Recent Developments in Multiple Comparison Procedures
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 47 (2004) 129–142

Compatible confidence intervals for intersection union tests involving two hypotheses*

Klaus Strassburger[1], Frank Bretz[2], and Yosef Hochberg[3]

German Diabetes Research Institute, University of Hannover and Tel Aviv University

Abstract: The intersection union test is a standard test in situations where the rejection of all elements of a set of k hypotheses is required. In particular, the intersection union test is known to be uniformly most powerful within a certain class of monotone level$-\alpha$ tests. In this article we consider the special case of $k = 2$. We consider the problem of deriving simultaneous confidence intervals which are compatible with the associated test decisions. We apply the general partitioning principle of Finner and Strassburger (2002) to derive a general method to construct confidence intervals which are compatible to a given test. Several examples of partitioning the two-dimensional parameter space are given and their characteristics are discussed in detail. The methods in this paper are illustrated by two gold standard clinical trials, where a new treatment under investigation is compared to both a placebo group and a standard therapy.

1. Introduction

Let $\Theta = \mathbb{R}^2$ be the parameter space and let $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta$ be the parameter vector of interest. Define $H_i = \{\vartheta \in \Theta : \vartheta_i \leq 0\}, i = 1, 2$. We are interested in testing the composite null hypothesis $\vartheta \in H = H_1 \cup H_2$ against the alternative $\vartheta \in K = \Theta \setminus H$. Consider the following two motivating examples, which can be traced back to the above test problem. In drug combination trials a combination treatment AB has to show a relevant improvement with respect to its individual components A and B. Here, $\vartheta_1 = \mu_{AB} - \mu_A - \delta$ and $\vartheta_2 = \mu_{AB} - \mu_B - \delta$, where μ defines the corresponding location parameter and $\delta \geq 0$ is the relevance shift. In contrast, gold standard trials are designed to compare a new treatment T to a placebo P (when the goal is to show relevant superiority) and a standard treatment S (when it may be sufficient to show non-relevant inferiority). Here, $\vartheta_1 = \mu_T - \mu_P - \delta_1$ and $\vartheta_2 = \mu_T - \mu_S + \delta_2$, $\delta_1, \delta_2 \geq 0$.

Given this test problem, we are concerned about finding good confidence sets, which yield maximum information about the unknown ϑ.

In the following we assume that ϑ can be estimated by $\hat{\vartheta} = (\hat{\vartheta}_1, \hat{\vartheta}_2)'$ such that $\hat{\vartheta} \sim N_2(\vartheta, \Sigma)$, where the elements of the covariance matrix Σ are given by

*The work of Frank Bretz is supported by the Deutsche Forschungsgemeinschaft, grant BR 2202/1.

[1]Department of Biometrics and Epidemiology, German Diabetes Research Institute, Leibniz Institute at Heinrich-Heine-University Düsseldorf, Germany. e-mail: strass@ddfi.uni-duesseldorf.de

[2]Research Unit Bioinformatics, University of Hannover, Germany. e-mail: bretz@bioinf.uni-hannover.de

[3]Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Israel. e-mail: hochberg@post.tau.ac.il

Keywords and phrases: multiple hypotheses testing, min-test, partitioning principle, gold standard clinical trials, stepwise testing.

AMS 2000 subject classifications: primary 62F03; secondary 62J15.

$\sigma_{ii} = \sigma^2\lambda_i^2$ and $\sigma_{12} = \sigma^2\lambda_1\lambda_2\rho$, with known correlation $\rho \in (-1, 1)$. We assume that $\lambda_i,\ i = 1, 2$, are positive known constants which usually depend on the involved sample sizes and that σ can be estimated by $\hat{\sigma}$ independent to $\hat{\vartheta}$ such that $\nu\hat{\sigma}^2$ follows a central χ^2 distribution with degree of freedom ν. In case of $\vartheta = 0$ the corresponding standardized $t-$test statistic $t = (t_1, t_2)'$ with $t_i = \hat{\vartheta}_i/(\lambda_i\hat{\sigma}),\ i = 1, 2$, has a bivariate t-distribution with degree of freedom ν and a correlation matrix $R = (\rho_{ij})_{ij}$ with $\rho_{ii} = 1,\ i = 1, 2$ and $\rho_{12} = \rho$.

The classical procedure for the above test problem is the intersection union test (IUT), which requires that both H_i are rejected in order to reject H (Berger, 1982). The IUT is also known as min-test, as coined by Laska and Meisner (1989). Denote by c_1 the $(1-\alpha)-$quantile of the univariate $t_\nu-$distribution. The composite hypothesis H is rejected, iff $\min\{t_1, t_2\} > c_1$. Since the maximum probability of rejection for the IUT under H does not depend on ρ and is attained at $\vartheta = (0, \infty)$ and $\vartheta = (\infty, 0)$, c_1 is computed from the univariate marginal distributions.

