087 (0.059) | 0.310 (0.316) | 0.197 (0.166) | 0.189 (0.137) |
| 90 | 0.5 | 0.312 (0.542) | 0.124 (0.225) | 0.140 (0.133) | 0.528 (0.697) | 0.240 (0.384) | 0.260 (0.240) |
| 90 | 0.25 | 0.384 (0.689) | 0.224 (0.289) | 0.164 (0.173) | 0.560 (0.841) | 0.396 (0.484) | 0.292 (0.313) |
| 90 | 0 | 0.448 (0.663) | 0.192 (0.259) | 0.148 (0.163) | 0.616 (0.807) | 0.372 (0.455) | 0.240 (0.309) |
| 150 | 0.5 | 0.528 (0.780) | 0.220 (0.392) | 0.160 (0.228) | 0.688 (0.896) | 0.404 (0.632) | 0.256 (0.400) |
| 150 | 0.25 | 0.724 (0.904) | 0.320 (0.544) | 0.260 (0.304) | 0.824 (0.936) | 0.540 (0.740) | 0.408 (0.516) |
| 150 | 0 | 0.644 (0.920) | 0.308 (0.580) | 0.224 (0.248) | 0.800 (0.956) | 0.532 (0.728) | 0.404 (0.484) |

Table 5: *Simulated power of the bootstrap test (2.9) for the equivalence of two sigmoid Emax models defined in Scenario 2 with $\varepsilon = 1$. The numbers in brackets show the simulated power when fixing the Hill parameters at their true underlying values.*

Scenario 3: We now investigate the operating characteristics of the bootstrap test assuming three, two, one and no shared parameters. We set $\varepsilon = 1$, $\alpha = 0.05$ and compare again two sigmoid Emax models. The true placebo response is chosen as $\beta_{1,1} = \beta_{2,1} = 0$. The reference model m_1 is specified by $(\beta_{1,2}, \beta_{1,3}, \beta_{1,4}) = (5, 2, 1.3)$. The second model is specified by $(\beta_{2,2}, \beta_{2,3}, \beta_{1,4}) = (5, 2, \kappa)$, where $\kappa \in (1.3, 3]$ is chosen such that the maximum distances with respect to m_1 lie between 0 and 2; see the resulting curves plotted in Figure 2(a) for $\kappa = 1.5, 1.7, 2, 2.5, 3$. Consequently the parameters specifying the two models only differ in ED_{50} parameters $\tilde{\beta}_{1,4}$ and $\tilde{\beta}_{2,4}$ and the maximum distance is determined by the choice of κ . As all other parameters are the same for the two models, we can compare the bootstrap test assuming three, two, one and no shared parameter. Note that we do not consider the case of identical models (i.e. $\kappa = 1.3$) because of the discontinuity of power at $d_\infty = 0$ described under Scenario 2. The dose range is given by $\mathcal{D} = [0, 10]$ with 5 different dose levels $d_{\ell,1} = 0$, $d_{\ell,2} = 1$, $d_{\ell,3} = 2$, $d_{\ell,4} = 5$, $d_{\ell,5} = 10$, $\ell = 1, 2$. We create $n_{\ell,i} = 35$ observations at each dose level for each group according to (3.1), which results in a total sample size of $n = n_1 + n_2 = 350$. Finally, we choose $\sigma_\ell^2 = 2$, $\ell = 1, 2$.

In Figure 2(b) we plot the proportion of rejections in dependence of the true maximum absolute difference $d_\infty \in (0, 2]$. Under the null hypothesis $d_\infty \geq 1$ all four tests control their level, as the proportion of rejections is smaller than or equal to $\alpha = 0.05$ within

