Assessing the similarity of dose response and target doses in two non-overlapping subgroups

Frank Bretz[1],[*], Kathrin Möllenhoff[2], Wei Liu[3], Matthias Trampisch[4], Holger Dette[2]

We consider two problems that are becoming more and more frequent in dose finding studies. First, we assess the similarity of two non-linear regression models for two non-overlapping subgroups of patients over a restricted covariate space. To this end, we derive a confidence interval for the maximum difference between the two given models. If this confidence interval excludes the equivalence margins, similarity of dose response can be claimed. Second, we address the problem of demonstrating the similarity of two target doses for two non-overlapping subgroups, using again a confidence interval based approach. The proposed methods are illustrated with a real case study and their operating characteristics (coverage probabilities, Type I error rates, power) are investigated via simulation.
Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: equivalence testing, multiregional trial, target dose estimation, subgroup analyses

1. Introduction

Establishing dose response information and defining optimal dosing regimens is a fundamental step in the investigation of any new compound, be it a medicinal drug, an herbicide or fertilizer, a molecular entity, an environmental toxin, or an industrial chemical [1]. This has been recognized for many years, especially in the drug development arena, where patients will be exposed to a medicinal drug once it has been released on the market [2]. An indication of the importance of properly conducted dose response studies is the early publication of the tripartite ICH-E4 guideline, which gives recommendations on the design and conduct of studies to assess the relationship between doses, blood levels and clinical response throughout the clinical development of a new drug [3].

Very often clinical trials are analyzed beyond the primary study objectives by assessing efficacy and safety profiles in clinically relevant subgroups, such as different gender, age classes, grades of disease severity, etc. A rising area of particular importance are global clinical trials, which are run in different countries and potentially serve different submissions. For example, many pharmaceutical companies focus on running global clinical trials that include a major

[1] *Novartis Pharma AG, CH-4002 Basel, Switzerland*

[2] *Department of Mathematics, Ruhr-Universität Bochum, Germany*

[3] *S3RI and School of Mathematics, University of Southampton, SO17 1TB, UK*

[4] *Boehringer Ingelheim Pharma, Global Biom. and Clin. Appl./BDM, Germany*

[*]*Correspondence to: Frank Bretz, Novartis Pharma AG, Statistical Methodology, WSJ-27.1.005, CH-4002 Basel, Switzerland. Phone: + 41 61 324 4064, Fax: + 41 61 324 3039, Email: frank.bretz@novartis.com*

Japanese subpopulation for later regulatory submission in Japan. A natural question is then whether the dose response results for the Japanese and the non-Japanese populations are consistent [4, 5].

To illustrate the general problem, assume that we are interested in assessing similarity for (a) two dose response curves or for (b) two same target doses, say for male/female or Japanese/non-Japanese patients. For question (a) we thus want to show that the maximum difference in response between two (potentially different) non-linear parametric regression models is smaller than a pre-specified margin. Figure $1a$ displays an example, where the two dose response curves follow an Emax and a logistic model and the maximum response difference over the dose range is indicated by the arrow. For question (b) we want to show that two same target doses do not differ relevantly. Figure $1b$ displays the minimum effective dose (MED) derived from the two previous dose response models. Here, the MED is defined as the smallest dose which demonstrates a clinically relevant benefit over placebo, as indicated by the horizontal line in Figure $1b$. If we succeed in demonstrating either (a) or (b), evidence is provided that the difference in response over the entire dose range under investigation or the the two target doses differ at most marginally. In practice, such a result may provide sufficient evidence that the same dose can be administered in both subgroups (i.e., the doses for males/females or Japanese/non-Japanese patients are the same).
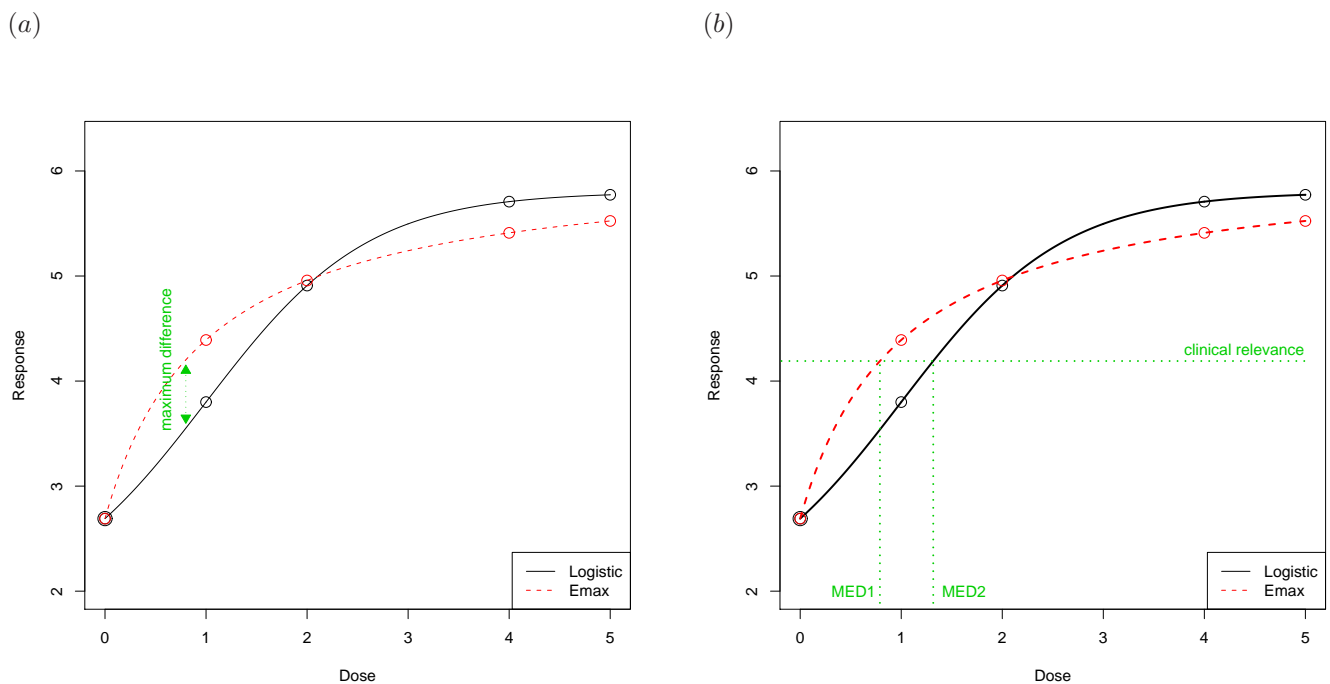
(a)

(b)

Figure 1. Assessing similarity for (a) two dose response curves and (b) two same target doses.

Demonstrating similarity of target doses or dose response curves in each of several subgroups has not been addressed in much detail so far in the literature. One exception is [6], who proposed a non-standard bootstrap approach for question (a) which addresses the specific form of the interval hypotheses. In particular, data has to be generated under the null hypothesis using constrained maximum likelihood estimates. In this paper we consider different methods to address both questions (a) and (b). In Section 2, we derive a confidence interval for the maximum difference between the two given non-linear regression models over the entire covariate space of interest using the results from [7] to address problem (a). If this confidence interval excludes the equivalence margins, similarity of dose response can be claimed. In Section 3, we consider asymptotic methods to derive confidence intervals for the difference between two same target doses to address problem (b). Again, if such a confidence interval excludes a pre-specified relevance margin, similarity in dose can be claimed. In Section 4 we provide some concluding remarks. Technical details are left for the Appendix.

2. Assessing similarity of two dose response curves

We consider the non-linear regression models

$$Y_{\ell,i,j} = m_\ell(\vartheta_\ell, d_{\ell,i}) + \epsilon_{\ell,i,j}, \; j = 1, \ldots, n_{\ell,i}, \; i = 1, \ldots, k_\ell, \; \ell = 1, 2, \; d_{\ell,i} \in \mathcal{D}, \tag{1}$$

where $Y_{\ell,i,j}$ denotes the jth observed response at the ith dose level $d_{\ell,i}$ under the ℓth dose response model m_ℓ. The error terms $\epsilon_{\ell,i,j}$ are assumed to be independent and identically distributed with expectation 0 and variance σ_ℓ^2. 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 ith dose level ($i = 1, \ldots k_\ell$, $\ell = 1, 2$). We further assume that for both regression models the different dose levels are attained on the same (restricted) covariate region $\mathcal{D}$. For the purpose of this paper, we assume $\mathcal{D}$ to be the dose range under investigation, although the results in this section can be generalized to include other covariates. The functions m_1 and m_2 in (1) denote the (non-linear) regression models with fixed but unknown p_1- and p_2-dimensional parameter vectors ϑ_1 and ϑ_2, respectively. Note that both the regression models m_1 and m_2 and the parameters ϑ_1 and ϑ_2 may be different; in particular, the design matrices for the two regression models may be unequal. This implies that we do not assume the same doses to be investigated for $\ell = 1, 2$ and that the sample sizes n_ℓ can be unequal. We refer to [8] for an overview of several linear and non-linear regression models commonly employed in clinical studies.

2.1. Methodology

In the following we derive a confidence interval for the maximum absolute absolute difference between the two given non-linear regression models m_1 and m_2 over the entire covariate space $\mathcal{D}$ based on results from [7]. Concluding, we will use this confidence interval in order to derive a test demonstrating similarity of the two dose response curves.

