Taylor & Francis
Taylor & Francis Group

POINT AND INTERVAL ESTIMATORS OF THE TARGET DOSE IN CLINICAL DOSE-FINDING STUDIES WITH ACTIVE CONTROL

H.-J. Helms[1], N. Benda[2], and T. Friede[1]

[1]*Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee, Göttingen, Germany*
[2]*Federal Institute for Drugs and Medical Devices, Kurt-Georg-Kiesinger-Allee, Bonn, Germany*

In a clinical dose finding study with active control a new drug with several dose levels is compared with an active comparator drug. The main focus of such studies often lies on the estimation of a target dose that leads to the same efficacy as the control. This article investigates the finite sample properties of the maximum likelihood estimation of the target dose and compares several approaches for constructing corresponding confidence intervals under the assumption of a linear dose-response curve and normal error terms. Furthermore, the impact of deviations from the model assumptions regarding the error distribution is explored.

1. INTRODUCTION

Phase II dose-finding studies aim at the characterization of the dose response relationship of an investigational drug for which the principle proof of concept has been shown, as well as the estimation of a target dose that leads to a predefined response. The investigation of dose response relation and the subsequent selection of a target dose is considered to be a pre-requisite for the transition to a confirmatory pivotal study in Phase III. Several approaches have been proposed for the efficacious planning and analysis of such a dose-finding study. Bretz et al. (2005) and Pinheiro et al. (2006a) proposed a methodology that combines formal hypothesis testing for dose response with flexible modeling of the dose response relationship and estimating a target dose, i.e., a minimum effective dose that produces a clinical relevant effect. For a more detailed discussion we refer to Bretz et al. (2008) and Pinheiro et al. (2006b). The estimation of a minimum effective dose based on a given model can be regarded as a calibration

problem, a reverse process to regression, i.e., the estimation of a value for an independent variable that yields an expected outcome for the dependent variable equal to a predefined value. There is an extensive literature not only on these calibration problems, e.g., in quality control, but also related to dose estimation, e.g. Hsu and Berger (1999), Tamhane and Logan (2002), Morales et al. (2006), and Budtz-Jørgensen (2007). Although these approaches focus on a target dose that is defined by a specific effect difference to placebo, there is an increasing interest in describing the dose response relative to an active comparator to be included in the dose-finding trial which, however, has found little attention in the literature with the exception of Källén and Larsson (1999) and Dette et al. (2014). In addition, in some situations the use of the placebo may be unethical even in early stages of the development.

If an active comparator study is planned in Phase II, the interest lies in the determination of a dose range that ensures superiority over the competitor. If, on one hand, the probability of an adverse drug reaction is expected to increase with a higher dose, and, on the other, the efficacy is expected to improve with increasing dose, the dose that leads to the same efficacy as the comparator needs to be defined as precise as possible. Only if this information is available, the confirmatory clinical trials in Phase III can properly be designed. Also, manufactural standards should be considered in the precision requirements on the target dose estimation. Whereas this paper refers to a target dose defined by the dose that leads to the same mean efficacy as the active comparator, the application of a more general definition given by a predefined difference to the control is straightforward.

The precision, with which the target dose is estimated, depends on the dose-response shape, especially the steepness of the dose response as well as on the residual variance and the mean response of the control. Focusing on Phase II dose-finding studies with active control and a linear dose-response function the properties of the maximum likelihood estimator (MLE) of the target dose and several methods for confidence intervals of the target dose were investigated in extensive simulation studies. The investigated confidence interval approaches include an approximation of the standard error (SE) by the Δ-method (Ferguson, 1996), an extension of the method by Fisch and Strehlau (1993) to repeated measures, a parametric bootstrap (Efron, 1979), and a profile likelihood method (Pawitan, 2001). The MLE is characterized in terms of bias, variance, and mean squared error (MSE) whereas coverage probabilities and interval lengths are presented for the confidence intervals. Further the impact of the residual error assumptions will be analyzed by sensitivity analysis to investigate a potential robustness of one of the four methods. In particular, the question of which method is the most useful in terms of coverage probability, interval weights, and numerical effort.

To motivate the settings for the simulation studies, two examples of Phase II dose-finding studies with active control are presented briefly in what follows. Krum et al. (1998) report a study investigating the effect of an endothelin-receptor antagonist called bosentan on blood pressure in patients with essential hypertension. In this study 293 patients were randomly assigned to receive placebo or one of four oral doses of bosentan (100, 500, 1000, 2000 mg per day) or the angiotensin-converting-enzyme inhibitor enalapril (20 mg once daily) as active control for 4 weeks. The second example is a Phase II study presented in Chapple et al. (2004), which investigated the effect of solifenacin on patients with symptomatic idiopathic detrusor overactivity. A total of 255 patients were randomized to receive placebo or one of four doses of solifenacin (2.5, 5,

10, or 20 mg per day), or tolterodine 2 mg twice daily as active control. The primary endpoints were voids/24 h and mean volume voided (ml).

The paper is organized as follows. In Section 2 notation and statistical model are introduced, and an overview over the different methods is presented. In Section 3 the properties of the point estimator of the target dose including bias, variance, and MSE are investigated by deriving approximations for the expressions which are subsequently compared to exact values obtained through simulations for a variety of scenarios. In Section 4 the properties of the various confidence intervals of the target dose are investigated. Therefore, the coverage probability as well as the median interval length are simulated. In Section 5 the robustness of the methods for point and interval estimation of the target dose to deviations from the model assumptions such as normality of the error terms are explored. Finally, we close with a brief discussion and some recommendations.

2. NOTATION, MODEL, AND METHODS

2.1. Notation and Statistical Model

In this paper bold small letters such as $\boldsymbol{\alpha}$ or $\boldsymbol{a}$ define a vector and bold capital letters such as $\boldsymbol{\Sigma}$ or $\boldsymbol{A}$ a matrix. Also $\boldsymbol{A}'$ is the transpose of $\boldsymbol{A}$ and the Kronecker sum of two matrices is defined as $\mathbf{A} \oplus \boldsymbol{B}$. Further $\sim$ defines "distributed as", $\xrightarrow{\mathcal{D}}$ "convergence in distribution" and $\dot{\sim}$ defines "approximately distributed as."

With k groups the dose levels $d_1, \ldots, d_k$, and an active control (ac) the random variable of the jth person in the ith dose level can be written as

$$\boldsymbol{Y}_{ij} = f_{\boldsymbol{\theta}}(d_i) + \epsilon_{ij} \; i = 1, \ldots, k \; j = 1, \ldots, n_i,$$

where $f_{\boldsymbol{\theta}}(d_i)$ denotes the linear mean function, i.e., $f_{\boldsymbol{\theta}}(d_i) = \theta_0 + \theta_1 d_i$, and normally distributed error terms $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \; \forall i = 1, \ldots, k, j = 1, \ldots, n_i$. The random variables of the active control can be written as

$$\boldsymbol{Y}_{ac,j} = \mu + \epsilon_{ac,j} \; j = 1, \ldots, n_{ac},$$

with expected value μ and error terms $\epsilon_{ac,j} \sim \mathcal{N}(0, \sigma^2) \forall j = 1, \ldots, n_{ac}$. Let $n_d = \sum_{i-1}^{k} n_i$ be the sample size of all dose levels and $N = n_d + n_{ac}$ the total sample size. The linear model can be written as

$$\begin{aligned} \boldsymbol{Y}_d = (Y_{11}, \ldots, Y_{k,n_k})' &= \begin{pmatrix} \mathbf{1}'_{n_1} \cdots \mathbf{1}'_{n_k} \\ d_1 \mathbf{1}'_{n_1} \cdots d_k \mathbf{1}'_{n_k} \end{pmatrix}' \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + (\epsilon_{11}, \ldots, \epsilon_{k,n_k})' \\ &= \boldsymbol{X}\boldsymbol{\theta} + \epsilon_d, \\ \boldsymbol{Y}_{ac} = (Y_{ac,1}, \ldots, Y_{ac,n_{ac}})' &= \mu \cdot \mathbf{1}_{n_{ac}} + \epsilon_{ac}. \end{aligned} \tag{1}$$

To determine the target dose d^* the equation $f_{\theta}(d^*) = \theta_0 + \theta_1 d^* = \mu$ must be solved. In general the parameters of the dose-response curve, i.e., θ_0 and θ_1, as well as the expected value of the active control μ are unknown and must be estimated from the data. In the next section the necessary parameter estimators will be presented.