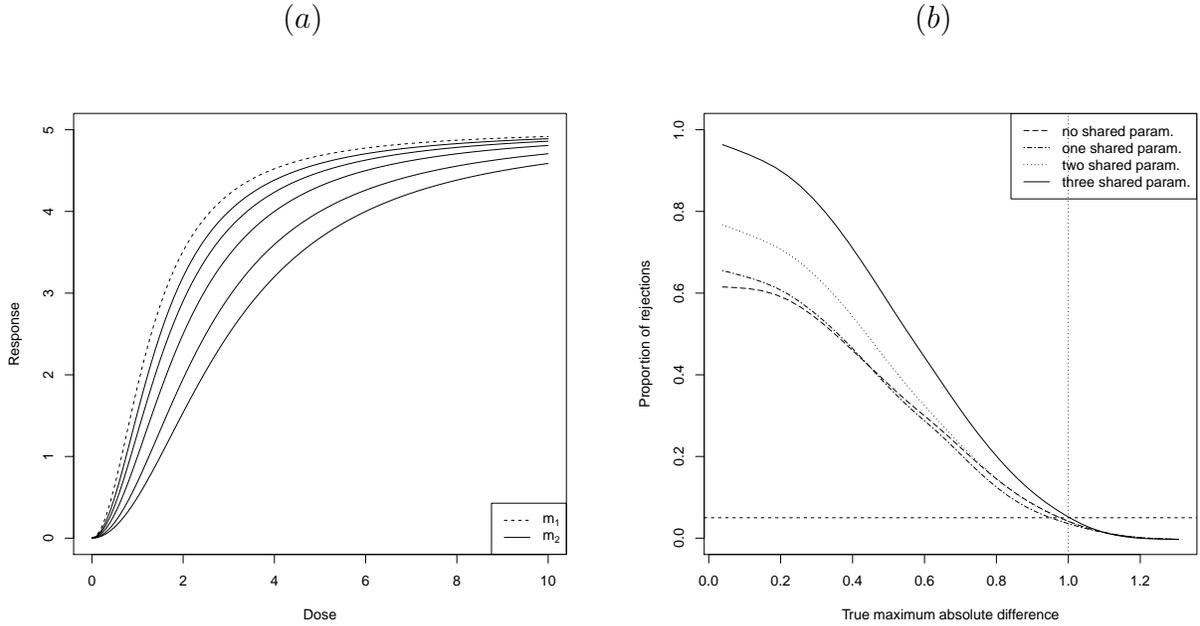


Figure 2: (a) Graphical illustration of the regression functions m_1 and m_2 with $\kappa = 1.5, 1.7, 2, 2.5, 3$ in Scenario 3. (b) Proportion of rejections in dependence of the true maximum absolute difference d_∞ for four different tests assuming one, two, three and no shared parameters for $\varepsilon = 1$ (vertical dotted line) and $\alpha = 0.05$ (horizontal dashed line).

simulation errors. Looking at the region $d_\infty < 1$, we observe that the test assuming three shared parameters has the highest power among all four tests, followed by the test assuming two shared parameters. The difference between the tests assuming one and no shared parameter is rather small. Concluding, the more parameters can be assumed to be common for the two regression curves the higher is the power of the test. Note, however, that strictly speaking the hypotheses (2.4) are different when assuming three, two, one and no shared parameters and that the perceived power gain when assuming more shared parameters comes at the cost of making additional assumptions that need to be verified in practice, as illustrated with the clinical trial example in Section 4.

4 Clinical trial example

We now illustrate the proposed method with a multi-regional clinical trial example. The objective of this trial is to evaluate the dose response relationships in Caucasian and Japanese patients and assess their similarity. Based on data from previous clinical trials investigating a drug with a similar mode of action, it is reasonable to assume a similar response

to placebo and a common maximum treatment effect in both populations, with the main difference expected to be in a different onset of treatment effect. Using the sigmoid Emax model (3.2), these considerations thus lead to different ED_{50} and Hill parameters for the two dose response curves. Because the trial is still at its design stage, we simulate data based on the trial assumptions. To maintain confidentiality, we scale the actual doses to lie within the $[0, 15]$ interval. These limitations do not change the utility of the calculations below.

We assume 60 Japanese and 240 Caucasian patients, resulting in 300 patients overall. Patients from both populations are randomized to receive either placebo (dose level 0) or one of three active dose levels, namely 1, 3, 15 for the Japanese and 0.5, 9 and 15 for the Caucasian patients. Assuming equal allocation of patients within each population, we thus have 75, 60, 15, 15, 60, and 75 patients randomized to the dose levels 0, 0.5, 1, 3, 9 and 15, respectively. The response variable is assumed to be normally distributed and larger values indicate a better outcome. Pharmacological and clinical considerations suggest the use of the (three-parameter) Emax model with the Hill parameter fixed at 1. Later on we relax this assumption as part of a sensitivity analysis. The R code for this example and all other calculations in this paper is available from the authors upon request.