Let $U(Y_1, Y_2, d)$ denote a $1 - \alpha$ pointwise upper confidence bound on the difference curve $m_2(\vartheta_2, d) - m_1(\vartheta_1, d)$, i.e. $P\{m_2(\vartheta_2, d) - m_1(\vartheta_1, d) \le U(Y_1, Y_2, d)\} \ge 1 - \alpha$ for all $d \in \mathcal{D}$, where α denotes the pre-specified significance level and Y_ℓ the vector of observations from group $\ell = 1, 2$. Similarly, let $L(Y_1, Y_2, d)$ denote a $1 - \alpha$ pointwise lower confidence bound on $m_2(\vartheta_2, d) - m_1(\vartheta_1, d)$. Using these pointwise confidence bounds we can deduce a confidence interval for the maximum absolute difference between the two models $\max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)|$ over the region $\mathcal{D}$, that is

$$P\left\{\max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)| \le \max\left\{\max_{d \in \mathcal{D}} U(Y_1, Y_2, d), -\min_{d \in \mathcal{D}} L(Y_1, Y_2, d)\right\}\right\} \ge 1 - \alpha. \tag{2}$$

The proof is given in Appendix A.

For moderate sample sizes the pointwise confidence bounds $U(Y_1, Y_2, d)$ and $L(Y_1, Y_2, d)$ can be derived from the delta method [9]. Let $u_{1-\alpha}$ denote the $1 - \alpha$ quantile of the standard normal distribution. Then,

$$U(Y_1, Y_2, d) = m_2(\hat{\vartheta}_2, d) - m_1(\hat{\vartheta}_1, d) + u_{1-\alpha}\hat{\rho}(d)$$

and

$$L(Y_1, Y_2, d) = m_2(\hat{\vartheta}_2, d) - m_1(\hat{\vartheta}_1, d) - u_{1-\alpha}\hat{\rho}(d)$$

are the desired $1 - \alpha$ asymptotic pointwise upper and lower confidence bounds, respectively, for $m_2(\vartheta_2, d) - m_1(\vartheta_1, d)$. Here, $\hat{\vartheta}_\ell$ denotes the non-linear least squares estimate of ϑ_ℓ and $\rho^2(d) = \frac{\sigma_1^2}{n_1} g_1^T \Sigma_1 g_1 + \frac{\sigma_2^2}{n_2} g_2^T \Sigma_2 g_2$, where $g_\ell = \frac{\partial m_\ell(\vartheta_\ell, d)}{\partial \vartheta_\ell}$ denotes the gradient of m_ℓ, $\ell = 1, 2$ and the estimate $\hat{\rho}$ arises from ρ by replacing the true parameters ϑ_ℓ and σ_ℓ by their estimates $\hat{\vartheta}_\ell$ and $\hat{\sigma}_\ell$, respectively, in all quantities ($\ell = 1, 2$). The matrix $\frac{\hat{\sigma}_\ell^2}{n_\ell}\hat{\Sigma}_\ell = \frac{\hat{\sigma}_\ell^2}{n_\ell}\sum_{i=1}^{k_\ell}\frac{n_{\ell,i}}{n_\ell}\frac{\partial}{\partial\vartheta_\ell}m_\ell(x_{\ell,i}, \hat{\vartheta}_\ell)\left(\frac{\partial}{\partial\vartheta_\ell}m_\ell(x_{\ell,i}, \hat{\vartheta}_\ell)\right)^T$ is a consistent estimator of the covariance matrix of ϑ_ℓ ($\ell = 1, 2$). In practice, all major statistical software packages can be used to compute these quantities [10, 11].

Next we are interested in demonstrating that the maximum absolute difference in response between the two regression models in (1) over the covariate space $\mathcal{D}$ is not larger than a pre-specified margin $\delta > 0$. Formally, we test the null hypothesis

$$H : \max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)| \ge \delta \tag{3}$$

against the alternative hypothesis

$$K : \max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)| < \delta. \tag{4}$$

Consequently, using the confidence interval (2), equivalence is claimed if

$$\max \left\{ \max_{d \in \mathcal{D}} U(Y_1, Y_2, d), -\min_{d \in \mathcal{D}} L(Y_1, Y_2, d) \right\} < \delta.$$

Thus, we reject the null hypothesis H at level α and assume similarity of m_1 and m_2 if

$$-\delta < \min_{d \in \mathcal{D}} L(Y_1, Y_2, d) \quad \text{and} \quad \max_{d \in \mathcal{D}} U(Y_1, Y_2, d) < \delta. \tag{5}$$

2.2. Case study

To illustrate the methodology described in Section 2.1, we consider a dose finding trial for a weight loss drug given to patients suffering from overweight or obesity. This trial aims at comparing the dose response relationship for two regimens, namely a once-daily (o.d.) and a twice-daily (b.i.d.) application of the drug. Note that in this trial the primary objective is not to apply a joint model that includes both regimen, but rather treat both regimen separately and assess the similarity of dose response. Because this study has not been completed yet, we simulate data based on the assumptions made at the design state of the trial. For confidentiality reasons, we use blinded dose levels and all chosen dose levels denote the total daily dose. The limitations do not change the utility of the calculations below.

In this trial, the dose levels for the o.d. and b.i.d. regimens are given by 0.033, 0.1 and 1 and 0.067, 0.3, 1, respectively. Patients are thus randomized to receive either placebo or one of the six active treatments. In total, we assume that 350 patients are allocated equally across the seven arms, resulting in a sample size of 50 patients per treatment group. The primary endpoint of the study was the percentage of weight loss after a treatment duration of 20 weeks, with smaller values corresponding to a better treatment effect.

We used the `nls` function in R to compute the non-linear least squares estimates $\hat{\vartheta}_\ell$ of ϑ_ℓ and the standard errors necessary for calculating $U(Y_1, Y_2, d)$ and $L(Y_1, Y_2, d)$ from Section 2.1. The R code for this example and all other calculations in this paper is available from the authors upon request.

For this example, we fitted two Emax models: $m_1(\vartheta_1, d) = \vartheta_{1,1} + \vartheta_{1,2} \frac{d}{\vartheta_{1,3}+d}$ for the o.d. regimen and $m_2(\vartheta_2, d) = \vartheta_{2,1} + \vartheta_{2,2} \frac{d}{\vartheta_{2,3}+d}$ for the b.i.d. regimen, where $\vartheta_1 = (\vartheta_{1,1}, \vartheta_{1,2}, \vartheta_{1,3})$ and $\vartheta_2 = (\vartheta_{2,1}, \vartheta_{2,2}, \vartheta_{2,3})$. For the data set at hand, $\hat{\vartheta}_1 = (0.03, -5.17, 7.94)$ and $\hat{\vartheta}_2 = (-0.09, -6.56, 31.24)$. Figure 2a displays the fitted dose response models $m_1(\hat{\vartheta}_1, d)$ and $m_2(\hat{\vartheta}_2, d)$, $d \in [0, 1]$, together with the individual observations, where the y-axis is truncated to $[-7, 1]$ for better readability. Figure 2b displays the difference $m_2(\hat{\vartheta}_2, d) - m_1(\hat{\vartheta}_1, d)$ together with the associated 90% pointwise confidence intervals for each dose $d \in [0, 1]$. The maximum upper confidence bound for $\alpha = 0.1$ is $\max_{d \in \mathcal{D}} U(Y_1, Y_2, d) = 2.137$ at dose $d = 0.08$ and the minimum lower confidence bound is $\min_{d \in \mathcal{D}} L(Y_1, Y_2, d) = -1.848$ at the maximum dose $d = 1$. That is, the maximum difference in response between the two regimens over the dose range $\mathcal{D} = [0, 1]$ lies between -1.848 and 2.137. Therefore, similarity of the dose response curves can be claimed at level $\alpha = 0.1$ as long as δ is larger than 2.137, according to (5).

2.3. Simulations

We conducted a simulation study to investigate the operating characteristics of the method described in Section 2.1. We investigated coverage probabilities of the confidence intervals as well as Type I error rates and power of the test (5) for different scenarios. To simplify the simulations, we assumed balanced designs and that dose is the only covariate. For all simulations below, we generated data as follows:

Step 1: Specify the models m_1, m_2, their parameters ϑ_1, ϑ_2, a common variance σ^2 and the actual dose levels $d_{\ell,i}$.