2.2. Point Estimator of the Target Dose *d**

Following standard theory for linear models (Mattai and Provost, 1992) we summarize briefly the parameter estimators in the linear model before deriving an estimator of *d**. The expected value of the active control can be estimated by using the mean value of the active control

$$\hat{\mu} = \bar{Y}_{ac\cdot} = \frac{1}{n_{ac}} \sum_{j=1}^{n_{ac}} Y_{ac,j} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n_{ac}}\right).$$

The parameter estimators of the linear dose response function can be written as

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_d \sim \mathcal{N}\left(\boldsymbol{\theta}, \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}\right).$$

Furthermore, the variance of the error terms is in practice unknown and must be estimated. Therefore, $\boldsymbol{X}_t = (\boldsymbol{X} \oplus 1_{n_{ac}})$ and $\boldsymbol{Y} = (\boldsymbol{Y}'_d \, \boldsymbol{Y}'_{ac})'$ are defined so that the variance can be estimated by

$$\widehat{\sigma^2} = \frac{1}{N-3} \boldsymbol{Y}'\left(\boldsymbol{I}_N - \boldsymbol{X}'_t(\boldsymbol{X}'_t \, \boldsymbol{X}_t)^{-t}\boldsymbol{X}_t\right)\boldsymbol{Y} \sim \frac{\sigma^2}{N-3} \cdot \chi^2_{N-3}.$$

If these estimators are used as plug-in estimators in the maximum likelihood equation of the target dose, the estimator of the target dose can be written as

$$\widehat{d^*} = \frac{\left(\hat{\mu} - \hat{\theta}_0\right)}{\hat{\theta}_1}. \tag{2}$$

In the next section several methods to compute a confidence interval for the target dose will be presented.

2.3. Confidence Intervals for the Target Dose *d**

2.3.1. Approximation of the Standard Error by the Δ-method. All
parameter estimators which are needed to determine *d** are normally distributed and can be written in a joint distribution

$$\hat{\boldsymbol{\theta}}_{d^*} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \hat{\mu} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\theta}_{d^*}, \boldsymbol{\Sigma}),$$

with expected value $\boldsymbol{\theta}_{d*} = (\theta_0 \ \ \theta_1 \ \ \mu)'$ and covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \cdot ((\boldsymbol{X}'\boldsymbol{X})^{-1} \oplus 1/n_{ac})$. The target dose *d** can then be written as a function $g(\boldsymbol{\theta}_{d*})$, i.e., $d^* = g(\boldsymbol{\theta}_{d*}) = (\mu - \theta_0)/\theta_1$. Then the gradient $\Delta g(.)$ of $g(.)$ is given by

$$\Delta g(\boldsymbol{\theta}_{d^*}) = \left(-\frac{1}{\theta_1} \quad -\frac{(\mu-\theta_0)}{\theta_1^2} \quad \frac{1}{\theta_1} \right).$$

From Cramér's theorem (Ferguson, 1996) with the use of the first-order Taylor approximation it can be shown that

$$\sqrt{N}\left(\widehat{d}^* - d^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, N \cdot \tau^2),$$

with

$$\tau^2 = \mathrm{Var}\left(\widehat{d}^*\right) = \Delta g(\boldsymbol{\theta}_{d^*})\boldsymbol{\Sigma}\Delta g(\boldsymbol{\theta}_{d^*})'$$

$$= \frac{1}{N}\frac{\sigma^2}{\theta_1^2}\left[\frac{\left(\sum w_i[d_i - d^*]^2\right)}{\left((\sum w_i)(\sum w_i d_i^2) - (\sum w_i d_i)^2\right)} + \frac{1}{w_{ac}} \right],$$

$w_i = n_i/N \; \forall i = 1, \ldots, k$ the weights of the ith dose level, $w_d = n_d/N$ the weight of all dose levels and $w_{ac} = n_{ac}/N$ the weights of the active control (Dette et al., 2008). Furthermore, let $\widehat{\tau^2}$ be the plug-in estimator of the unknown variance τ^2 so that a $(1 - \alpha)$-confidence interval of the target dose can be constructed as

$$CI_\Delta = \left[\widehat{d}^* - u_{1-\frac{\alpha}{2}}\sqrt{\widehat{\tau^2}}, \; \widehat{d}^* + u_{1-\frac{\alpha}{2}}\sqrt{\widehat{\tau^2}} \right],$$

with $u_{1-\frac{\alpha}{2}}$ the $(1 - \alpha/2)$-quantile of the standard normal distribution.

2.3.2. An Extension of the Method by Fisch and Strehlau (1993).

An alternative way to determine a confidence interval for the target dose d^* is an extension of the method by Fisch and Strehlau (1993) to repeated measures which was originally proposed for calibration problems without replications. Therefore, the expected value of the response of the active control is replaced by $\mu = \theta_0 + \theta_1 d^*$ in (1) which leads to a nonlinear model. Instead of performing a standard nonlinear regression the maximum likelihood equation of the estimator of the target dose $\widehat{d}^*$ can be rewritten in the following form:

$$\overline{Y}_{ac\cdot} - \widehat{\theta}_0 - \widehat{\theta}_1 d = 0$$

and the dose d should be considered as fixed, but unknown. All necessary estimators were defined in Section 2.2 and are identical with the maximum likelihood based estimators. The left side of the equation can be defined as a new random variable $V(d)$ with the following exact distribution:

$$V(d) = \overline{Y}_{ac\cdot} - \widehat{\theta}_0 - \widehat{\theta}_1 d = \left(\theta_0 - \widehat{\theta}_0\right) + \left(\theta_1 - \widehat{\theta}_1\right) \cdot d + \overline{\epsilon}_{\cdot ac} \sim \mathcal{N}\left(0, \tau(d)^2\right),$$

because μ can be written as $\mu = \theta_0 + \theta_1 d$. The variance $\tau(d)^2$ of the linear combination of normally distributed random variables can be written as

$$\tau(d)^2 = \text{Var}\left(\widehat{\theta}_0 + \widehat{\theta}_1 \cdot d\right) + \text{Var}(\overline{\epsilon}_{\cdot ac}) = \sigma^2 \cdot \left[\frac{1}{n_{ac}} + \frac{1}{n_d} + (d - \overline{d})^2/S_{dd}\right],$$

with $\overline{d} = \frac{1}{n_d}\sum_{i=1}^{k} n_i d_i$ and $S_{dd} = \sum_{i=1}^{k} n_i(d_i - \overline{d})^2$. Because d was considered as fixed, the only unknown variable in $\tau(d)^2$ is the error term σ^2 which can be estimated by $\widehat{\sigma^2}$ presented in Section 2.2. Finally the new variable $W(d)$ is defined as

$$W(d) = (V(d)/\tau(d))(\sigma^2/\widehat{\sigma}^2)^{-1/2} = \frac{V}{\left(\widehat{\sigma} \cdot \left[\frac{1}{n_{ac}} + \frac{1}{n_d} + (d - \overline{d})^2/S_{dd}\right]\right)} \sim t_{N-3}.$$

The t-distribution of the variable $W(d)$ only depends on the total sample size N. With the knowledge of the distribution of $W(d)$ a confidence interval of the target dose d^* can be computed by finding the limits of the confidence interval of $W(d)$ depending on d. This leads to the following $(1 - \alpha)$-confidence interval with $w(d)$ the realization of $W(d)$:

$$CI_{F\&S} = \left\{d| - t_{N-3,1-\frac{\alpha}{2}} \le w(d) \le t_{N-3,1-\frac{\alpha}{2}}\right\}.$$

This method does not always lead to finite confidence intervals, but the conditions for the appearance of infinite confidence intervals can be found in Fisch and Strehlau (1993).

