Informative Simultaneous Confidence Intervals in Hierarchical Testing*

S. Schmidt; W. Brannath
University of Bremen, Faculty of Mathematics and Computer Sciences, Bremen, Germany

Keywords
Confidence interval, fixed-sequence test, family wise error rate, simultaneous coverage probability

Summary
Background and Objectives: In clinical trials involving multiple tests it is often difficult to obtain informative simultaneous confidence intervals (SCIs). In particular in hierarchical testing, no quantification of effects is possible for the first tested (and most important) hypothesis after its rejection. Our goal is a construction of SCIs that are always informative.

Methods: We present an approach where the level is split after rejection of each hypothesis to obtain an informative confidence bound. The splitting weights are continuous functions of the parameters. Our method is realizable by a simple algorithm and is illustrated by an intuitive graphical representation.

Results: We show theoretically and by an example that the new SCIs always provide information when a hypothesis is rejected. The power to reject the first hypothesis is not smaller than for the classical fixed-sequence procedure. The price for the extra information is a small power loss in the hypotheses proceeding the most important one.

Conclusions: Given the substantial gain in information, a small loss of power for the non-primary hypotheses seems often acceptable. Especially in the context of non-inferiority trials, this method is a useful alternative. The flexibility in the choice of the weight functions makes the procedure attractive for applications.

Correspondence to:
Sylvia Schmidt
Kompetenzzentrum für Klinische Studien Bremen
Linzer Str. 4
28359 Bremen
Germany
E-mail: sylviaschmidt@math.uni-bremen.de

doi: 10.3414/ME13-01-0140
received: December 9, 2013
accepted: March 26, 2014
epub ahead of print: June 27, 2014

1. Introduction
Consider a clinical trial situation where you want to make more than one confirmative assertion at the same time. For example, you want to show non-inferiority of a new drug over a standard therapy with regard to the primary endpoint as well as superiority for an important secondary endpoint. The non-inferiority claim in the primary endpoint is required for the licensing of the drug. The superiority claim in the secondary endpoint would confirm an additional benefit of the experimental treatment. Then you have a multiple testing problem, and according to the ICH guidelines E9 and the CPMP document “Points to consider on multiplicity issues in clinical trials”, you need a testing procedure that controls the familywise error rate \( P_{FW} \) (at least one false rejection) for any true parameter constellation.

There exist many multiple test procedures (see, e.g., [1] or [2] for a summary), which can be divided into single-step tests and stepwise tests. A special stepwise test is the fixed-sequence procedure [4], where the hypotheses have a prespecified order and are tested one after another stopping whenever a hypothesis could not be rejected. This is useful in situations where testing \( H_2 \) (in our example the superiority null hypothesis in the secondary endpoint) makes only sense after \( H_1 \) (in the example the non-inferiority null hypothesis in the primary endpoint) has been rejected. In this case, no adjustment of significance level is necessary to control the familywise error rate, because the whole level can be “shifted” after rejection of a hypothesis. Such a procedure is often denoted also as hierarchical test procedure.

The idea of transferring significance levels between hypotheses can be well illustrated by using graphs like in Figure 1 for a hierarchical test of two hypotheses. Bretz et al. [5] introduced a large class of multiple tests, including many Bonferroni-like and Gatekeeping procedures, that are represented by such graphs. We will explain and extend this illustration in this paper to develop a new method for simultaneous confidence intervals in trials where the fixed-sequence procedure is a natural starting point.

In non-inferiority trials it is common (and suggested by the ICH E9 guideline) to test non-inferiority by a one-sided (or two-sided) confidence interval. This permits to improve the claim with regard to the non-inferiority margin and has the additional advantage to permit a superiority claim if the lower confidence bound is positive, see e.g. the CPMP-document “Points to consider on switching between superiority and non-inferiority” of the EMA. A one-sided simultaneous confidence interval (SCI) is given by its lower bounds \( L_i \), i.e., SCI = \( (L_1, \infty) \times \ldots \times (L_m, \infty) \) where \( m \) is the

* Supplementary material published on our website www.methods-online.com
number of parameters of interest. One-sided SCIs are constructed in a way that their coverage probability \( P_\alpha(\theta \in \text{SCI}) = P_\alpha(\theta > L_1, \ldots, \theta_m > L_m) \) exceeds a prespecified limit \( 1 - \alpha \) for any true parameter \( \theta \).