In Figure 3 we display the fitted dose response models $m_1(d, \hat{\beta}_1)$ and $m_2(d, \hat{\beta}_2)$ for the Japanese and Caucasian patients, respectively, together with the individual observations, where $d \in [0, 15]$ and the y -axis is truncated to $[-1, 6]$ for better readability. The parameter estimates from the two separate model fits are given by $\hat{\beta}_1 = (-0.195, 4.751, 11.991)$ and $\hat{\beta}_2 = (-0.002, 5.676, 33.887)$. The observed differences for the placebo response and the maximum treatment effect are given by $|\hat{\beta}_{1,1} - \hat{\beta}_{2,1}| = 0.193$ and $|\hat{\beta}_{1,2} - \hat{\beta}_{2,2}| = 0.925$, respectively, and thus relatively small, as it also transpires from the plots in Figure 3. To corroborate this empirical observation, we formally test whether the assumption of shared parameters is plausible by applying the equivalence test described in Section 2.3 on the data set under consideration. We choose the threshold $\delta = 1.5$ and therefore test the null hypothesis $K_0 : \max_{i=1,2} |\beta_{1,i} - \beta_{2,i}| \geq 1.5$ against the alternative $K_1 : \max_{i=1,2} |\beta_{1,i} - \beta_{2,i}| < 1.5$. Applying the test (2.21) for $\alpha = 0.05$, we obtain $\hat{\Omega}_{11} = 3127.91$ and $\hat{\Omega}_{22} = 10748.27$ and therefore $\delta - t_{1-\alpha, n-2} \left(\frac{\hat{\Omega}_{11}}{n(n-2)} \right)^{1/2} = 1.191$ and $\delta - t_{1-\alpha, n-2} \left(\frac{\hat{\Omega}_{22}}{n(n-2)} \right)^{1/2} = 0.928$, respectively. We can thus reject K_0 at the relatively stringent 5% level and conclude equivalence of the two parameters, which justifies using the bootstrap test (2.1) with shared parameters.

We now evaluate the similarity of the dose response curves for the Japanese and Caucasian patients, assuming the same placebo and maximum treatment effect. In order to compute the non-linear least squares estimates in model (2.1) with (2.2) we formulate the objective

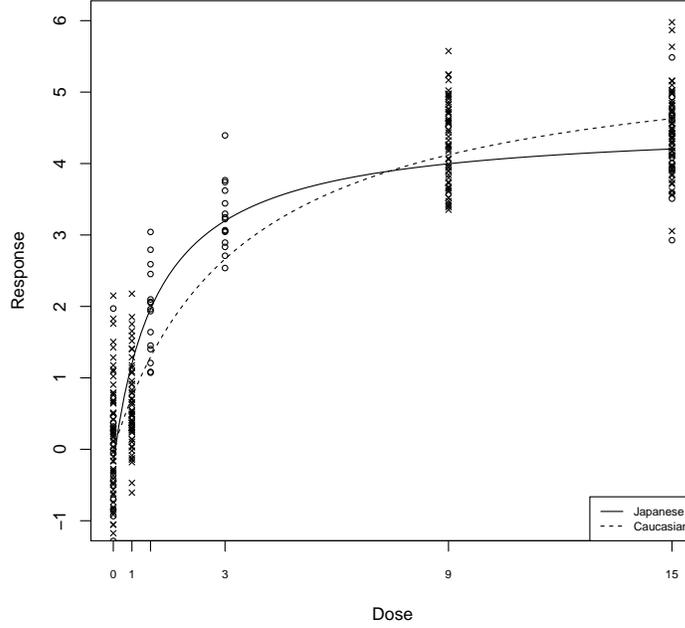


Figure 3: *Fitted Emax model m_1 (m_2) for the Japanese (Caucasians) patients given by the solid (dashed) line with observations marked by “o” (“x”).*

function of the minimization step as

$$\sum_{i=1}^4 \sum_{j=1}^{60} \left(Y_{1,i,j} - \left(\beta_{0,1} + \frac{\beta_{0,2} d_{1,i,j}}{\tilde{\beta}_{1,3} + d_{1,i,j}} \right) \right)^2 + \sum_{i=1}^4 \sum_{j=1}^{240} \left(Y_{2,i,j} - \left(\beta_{0,1} + \frac{\beta_{0,2} d_{2,i,j}}{\tilde{\beta}_{2,3} + d_{2,i,j}} \right) \right)^2.$$

Here, $\beta_{0,1}$ denotes the (shared) placebo effect, $\beta_{0,2}$ the (shared) maximum treatment effect E_{max} , and $\tilde{\beta}_{1,3}$ and $\tilde{\beta}_{2,3}$ the ED_{50} parameters of the two models. Using the `auglag` function from the `alabama` package Varadhan (2014) to solve the above optimization problem, we obtain the parameter estimates $\hat{\beta}_{0,1} = -0.064$ (0.074), $\hat{\beta}_{0,2} = 5.366$ (0.137), $\hat{\beta}_{1,3} = 19.400$ (2.634) and $\hat{\beta}_{2,3} = 25.681$ (3.256). In brackets we report the associated standard errors, which have to be calculated manually based on (5.5). The estimates for the population variances are $\hat{\sigma}_1^2 = 0.508$ and $\hat{\sigma}_2^2 = 0.455$. The observed maximum difference between both curves over the investigated dose range $[0, 15]$ is $\hat{d}_\infty = 0.376$, attained at dose 2.23. We apply the bootstrap test (2.9) using $B = 1'000$ bootstrap replications. Setting $\varepsilon = 0.7$ for the equivalence margin in (2.4), we obtain the quantile $q_{0.05} = 0.438$ for $\alpha = 0.05$. Thus, we reject the null hypothesis (2.4) at the 5% significance level and conclude that the dose response curves for the Japanese and Caucasian populations are similar, under the shared parameter assumption. Alternatively, we can calculate the p -value