2.3.3. Semiparametric Bootstrap.
All estimators $(\widehat{\mu}, \widehat{\theta}_0, \text{ and } \widehat{\theta}_1)$ which are necessary to estimate the target dose d^* are known with their explicit distribution functions. Hence, a parametric bootstrap (Efron, 1979) can be used to construct a confidence interval for the target dose d^*. Therefore, the bootstrap estimates have to be generated by simulation based on the parameter estimators gained from the data

$$\widehat{\boldsymbol{\theta}}^b = \widehat{\boldsymbol{\theta}} + \boldsymbol{Z}_\theta, \ \boldsymbol{Z}_\theta \sim \mathcal{N}\left(\boldsymbol{0}, \widehat{\sigma^2}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \text{ and}$$

$$\widehat{\mu}^b = \widehat{\mu} + Z_\mu, \ Z_\mu \sim \mathcal{N}\left(0, \frac{\widehat{\sigma^2}}{n_{ac}}\right).$$

Then the bootstrapped estimator of the target dose can be computed as

$$\widehat{d}^{*b} = \frac{\left(\widehat{\mu}^b - \widehat{\theta}_0^b\right)}{\widehat{\theta}_1^b}.$$

This simulation step is repeated n_{boot} times resulting in the vector $\widehat{\boldsymbol{d}}^{*b}$ of all bootstrapped parameter estimators. The confidence interval is then computed from the ordered vector of the bootstrapped target dose estimators by using, for example, the quantile method. The $n_{boot} \cdot \frac{\alpha}{2}$th and the $n_{boot} \cdot (1 - \frac{\alpha}{2})$th component of the sorted vector $\widehat{\boldsymbol{d}}^{*b}$ are selected to construct the $(1-\alpha)$-confidence interval

$$CI_b = \left[\widehat{d}^{*b}\left(n_{boot} \cdot \frac{\alpha}{2}\right), \widehat{d}^{*b}\left(n_{boot} \cdot (1-\frac{\alpha}{2})\right)\right].$$

2.3.4. Profile Likelihood. The idea of this approach is to set the target dose d^* to a fixed value and to optimize the maximum likelihood equation depending on this fixed target dose (Pawitan, 2001). The likelihood is given by

$$
\begin{aligned}
L &= L(d^*, \mu, \boldsymbol{\theta}, \sigma | \boldsymbol{Y}_d, \boldsymbol{Y}_{ac}) \\
&= (\gamma)^N \exp\left[-\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \boldsymbol{X}_i \boldsymbol{\theta})^2 + \sum_{j=1}^{n_{ac}} (Y_{ac,j} - \mu)^2}{2\sigma^2} \right] \\
&= (\gamma)^N \exp\left[-\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \boldsymbol{X}_i \boldsymbol{\theta})^2 + \sum_{j=1}^{n_{ac}} (Y_{ac,j} - \theta_0 - \theta_1 d^*)^2}{2\sigma^2} \right],
\end{aligned}
$$

with $\boldsymbol{X}_i$ the ith row of the matrix $\boldsymbol{X}$ in (1) and $\gamma = 1/\sqrt{2\pi\sigma^2}$. If the target dose is set to a fixed and known value d the parameter estimates can be computed depending on d so that the profile likelihood is defined as

$$
L(d) = \max_{\boldsymbol{\theta}, \sigma} L(d, \boldsymbol{\theta}, \sigma).
$$

For convenience the linear model with fixed d can be written as

$$
\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_d \\ \boldsymbol{Y}_{ac} \end{pmatrix} = \begin{pmatrix} \boldsymbol{1}_{n_1} & d_1 \boldsymbol{1}_{n_1} \\ \vdots & \vdots \\ \boldsymbol{1}_{n_k} & d_k \boldsymbol{1}_{n_k} \\ \boldsymbol{1}_{n_{ac}} & d \boldsymbol{1}_{n_{ac}} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_d \\ \epsilon_{ac} \end{pmatrix}
$$

$$
\boldsymbol{Y} = \boldsymbol{X}_p(d)\boldsymbol{\theta} + \epsilon.
$$

With this notation and standard methods for linear models (Mattai and Provost, 1992) the MLE for $\boldsymbol{\theta}$ and σ^2 depending on d can be constructed as

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_p(d) &= \left(\boldsymbol{X}_p(d)' \boldsymbol{X}_p(d) \right)^{-1} \boldsymbol{X}_p(d)' \boldsymbol{Y} \text{ and} \\
\widehat{\sigma}_p^2(d) &= \frac{1}{N-2} \cdot \boldsymbol{Y}' \left(I_N - \boldsymbol{X}_p(d)(\boldsymbol{X}_p(d)' \boldsymbol{X}_p(d))^{-1} \boldsymbol{X}_p(d)' \right) \boldsymbol{Y} \\
&= \frac{1}{N-2} \sum_{i=1}^{N} \left(Y_i - \boldsymbol{X}_{p,i}(d) \cdot \widehat{\boldsymbol{\theta}}_p(d) \right)^2,
\end{aligned}
$$

with $\boldsymbol{X}_{p,i}(d)$ the ith row of the matrix $\boldsymbol{X}_p(d)$. Then the profile likelihood equation of the target dose d can be written as

$$L(d) = \frac{1}{\left(\sqrt{2\pi\widehat{\sigma}_p^2(d)}\right)^N} \cdot \exp\left[-\frac{\sum_{i=1}^{N}\left(Y_i - \boldsymbol{X}_{p,i}(d)\hat{\boldsymbol{\theta}}_p(d)^2\right)}{2\widehat{\sigma}_p^2(d)}\right]$$

$$= \frac{1}{\left(\sqrt{2\pi\widehat{\sigma}_p^2(d)}\right)^N} \cdot \exp\left[-\frac{(N-2)}{2}\right].$$

The maximum $L_{\max} = \max\limits_{d} L(d)$ of the profile likelihood is at $\widehat{d^*} = (\widehat{\mu} - \widehat{\theta}_0)/\widehat{\theta}_1$ so that $L_{\max} = L(\widehat{d^*})$. The confidence interval for the target dose can be constructed by using the ratio $L(d)/L(\widehat{d^*}) \in [0,1]$. Then the Wilks likelihood ratio statistic W (see, e.g., Pawitan (2001)) has the following distribution under $H_0 : d = d^*$:

$$W = 2\log\frac{L\left(\widehat{d^*}\right)}{L(d)} \xrightarrow{\mathcal{D}} \chi_1^2$$

and it follows $P(L(d)/L(\widehat{d^*}) > c) = 1 - \alpha$ with $c = e^{-\frac{1}{2}\chi_{1,(1-\alpha)}^2}$. This generates a confidence interval for the target dose which is given by

$$CI_P = \left\{ d \mid \frac{L(d)}{L\left(\widehat{d^*}\right)} > c \right\}.$$

For example a 95%-confidence interval contains all d for which the condition $L(d)/L(\widehat{d^*}) > 0.1465$ holds.