Given a multiple test, an SCI is an extension of this test if the individual SCIs are consistent with the test decisions in the sense that each interval excludes a null hypothesis if and only if the multiple test rejects it. For example, the one-sided null hypothesis \( H_i^0 : \theta_i \leq 0 \) is rejected if and only if the corresponding confidence interval \( (L_i, \infty) \) and the null-hypothesis interval \( (-\infty, 0] \) are disjoint, i.e., if \( L_i \geq 0 \). Moreover, the confidence bound is informative if it does not stick to the border of the null-hypothesis interval, i.e., if \( L_i > 0 \). Otherwise, if the confidence interval is \((0, \infty)\), then no values of the alternative hypothesis would be excluded. This would make the confidence interval superfluous, because the “confidence interval” \((0, \infty)\) could also be concluded from the rejection of the null hypothesis \((-\infty, 0]\) and hence the SCI does not provide any additional information to the hypothesis test. For instance, in a non-inferiority trial, an improvement of the non-inferiority margin or a superiority claim requires that the lower confidence bound is strictly above the non-inferiority margin.

Single-step tests have a canonical extension to consistent simultaneous confidence intervals, but this is not straightforward for stepwise intervals. Some approaches were made by Stefansson, Kim and Hsu [6], Hsu and Berger [7] and, more recently, by Strassburger and Bretz [8] and Guilbaud [9], for a large class of Bonferroni-type stepdown tests including also the fixed-sequence procedure. A major problem of all approaches is that they only lead to informative rejections if all hypotheses can be rejected. Guilbaud [10] was able to improve this property for test procedures that are not \( \alpha \)-exhaustive, like the fallback procedure in [11], but there still remains a chance of non-informative confidence bounds. Since the fixed-sequence test is \( \alpha \)-exhaustive, no such improvement is possible in this case.

The method proposed in [12] leads to SCIs which are uniformly more informative than the Bonferroni intervals and the consistent SCIs for the Holm procedure [13] as constructed in [8]. An extension to the improvement of general union intersection tests with respect to informative rejections is also given in [12]. In this paper we extend the construction method of [12] to the fixed-sequence testing setting. More precisely, we present an approach that modifies the fixed-sequence procedure in a way that the corresponding SCIs are always informative for all hypotheses. The basic idea of our approach is that we do not shift the whole level of a rejected hypothesis to the next one, but keep some of the level for the construction of a positive confidence bound. A similar idea of error splitting for a fixed-sequence procedure in dose-finding studies was proposed by [14]. In order to loose not too much power for the following hypothesis, we specify the amount of reserved level in dependence of the observation. This is achieved by introducing individual weight functions that are continuous in the parameters. The resulting test procedure and its extension to SCIs can then be computed by an explicit and simple stepwise algorithm.

Our method will have no loss of power for the first hypothesis compared to the fixed-sequence procedure. The gain in information will be payed by a small loss in power of the second and following hypotheses, which is in general acceptable, since the first hypothesis is considered to be of greatest interest. An optimal trade-off between information gain and power loss can be found by variation of the underlying weight function.

In the next section, we introduce our idea with the help of graphical representations, starting with two hypotheses and concluding with a stepwise algorithm for the general case. In the Appendix, we give an alternative construction based on the idea of [12], which is less intuitive but demonstrates that the significance level is maintained by our method. In Section 3, we show how our new approach could be applied to a non-inferiority trial with an important secondary endpoint and give some simulation results. Finally, Section 4 gives a summary as well as an outlook on further extensions and research for this method.

### 2. Methods

#### 2.1 Notation

From now on, we will consider only superioriy null hypotheses, i.e., null hypotheses of the form \( H_i^0 : \theta_i \leq 0 \). The general case, required for instance for a non-inferiority null hypothesis (see also Section 3), is obtained by a simple parameter shift. For convenience, we will only discuss one-sided hypotheses in this paper.