$\frac{1}{B} \sum_{i=1}^B I(d_{\infty}^{*(i)} \leq \hat{d}_{\infty}) = 0.023$ for the bootstrap test and obtain the same test decision at level $\alpha = 0.05$. For illustration purposes we also apply the bootstrap test (2.9) but without shared parameters (yet under the assumption of a fixed Hill parameter). Accordingly, we obtain a considerably larger p -value of 0.458, which supports our findings from Scenario 3 in Section 3 about the loss in power when no shared parameters are assumed. In this case the observed maximum distance is $\hat{d}_{\infty} = 0.706$, attained at dose 1.42, and the quantile of the bootstrap distribution is $q_{0.05} = 0.449$.

Finally, we perform a sensitivity analysis to investigate the assumption of the Hill parameter being equal to 1. As part of this analysis we repeat the model fit and the bootstrap test using the sigmoid Emax model (3.2) where the Hill parameter is now part of the estimation. The parameter estimates (standard errors in brackets) are $\hat{\beta}_{0,1} = 0.037$ (0.082), $\hat{\beta}_{0,2} = 4.544$ (0.218), $\hat{\beta}_{1,3} = 1.05$ (0.229), $\hat{\beta}_{1,4} = 13.542$ (2.095), $\hat{\beta}_{2,3} = 1.650$ (0.331) and $\hat{\beta}_{2,4} = 16.558$ (4.521). Now, the maximum distance between the curves is $\hat{d}_{\infty} = 0.640$, attained at dose 0.6. It turns out that the standard errors of the estimates are slightly higher which is in line with the results shown in the simulation studies in Section 3. Performing again the bootstrap test with two shared parameters results the quantile $q_{0.05} = 0.429$ and the p -value 0.285. Consequently, we cannot reject the null hypothesis in this case. In conclusion, fixing the Hill parameter to 1 and assuming both the placebo effect and the maximum treatment effect to be the same in both populations clearly results in the most powerful procedure. We can demonstrate equivalence at the significance level of $\alpha = 0.05$, whereas in case of estimating both models separately (i.e. no shared parameters) or including the Hill parameter in the estimation we obtain considerably larger p -values.

5 Conclusion

In this paper we developed a new test for the equivalence of two regression curves when it is reasonable to assume that some model parameters are the same. Our approach is based on an estimate of the maximum deviation between the two curves, where critical values are obtained by a novel constraint bootstrap procedure. We demonstrated that the new test controls its level properly and is consistent.

We investigated the finite sample properties of the proposed procedure using extensive simulations and observed that the Type I error rate is controlled in all scenarios under consideration, even for sample sizes as small as 6 patients per dose level. Further, we concluded that the test reaches a reasonable power that increases with larger sample sizes. In particular, we demonstrated that the power of tests for the equivalence of curves can be improved substantially by using the additional information of common parameters in the

two regression curves. This effect could also be observed in the clinical trial example, which showed the power advantage of the bootstrap test (2.9) if the underlying assumptions are well justified. Relaxing those assumptions may lead to more robust conclusions, but only at a cost of a loss in power.

An interesting extension of the proposed methodology arises from the need to include covariates in clinical trial practice. Covariates can be continuous (e.g. age or body mass index), categorical (e.g. disease status or race), or binary (e.g. gender or smoking yes/no), possibly changing over time. These cases may have to be treated differently and we leave this problem for future research. Another area of research could be the assessment of similarity in two nested populations, thus relaxing the assumption of independence between the observations. In our multi-regional clinical trial example we compared the Japanese with Caucasian patients. It will be interesting and relevant to clinical trials to explore the development of the proposed methods when comparing the Japanese with an overall population that includes Japanese and Caucasian patients. Again, we leave this topic for future research.

Acknowledgments This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt T1) of the German Research Foundation (DFG) Parts of this manuscript were written while Frank Bretz was on a sabbatical leave at University of Canterbury in Christchurch, New Zealand. He would like to thank Dr. Daniel Gerhard for his support.

References

- Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, 24:295–300.
- Bretz, F., M. K., Dette, H., Liu, W., and Trampisch, M. (2016). Assessing the similarity of dose response and target doses in two non-overlapping subgroups. *arXiv:1607.05424*.
- Collignon, O., Moellenhoff, K., and Dette, H. (2018). Equivalence analyses of dissolution profiles with the mahalanobis distance: a regulatory perspective and a comparison with a parametric maximum deviation-based approach. *Biometrical Journal*.
- Dette, H., Möllenhoff, K., Volgushev, S., and Bretz, F. (2018). Equivalence of regression curves. *Journal of the American Statistical Association*, 113:711–729.
- Feng, L., Zou, C., Wang, Z., and Zhu, L. (2015). Robust comparison of regression curves. *TEST*, 24(1):185–204.
- Gabrielsson, J. and Weiner, D. (2007). *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. Swedish Pharmaceutical Press, Stockholm, 4th edition.