3. PROPERTIES OF THE POINT ESTIMATOR OF d^*

In this section the properties of the point estimator $\widehat{d^*}$ of the target dose will be investigated through second-order Taylor approximation as well as simulation studies. The focus will lie on the assessment of the bias as well as the MSE.

It is not possible to calculate the exact expected value of $\widehat{d^*}$ in (2) because it is a ratio of dependent random variables. Therefore, an approximation with second-order Taylor approximation will be exploited here. As already described in Section 2.3.1, the target dose d^* can be written as a function $g(\boldsymbol{\theta}_{d^*})$ of the parameter vector $\boldsymbol{\theta}_{d^*}$. For the calculations the gradient $\boldsymbol{\Delta}$ and the Hessian matrix $\boldsymbol{H}$ of the function g are required, which are given by

$$\Delta g(\boldsymbol{\theta}_{d^*}) = \left(-\frac{1}{\theta_1}, \quad -\frac{(\mu-\theta_0)}{\theta_1^2}, \quad \frac{1}{\theta_1} \right), \quad \boldsymbol{H}(\boldsymbol{\theta}_{d^*}) = \begin{pmatrix} 0 & \frac{1}{\theta_1^2} & 0 \\ \frac{1}{\theta_1^2} & \frac{2\cdot(\mu-\theta_0)}{\theta_1^3} & -\frac{1}{\theta_1^2} \\ 0 & -\frac{1}{\theta_1^2} & 0 \end{pmatrix}.$$

With these the point estimator of the target dose can be approximated as

Table 1 Simulation scenarios for different variances σ^2, different slope levels θ_1, corresponding target doses d^*, and various sample sizes n per group

σ^2	θ_1	d^*	n
1	1.25	0.8	10,12,…, 30, 40,…, 100
	2	0.5	10,12,…, 30, 40,…, 100
	5	0.2	10,12,…, 30, 40,…, 100
2	1.25	0.8	10,12,…, 30, 40,…, 100
	2	0.5	10,12,…, 30, 40,…, 100
	5	0.2	10,12,…, 30, 40,…, 100

$$\widehat{d^*} \approx g(\boldsymbol{\theta}_{d^*}) + \Delta g(\boldsymbol{\theta}_{d^*})' \left(\widehat{\boldsymbol{\theta}}_{d^*} - \boldsymbol{\theta}_{d^*} \right) + \frac{1}{2} \left(\widehat{\boldsymbol{\theta}}_{d^*} - \boldsymbol{\theta}_{d^*} \right)' \boldsymbol{H}(\boldsymbol{\theta}_{d^*}) \left(\widehat{\boldsymbol{\theta}}_{d^*} - \boldsymbol{\theta}_{d^*} \right).$$

Hence, the expected value $E(\widehat{d^*})$ is approximated by $d^* + \frac{1}{2} tr(\boldsymbol{H\Sigma})$. Then an approximation of the bias of $\widehat{d^*}$ is given by

$$\frac{1}{2} \text{tr}(\boldsymbol{H\Sigma}) = \frac{\sigma^2}{N\theta_1^2} \frac{\left[\sum_{i=1}^{k} w_i[d^* - d_i] \right]}{\left(\left(\sum_{i=1}^{k} w_i \right) \left(\sum_{i=1}^{k} w_i d_i^2 \right) - \left(\sum_{i=1}^{k} w_i d_i \right)^2 \right)}.$$

If $d^* = (\sum_{i=1}^{k} w_i d_i)/(1 - w_{ac})$, the approximate bias is equal to 0. Also the approximate bias is increasing with larger error term variance σ^2 and is decreasing for larger total sample sizes N and for larger slopes of the dose-response function θ_1.

The properties of the point estimator of the target dose and its approximation described above are evaluated in a simulation study motivated by the example studies described in Section 1. We consider studies with $k = 5$ dose levels which are equidistant on [0, 1], and $d_1 = 0$, and $d_5 = 1$. The sample sizes $n_i = n$ per dose group for $i = 1, \ldots, k$, the sample size of the active control $n_{ac} = 2n$, the mean effect of the active control is set to $\mu = 1$, and the intercept of the linear response function is $\theta_0 = 0$. In all simulations the number of replications is $n_{sim} = 10,000$ and the number of bootstrap simulations per simulation run is $n_{boot} = 10,000$. Table 1 summarizes the considered simulation scenarios.

In Fig. 1 it can be seen that the derived approximation of the bias based on the real values explains the trend of the simulated bias quite well. However, there are limitations to the use of the approximation. In various situations, especially for small and moderate sample sizes, the estimated bias approximation is way too large compared to the estimator of the target dose. Therefore, we refrain from using the approximation for bias correction. Also, as we will see below, the bias is fairly small in comparison to the variance. The MSE is given by

$$MSE = \text{E}\left(\widehat{d^*} - d^* \right)^2 = \text{Var}\left(\widehat{d^*} \right) + \left[\text{Bias}\left(\widehat{d^*} \right) \right]^2.$$

The variance and the bias could not be calculated directly and therefore the same second-order Taylor approximation as in Section 3 was used to approximate the $\text{Var}\left(\widehat{d^*} \right)$ which leads to
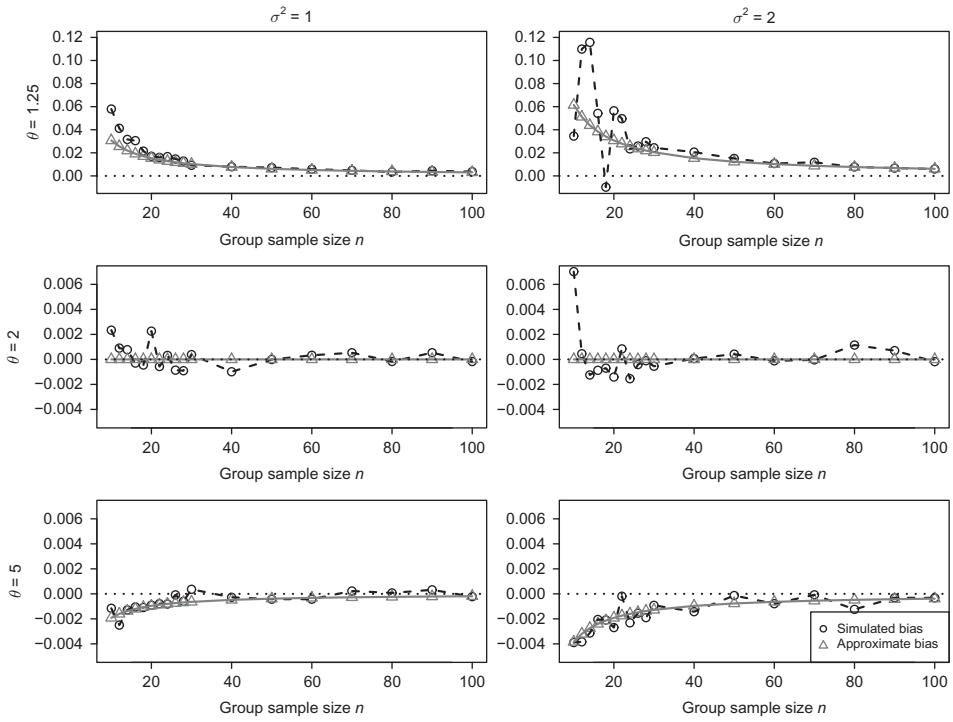
Figure 1 Approximated and simulated bias of the estimator of the target dose for various variances σ^2, slopes θ_1, and sample sizes n per group.