Our construction of simultaneous confidence intervals will be based on a test for each individual hypothesis \( H_i^0 : \theta_i \leq \mu_i \) for real \( \mu_i \) and \( i = 1, \ldots, m \). We assume that these tests can be expressed in terms of individual p-values \( p_i(\mu_i) \), i.e., \( H_i^{\mu_i} \) is rejected at level \( \gamma \) if and only if \( p_i(\mu_i) \leq \gamma \). A typical example are p-values from asymptotic z-tests \( p_i(\mu_i) = 1 - 0.5 \Phi(\hat{\theta}_i - \mu_i / s_i) \), where \( \hat{\theta}_i \) is asymptotically normal with mean \( \theta_i \) and standard error \( s_i \). We will always assume the very natural property that the individual p-values are continuously increasing from 0 to 1 in the parameter \( \mu_i \). Under this natural assumption, one can find a unique solution of the equation \( p_i(\mu_i) = y \) for any \( y \in (0, 1) \). We will denote this solution by \( \gamma_i^{-1}(y) \). It gives a lower confidence bound for \( \mu_i \) with coverage probability \( 1 - y \). For example, in the case of z-tests, we simply obtain \( \gamma_i^{-1}(y) = \hat{\theta}_i - s_i \Phi^{-1}(1 - y) \).

#### 2.2 Graphical Description of Confidence Interval Construction

We extend the graphical representation introduced in [5] to include also the construction of confidence intervals. Figure 2a illustrates this new description for the case of only one hypothesis. It shows not only the null hypothesis of interest \( H_0^{10} \), but the whole nested family of individual hypotheses \( H_0^{\mu_i} \), that are all tested at significance level \( \alpha \). Due to the monotonicity of the individual p-values, this corresponds...
to a fixed-sequence test within the family. The lower confidence bound is then determined as \( L_1 = p_i^{-1}(\alpha) \).

Consider now the case of testing and constructing SCIs for two hypotheses with the fixed-sequence procedure. The graphical description of the SCIs suggested in [7] is shown in Figure 2b, which is a natural extension of Figure 1 to SCIs. It shows which nested family of hypotheses is tested in which order. The individual hypotheses of each nested family are all tested at the significance level assigned to this family. Let us describe this procedure in detail. In the first step we test all hypotheses \( H_i \) for \( \mu_1 \leq 0 \) at level \( \alpha \). This provides the lower confidence bound \( L_1 = \min \{ p_i^{-1}(\alpha), 0 \} \). If \( H_i \) is rejected and thereby all \( H_i \) for \( \mu_1 \leq 0 \) are rejected, i.e., \( L_1 = 0 \), then we pass over the level \( \alpha \) to the second family of hypotheses and test all \( H_i \) for \( \mu_2 \in \mathbb{R} \) at level \( \alpha \). This leads to the lower confidence bound \( L_2 = p_2^{-1}(\alpha) \). If in the first step \( H_i \) is accepted, i.e., \( L_1 < 0 \), no level is passed to the second family and none of its hypotheses can be tested. Hence, we obtain \( L_2 = -\infty \). One can easily verify that \( L_i \geq 0 \) if and only if the fixed-sequence test rejects \( H_i \).

### 2.3 Informative Fixed-sequence Confidence Bounds

Comparing Figure 2a and Figure 2b, one sees that with the fixed-sequence procedure, only the nested family \( H_i \) for non-positive \( \mu_1 \) is considered. This is necessary, because the significance level can be shifted to the second family of hypotheses only if all individual hypotheses of the first family are rejected. This means that, in a sense, no significance level is left over for the hypotheses \( H_i \) for \( \mu_1 > 0 \). As a consequence, the confidence bound \( L_i \) can never be positive, which is a serious problem, since we aim to obtain more detailed information about the size of the effect \( \theta_i \), especially in the case where \( H_i \) is rejected.

This deficiency may be removed by splitting up the level available after rejection of \( H_i \) and passing a part of the level to the family \( \{ H_i \} \) after rejection of \( H_1 \) and \( H_2 \). The idea is represented in Figure 3 where the fraction \( 0 < q < 1 \) is reserved for testing positive \( \mu_1 \). In the case that \( H_i \) is rejected, we test \( H_i^{\mu_1} \) for all \( \mu_1 > 0 \) at level \( qa \) and pass to the family \( \{ H_i^{\mu_2} \} \) only the level \( (1 - q) \alpha \). This provides the lower bounds \( L_i = 0 \) if \( p_i^{-1}(qa) \), 0 and \( L_2 = p_2^{-1}(1 - q) \alpha \). If \( H_i \) is accepted, no level is passed over and so, as before, \( L_i = p_i^{-1}(\alpha) < 0 \) and \( L_2 = -\infty \).