- Gsteiger, S., Bretz, F., and Liu, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics*, 21(4):708–725.
- Hoffelder, T. (2018). Equivalence analyses of dissolution profiles with the mahalanobis distance. *Biometrical Journal*.
- ICH (2017). E17-general principles for planning and design of multi-regional clinical trials. available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E17/E17EWG_Step4_2017_1116.pdf.
- Liu, W., Bretz, F., Hayter, A. J., and Wynn, H. P. (2009). Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics*, 65(4):1279–1287.
- Macdougall, J. (2006). Analysis of dose-response studies-e max model. In *Dose finding in drug development*, pages 127–145. Springer.
- Malinowski, H. J., Westelinck, A., Sato, J., and Ong, T. (2008). Same drug, different dosing: differences in dosing for drugs approved in the united states, europe, and japan. *The Journal of Clinical Pharmacology*, 48(8):900–908.
- Mielke, T. (2016). Study designs for the estimation of the hill parameter in sigmoidal response models. In *mODa 11-Advances in Model-Oriented Design and Analysis*, pages 183–190. Springer.
- Moellenhoff, K., Dette, H., Kotzagiorgis, E., Volgushev, S., and Collignon, O. (2018). Regulatory assessment of drug dissolution profiles comparability via maximum deviation. *Statistics in medicine*.
- Pinheiro, J., Bretz, F., and Branson, M. (2006). Analysis of dose-response studies-modeling approaches. In *Dose finding in drug development*, pages 146–171. Springer.
- Thomas, N., Sweeney, K., and Somayaji, V. (2014). Meta-analysis of clinical dose-response in a large drug development portfolio. *Statistics in Biopharmaceutical Research*, 6:302–317.
- Ting, N. (2006). *Dose finding in drug development*. Springer Science & Business Media.
- van der Vaart, A. W. (2000). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge, Cambridge University Press.
- Varadhan, R. (2014). Constrained nonlinear optimization. R package version 2011.9-1, available at <http://cran.r-project.org/web/packages/alabama/index.html>.
- Wang, W., Gene Hwang, J., and Dasgupta, A. (1999). Statistical tests for multivariate bioequivalence. *Biometrika*, 86(2):395–402.

6 Appendix: Proof of Theorem 2.1 and Remark 2.2

The proof of the theoretical results of this paper proceeds in several steps. First we state the assumptions under which the statements hold (Section 6.1). Second, we derive the asymptotic distribution of the parameter estimates in models with common parameters (Section 6.2). In Section 6.3 we derive a result on the weak convergence of a stochastic process, from which the proof of Theorem 2.1 and Remark 2.2 can be derived (see Section 6.4).

6.1 Assumptions

For the theoretical results of this paper we make the same assumptions as in Dette et al. (2018).

1. The errors $\eta_{\ell,i,j}$ are independent, have finite variance σ_ℓ^2 and expectation zero.
2. The covariate region $\mathcal{D} \subset \mathbb{R}^d$ is compact and the number and location of dose levels k_ℓ does not depend on n_ℓ , $\ell = 1, 2$.
3. All estimates of the parameters β_1, β_2 are computed over compact sets $B_1 \subset \mathbb{R}^{p_1}$ and $B_2 \subset \mathbb{R}^{p_2}$.
4. The regression functions m_1 and m_2 are twice continuously differentiable with respect to the parameters for all b_1, b_2 in neighbourhoods of the true parameters β_1, β_2 and all $d \in \mathcal{D}$. The functions $(d, b_\ell) \mapsto m_\ell(d, b_\ell)$ and their first two derivatives are continuous on $\mathcal{D} \times B_\ell$.
5. The gradients with respect to the parameters are uniformly bounded, that is $\sup_{d \in \mathcal{D}} \left\| \frac{\partial}{\partial \beta_\ell} m_\ell(d, \beta_\ell) \right\| < \infty$, $\ell = 1, 2$.
6. Defining $\psi_{a,\ell}^{(n)}(b) := \sum_{i=1}^{k_\ell} \frac{n_{\ell,i}}{n_\ell} (m_\ell(d_{\ell,i}, a) - m_\ell(d_{\ell,i}, b))^2$, we assume that for any $u > 0$ there exists a constant $v_{u,\ell} > 0$ such that

$$\liminf_{n_1, n_2 \rightarrow \infty} \inf_{a \in B_\ell} \inf_{|b-a| \geq u} \psi_{a,\ell}^{(n)}(b) \geq v_{u,\ell} \quad \ell = 1, 2.$$