$$\text{Var}\left(\widehat{d^*}\right) \approx \Delta g(\boldsymbol{\theta}_{d^*})\boldsymbol{\Sigma}\Delta g(\boldsymbol{\theta}_{d^*})' + \frac{1}{2}\text{tr}\left([\boldsymbol{H}\boldsymbol{\Sigma}]^2\right).$$

Hence, the approximate MSE can be rewritten as

$$MSE \approx \Delta g(\boldsymbol{\theta}_{d^*})\boldsymbol{\Sigma}\Delta g(\boldsymbol{\theta}_{d^*})' + \frac{1}{2}\text{tr}\left([\boldsymbol{H}\boldsymbol{\Sigma}]^2\right) + \frac{1}{4}[\text{tr}(\boldsymbol{H}\boldsymbol{\Sigma})]^2.$$

In the simulation studies the MSE could be estimated by

$$\widehat{MSE} = \frac{1}{n_{sim} - 1}\sum_{i=1}^{n_{sim}}\left[\widehat{d^*}_{sim,i} - d^*\right]^2,$$

with $\widehat{d^*}_{sim,i}$ denoting the simulated target dose estimator in the ith simulation run. However, this estimator is highly sensitive to outliers which occur in several simulation runs with small group sample sizes. Therefore, instead of $\widehat{MSE}$ the median of the simulated squared differences $[\hat{d}^*_{sim,i} - d^*]^2$ can be used as an approximation of $\widehat{MSE}$ which is more robust to outliers. This is possible because $(\hat{d}^* - d^*)^2 \dot{\sim} gU$ with $U \sim \chi^2_1$ and $g = MSE$. So the expected value of U is $E(U) = 1$ and the $\text{Median}(U) \approx 1 \cdot (1 - 2/9)^3$. Hence, $E(gU) = gE(U) \approx g \cdot \text{Median}(U) \cdot (1 - 2/9)^{-3}$. With this approach the estimator of
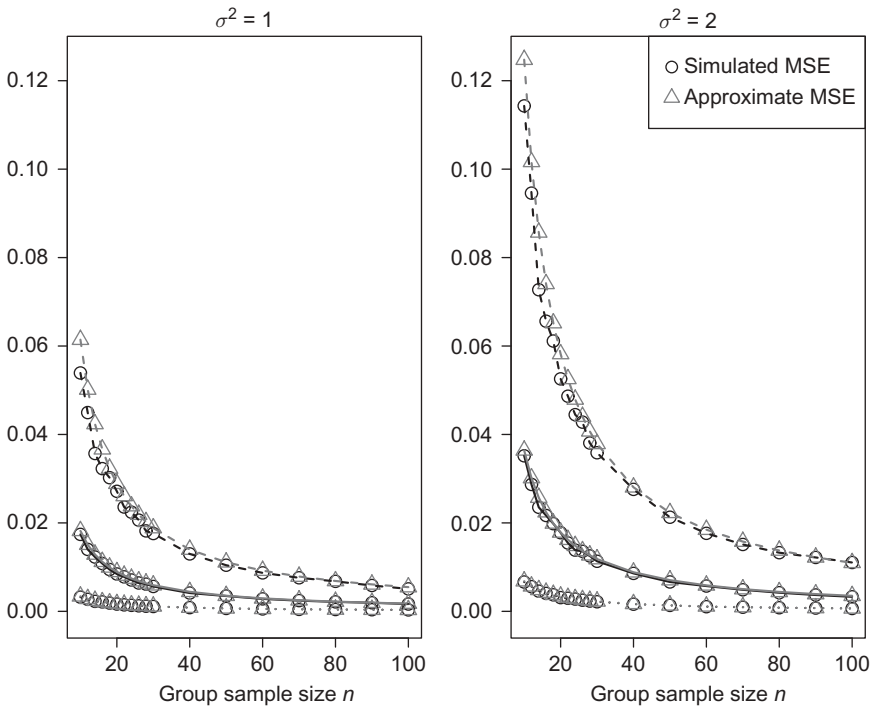
Figure 2 Mean squared error (MSE) of the estimator of the target dose shown by the simulated MSE and the approximate MSE for different slopes $\theta_1 = 1.25$ (- - -), 2 (—), 5 (…), various variances σ^2, and sample sizes n per group.

the MSE is not influenced by simulation caused outliers and extremes and therefore will be used in the following simulation study.

The MSE of the estimator of the target dose was simulated for the scenarios described in Table 1 and compared to the approximate MSE. In Fig. 2 it can be seen that the MSE, as with the bias above, depends on the error term variance σ^2, the slope θ_1, and the sample size n per dose group. The MSE is decreasing with larger n and larger slopes θ_1 and is increasing with larger error term variances σ^2. These numerical investigations demonstrate that the estimator of the target dose is not unbiased, but that the bias is relatively small for moderate sample sizes and not really a concern for estimating the target dose.

4. PROPERTIES OF THE CONFIDENCE INTERVALS OF THE TARGET DOSE

In this section the performance of the different methods for constructing a confidence interval will be assessed through simulation studies. The setting of the linear model in the simulation studies is the same as in Section 3. To investigate the behavior of the different interval methods the scenarios described in Table 1 will be investigated at a 95%-confidence level. In Fig. 3 it can be seen that the method of Fisch and Stehlau (*F&S*), the semi-parametric bootstrap (*Bootstrap*), and the profile likelihood (*PL*) perform
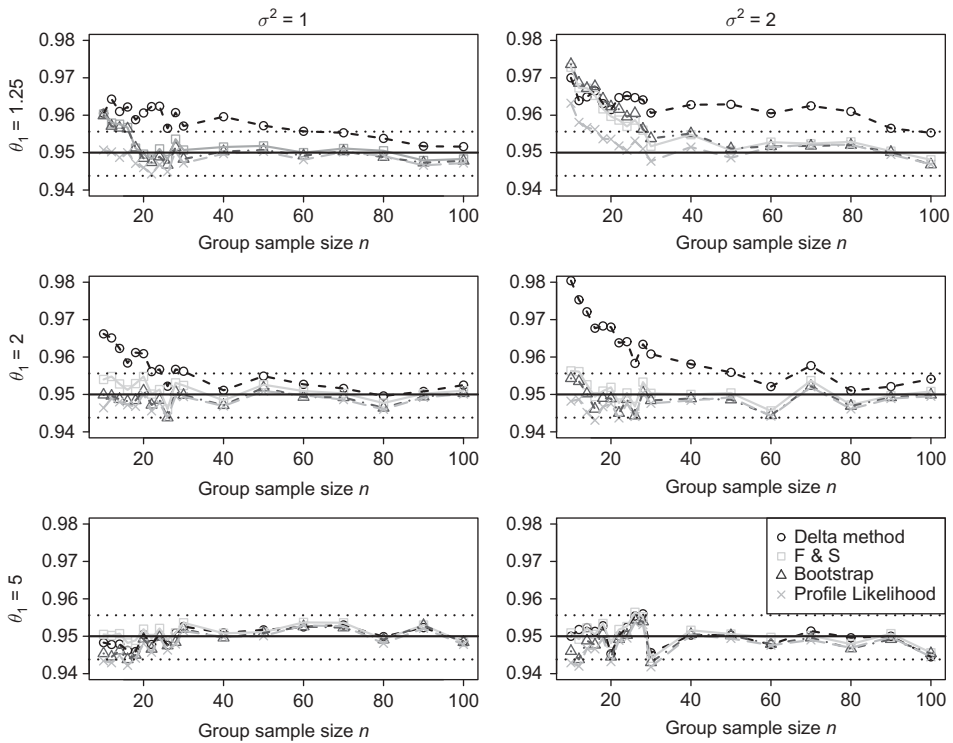
Figure 3 Coverage probability of the Δ-method, the method by Fisch and Strehlau ($F\&S$), the parametric bootstrap, and the profile likelihood for various variances σ^2, slopes θ_1, and group sample sizes n. The dotted lines indicate the simulation error with 99% probability.