Note that the level available to test the informative hypotheses \( H_i^{\mu_1} \) for \( \mu_1 > 0 \) is only \( qa \) and hence strictly smaller than the original level \( \alpha \) used for the null hypothesis \( H_0 \). Therefore, in case of rejection of \( H_i^{\mu_1} \), the final confidence bound \( L_i \) has to be defined as the maximum of \( p_i^{-1}(qa) \) with 0. Indeed, it may happen that \( p_i^{-1}(\alpha) \geq 0 \) and \( p_i^{-1}(qa) < 0 \), namely when \( qa < p_i(0) \leq \alpha \). Since the rejection of \( H_i^{\mu_1} \) is valid independently of the test results in the family \( \{ H_i \} \), we can report \( L_i = 0 \) also when \( qa < p_i(0) \). Note that in this case the lower bound \( L_i \) equals 0 and hence does not provide any additional information to the shear hypothesis test. As a consequence the approach does not completely remove the possibility of non-informative lower confidence bounds.

We will now add two generalizations to the described proceeding. First, we consider \( m \) hypotheses instead of only two. With the second generalization we will avoid non-informative confidence bounds completely. Instead of a constant \( q \), we choose the proportion spent for positive parameter values in dependence of their size.

Figure 4 shows the graph for the modified procedure. To understand the graph we start explaining the first three steps. If \( H_i^{0} \) is accepted then the procedure stops with reporting \( L_1 = p_i^{-1}(\alpha) \) and \( L_2 = -\infty \) for \( i = 2, \ldots, m \). This is the same as with the fixed-sequence SCIs in Figure 2b. If \( H_i^{0} \) can be rejected then we test each \( H_i^{\mu_1} \) for \( \mu_1 > 0 \) at level \( q^{-1} \alpha \). The resulting confidence bound \( L_i \) is the unique solution of the equation \( p_i(\mu_i) = q^{-1} \alpha \). The solution is unique because \( p_i(\mu_i) \) is increasing and \( q^{-1} \alpha \) is decreasing in \( \mu_i \), recall that \( 0 < q < 1 \). Moreover, \( H_i^{\mu_1} \) is rejected for all \( \mu_1 \leq L_i \) and accepted for all \( \mu_1 > L_i \), hence \( L_i = \inf \{ \mu_i > 0 : p_i(\mu_i) > q^{-1} \alpha \} \). We then pass over the constant level \( q^{-1} \alpha \) to the family \( \{ H_i^{\mu_2} \} \) for \( \mu_2 \leq 0 \). The full procedure can be summarized in terms of the following stepwise algorithm:

**Algorithm 1** (Informative fixed-sequence SCI). Fix \( q \in (0, 1) \), let \( \hat{\theta}_i := 1 \) and do for \( i = 1, \ldots, m \):

1. When \( i < m \) and \( \hat{\theta}_{i-1} = \hat{\theta}_i \): accept \( H_i^{\mu_1} \), let \( L_i = p_i^{-1}(\hat{\theta}_i \alpha) \) and \( L_{i+1} = \ldots = L_m = -\infty \) and stop.
2. If \( p_i(0) \leq \hat{\theta}_i \alpha \): reject \( H_i^{\mu_1} \) and let \( L_i \geq 0 \) be the solution of \( p_i(\mu_i) = q^{-1} \hat{\theta}_i \alpha \). Define \( \hat{\theta}_{i+1} = (1 - q^{-1}) \alpha \hat{\theta}_i \) and proceed with Step 3.
3. When \( i = m \) and let \( L_m = p_m^{-1}(\hat{\theta}_m \alpha) \).