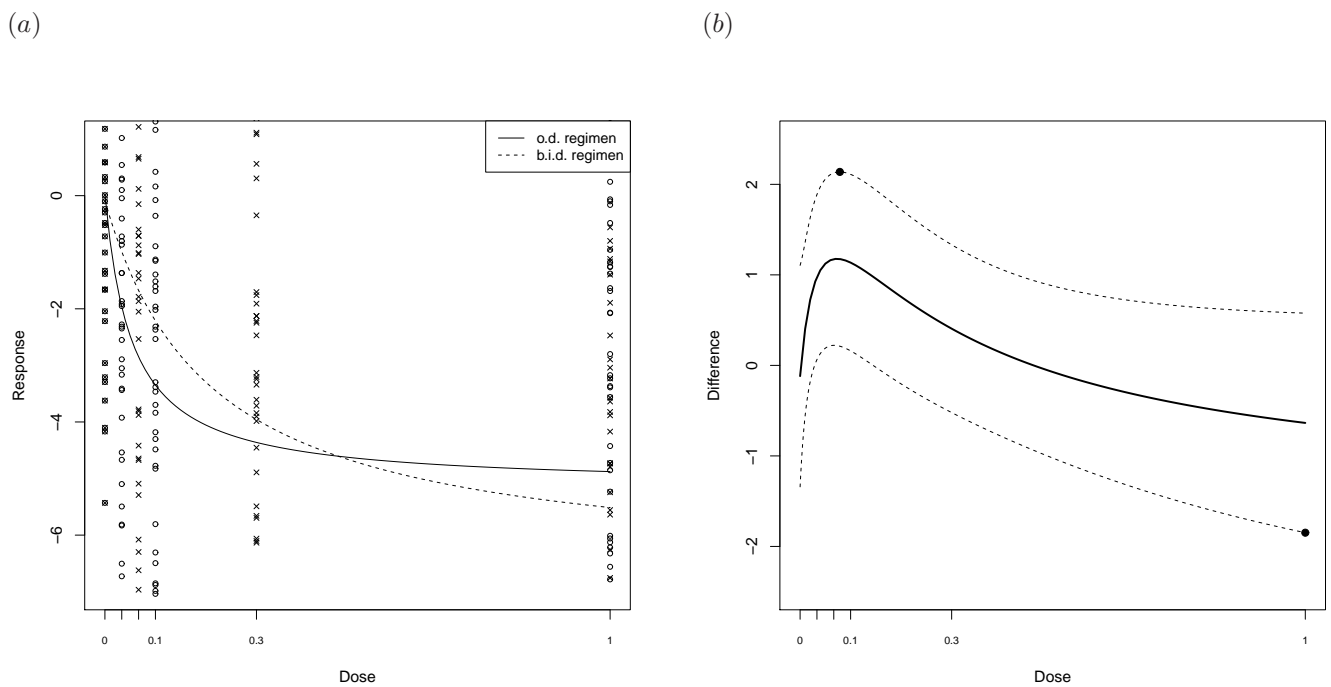**Step 2:** Generate $n_{l,i}$ values $m_\ell(\vartheta_\ell, d_{\ell,i})$ at each dose $d_{\ell,i}$.

Figure 2. Plots for the weight loss case study. (a) The fitted Emax model m_1 (m_2) for the o.d. (b.i.d.) regimen is given by the solid (dashed) line with observations marked by "x" ("o"). (b) Mean difference curve with associated pointwise 90% confidence bounds. Bold dots denote the maximum upper and minimum lower confidence bound over $\mathcal{D} = [0, 1]$.

Step 3: Generate normally distributed residual errors $\epsilon_{\ell,i,j} \sim N(0, \sigma^2)$ and use the final response data

$$Y_{\ell,i,j} = m_\ell(\vartheta_\ell, d_{\ell,i}) + \epsilon_{\ell,i,j}, \qquad j = 1, \ldots, n_{\ell,i}, \ i = 1, \ldots k_\ell, \ \ell = 1, 2. \tag{6}$$

This procedure is repeated in $10,000$ simulation runs.

2.3.1. Coverage probabilities In the following we report the coverage probabilities of the confidence intervals for the maximum absolute difference derived in (2) under two different scenarios.

Scenario 1 We start with the comparison of a linear and a quadratic model. More specifically, we chose the linear model $m_1(d) = d$ and the quadratic model $m_2(d) = 3\delta_1 + (1 - 4\delta_1)d + \delta_1 d^2$, $d \in [1, 3]$; see Figure 3a for $\delta_1 = 1$. We assumed identical dose levels $d_{\ell,i} = i$, $i = 1, 2, 3$ for both regression models $\ell = 1, 2$. Consequently, the two curves coincide at the two boundary doses $d = 1, 3$, and the maximum difference δ_1 occurs at dose $d = 2$. For each configuration of $\sigma^2 = 1, 2, 3$ and $\delta_1 = 1, 2, 3$ we used (6) to simulate $n_{\ell,i} = 10(50)$ observations at each dose level $d_{\ell,i}$, resulting in $n_\ell = 30(150)$, $\ell = 1, 2$.

The left side of Table 1 displays the coverage probabilities for $\alpha = 0.05, 0.1$. We observe that the nominal level of $1 - \alpha$ is reached in all cases under consideration, which confirms the theoretical result derived in (2). The confidence intervals are more accurate for larger sample sizes and smaller variances, because we used the asymptotic quantiles from the normal distribution. If, instead, we select the quantiles from the t distribution, the simulated coverage probabilities are closer to the nominal $1 - \alpha$ level (results not shown here). Note that the confidence bounds perform better for larger values of δ_1. This effect can be explained by considering the proof given in Appendix A. For smaller values of the true maximum absolute distance, given by δ_1 in this case, the probability increases that $m_2(\hat{\vartheta}_2, d_0) - m_1(\hat{\vartheta}_1, d_0) > 0$ for $d_0 \in \mathcal{D}$ and therefore the maximum $\max \{ U(Y_1, Y_2, d_0), -L(Y_1, Y_2, d_0) \}$ is more likely attained in $U(Y_1, Y_2, d)$. Because $m_2(\vartheta_2, d) - m_1(\vartheta_1, d) < 0$ for all choices of δ and for all d, inequality $(*)$ in Appendix A is strict and the coverage probability of the confidence interval becomes larger. For increasing sample sizes this effect disappears as the

(a) Scenario 1   (b) Scenario 2
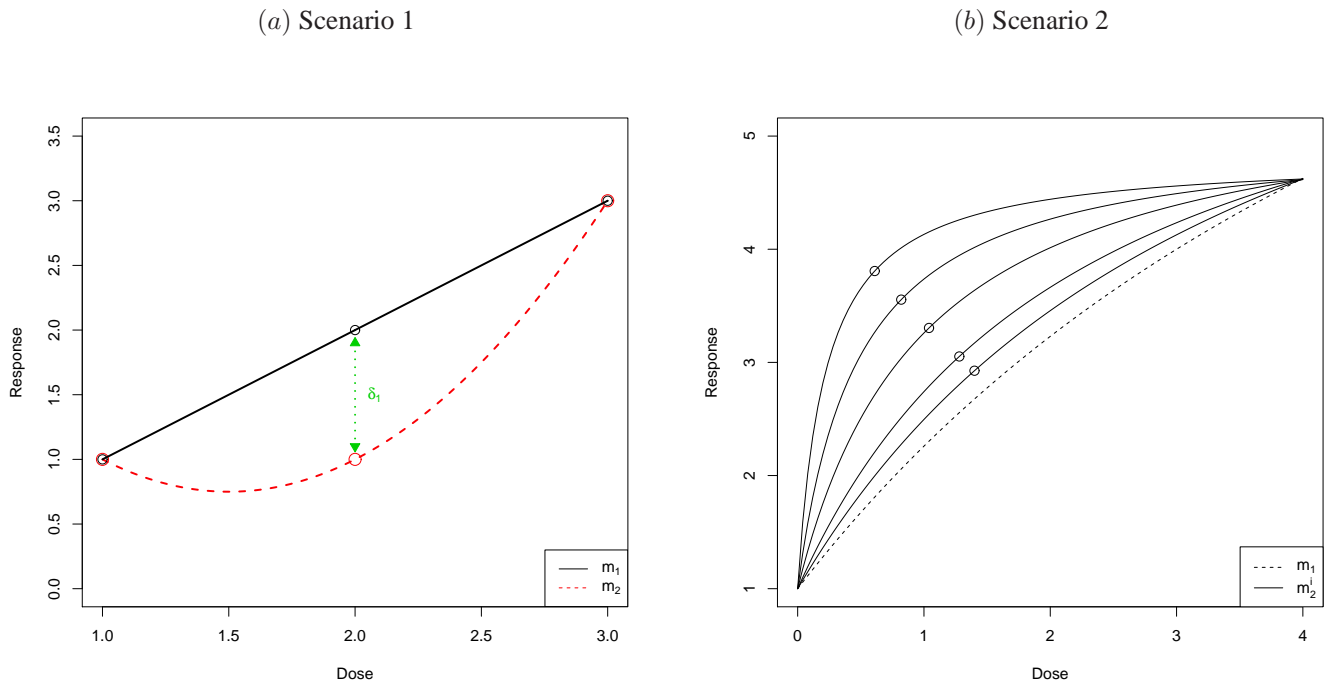

Figure 3. Graphical illustration of the two scenarios used for the simulations. Open dots in the left panel indicate the actual dose levels. In the right panel they indicate the doses where the maximum distance to the reference curve m_1 (dashed line) is observed.

estimation of $U(Y_1, Y_2, d)$ and $L(Y_1, Y_2, d)$ becomes more precise. However, asymptotically the coverage probability is at least $1 - \alpha$ under all configurations.

		Coverage probabilities				Type I error rates			
		$\alpha = 0.05$		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.1$	
δ_1	σ^2	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$
1	1	0.987	0.950	0.953	0.906	0.012	0.050	0.046	0.095
1	2	0.999	0.956	0.991	0.906	0.001	0.042	0.009	0.088
1	3	1.000	0.971	0.999	0.923	0.000	0.031	0.001	0.077
2	1	0.949	0.952	0.901	0.907	0.047	0.049	0.096	0.105
2	2	0.960	0.951	0.913	0.901	0.039	0.048	0.079	0.095
2	3	0.977	0.950	0.936	0.902	0.025	0.047	0.065	0.097
3	1	0.951	0.954	0.906	0.908	0.053	0.048	0.102	0.100
3	2	0.952	0.954	0.905	0.907	0.048	0.047	0.094	0.099
3	3	0.949	0.952	0.900	0.903	0.052	0.049	0.098	0.099

Table 1. Simulated coverage probabilities and Type I error rates for different configurations of δ_1, σ^2, α, and n_ℓ under Scenario 1.