quite well under nearly all scenarios. Furthermore, it can be seen that especially for small sample sizes the Δ-method (*Delta method*) is relatively conservative compared to the other three methods. To assess the performance of the different methods for the nominal coverage probability of 95% the dotted lines in Figs. 3 and 6 indicate the simulation error and were derived as $0.95 \pm u_{99.5}\sqrt{0.95 \cdot 0.05}/\sqrt{n_{sim}}$. This is roughly ± 0.005 which has been considered as a practical irrelevant deviation from the nominal level (Friede et al., 2007).

Now the median length of the different intervals will be investigated. Because a closed expression of the confidence interval only exists for the Δ-method (see Section 2.3.1) the length will be investigated via simulation. Therefore, the scenarios defined in Table 1 will be simulated under the same conditions as described in Section 3. In Fig. 4 it can be seen that the confidence interval length is relatively similar for all four methods. But in all scenarios the median length of the Δ-method is a little bit smaller than the length for the other three methods. This is possible even if the Δ-method is more conservative than the other three methods because the Δ-method generate symmetric confidence intervals and all other methods generates asymmetric confidence intervals. This will be explained in more detail in a simulated example next.
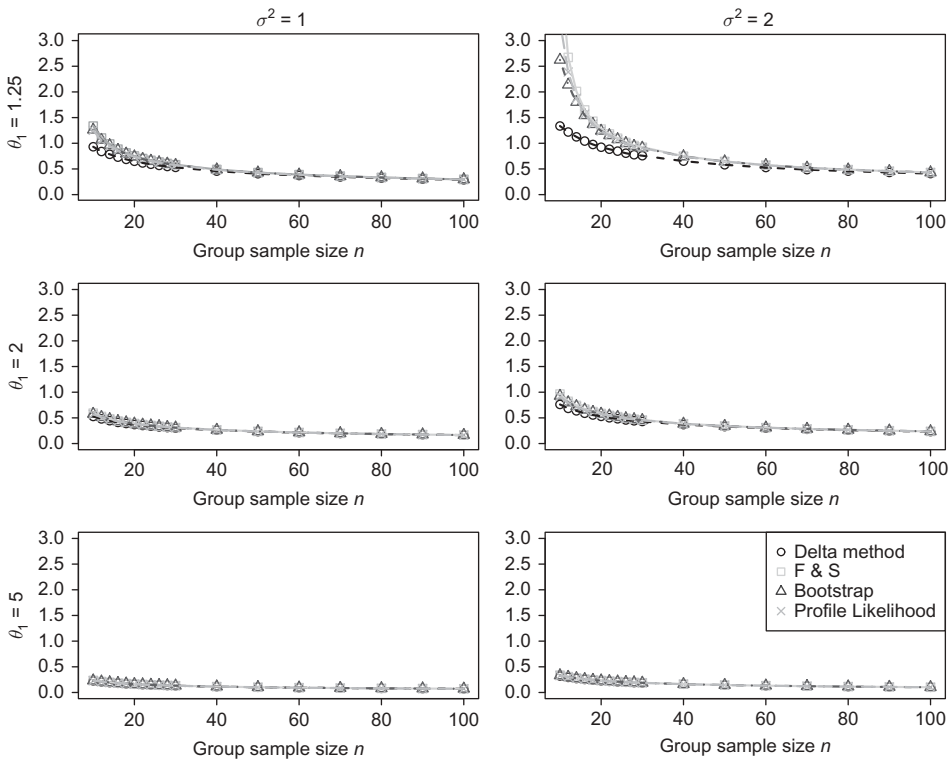
Figure 4 Median interval length of the Δ-method, the method by Fisch and Strehlau ($F\&S$), the parametric bootstrap, and the profile likelihood for various variances σ^2, slopes θ_1, and group sample sizes n.

To explain the behavior of the median length of the confidence intervals the scenario $\theta_1 = 1.25$, $\sigma^2 = 1$ with group sample size $n = 20$ will be investigated. Therefore, the simulated lengths were split in upper and lower confidence interval halves. The results of these simulations are presented in Fig. 5. It can be seen that the median length of the lower confidence interval half of the Δ-method is larger than the median length of the other methods, but on the other hand the median length of the upper confidence interval half is larger for the method of Fisch and Strehlau($F\&S$), the bootstrap, and the profile likelihood. The reason for this is that the Δ-method, in contrast to the other methods, generates symmetric confidence intervals. This leads to a slightly larger lower confidence interval half and a smaller upper confidence interval half.

5. ROBUSTNESS AGAINST NON-NORMAL RESIDUALS

In this section the robustness of the methods to deviations from the assumption of normally distributed error terms will be investigated. Therefore, different non-normal error terms will be simulated in the linear model for the scenarios explained
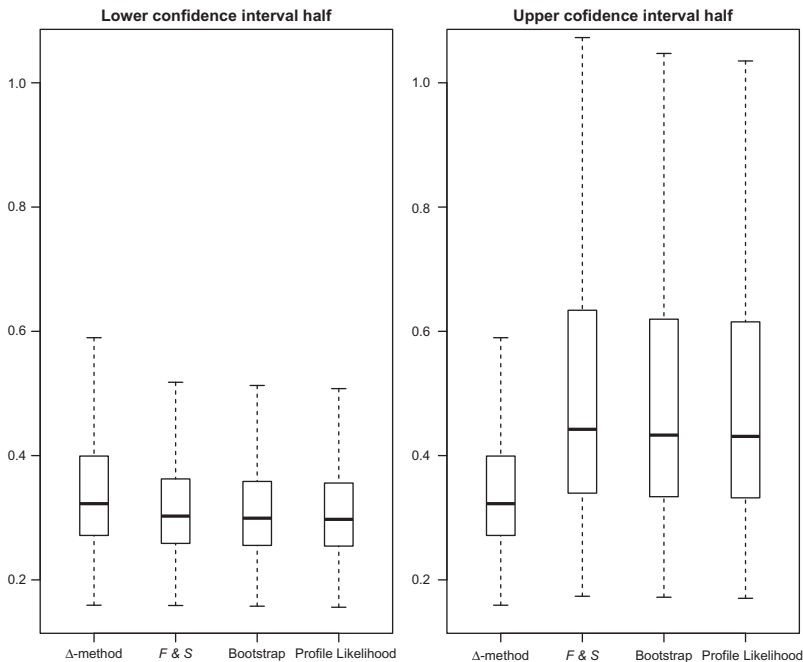
Figure 5 Length of the lower and upper interval of the Δ-method, the method by Fisch and Strehlau ($F\&S$), the parametric bootstrap and the profile likelihood for variance $\sigma^2 = 1$, slope $\theta_1 = 1.25$, and sample size $n = 20$ per group.

in Table 1. The methods described in the sections above were used to estimate the target dose and to construct intervals under the assumption of normally distributed error terms. As non-normal distributions the t-distribution with 4 degrees of freedom as well as a log-normal distribution were used. All error terms were scaled so that the expected value is equal to 0 and the variance is equal to 1. In Fig. 6 it can be seen, that for t and log-normal distributions the methods perform quite well and that there is basically no difference in the coverage probabilities compared to normally distributed error terms. Furthermore, there is a slight decrease in median interval length for all methods except of the Δ-method. Overall, it appears that these methods are quite robust to the deviations from the distributional assumption of the error term. This might be explained by the fact that the used total sample size, with a minimum of $N = 70$, is large enough for large sample approximations by the central limit theorem.