We show next that, for every \( q \in (0, 1) \), the algorithm produces non-informative confidence bounds with probability 0 when \( p_i(\mu) \) has a continuous distribution. Without a continuous distribution even the classical bound \( p_i^{-1}(\alpha) \) may be non-informative with positive probability. When the distribution is continuous, then we have \( p_i(0) < \hat{\theta}_i \alpha \) whenever \( H_i^{\mu_1} \) is rejected by Algorithm 1. Since \( p_i(\mu) / q^{-1} \alpha \) is continu-
and increasing in \( \mu_i \) and \( p_i(0)/q^0 = p_i(0) < \alpha \) we have that \( p_i(\mu_i)/q^0 < \alpha \) also for some \( \mu_i > 0 \). Hence, \( L_i = \inf \{ \mu_i > 0: p_i(\mu_i) > q^0/\alpha \} \) is strictly positive.

The choice of the parameter \( q \in (0, 1) \) quantifying the level kept for testing the informative hypotheses \( H^q_i \) with \( \mu_i > 0 \) will depend on the investigator’s goals in a concrete situation. In Section 3, we give one explicit possibility of calculating \( q \) by means of power considerations. In general, one has to keep in mind that, the smaller \( q \) is, the more similar is the algorithm to the original fixed-sequence procedure, where an informative confidence interval is only available for the last hypothesis. If \( q \rightarrow 1 \), the algorithm converges to the exclusive test of \( H_i^0 \) at full level \( \alpha \) with the canonical corresponding confidence interval. The fine tuning of priorities between rejecting \( H^q_{i+1} \) and collecting information about \( \theta_i \) can be done by choosing the size of the value \( q \). Moreover, individual proportions \( q_1, \ldots, q_m \) can be defined in each step. More generally, one could choose any decreasing function \( g_i(\mu_i) \) with \( g_i(0) = 1 \) and \( g_i(\mu_i) \rightarrow 0 \) as \( \mu_i \rightarrow \infty \) instead of \( q^\mu \). This is in the spirit of the idea of penalizing functions introduced in [12]. Indeed, one can define the same procedure via the method applied there, depending on an increasing penalizing function \( \lambda_i \). Then \( q^\mu \) corresponds to \( \lambda_i(\mu_i)^{-1} \). This alternative construction also ensures that the SCIs built by Algorithm 1 have coverage probability of at least \( 1 - \alpha \). Since it is a rather technical method, it is presented in the Appendix.

### 3. Example

As in the introduction, we consider a trial with an experimental and standard therapy and with one primary and one key secondary endpoint. First, non-inferiority is tested for the primary endpoint. Usually we are interested in obtaining more information on this endpoint, namely in strengthening the non-inferiority claim and proving superiority, if the effect of the new therapy is strong enough. After rejecting the non-inferiority hypothesis, we would also be interested in showing superiority with regard to the key secondary endpoint.

The main hypotheses of interest are \( H_i^N: \theta_i \leq -\delta_N \) (non-inferiority in the primary endpoint), \( H_i^S: \theta_i \leq 0 \) (superiority in the primary endpoint) and \( H_i^2: \theta_i \leq 0 \) (superiority in the secondary endpoint). A classical fixed-sequence test of the ordered hypotheses \( H_i^N \), \( H_i^S \), \( H_i^2 \) has the disadvantage that \( H_i^2 \) cannot be rejected when \( H_i^S \) is accepted. A similar problem occurs with the ordering \( H_i^N \), \( H_i^S \), \( H_i^2 \). For this reason we follow an alternative approach and split after rejection of \( H_i^N \) the level \( \alpha \) into \( \alpha_1 = qa \) and \( \alpha_2 = (1 - q)\alpha \) for some \( 0 < q < 1 \) and test each \( H_i^S \) at level \( \alpha/2 \) (Figure 5a).

Assume estimates \( \bar{X} \) for \( \theta_i \) that are (asymptotically) normal with mean \( \theta_i \) and standard deviation \( \text{SE}_i \). We use the p-values \( p_i(\mu_i) = 1 - \Phi((\bar{X} - \mu_i)/\text{SE}_i) \) for \( H_i^{\mu_i}: \theta_i \leq \mu_i \), where \( \Phi \) is the standard normal distribution function. Let us further assume that we wish to show non-inferiority with power \( 1 - \beta \) when both treatments are equivalent (\( \theta_1 = 0 \)). Since non-inferiority is tested at level \( \alpha \), the power constraint can be achieved by fixing the sample size (or information) such that \( \text{SE}_i = \delta_i/(\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)) \). Assuming \( \alpha = 0.025 \), \( \beta = 0.1 \) and \( \delta_N = 1 \), we obtain \( \text{SE}_i = 0.3085 \).