Scenario 2   We now consider the comparison of two different Emax models, where the maximum distances with respect to the same reference model is 0.25, 0.5, 1, 1.5 and 2. More specifically, we compared the reference Emax model $m_1(d) = 1 + \frac{9.70d}{6.70+d}$ with

$$m_2^1(d) = 1 + \frac{6.88d}{3.60 + d}, \ m_2^2(d) = 1 + \frac{5.66d}{2.25 + d}, \ m_2^3(d) = 1 + \frac{4.52d}{1 + d}, \ m_2^4(d) = 1 + \frac{4.05d}{0.48 + d}, \ m_2^5(d) = 1 + \frac{3.82d}{0.22 + d}, \tag{7}$$

where the dose range is given by $\mathcal{D} = [0, 4]$. Note that the placebo response at $d = 0$ is 1 and the response at the highest dose $d = 4$ is 4.62 for all five models; see Figure 3b. The difference curve is given by $m_2^h(\vartheta_2^h, d) - m_1(\vartheta_1, d)$ for each $h = 1, 2, 3, 4, 5$. Note that the dose which produces the maximum difference is different for each h. More precisely, these doses are given by 1.4, 1.28, 1.04, 0.82 and 0.61 for $h = 1, \ldots, 5$; see again Figure 3b. The maximum absolute distance attained at each of these doses is denoted by $d_\infty = \max_{d \in \mathcal{D}} |m_2^h(\vartheta_2^h, d) - m_1(\vartheta_1, d)|$. We assumed 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 $\sigma^2 = 1, 2, 3$ and $d_\infty = 0.25, 0.5, 1, 1.5, 2$, we used (6) to simulate $n_{\ell,i} = 30$ observations at each dose level $d_{\ell,i}$, resulting in $n_\ell = 150$, $\ell = 1, 2$.

The left side of Table 2 displays the coverage probabilities for $\alpha = 0.05, 0.1$. As already seen under Scenario 1, the confidence intervals are more accurate for smaller variances (and larger sample sizes, results not shown here) and for increasing values of d_∞. As before, asymptotically the coverage probability is at least $1 - \alpha$ under all configurations investigated here.

(m_1, m_2)	d_∞	σ^2	Coverage probabilities		Type I error rates	
			$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
(m_1, m_2^1)	0.25	1	1.000	1.000	0.000	0.000
		2	1.000	1.000	0.000	0.000
		3	1.000	1.000	0.000	0.000
(m_1, m_2^2)	0.5	1	0.994	0.960	0.006	0.040
		2	1.000	0.993	0.000	0.007
		3	1.000	1.000	0.000	0.000
(m_1, m_2^3)	1	1	0.954	0.893	0.036	0.107
		2	0.963	0.903	0.047	0.097
		3	0.983	0.942	0.015	0.058
(m_1, m_2^4)	1.5	1	0.952	0.899	0.048	0.101
		2	0.962	0.913	0.038	0.087
		3	0.949	0.897	0.051	0.103
(m_1, m_2^5)	2	1	0.945	0.902	0.055	0.098
		2	0.942	0.889	0.068	0.118
		3	0.941	0.896	0.065	0.116

Table 2. Simulated coverage probabilities and Type I error rates for different model choices and configurations of σ^2 and α under Scenario 2, for $n_\ell = 150$, $\ell = 1, 2$.

2.3.2. Type 1 error rates For the Type I error rate simulations we investigated the two scenarios from Figure 3 for each configuration of $\alpha = 0.05, 0.1$ and $\sigma^2 = 1, 2, 3$. Further, we set $\delta = d_\infty$ in (3). For a fixed configuration, we generated data according to (6), fit both models, performed the hypothesis test (5) and counted the proportion of rejecting the null hypothesis H. Note that due to the choice of δ both Scenarios 1 and 2 belong to the null hypothesis H defined in (3). Thus, rejecting H would be a Type I error, i.e. we would erroneously claim similarity of the two dose response curves.

The right side of Table 1 displays the simulated Type I error rates under Scenario 1. We observe that the simulated Type I error rate is bounded by the nominal significance level α for all configurations investigated here, indicating that the hypothesis test (5) is indeed a level-α test, even under total sample sizes as small as 30. Note also that the significance level is actually well exhausted under many configurations. For small sample sizes and small values of δ the test becomes conservative, matching the observed performance of the confidence bounds shown in the left side of Table 1. Again, this conservatism disappears for large sample sizes.

The right side of Table 2 displays the simulated Type I error rates under Scenario 2. As before, the simulated Type I error rate is bounded by the nominal significance level α under all configurations. However, we observe that the test is very conservative for small values of d_∞, as already expected from the previously reported results on the coverage probabilities.

2.3.3. Power We now consider testing the null hypothesis H in (3) for $\delta = 1$, where in fact the maximum difference is smaller than 1. We start with the models chosen under Scenario 1 in Section 2.3.1. That is, we consider again the linear model $m_1(d) = d$ and compare it with the quadratic model $m_2(d, \delta_1) = 3\delta_1 + (1 - 4\delta_1)d + \delta_1 d^2$ for $d \in [1, 3]$ and under different values in the alternative, $\delta_1 \in \{0, 0.25, 0.5, 0.75, 0.9\}$; see Figure 4. The dose levels remain the same as under Scenario 1. For each configuration of $\sigma^2 = 1, 2, 3$ and $\delta_1 = 0, 0.25, 0.5, 0.75, 0.9$, we used (6) to simulate $n = 10(30, 50)$ observations under m_1 and m_2 at each dose level $d_{\ell,i}$, resulting in $n_\ell = 30(90, 150)$, $\ell = 1, 2$. Table 3 summarizes the results for $\alpha = 0.05, 0.1$. The power increases with decreasing values of δ_1. For large values of σ^2 the power remains small, even for $\delta_1 = 0$. In these cases we need larger sample sizes n_ℓ in order to achieve reliable results, as otherwise, due to the large variances, the confidence intervals in (2) become too wide and hence the test very conservative.
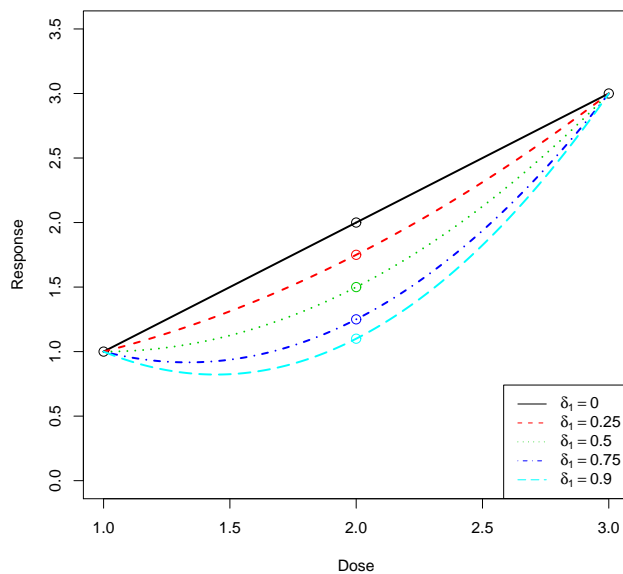

Figure 4. Graphical illustration of Scenario 1 used for the power simulations. Open dots indicate the actual dose levels.

Regarding Scenario 2, we tested the null hypothesis H in (3) using $\delta = 1$ and generating data under the models m_1, m_2^1 and m_2^2 defined in (7). Hence we simulate the performance of the test under the alternative K in (2.1) for different choices of σ and α. For the sake of brevity we restrict ourselves again to a fixed sample size of $n_\ell = 150$, $\ell = 1, 2$. Table 4 displays the simulated power. We observe that the test achieves high power, even for larger variances. However, the power decreases for an increasing true maximum distance between the models and for increasing variances.

2.4. Placebo-adjusted modeling

So far we assessed the similarity of two dose response models in terms of the maximum difference over the dose range under investigation. Sometimes one might be interested in adjusting for the placebo response, that is, the treatment effect relative to the placebo response, before comparing two dose response curves. In this case one has to modify the results from Section 2.1 as follows. Different to model (1), we consider the placebo-adjusted responses

$$Y_{\ell,i,j} = m_\ell(\vartheta_\ell, d_{\ell,i}) - m_\ell(\vartheta_\ell, 0) + \epsilon_{\ell,i,j}, \quad j = 1, \dots, n_{\ell,i}, \ i = 1, \dots k_\ell, \ \ell = 1, 2, \ d_{\ell,i} \in \mathcal{D}.$$

δ_1	σ^2	$\alpha = 0.05$			$\alpha = 0.1$		
		$n_\ell = 30$	$n_\ell = 90$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 90$	$n_\ell = 150$
0.00	1	0.211	0.966	0.999	0.426	0.988	0.999
0.25	1	0.170	0.939	0.997	0.377	0.974	0.999
0.50	1	0.102	0.731	0.917	0.268	0.843	0.958
0.75	1	0.046	0.306	0.444	0.143	0.433	0.583
0.90	1	0.023	0.111	0.144	0.074	0.195	0.245
0.00	2	0.000	0.052	0.390	0.000	0.218	0.614
0.25	2	0.000	0.044	0.326	0.000	0.194	0.558
0.50	2	0.000	0.028	0.199	0.000	0.130	0.393
0.75	2	0.000	0.013	0.076	0.000	0.071	0.186
0.90	2	0.000	0.007	0.036	0.000	0.043	0.101
0.00	3	0.000	0.000	0.003	0.000	0.002	0.075
0.25	3	0.000	0.000	0.002	0.000	0.002	0.066
0.50	3	0.000	0.000	0.002	0.000	0.001	0.050
0.75	3	0.000	0.000	0.001	0.000	0.001	0.023
0.90	3	0.000	0.000	0.001	0.000	0.000	0.015

Table 3. Simulated power for $\delta = 1$ and different configurations of δ_1, σ^2, α, and n_ℓ in Scenario 1.