6. REAL DATA EXAMPLE

In this section the methods introduced above are illustrated by a dose-finding study including placebo, four dose levels of Solifenacin (0, 2.5, 5, 10, 20 mg) and 2 mg Tolterodine as active control with sample sizes $\boldsymbol{n} = (n_1,\ldots, n_5)' = (36, 40, 37, 33, 34)$ and $n_{ac} = 37$ (Chapple et al., 2004). The primary endpoint is the reduction in "Voids/24 h" after 6 weeks from baseline. For the different dose levels and the active control only the
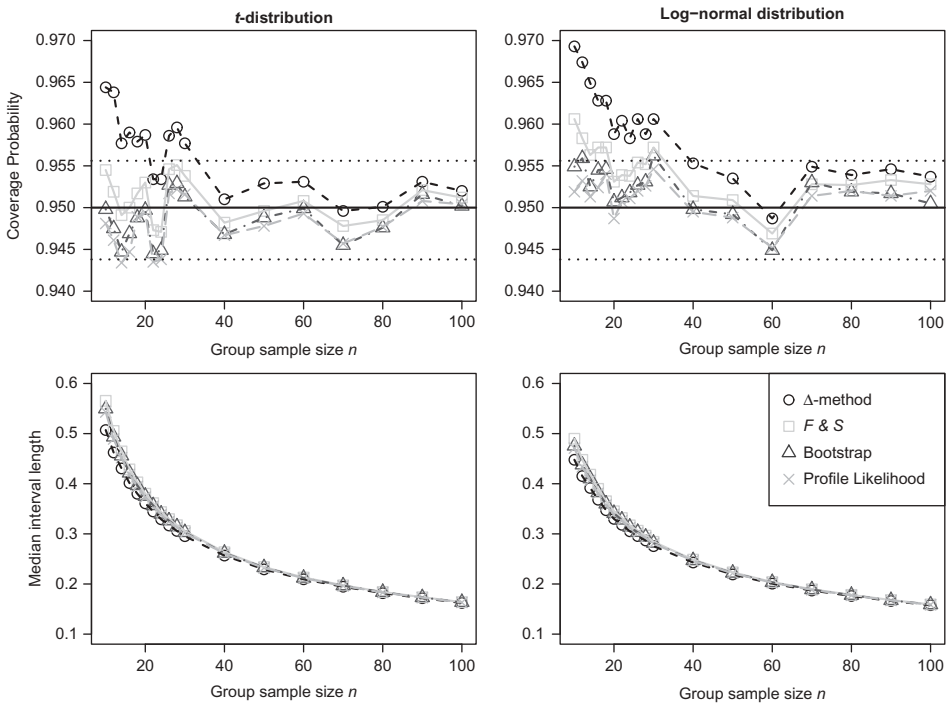
Figure 6 Coverage probability and median interval length of the Δ-method, the method by Fisch and Strehlau (*F&S*), the parametric bootstrap, and the profile likelihood for *t*-distributed and log-normal distributed error terms with variance $\sigma^2 = 1$, slope $\theta_1 = 2$, and various group sample sizes n. The dotted lines indicate the simulation error (99% probability interval).
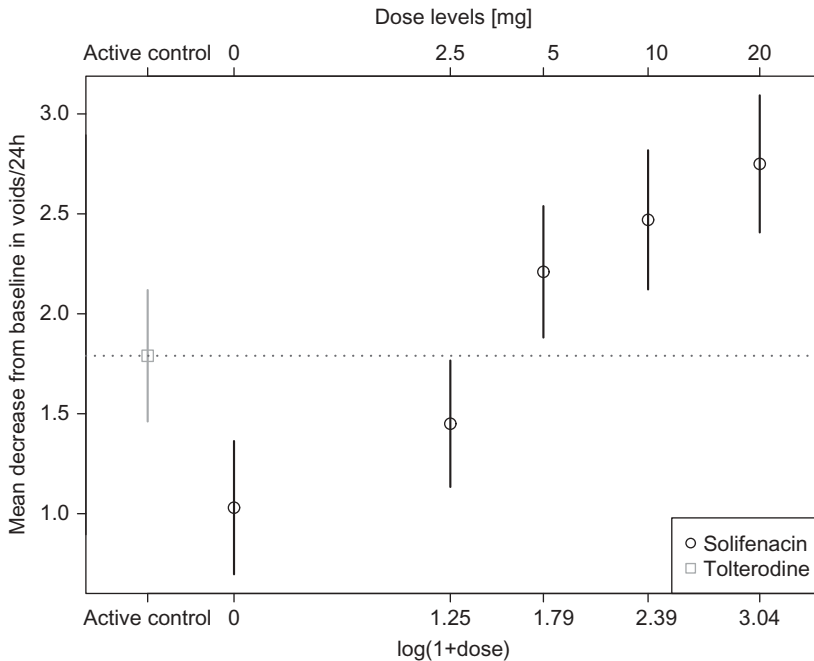

Figure 7 Voids/24 h decrease from baseline of the experimental drug dose levels and active control at the end of the study displayed as mean ± standard error (SE) reported in Chapple et al. (2004).

Table 2 The estimated target dose and the corresponding confidence intervals (CI) of the Δ-method, the method of Fisch and Strehlau (*F&S*), the parametric bootstrap (*Bootstrap*), and the profile likelihood approach (*PLH*) for the data presented in Fig. 7.

$\hat{d}^*$	CI-approach	95%-CI	CI-length
3.011	Δ-method	[0.236, 12.001]	11.765
	F&S	[0.000, 13.237]	13.237
	Bootstrap	[0.000, 12.671]	12.671
	PLH	[0.000, 13.018]	13.018

mean values of the reduction in "Voids/24 h" are reported in Chapple et al. (2004). The standard deviation (SD) of the error terms is not displayed but can be calculated using the reported p-values of the test statistics which lead to $\sigma \in [1.9, 2.5]$. We used $\sigma = 2$ to evaluate this example. For the log(1+dose) the results of the study are summarized in Fig. 7 as mean responses and SEs. Even though the individual patient data are not available it is possible to calculate the estimator of the target dose as well as the confidence intervals of the Δ-method, the method of *F&S*, and the parametric bootstrap (*Bootstrap*) which only need the sample sizes, the mean responses, and the SD by using the SAS macro "*DF_AC_LIN_MEAN*." The results of the confidence interval methods are shown in Table 2 for the target dose estimator $\hat{d}^* = 3.011$ mg. For some of the confidence intervals the lower interval limit was set to zero to guarantee a positive dose range. Further it is not possible to calculate the profile likelihood based interval on basis of the mean values alone. Therefore, a dataset was generated with identical mean responses and SD. It is somewhat surprising that with a total sample size of $N = 217$ it is only possible to exclude the maximum dose as potential target dose. The SAS macros for the linear dose-finding problem based on the raw data (*DF_AC_LIN*) as well as based on the mean values as presented here are available in the supplementary material.