Assume further that we wish to show superiority in the primary endpoint with power \( 1 - \beta \) if the experimental treatment is superior to the control by the effect \( \delta_s = 1.5 > \delta_N = 1 \). Accordingly, we choose the level \( \alpha_1 \) such that this goal is achieved with the previously calculated sample size (information). This leads to

\[
\alpha_1 = \Phi(\Phi^{-1}(1 - \beta) - \text{SE}_i \cdot \delta_s) = 0.00017.
\]

Superiority in the secondary endpoint can then be tested at level \( \alpha_2 = \alpha - \alpha_1 = 0.02483 \).

Since it is recommendable to simultaneously test non-inferiority and superiority by confidence intervals (which allows us to strengthen the non-inferiority claim also when superiority is missed), we wish to extend the above multiple testing procedure.
to SCIs for \( \theta_1 \) and \( \theta_2 \). This can be achieved by applying the approach in Figure 3, namely by assigning a fixed weight \( q \) to the family \( \{ H_{\theta_1}^0 \colon \mu_1 > -\delta_N \} \) and \( 1 - q \) to the family \( \{ H_{\theta_2}^0 \colon \mu_2 \in \mathbb{R} \} \). If we choose \( q \) such that \( q\alpha = \alpha_1 \), i.e. \( q = 0.0068 \), then the confidence bound \( L_1 \) meets the desired power constraints for \( H_{\theta_1}^0 \) and \( H_{\delta}^0 \).

Alternatively, we can apply Algorithm 1 to the families \( \{ H_{\theta_1}^0 \colon \mu_1 > -\delta_N \} \) and \( \{ H_{\theta_2}^0 \colon \mu_2 \in \mathbb{R} \} \) as indicated in Figure 5b. Note that we now assign the level \( q\alpha = \alpha_1 \) to \( H_{\theta_1}^0 \) for \( \mu_1 > -\delta_N \) since we need to shift \( \theta_1 \) by \( -\delta_N \). The above power constraints are then met if \( q^{1/\alpha_1} = \alpha_1 \), i.e. for \( q = (\alpha_1 / \alpha)^{1/\alpha_1} = 0.0068 \), the same \( q \) as for the approach with fixed splitting weight because \( \delta_N = 1 \).

Table 1 shows the confidence bounds of both methods for three possible sample points. In the first two sample points, the p-value for \( H_{\theta_1}^0 \) is 0.0002. This corresponds to a p-value of 0.37 for the superiority test in the primary endpoint. Hence, for the primary endpoint, non-inferiority can clearly be claimed but not superiority. The informative confidence bound for \( \theta_1 \) is \( L_1 = -0.69 \) and improves the preplanned non-inferiority margin by more than 30%. For comparison, if we wanted to make confirmatory assertions only for the primary endpoint, we would obtain the full level confidence bound \( L_1 = -0.5 \). Note that the approach with a fixed weight leads to the non-inferiority margin \( -1 \) as lower confidence bound (because \( p_1(-\delta_N > \alpha_1) \)) and thus provides no extra information, although the p-value of the non-inferiority test is rather small.

In the first example in Table 1, both confidence bounds \( L_2 \) are positive, hence superiority is shown with both approaches. However, we pay a prize in terms of the size of \( L_2 \). The second example shows a more critical scenario where only the simpler approach allows to verify superiority in the secondary endpoint. The third example shows that the informative \( L_2 \) is not always smaller than the one from the fixed weighting approach. It will be larger (here only by a negligible amount) whenever superiority is shown in the primary endpoint, because the weight passed over to the secondary endpoint is then at least as large as with the fixed weighting approach.