(m_1, m_2)	d_∞	σ^2	$\alpha = 0.05$	$\alpha = 0.1$
(m_1, m_1)	0	1	0.986	0.996
(m_1, m_2^1)	0.25	1	0.980	0.992
(m_1, m_2^2)	0.5	1	0.871	0.938
(m_1, m_1)	0	2	0.719	0.873
(m_1, m_2^1)	0.25	2	0.657	0.833
(m_1, m_2^2)	0.5	2	0.442	0.655
(m_1, m_1)	0	3	0.350	0.622
(m_1, m_2^1)	0.25	3	0.286	0.553
(m_1, m_2^2)	0.5	3	0.183	0.400

Table 4. Simulated power for different model choices and configurations of σ^2 and α under Scenario 2, for $\delta = 1$ and $n_\ell = 150, l = 1, 2$.

The confidence interval for the maximum absolute difference between the placebo-adjusted curves is then given by

$$P\left\{\max_{d \in \mathcal{D}}|(m_2(\vartheta_2, d) - m_2(\vartheta_2, 0)) - (m_1(\vartheta_1, d) - m_1(\vartheta_1, 0))| \leq \max\left\{\max_{d \in \mathcal{D}} U'(Y_1, Y_2, d), -\min_{d \in \mathcal{D}} L'(Y_1, Y_2, d)\right\}\right\} \geq 1 - \alpha,$$

where $U'(Y_1, Y_2, d)$ and $L'(Y_1, Y_2, d)$ denote the pointwise confidence bounds for the placebo-adjusted difference derived by the delta method. For example,

$$U'(Y_1, Y_2, d) = (m_2(\hat{\vartheta}_2, d) - m_2(\hat{\vartheta}_2, 0)) - (m_1(\hat{\vartheta}_1, d) - m_1(\hat{\vartheta}_1, 0)) + u_{1-\alpha}\hat{\rho}'(d),$$

where $\hat{\rho}'(d)$ is calculated for the difference of two placebo-adjusted dose response curves. Proceeding, the null hypothesis of interest becomes

$$H': \max_{d \in \mathcal{D}}|(m_2(\vartheta_2, d) - m_2(\vartheta_2, 0)) - (m_1(\vartheta_1, d) - m_1(\vartheta_1, 0))| \geq \delta.$$

Statist. Med. **0000**, 00 1–17
Prepared using **simauth.cls**

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org   9

and following (5) we reject H' if

$$-\delta < \min_{d \in \mathcal{D}} L'(Y_1, Y_2, d) \quad \text{and} \quad \max_{d \in \mathcal{D}} U'(Y_1, Y_2, d) < \delta. \tag{8}$$

To illustrate this methodology, we revisit the weight loss case study from Section 2.2. The individual model fits remain the same, i.e. $m_1(\vartheta_1, d) = 0.03 - 5.17\frac{d}{7.94+d}$ for the o.d. regimen and $m_2(\vartheta_2, d) = -0.09 - 6.56\frac{d}{31.24+d}$ for the b.i.d. regimen. Figure 5a displays the placebo-adjusted model fits $m_1(\hat{\vartheta}_1, d) - m_1(\hat{\vartheta}_1, 0)$ and $m_2(\hat{\vartheta}_2, d) - m_2(\hat{\vartheta}_2, 0)$, $d \in [0, 1]$, together with the individual observations, where only the range $[-7, 1]$ is displayed on the vertical axis for better readability. Figure 5b displays the difference $(m_2(\hat{\vartheta}_2, d) - m_2(\hat{\vartheta}_2, 0)) - (m_1(\hat{\vartheta}_1, d) - m_1(\hat{\vartheta}_1, 0))$ together with the associated 90% pointwise confidence intervals for each dose $d \in [0, 1]$. In this example, the estimated placebo effects are very similar compared with the original fits. Thus, the placebo-adjusted difference curve and its confidence bounds do not change much compared with the previous results of Section 2.2, see Figure 2. The maximum upper confidence bound for $\alpha = 0.1$ is $\max_{d \in \mathcal{D}} U'(Y_1, Y_2, d) = 2.255$, again observed at dose $d = 0.08$, and the minimum lower confidence bound is $\min_{d \in \mathcal{D}} L'(Y_1, Y_2, d) = -1.730$ at dose $d = 1$. That is, the maximum placebo-adjusted difference between the two regimens over the dose range $\mathcal{D} = [0, 1]$ lies between -1.730 and 2.255. Therefore, similarity of the placebo-adjusted dose response curves can be claimed according to (8) as long as δ is larger than 2.255.
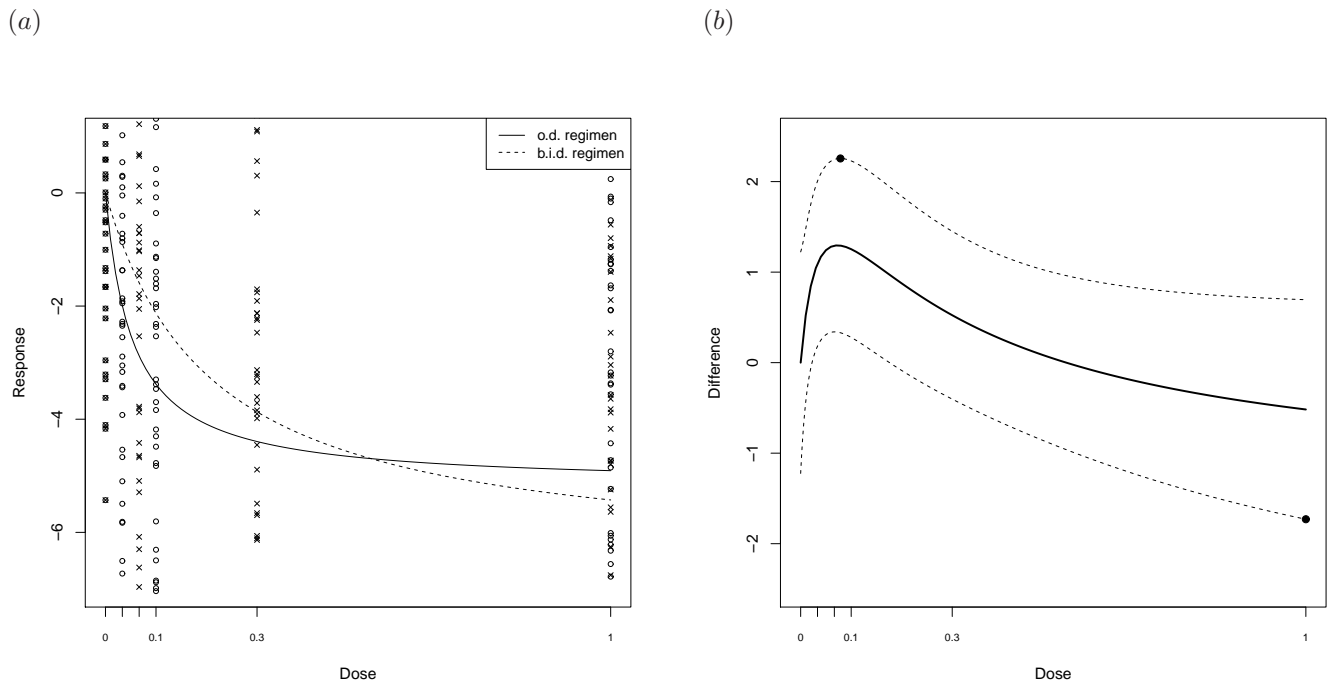
(a) (b)

Figure 5. Placebo-adjusted plots for the weight loss case study. (a) The placebo-adjusted Emax model fit m_1 (m_2) for the o.d. (b.i.d.) regimen is given by the solid (dashed) line with observations marked by "x" ("o"). (b) Mean difference curve with associated pointwise 90% confidence bounds. Bold dots denote the maximum upper and minimum lower confidence bound over $\mathcal{D} = [0, 1]$.

3. Assessing the similarity of two target doses

This section focuses on assessing the similarity of two target doses. We consider the difference between the minimum effective doses (MEDs) of two dose response curves from two non-overlapping subgroups. We derive confidence intervals and statistical tests to decide at a given level α whether the absolute difference of two MEDs is smaller than a prespecified margin η. Furthermore, we illustrate the proposed methodology by revisiting the case study from 2.2 and investigate its operating characteristics.