7. CONCLUSIONS

In this paper we presented an MLE of the target dose and several approaches for confidence intervals of the target dose in clinical dose-finding studies with active control. We found that the MLE $\hat{d}^*$ of the target dose is slightly biased and that the size of the bias depends on the design of the dose-finding study. If the target dose $d^* = (\sum_{i=1}^{k} w_i d_i)/(1 - w_{ac})$, then the bias is relatively small. Also the total sample size *N*, the strength of the dose relationship (the slope θ_1), and the variance σ^2 of the error term have an influence on the size of the bias. For larger variances σ^2 the bias is increasing and for larger slopes the bias is decreasing. With larger sample sizes *N* the bias approaches zero.

Because the bias and the MSE could not be calculated exactly a second-order Taylor approximation was used instead. These approximations perform overall well and were compared with the simulated bias and MSE in numerous scenarios. We showed that the normal approximation based on the Δ-method is the simplest way to construct a confidence interval of the target dose, but the coverage probability is fairly conservative for small sample sizes. The method by *F&S* requires numerical optimization and is therefore

computationally more involved than the Δ-method. In terms of coverage probability however, the method by *F&S* outperforms the Δ-method. In nearly all scenarios the coverage probability is within the simulation error from the nominal level. One disadvantage of the method by *F&S* is that it cannot be easily extended to nonlinear models. The semi-parametric bootstrap is slightly more computer intensive than the other methods because the bootstrapped parameter estimates have to be generated n_{boot} times. In terms of coverage probability performs the bootstrap quite well in most of the scenarios considered. The profile likelihood performs quite similar to the method by *F&S*, and the semi-parametric bootstrap, but tends to be slightly more liberal compared to the other methods. The median confidence interval length is quite similar for all methods except of the Δ-method, there the median length is in all scenarios a little bit smaller than in the other methods. One reason for that is that the Δ-method always generates symmetric intervals whereas the other methods yield asymmetric intervals. Even if the median confidence interval is shorter for the Δ-method, it does not mean that the Δ-method is superior over the other methods in general. As Lehmann (1959) argues "short intervals are desirable when they cover the true parameter value but not necessarily otherwise". Therefore, we selected first the methods which hold the coverage probability well and from this subset the method with the shortest interval length would be used.

As illustrated in the supplementary material the results in terms of coverage probability are quite similar for designs with four or even three dose levels instead of five. Especially in the last design with only three dose levels it is quite difficult to use higher or more complex models than the linear model because of the small amount of information on the dose-response curve. To analyze the influence of a non-normal error term in the linear model a *t*-distribution as well as a log-normal distribution were simulated. We found that the non-normal error term has basically no influence on the coverage probability and on the median interval length. Therefore linear models or transformed models (linearization, log-linear dose response) with non-normal error terms can be evaluated quite well with the methods described in this paper, for moderate group sample sizes $n > 10$.

To our knowledge up to now a systematic comparison of the various approaches for confidence intervals of the target dose in dose-finding studies with active control is lacking. Here we presented such a comparison for linear models because these models are of interest in some situations and are subject of recent research (Demidenko et al., 2013). They are used in a wide range of applications such as transformation techniques into linear models (see, for example, Neter et al. (1996) and Box and Cox (1964)), generalized linear models via link function (McCullagh and Nelder, 1989), as well as regression situations with low information of the dose relationship. In these situations the MLEs of more complex models might not converge and the model has to be simplified, which is described in more detail by Jones et al. (2011) and Kirby et al. (2011).

We assumed variance homogeneity, i.e., the variance of the error terms to be the same for all dose levels as well as for the active control. One reason for this assumption was that in the case of equal variances and normally distributed error terms the least square estimators of the linear model theory are identical with the MLEs, resulting in a simple presentation of the various methods. For both approaches, least squares and maximum likelihood and all presented confidence interval methods, it is possible to use separate variance estimators for each dose level and/or only for the active control and the dose levels. This can be done straightforward but the MLH

estimators would differ from the least squares estimators and have to be computed numerically.

Of course for larger deviations from a linear dose-response curve or without a useful transformation it will be necessary to fit appropriate nonlinear models. The extension of some of these methods to nonlinear mean functions in the setting of active controlled dose-finding studies is subject to ongoing research with a focus on adaptive designs as presented in Jones et al. (2011) and Kirby et al. (2011).

SUPPLEMENTAL MATERIAL

Supplemental data for this article can be accessed on the publisher's website.

REFERENCES

Box, G. E. P., Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2):211–252.

Bretz, F., Hsu, J., Pinheiro, J., Liu, Y. (2008). Dose finding – a challenge in statistics. *Biometrical Journal* 50(4):480–504.

Bretz, F., Pinheiro, J. C., Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* 61(3):738–748.

Budtz-Jørgensen, E. (2007). Estimation of the benchmark dose by structural equation models. *Biostatistics* 8(4):675–688.

Chapple, C., Araño, P., Bosch, J., De Ridder, D., Kramer, A., Ridder, A. (2004). Solifenacin appears effective and well tolerated in patients with symptomatic idiopathic detrusor over-activity in a placebo- and tolterodine-controlled phase 2 dose-finding study. *BJU International* 93(1):71–77.

Demidenko, E., Williams, B. B., Flood, A. B., Swartz, H. M. (2013). Standard error of inverse prediction for dose-response relationship: approximate and exact statistical inference. *Statistics in Medicine* 32(12):2048–2061.

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.

Dette, H., Kiss, C., Benda, N., Bretz, F. (2014). Optimal designs for dose finding studies with an active control. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):265–295.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26.

Ferguson, T. (1996). *A Course in Large Sample Theory*. London, UK: Chapman & Hall.

Fisch, R. D., Strehlau, G. A. (1993). A simplified approach to calibration confidence sets. *The American Statistician* 47(3):168–171.

Friede, T., Mitchell, C., Müller-Velten, G. (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal* 49(6):903–916.

Hsu, J. C., Berger, R. L. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association*, 94(446): 468–482.

Jones, B., Layton, G., Richardson, H., Thomas, N. (2011). Model-based Bayesian adaptive dose-finding designs for a Phase II trial. *Statistics in Biopharmaceutical Research* 3(2):276–287.

Källén, A., Larsson, P. (1999). Dose response studies: how do we make them conclusive? *Statistics in Medicine* 18(6):629–641.

Kirby, S., Brain, P., Jones, B. (2011). Fitting emax models to clinical trial dose-response data. *Pharmaceutical Statistics* 10(2):143–149.

Krum, H., Viskoper, R. J., Lacourciere, Y., Budde, M., Charlon, V. (1998). The effect of an endothelin-receptor antagonist, bosentan, on blood pressure in patients with essential hypertension. *New England Journal of Medicine* 338(12):784–791.

Lehmann, E. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

Mattai, A. M., Provost, S. B. (1992). *Quadratic forms in random variables: Theory and applications*. Marcel New York: Dekker, INC.

McCullagh, P. Nelder, J. A. (1989). *Generalized linear models*, 2nd ed. London: Chapman & Hall.

Morales, K., Ibrahim, J., Chen, C., Ryan, L. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association* 101(473):9–17.

Neter, J., Kunter, M., Nachtsheim, C. (1996). *Applied Linear Statistical Models*, 4th ed. New York: McGraw-Hill Companies.

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and Inference Using Likelihood*. New York: Oxford University Press.

Pinheiro, J., Bornkamp, B., Bretz, F. (2006a). Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics* 16(5):639–656.

Pinheiro, J., Bretz, F., Branson, M. (2006b). *Analysis of Dose-Response Studies–Modeling Approaches*. Statistics for Biology and Health. New York: Springer.

Tamhane, A. C. Logan, B. R. (2002). Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *Journal of the American Statistical Association* 97(457):293–301.