6.2 Asymptotic properties of the OLS

In this section we derive the asymptotic normality of the parameter estimates in models with common parameters. Observing the definition of the OLS $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ in (2.3) we

obtain, by taking the partial derivatives, $\hat{\beta}$ by the necessary conditions

$$\begin{aligned}
0 &= \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(d_{\ell,i}, (\hat{\beta}_0, \hat{\beta}_\ell))) \left(\frac{\partial}{\partial \beta_0} m_\ell(d_{\ell,i}, \beta_0, \tilde{\beta}_\ell) \Big|_{(\beta_0, \tilde{\beta}_\ell) = (\hat{\beta}_0, \hat{\beta}_\ell)} \right)^T \\
0 &= \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(d_{\ell,i}, (\hat{\beta}_0, \hat{\beta}_\ell))) \left(\frac{\partial}{\partial \tilde{\beta}_\ell} m_\ell(d_{\ell,i}, \beta_0, \tilde{\beta}_\ell) \Big|_{(\beta_0, \tilde{\beta}_\ell) = (\hat{\beta}_0, \hat{\beta}_\ell)} \right)^T, \quad \ell = 1, 2. \quad (5.1)
\end{aligned}$$

Defining

$$\begin{aligned}
h_\ell(d_{\ell,i}) &= \left(\frac{\partial}{\partial \beta_0} m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell)) \right)^T, \quad \tilde{h}_\ell(d_{\ell,i}) = \left(\frac{\partial}{\partial \tilde{\beta}_\ell} m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell)) \right)^T \\
\hat{h}_\ell(d_{\ell,i}) &= \left(\frac{\partial}{\partial \beta_0} m_\ell(d_{\ell,i}, \beta_0, \tilde{\beta}_\ell) \Big|_{(\beta_0, \tilde{\beta}_\ell) = (\hat{\beta}_0, \hat{\beta}_\ell)} \right)^T, \quad \hat{\tilde{h}}_\ell(d_{\ell,i}) = \left(\frac{\partial}{\partial \tilde{\beta}_\ell} m_\ell(d_{\ell,i}, \beta_0, \tilde{\beta}_\ell) \Big|_{(\beta_0, \tilde{\beta}_\ell) = (\hat{\beta}_0, \hat{\beta}_\ell)} \right)^T,
\end{aligned}$$

for $\ell = 1, 2$ we can write

$$\begin{aligned}
\left(\frac{\partial}{\partial \beta} m_1(d_{1,i}, \beta_0, \tilde{\beta}_1) \Big|_{(\beta_0, \tilde{\beta}_1) = (\hat{\beta}_0, \hat{\beta}_1)} \right)^T &= (\hat{h}_1(d_{1,i}), \hat{\tilde{h}}_1(d_{1,i}), 0_{p-p'}) \\
\left(\frac{\partial}{\partial \beta} m_2(d_{2,i}, \beta_0, \tilde{\beta}_2) \Big|_{(\beta_0, \tilde{\beta}_2) = (\hat{\beta}_0, \hat{\beta}_2)} \right)^T &= (\hat{h}_2(d_{2,i}), 0_{p-p'}, \hat{\tilde{h}}_2(d_{2,i})),
\end{aligned}$$

where $0 = 0_{p-p'}$ denotes the zero vector in $\mathbb{R}^{p-p'}$. Therefore the equations in (5.1) can be summarized to

$$\begin{aligned}
0 &= \sum_{i=1}^{k_1} \sum_{j=1}^{n_{1,i}} (Y_{1,i,j} - m_1(d_{1,i}, (\hat{\beta}_0, \hat{\beta}_1))) (\hat{h}_1(d_{1,i}), \hat{\tilde{h}}_1(d_{1,i}), 0_{p-p'})^T \\
&+ \sum_{i=1}^{k_2} \sum_{j=1}^{n_{2,i}} (Y_{2,i,j} - m_2(d_{2,i}, (\hat{\beta}_0, \hat{\beta}_2))) (\hat{h}_2(d_{2,i}), 0_{p-p'}, \hat{\tilde{h}}_2(d_{2,i}))^T \\
&= \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(d_{\ell,i}, (\hat{\beta}_0, \hat{\beta}_\ell))) \left(\frac{\partial}{\partial \beta} m_\ell(d_{\ell,i}, \beta_0, \tilde{\beta}_\ell) \Big|_{(\beta_0, \tilde{\beta}_\ell) = (\hat{\beta}_0, \hat{\beta}_\ell)} \right)^T. \quad (5.2)
\end{aligned}$$