3.1. Methodology

Following Ruberg [1], the MED is defined as the smallest dose that produces a clinically relevant response Δ on top of the placebo effect (i.e. at dose $d = 0$). That is,

$$MED_\ell = MED_\ell(\vartheta_\ell) = \inf_{d \in \mathcal{D}} \{m_\ell(\vartheta_\ell, 0) < m_\ell(\vartheta_\ell, d) - \Delta\}, \ \ell = 1, 2. \tag{9}$$

From now on we assume strict monotonicity of the dose response curves m_ℓ such that (9) becomes

$$MED_\ell = MED_\ell(\vartheta_\ell) = m_\ell^{-1}(\vartheta_\ell, m_\ell(\vartheta_\ell, 0) + \Delta), \ell = 1, 2,$$

where the inverse is calculated with respect to d for fixed model parameters ϑ_1 and ϑ_2. Estimates for the MED are then given by

$$\widehat{MED}_\ell = m_\ell^{-1}(\hat{\vartheta}_\ell, m_\ell(\hat{\vartheta}_\ell, 0) + \Delta), \ell = 1, 2,$$

where $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$ are the non-linear least squares estimators for the true parameters. Due to the asymptotic normality of the estimates $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$, the estimated difference of the MEDs is approximately normal distributed [12]. To be more precise, the delta method [13] gives

$$\widehat{MED}_1 - \widehat{MED}_2 - (MED_1 - MED_2) \approx \mathcal{N}(0, \tau^2), \tag{10}$$

for

$$\tau^2 = \left(\frac{\partial}{\partial \vartheta_1} m_1^{-1}(\vartheta_1, \Delta_1)\right)^T \frac{\sigma_1^2}{n_1} \Sigma_1^{-1} \frac{\partial}{\partial \vartheta_1} m_1^{-1}(\vartheta_1, \Delta_1) + \left(\frac{\partial}{\partial \vartheta_2} m_2^{-1}(\vartheta_2, \Delta_2)\right)^T \frac{\sigma_2^2}{n_2} \Sigma_2^{-1} \frac{\partial}{\partial \vartheta_2} m_2^{-1}(\vartheta_2, \Delta_2)$$

and $\Delta_\ell = m_\ell(\vartheta_\ell, 0) + \Delta$, $\ell = 1, 2$. The variance τ^2 can be estimated by replacing ϑ_ℓ and Σ_ℓ by their estimates $\hat{\vartheta}_\ell$ and $\hat{\Sigma}_\ell$, $\ell = 1, 2$ (see Section 2.1). It then follows from (10) that a confidence interval for the difference of the MEDs is given by

$$P\left\{MED_1 - MED_2 \in \left[\widehat{MED}_1 - \widehat{MED}_2 - u_{1-\alpha/2}\hat{\tau}, \widehat{MED}_1 - \widehat{MED}_2 + u_{1-\alpha/2}\hat{\tau}\right]\right\} \overset{n \to \infty}{\longrightarrow} 1 - \alpha. \tag{11}$$

As we are interested in claiming similarity of two target doses, we consider the problem of testing

$$H'' \colon |MED_1 - MED_2| \geq \eta \qquad \text{against} \qquad K'' \colon |MED_1 - MED_2| < \eta. \tag{12}$$

In Appendix B we show that rejecting H'' if

$$|\widehat{MED}_1 - \widehat{MED}_2| < c, \tag{13}$$

gives an asymptotic α-level test, where c is the unique solution of the equation

$$\alpha = \Phi\left(\frac{c - \eta}{\hat{\tau}}\right) - \Phi\left(\frac{-c - \eta}{\hat{\tau}}\right). \tag{14}$$

Note that (14) can easily be solved by using Newton's algorithm [14].

3.2. Case study revisited

To illustrate the methodology in the previous subsection, we revisit the weight loss case study from Section 2.2. Recall the individual model fits $m_1(\hat{\vartheta}_1, d) = 0.03 - 5.17\frac{d}{7.94 + d}$ for the o.d. regimen and $m_2(\hat{\vartheta}_2, d) = -0.09 - 6.56\frac{d}{31.24 + d}$ for the b.i.d. regimen. We choose a clinically relevant difference of $\Delta = -3$, that is a weight loss of 3% compared to the placebo response is assumed to be a clinically relevant effect on top of the placebo response at dose $d = 0$. Therefore, $\widehat{MED}_1 = m_1^{-1}(\hat{\vartheta}_1, 0.03 - 3) = 0.073$, $\widehat{MED}_2 = m_2^{-1}(\hat{\vartheta}_2, -0.09 - 3) = 0.176$ and $\widehat{MED}_1 - \widehat{MED}_2 = -0.103$. Figure 6(a) displays the model fits $m_\ell(\hat{\vartheta}_\ell, d)$, together with the estimates $\widehat{MED}_\ell$, $\ell = 1, 2$.

The $1 - \alpha$ confidence interval for the true difference $MED_1 - MED_2$ is then given by $[-0.103 - u_{1-\alpha/2}0.119, -0.103 + u_{1-\alpha/2}0.119]$. For example, $MED_1 - MED_2 \in [-0.338, 0.133]$ for $\alpha = 0.05$ and $MED_1 - MED_2 \in [-0.300, 0.094]$ for $\alpha = 0.1$. Applying the test in (13) for $\alpha = 0.05$ allows us to claim similarity of the two MEDs whenever $\eta > 0.3$ as we have $c > 0.103 = |\widehat{MED}_1 - \widehat{MED}_2|$ in this case. Figure 6(b) displays the value of c as a function of η. For $\alpha = 0.1$ we obtain by similar calculations that η has to be larger than 0.255 in order to claim similarity.
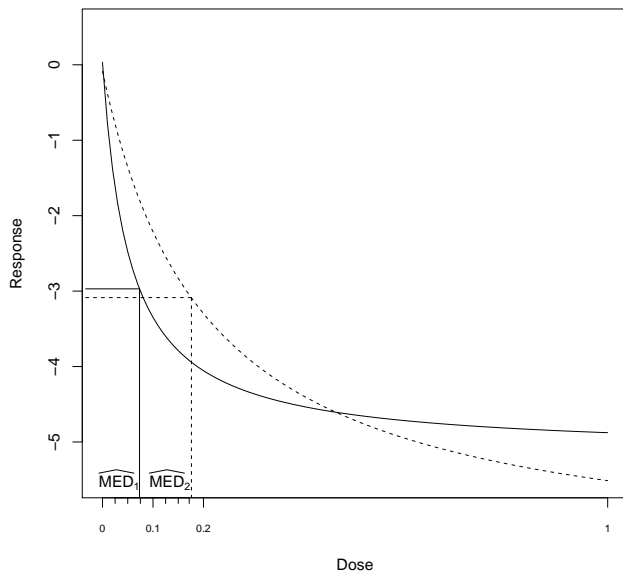
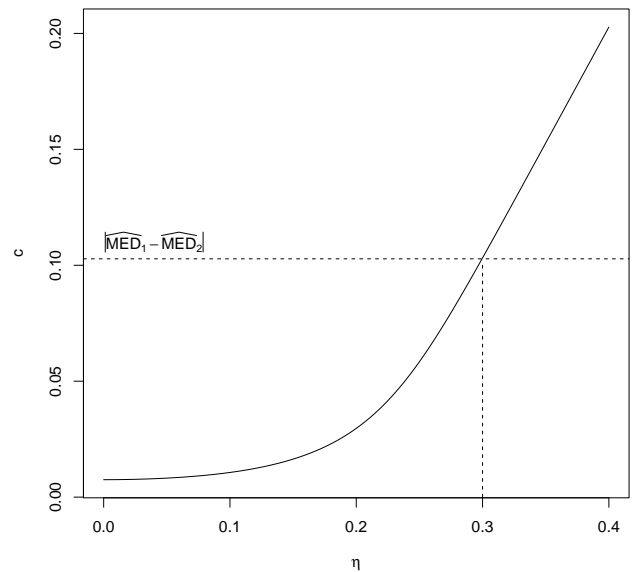(a)   (b)

Figure 6. Plots for the revisited weight loss case study. (*a*) The fitted Emax model m_1 (m_2) for the o.d. (b.i.d.) regimen is given by the solid (dashed) line, together with the estimated $MEDs$ for $\Delta = -3$. (*b*) Plot of c as a function of η. The dashed lines indicate the absolute difference of the MED estimates and the minimum choice of η in order to claim similarity for $\alpha = 0.05$.

3.3. Simulations

We now report the results of a simulation study to investigate the operating characteristics of the method described in Section 3.1. Adapting the data generation algorithm from Section 2.3, we investigated the coverage probabilities of the confidence intervals in (11) as well as the Type I error rates and power of the test (13) for different scenarios. All results were obtained using $10,000$ simulation runs.

3.3.1. Coverage probabilities

Scenario 3 We start with the comparison of two shifted Emax models $m_1(d, \vartheta_1) = \delta_1 + 5d/(1 + d)$ and $m_2(d, \vartheta_2) = 5d/(1 + d)$ over $\mathcal{D} = [0, 4]$, with identical dose levels $d_{\ell,i} = i - 1, i = 1, \ldots, 5$ for both regression models $\ell = 1, 2$; see Figure 7*a*. Because the models are shifted by the constant δ_1, the true difference $MED_1 - MED_2 = 0$ regardless of the value for Δ. For each configuration of $\sigma^2 = 1, 2$ and $\delta_1 = 1, 2, 3$ we used (6) to simulate $n_{\ell,i} = 6(30)$ observations at each dose level $d_{\ell,i}$, resulting in $n_\ell = 30(150)$, $\ell = 1, 2$.

The left side of Table 5 displays the coverage probabilities for $\alpha = 0.05, 0.1$. We observe that the coverage probability is at least $1 - \alpha$ under all configurations. The confidence intervals are more accurate for larger sample sizes and smaller variances, which confirms the asymptotic result from (11). Furthermore, the simulated differences between the MED estimates are very close to the true difference under all configurations (results not shown here).