To get a more complete picture about the behavior of the informative fixed-sequence SCIs compared to the simpler fixed weighted SCIs, we made a simulation for different values of the true effects \( \theta_1, \theta_2 \) in the two endpoints. The parameters were as above: \( \alpha = 0.025, \beta = \beta_1 = 0.1, \delta_N = 1, \delta_S = 1.5 \) so that \( q = 0.0068 \) in both approaches. We made 2500 simulation runs for each parameter combination of the \( \theta_i \) between 0 and 4.5. Figure 6 plots the resulting probabilities of interest. The left graphic shows the probability to obtain an informative rejection for the first endpoint.
For the informative SCI approach, \( P(L_1 > 0) = P(L_2 \geq 0) \), hence this line shows also the power to reject \( H_1^{\infty} \). It is 0.9 for \( \theta_1 = 0 \) by the planning assumptions and increases with increasing \( \theta_2 \). With the fixed weighting approach the probability \( P(L_1 > 0) \) is only 0.38 in \( \theta_1 = 0 \). This shows that the new approach gains a lot in information for the primary endpoint. In the right graphic, a comparison of the power \( P(L_2 \geq 0) \) for the secondary endpoint \( H_2^S \) is illustrated in dependence of the true effect \( \theta_2 \). We assume here that the estimates \( X_1 \) and \( X_2 \) are independent. If \( \theta_1 = 0 \), then the fixed weighting approach leads to somewhat more power than the informative SCI method. This power advantage decreases with increasing \( \theta_2 \) and becomes negligible for \( \theta_2 = 1.5 \). Altogether, one can say that the loss in power for the second hypothesis appears acceptable in view of the information gain in the more important first hypothesis.

We finally note that instead of planning the study for \( H_1^{\infty} \) and \( H_1^{S} \), we could plan for \( H_1^{\infty} \) and \( H_2^S \). This would require planning assumption for the secondary endpoint. Assuming the likely scenario that under these assumptions it is easier to show superiority for the secondary endpoint than non-inferiority for the primary one, we could again start with a sample size calculation for the non-inferiority hypothesis and then fix \( \alpha_2 < \alpha \) such that the same power is achieved for \( H_2^S \) under the given sample size. Given \( \alpha_1 = \alpha - \alpha_2 \) we would then determine \( g \) as above.

### 4. Discussion

We have proposed an approach to adapt the SCIs of fixed-sequence test procedures in a way that the confidence bounds always provide extra information on the parameter values of rejected null hypotheses. This is especially interesting in the context of non-inferiority trials, where informative confidence bounds allow us to improve the pre-planned non-inferiority margin and to test superiority. Our method is very much in the spirit of the penalized Bonferroni-like intervals in [12], where larger parameter values are tested at lower significance level for the possibility of assigning positive power also to alternatives close to the null hypothesis. We have illustrated that the weights need to depend continuously on the parameter values in order to completely avoid non-informative lower bounds. A similar observation was made in [12]. Starting with a fixed-sequence procedure, the price paid for the informative SCIs is a power loss for the hypotheses following the more important ones.

Our procedure can be illustrated by intuitive graphs and the computations can be done by a simple iterative algorithm. This makes the method attractive for applications. We have illustrated in a clinical trial example how the parameter dependent weight can be chosen based on power constraints for the specific null hypothesis (e.g. non-inferiority and superiority for the primary endpoint). We noted that the choice of the weighting functions could be much more general than presented in this paper where only exponential functions \( g^\theta \) for some \( 0 < \theta < 1 \) were considered. The method would work with arbitrary decreasing weight functions that can also be chosen individually for each parameter value.

The proposed algorithm for simultaneous confidence intervals is based on univariate \( \tilde{p} \)-values for hypotheses \( H_1^{\infty} : \theta \leq \mu \), that need to be monotone in \( \theta \). It thus works for many types of endpoint of interest like means, ratios or hazard rates, and may even work with non-parametric tests for suitable parameters like the median.

The analogy of our proofs in the Appendix with the methods of [12] suggests a broad generality of the ideas. We are convinced that they can also be extended to the general class of gatekeeping procedures introduced in [5]. An interesting application in this context is the recently considered situation of testing a composite endpoint and subsequently individual endpoints, where the hierarchical procedure may be suboptimal [15]. Further work need to be done here as well as on the analysis of the properties of the confidence bounds in dependence of the choice of \( g \) or other weight functions.

We finally note that we have only considered the case of one-sided hypotheses in this paper, which is the important case in the context of non-inferiority trials. Our method can be easily extended to two-sided hypotheses by intersecting two one-sided simultaneous confidence intervals with coverage probability \( 1 - \alpha/2 \).

### Acknowledgment

We thank the anonymous reviewers for their comments, which helped to improve the article. This work was supported by the DFG project BR 3737/1–1.

### References