A Taylor expansion now yields

$$\begin{aligned}
0 &= \sqrt{n}(\hat{\beta} - \beta) \frac{1}{n} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \left(\frac{\partial^2 m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial^2 \beta} \right)^T \eta_{\ell,i,j} - \frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \left(\frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \right)^T \\
&+ \frac{1}{\sqrt{n}} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \eta_{\ell,i,j} \left(\frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \right)^T + o_{\mathbb{P}}(1),
\end{aligned}$$

and therefore $\hat{\beta}$ can be linearized as

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \frac{1}{\sqrt{n}} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \eta_{\ell,i,j} \left(\frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \right)^T \quad (5.3) \\
&\cdot \left(-\frac{1}{n} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \left(\frac{\partial^2 m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial^2 \beta} \right)^T \eta_{\ell,i,j} - \frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \left(\frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \right)^T \right)^{-1} + o_{\mathbb{P}}(1).
\end{aligned}$$

Due to the strong law of large numbers it holds

$$\begin{aligned}
& -\frac{1}{n} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \left(\frac{\partial^2 m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial^2 \beta} \right)^T \eta_{\ell,i,j} - \frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \left(\frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \right)^T \\
& \xrightarrow{a.s.} \frac{1}{\lambda} \sum_{i=1}^{k_1} \zeta_{1,i} \frac{\partial m_1(d_{1,i}, (\beta_0, \tilde{\beta}_1))}{\partial \beta} \left(\frac{\partial m_1(d_{1,i}, (\beta_0, \tilde{\beta}_1))}{\partial \beta} \right)^T + \frac{\lambda-1}{\lambda} \sum_{i=1}^{k_2} \zeta_{2,i} \frac{\partial m_2(d_{2,i}, (\beta_0, \tilde{\beta}_2))}{\partial \beta} \left(\frac{\partial m_2(d_{2,i}, (\beta_0, \tilde{\beta}_2))}{\partial \beta} \right)^T \\
& = \tilde{\Sigma}_1 + \tilde{\Sigma}_2 =: \tilde{\Sigma}, \tag{5.4}
\end{aligned}$$

where the matrices $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ are defined by

$$\begin{aligned}
\tilde{\Sigma}_1 &= \frac{1}{\lambda} \sum_{i=1}^{k_1} \zeta_{1,i} (h_1(d_{1,i}), \tilde{h}_1(d_{1,i}), 0_{p-p'})^T (h_1(d_{1,i}), \tilde{h}_1(d_{1,i}), 0_{p-p'}) \\
\tilde{\Sigma}_2 &= \frac{\lambda-1}{\lambda} \sum_{i=1}^{k_2} \zeta_{2,i} (h_2(d_{2,i}), 0_{p-p'}, \tilde{h}_2(d_{2,i}))^T (h_2(d_{2,i}), 0_{p-p'}, \tilde{h}_2(d_{2,i})),
\end{aligned}$$

respectively. Therefore, using the representation in (5.3) and the result in (5.4) we obtain the asymptotic normality of the estimate, that is

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\Sigma}^{-1}(\sigma_1^2 \tilde{\Sigma}_1 + \sigma_2^2 \tilde{\Sigma}_2) \tilde{\Sigma}^{-1}). \tag{5.5}$$

6.3 Weak convergence of a stochastic process

The essential step in the proof is a result regarding the weak convergence of the process

$$\tilde{p}_n(d) = m_1(d, (\hat{\beta}_0, \hat{\beta}_1)) - m_2(d, (\hat{\beta}_0, \hat{\beta}_2)) - (m_1(d, (\beta_0, \tilde{\beta}_1)) - m_2(d, (\beta_0, \tilde{\beta}_2))).$$

We further define $\tilde{\Delta}(d, \beta) = \tilde{\Delta}(d, \beta_0, \tilde{\beta}_1, \tilde{\beta}_2) = \tilde{\Delta}(d, \beta_1, \beta_2) = m_1(d, \beta_1) - m_2(d, \beta_2)$, which yields $\tilde{p}_n(d) = \tilde{\Delta}(d, \hat{\beta}) - \tilde{\Delta}(d, \beta)$.

Proof. At first we derive a linearization of \tilde{p}_n by using the Taylor expansion $\tilde{\Delta}(d, \hat{\beta}) = \tilde{\Delta}(d, \beta) + \left(\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta} \right)^T (\hat{\beta} - \beta)^T + R(\beta)$, where $R(\beta)$ is a remainder term and due to (5.3) we have $R(\beta) = O_{\mathbb{P}}(\frac{1}{n})$. Therefore it holds $\sqrt{n} \tilde{p}_n(d) = \sqrt{n} \left(\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta} \right)^T (\hat{\beta} - \beta)^T + o_{\mathbb{P}}(1)$ uniformly with respect to $d \in \mathcal{D}$ (note that due to the assumptions $\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta}$ is continuous in $d \in \mathcal{D}$ and \mathcal{D} is a compact set) and with the result from (5.3) this can be written as

$$\sqrt{n} \tilde{p}_n(x) = \frac{\partial \tilde{\Delta}(x, \beta)}{\partial \beta} \frac{1}{\sqrt{n}} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \eta_{\ell,i,j} \frac{\partial m(x_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \tilde{\Sigma}^{-1} + o_{\mathbb{P}}(1). \tag{5.6}$$