Scenario 4 We now consider the comparison of the Emax model $m_1(d, \vartheta_1) = 1 + 4d/(2 + d)$ with the linear model $m_2(d, \vartheta_2) = 1 + 0.8d$ for the same set of doses as in Scenario 3. Note that the responses at doses $d = 0$ and $d = 3$ are the same in both models; see Figure 7*b*. For each configuration of $\sigma^2 = 1, 2, 3$ and $\Delta = 0.8, 1.6, 2.4$, we used again (6) to simulate $n_{\ell,i} = 6(30)$ observations at each dose level $d_{\ell,i}$, resulting in $n_\ell = 30(150)$, $\ell = 1, 2$.
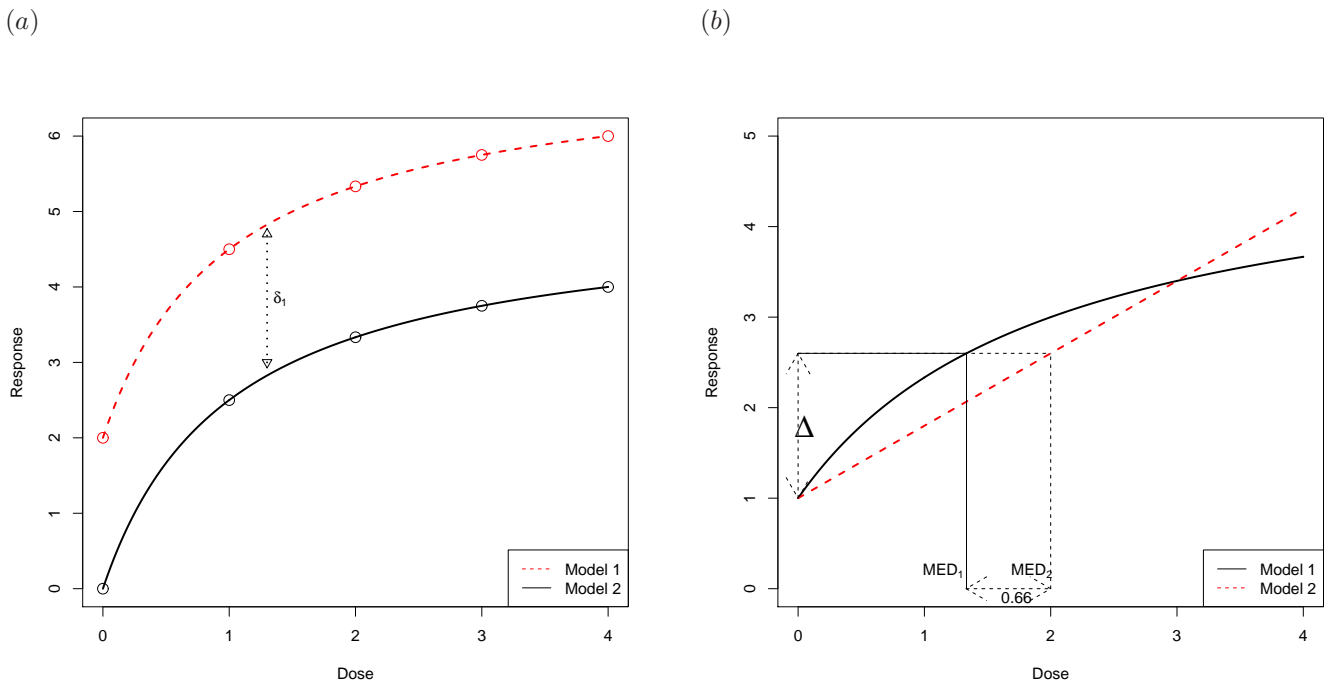
(a) (b)

Figure 7. Graphical illustration of Scenarios 3 and 4 used for the simulations. (a) displays the shifted Emax models with $\delta_1 = 2$. (b) displays the curves for Scenario 4, together with the MEDs corresponding to $\Delta = 1.6$.

		Coverage probabilities				Type I error rates			
		$\alpha = 0.05$		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.1$	
δ_1	σ^2	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$
1	1	0.979	0.959	0.941	0.907	0.050	0.050	0.103	0.103
2	1	0.982	0.958	0.945	0.909	0.053	0.048	0.105	0.105
3	1	0.980	0.961	0.946	0.908	0.053	0.052	0.099	0.105
1	2	0.996	0.967	0.977	0.917	0.049	0.051	0.104	0.101
2	2	0.996	0.968	0.978	0.922	0.052	0.049	0.103	0.101
3	2	0.995	0.966	0.976	0.916	0.045	0.049	0.100	0.099

Table 5. Simulated coverage probabilities and Type I error rates for different configurations of δ_1, σ^2, α, and n_ℓ under Scenario 3.

The left side of Table 6 displays the coverage probabilities for $\alpha = 0.05, 0.1$. As before, asymptotically the coverage probability is at least $1 - \alpha$ under all configurations investigated here, except for small sample sizes and $\Delta = 2.4$ (in which case the MEDs coincide). This is a direct consequence of the definition of the MED. Inverting an Emax model $m(\vartheta, d) = y = \vartheta_1 + \vartheta_2 d/(\vartheta_3 + d)$ gives $m_1^{-1}(\vartheta, y) = \vartheta_3(y - \vartheta_1)/(\vartheta_1 + \vartheta_3 - y)$. Therefore higher values of Δ result in being closer to the pole of m^{-1}, which is at $\vartheta_1 + \vartheta_2 = 5$ in this case. However, further calculations show that the results get better for larger sample sizes and the coverage probabilities converge quickly to their nominal values. Finally, the simulated differences between the MED estimates are very close to the true difference under all configurations, except again when the MEDs coincide (i.e. $\Delta = 2.4$; results not shown here).

3.3.2. Type 1 error rates For the Type I error rate simulations we investigated the two scenarios from Figure 7. We start with Scenario 3. Because $|MED_1 - MED_2| = 0$ for all values of Δ, we chose $\eta = 0$. For a fixed configuration

		Coverage probabilities				Type I error rates			
		$\alpha = 0.05$		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.1$	
Δ	σ^2	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$
0.8	1	0.964	0.965	0.929	0.908	0.036	0.025	0.077	0.068
1.6	1	0.953	0.946	0.922	0.903	0.051	0.040	0.098	0.087
2.4	1	0.920	0.949	0.877	0.916	0.069	0.057	0.137	0.116
0.8	2	0.989	0.968	0.960	0.928	0.050	0.026	0.101	0.061
1.6	2	0.967	0.956	0.936	0.913	0.053	0.045	0.111	0.088
2.4	2	0.918	0.932	0.870	0.901	0.069	0.060	0.143	0.124

Table 6. Simulated coverage probabilities and Type I error rates for different configurations of Δ, σ^2, α, and n_ℓ under Scenario 4.

of parameters, we generated data according to (6), fit both models, performed the hypothesis test (13) and counted the proportion of rejecting the null hypothesis H''. The right side of Table 5 displays the simulated Type I error rates under Scenario 3. We observe that the simulated Type I error rate is well exhausted at the nominal significance level α for all configurations investigated here, indicating that the hypothesis test (13) is indeed a level-α test, even under total sample sizes as small as 30.

The right side of Table 6 displays the simulated Type I error rates under Scenario 4. As before, the simulated Type I error rate is bounded by the nominal significance level α under almost all configurations. The test can be liberal small sample sizes and large values of Δ, matching the observed performance of the confidence bounds shown in the left side of Table 6. Again, this size inflation disappears for large sample sizes.

3.3.3. Power For the power simulations we again investigated the two scenarios from Figure 7 and start with Scenario 3. Because $|MED_1 - MED_2| = 0$ for all values of Δ, the power of the test depends only on the given threshold η. For the concrete simulations, we set $\Delta = 1$ and used $\delta_1 = 1$ for convenience. For each configuration of $\sigma^2 = 1, 2, 3$ and $\eta = 0.1, 0.2, 0.5, 1$, we used (6) to simulate $n = 10(30, 50)$ observations under m_1 and m_2 at each dose level $d_{\ell,i}$, resulting in $n_\ell = 30(90, 150)$, $\ell = 1, 2$. All configurations belong to the alternative in (12). Table 7 summarizes the results for $\alpha = 0.05, 0.1$. The power increases with increasing values of η. The power decreases for larger values of σ^2, especially for small values of η. In these cases we need larger sample sizes n_ℓ in order to achieve reliable results.

		$\alpha = 0.05$			$\alpha = 0.1$		
η	σ^2	$n_\ell = 30$	$n_\ell = 90$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 90$	$n_\ell = 150$
1	1	0.979	1.000	1.000	0.989	1.000	1.000
0.5	1	0.679	0.988	0.999	0.783	0.995	1.000
0.2	1	0.116	0.364	0.641	0.226	0.543	0.784
0.1	1	0.061	0.086	0.123	0.123	0.179	0.232
1	2	0.823	0.997	1.000	0.893	0.999	1.000
0.5	2	0.400	0.853	0.975	0.524	0.915	0.987
0.2	2	0.078	0.167	0.283	0.163	0.288	0.462
0.1	2	0.055	0.066	0.077	0.109	0.132	0.156

Table 7. Simulated power for different configurations of η, σ^2, α, and n_ℓ in Scenario 3.