Observing (5.3) we define

$$\tilde{G}_n(d) = \left(\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta} \right)^T (\hat{\beta} - \beta)^T = \left(\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta} \right)^T \left(\frac{1}{n} \sum_{\ell=1}^2 \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \eta_{\ell,i,j} \frac{\partial m_\ell(d_{\ell,i}, (\beta_0, \tilde{\beta}_\ell))}{\partial \beta} \tilde{\Sigma}^{-1} \right)^T, \tag{5.7}$$

and obtain

$$\sqrt{n}(\tilde{p}_n(d) - \tilde{G}_n(d)) = o_{\mathbb{P}}(1). \quad (5.8)$$

We now define $f(d) := \left(\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta}\right)^T \Sigma^{-1/2}$, and we consider the map

$$\tilde{\Phi} : \begin{cases} \mathbb{R}^{p_1+p_2-p'} \rightarrow \ell_{\infty}(\mathcal{D}) \\ d \mapsto (\Phi(d) : d \mapsto f(d) \cdot d) \end{cases}$$

Obviously, using the same arguments as before, $\frac{\partial \tilde{\Delta}(d, \beta)}{\partial \beta}$ is continuous and therefore the same holds for $\tilde{\Phi}$. Consequently, (5.5) and the Continuous Mapping Theorem (see for example van der Vaart (2000)[p.7 f.]) yield

$$\{\sqrt{n}\tilde{G}_n(d)\}_{d \in \mathcal{D}} = \{\tilde{\Phi}(\sqrt{n}\Sigma^{1/2}(\hat{\beta} - \beta)^T)(d)\}_{d \in \mathcal{D}} \xrightarrow{\mathcal{D}} \{f(d)Z\}_{d \in \mathcal{D}} \stackrel{\mathcal{D}}{=} \{\tilde{G}(d)\}_{d \in \mathcal{D}},$$

uniformly with respect to $d \in \mathcal{D}$, where $Z \sim \mathcal{N}(0, I_{p_1+p_2-p'})$, which proves the assertion. \square

6.4 Proof of Theorem 2.1 and Remark 2.2

The assertion of Theorem 2.1 now follows by the same arguments as given in the proof of Theorem 4 in Dette et al. (2018). In particular it can be shown that

$$\sqrt{n}(\tilde{d}_{\infty} - d_{\infty}) \xrightarrow{\mathcal{D}} \mathcal{Z} := \max \left\{ \max_{d \in \mathcal{E}^+} \tilde{G}(d), \max_{d \in \mathcal{E}^-} (-\tilde{G}(d)) \right\},$$

where $\mathcal{E}^{\pm} = \{d \in \mathcal{D} \mid m_1(d, \beta_1) - m_2(d, \beta_2) = \pm d_{\infty}\}$ denote the sets of extremal points, that is those points, where the unknown difference $m_1 - m_2$ attains its maximum absolute deviation. The details are omitted for the sake of brevity.

Similarly, note that in the situation of a common placebo group as described in Remark 2.2 the estimates are obtained by minimizing the sum of squares in (2.13). As $m_{\ell}^0(0, \tilde{\beta}_{\ell}^0) = 0$, $\ell = 1, 2$, this function is the same as the one which is obtained allocating the observations at placebo arbitrarily to the two groups. More precisely, we can also write the sum of squares in (2.13) as

$$\sum_{\ell=1}^2 \sum_{i=1}^{k_{\ell}} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - (b_{0,1} + \tilde{b}_{\ell,1} \cdot m_{\ell}^0(d_{\ell,i}, \tilde{b}_{\ell}^0)))^2,$$

where $Y_{1,1,j} = Y_{0,j}$ ($j = 1, \dots, n_{1,1}$) and $Y_{2,1,j} = Y_{0,j}$ ($j = n_{1,1} + 1, \dots, n_{1,1} + n_{2,1} = n_0$). This corresponds to the minimization of the sum of squares in (2.3) with a common intercept $b_{0,1}$, and consequently this situation can be treated in the same way as considering two different placebo groups with a common intercept $b_{0,1}$. By the same arguments as given in Appendix

6.2 the corresponding estimates are asymptotically normal distributed (if $n_{1,1}/n \rightarrow c_1$ and $n_{2,1}/n \rightarrow c_2$ for some constants $c_1, c_2, \in (0, 1)$), and the proof in Section 6.3 shows that Theorem 2.1 remains valid in the model with a common placebo group. As a consequence we obtain the claim in Remark 2.2