For the final set of simulations, we revisit Scenario 4 and investigate the power for different values of σ^2 and Δ. We set $\eta = 0.8$ and $n_\ell = 30, 150$ for $\ell = 1, 2$ and summarize the results in Table 8. In alignment with all former results, the

performance of the test is worse in case of $\Delta = 2.4$ due to the already mentioned numerical problems when calculating the MEDs. In general, the power increases with increasing sample sizes and decreasing variances under all observed configurations. The power tends to 1 for $n \to \infty$.

Δ	σ^2	$\alpha = 0.05$		$\alpha = 0.1$	
		$n_\ell = 30$	$n_\ell = 150$	$n_\ell = 30$	$n_\ell = 150$
0.4	1	0.914	1.000	0.958	1.000
0.8	1	0.116	0.625	0.261	0.778
1.6	1	0.057	0.080	0.118	0.152
2.4	1	0.090	0.118	0.163	0.233
0.4	2	0.668	0.999	0.806	0.999
0.8	2	0.089	0.324	0.183	0.523
1.6	2	0.058	0.060	0.116	0.122
2.4	2	0.081	0.093	0.165	0.189

Table 8. Simulated power for different configurations of Δ, σ^2, α, and n_ℓ in Scenario 4.

4. Conclusions

The choice of the equivalence margins δ and η in (3) and (12), respectively, is a delicate problem. This choice has to be made by clinical experts, possibly with input from statisticians and other quantitative scientists, and depends on the particular application. In the specific problem of demonstrating bioequivalence guidance documents are available. For example, [15] discusses how the thresholds for bioequivalence hypotheses of the form considered in this paper can be defined in various settings. For the comparison of curves as considered in this paper we refer to Appendix 1 of [15], with emphasis on dissolution profiles on the basis of specific measures.

In this paper we investigated the problem of assessing similarity of dose response curves or target doses for two non-overlapping subgroups of patients for normally distributed responses. We leave several extensions of this basic problem for further research. For example, in many dose finding trials the assumption of normally distributed responses is hard to justify and exponential families for modeling the distribution of the responses have been considered in others settings [16]. Also, in certain applications it may be necessary to demonstrate similarity of dose response curves or target doses for more than two non-overlapping subgroups (such as more than two geographic regions or age classes). Finally, the research of this paper was motivated by the need of comparing the dose response information from males with female or Japanese with non-Japanese patients. In practice, one may equally be interested in comparing males or Japanese with the overall population rather than the complementary subgroup. These cases are more difficult to handle because the data for the specific subgroup of interest is also included the overall population, thus introducing correlations through nested data structure that need special attention. Again, we leave this topic for future research.

References

1. Ruberg, S.J. (1995). Dose response studies I. Some design considerations. *Journal of Biopharmaceutical Statistics*, **5**: 1–14.

2. Bretz F., Hsu J., Pinheiro J., Liu Y. (2008). Dose Finding – A Challenge in Statistics. *Biometrical Journal*, 50(4): 480–504.

3. ICH Harmonized Tripartite Guideline. (1994). Topic E4: Dose-response information to support drug registration. Available at http://www.ich.org/LOB/media/MEDIA480.pdf

4. Malinowski, H.J., Westelinck, A., Sato, J., Ong, T. (2008). Same Drug, Different Dosing: Differences in Dosing for Drugs Approved in the United States, Europe, and Japan. J. Clin. Pharmacol., 48, 900–908.

5. Ueasaka, H. (2009). Statistical Issues in global drug development with a close look at Japan and the Asian region. Invited presentation at the 57th Session of the International Statistical Institute. Available at http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0248.pdf

6. Dette, H., Moellenhoff, K., Volgushev, S., Bretz, F. (2016) Equivalence of regression curves. Available at http://arxiv.org/pdf/1505.05266v3.pdf

7. Liu, W., Bretz, F., Hayter, A.J., Wynn, H.P. (2009). Assessing Nonsuperiority, Noninferiority, or Equivalence When Comparing Two Regression Models Over a Restricted Covariate Region. *Biometrics*, **65**(4):1279–1287.

8. Pinheiro, J., Bretz, F., Branson, M. (2006). Analysis of dose response studies: Modeling approaches. Pp. 146–171 in Ting, N. (ed.) *Dose finding in drug development*, Springer, New York.

9. Serfling, R. (1980). Approximation Theorems of Mathematical Statistics. Wiley.

10. Pinheiro, J. C., Bates, D. M. (2000) Mixed-effects models in S and S-PLUS. *Statistics and Computing.* Springer: Heidelberg.

11. Bornkamp, B., Pinheiro,J. and Bretz, F. (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

12. Dette, H., Bretz, F., Pepelyshev, A., Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, 103(483), 1225-1237

13. Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. New York: Wiley.

14. Deuflhard, P. (2001). Newton methods for nonlinear problems: affine invariance and adaptive algorithms. Springer.

15. European Medical Agency (2010). Guideline on the investigation of bioequivalence. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf

16. Dette, H., Kettelhake, K., Bretz, F. (2014) Designing dose finding studies with an active control for exponential families. Available at http://arxiv.org/pdf/1410.8688v1.pdf

17. Lehmann, E.L., Casella, G. (1986). Testing statistical hypotheses. Springer.

A. Derivation of the confidence interval for the maximum absolute difference of two models

In the following we prove equation (2) from Section 2.1 To this end, let $d_0 \in \mathcal{D}$ such that

$$\max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)| = |m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0)|.$$

Hence

$$P\Big\{ \max_{d \in \mathcal{D}} |m_2(\vartheta_2, d) - m_1(\vartheta_1, d)| \leq \max \Big\{ \max_{d \in \mathcal{D}} U(Y_1, Y_2, d), - \min_{d \in \mathcal{D}} L(Y_1, Y_2, d) \Big\} \Big\}$$

$$= P\Big\{ |m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0)| \leq \max \Big\{ \max_{d \in \mathcal{D}} U(Y_1, Y_2, d), - \min_{d \in \mathcal{D}} L(Y_1, Y_2, d) \Big\} \Big\}$$

$$\geq P\Big\{ |m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0)| \leq \max \Big\{ U(Y_1, Y_2, d_0), -L(Y_1, Y_2, d_0) \Big\} \Big\}.$$

Now we distinguish two cases. If $m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0) \geq 0$ we have

$$P\Big\{ |m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0)| \leq \max \Big\{ U(Y_1, Y_2, d_0), -L(Y_1, Y_2, d_0) \Big\} \Big\}$$

$$\geq P\Big\{ m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0) \leq U(Y_1, Y_2, d_0) \Big\}$$

$$\geq 1 - \alpha,$$

as $U(Y_1, Y_2, d)$ is a $1 - \alpha$ pointwise upper confidence bound on $m_2(\vartheta_2, d) - m_1(\vartheta_1, d)$. Otherwise, $m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0) \leq 0$ and the same argument applies to $L(Y_1, Y_2, d)$, yielding

$$P\Big\{ |m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0)| \leq \max \Big\{ U(Y_1, Y_2, d_0), -L(Y_1, Y_2, d_0) \Big\} \Big\}$$

$$\overset{(*)}{\geq} P\Big\{ m_2(\vartheta_2, d_0) - m_1(\vartheta_1, d_0) \geq L(Y_1, Y_2, d_0) \Big\}$$

$$\geq 1 - \alpha.$$

B. Derivation of the test for similarity of two target doses

We show that the test from Section 3.1 has asymptotic level α. That is, under the null hypothesis

$$\lim_{n_1, n_2 \to \infty} P\left(|\widehat{MED}_1 - \widehat{MED}_2| \leq c \right) \leq \alpha. \tag{15}$$

Let the parameter space $\Theta = \Theta_0 \cup \Theta_1$ be partitioned in two disjoint sets referring to the hypotheses H'' and K'', respectively and c be given as the unique solution of the equation (14). Let $t = MED_1 - MED_2$ and denote the power function of the test by

$$G_{n_1,n_2}(\theta) = P_\theta\left(\left|\hat{t}\right| < c\right) \qquad (16)$$

for $\theta \in \Theta$. The assertion (15) is then equivalent to

$$\lim_{n_1,n_2 \to \infty} G_{n_1,n_2}(\theta) \leq \alpha \qquad \text{forall } \theta \in \Theta_0. \qquad (17)$$

Because $\hat{t}$ is the test statistic, we have

$$
\begin{aligned}
G_{n_1,n_2}(t) &= P_t(|\hat{t}| \leq c) = P_t(-c \leq \hat{t} \leq c) = P_t\left(\frac{-c-t}{\hat{\tau}} \leq \frac{\hat{t}-t}{\hat{\tau}} \leq \frac{c-t}{\hat{\tau}}\right) \stackrel{n_1,n_2 \to \infty}{\longrightarrow} \Phi\left(\frac{c-t}{\tau}\right) - \Phi\left(\frac{-c-t}{\tau}\right) \\
&=: \tilde{G}(t)
\end{aligned}
$$

Therefore the (asymptotic) power function $\tilde{G}$ coincides with the power of the test for the hypotheses $H\colon |t| = |MED_1 - MED_2| \geq \eta$ against $K\colon |t| = |MED_1 - MED_2| < \eta$ for normally distributed data $\left(\sim \mathcal{N}(\eta, \tau^2)\right)$. Hence, considering Lehmann et al. [17, p.81], this test is the uniformly most powerful unbiased test of given size α for the test problem (13) and the constant c is then determined by the condition $\tilde{G}(\eta) = \tilde{G}(-\eta) = \alpha$. This yields the choice of c in (14).