As an example for the IUT, consider the following gold standard trial. Lange et al. (1998) described a clinical trial to show therapeutic equivalence of horse chestnut extract (HCSE) and compression treatment in patients with chronic venous insufficiency. Compression was regarded as standard therapy. The study included a placebo group as third arm, thus leading to a three-armed gold standard trial as described above. The primary endpoint was oedema reduction (ml) after 12 weeks. Table 1 shows the summary data of this trial. The relevance shifts were $\delta_1 = \delta_2 = 50$ml. Let $\bar{X}_P$, $\bar{X}_T$, $\bar{X}_S$ denote the sample means and n_P, n_T, n_S the corresponding sample sizes of the three treatments. Then $\hat{\vartheta}_1 = \bar{X}_T - \bar{X}_P - \delta_1$ and $\hat{\vartheta}_2 = \bar{X}_T - \bar{X}_S + \delta_2$. In this case the correlation ρ equals $[(1+n_T/n_P)(1+n_T/n_S)]^{-\frac{1}{2}}$ and the standardizing constants λ_i are given by $\lambda_1 = \sqrt{n_T^{-1} + n_P^{-1}}$ and $\lambda_2 = \sqrt{n_T^{-1} + n_S^{-1}}$.

To estimate σ the common pooled variance estimator is used, yielding $\hat{\sigma} = 97.7$ such that the t-test statistic are given by $t = (0.21, 3.36)'$. The decision of the IUT is based on the 95% quantile $c_1 = 1.65$ of the univariate t_{237} distribution. Since $t_1 = 0.21 < 1.65 = c_1$, the classical IUT would not reject H, even though HCSE were at most non-relevantly inferior compared with compression ($t_2 = 3.36$). Thus the study failed to achieve its aim and the final decision is the unsatisfactory statement "$\vartheta \in \Theta$". Any further conclusion could only be drawn at the exploratory level. This is a general drawback of the IUT, since it does not provide any intermediate results, if one of the elementary hypotheses remains not rejected. The focus of this paper is to provide additional information about the parameters ϑ_1 and ϑ_2 by deriving compatible confidence intervals for the IUT with two hypotheses.

Standard techniques to deduce confidence intervals from a given test fail in case of the IUT. A common technique used in shift models is to test each $\eta \in \Theta$ with acceptance region $\mathcal{A}_\eta = \mathcal{A}_0 + \eta$, where $\mathcal{A}_0$ is the acceptance region of the test at hand. The set of all non rejected η's then yields a $(1 - \alpha)$-confidence set for ϑ. Usually this technique provides good practical confidence sets, but in case of the IUT, where $\mathcal{A}_0 = \{x \in \mathbb{R}^2 : \min_{i=1,2}\{x_i/(\lambda_i\hat{\sigma})\} \leq c_1\}$, we end up with the confidence set $\Theta \setminus \{\eta : \eta_i < \hat{\vartheta}_i - c_1\lambda_i\hat{\sigma}, i = 1, 2\}$. This set is neither consistent

Table 1: Summary data for the HCSE clinical trial

	Placebo (P)	HCSE (T)	Compression (S)
Sample size	46	95	99
Mean	-9.8	43.8	46.7
Std. dev.	100.1	111.1	81.6

with the IUT nor does it provide lower confidence bounds. Even worse the same confidence set could have been guaranteed with a smaller critical value c_2 being the solution of $P(t_1 > c_2, t_2 > c_2 | \vartheta = 0) = \alpha$.

One way to overcome the disadvantages of the IUT is to priorize the hypotheses and test them sequentially in a pre-determined order as described by Hsu and Berger (1999) in connection with some dose response problems and toxicity studies. Let $A_i = \hat{\vartheta}_i - c_1 \lambda_i \hat{\sigma}, i = 1, 2$. Denote further by $H_{(i)}, i = 1, 2$, the ordered hypotheses such that $H_{(1)}$ is tested first. Then the stepwise confidence intervals of Hsu and Berger (1999) provide the intermediate information $\vartheta_{(1)} > A_{(1)}$ if $H_{(1)}$ can not be rejected and $\vartheta_{(1)} > 0$, $\vartheta_{(2)} > A_{(2)}$ if $H_{(1)}$ can be rejected but $H_{(2)}$ is retained. If both hypotheses are rejected, one concludes that $\min\{\vartheta_1, \vartheta_2\} > \min\{A_1, A_2\}$.

In this paper we make use of a construction method similar to that of Hsu and Berger (1999) by applying the general partitioning principle of Finner and Strassburger (2002). As seen later in the paper, the stepwise confidence intervals of Hsu and Berger (1999) turn out to be one choice within a wide class of strategies for constructing confidence intervals which are compatible with the IUT. Here, compatibility of a level-$(1-\alpha)$ confidence set C with a level-α test ψ for H is taken to mean that the events $\{\psi = 1\}$ and $\{C \subseteq K\}$ coincide for all possible parameter configurations with probability one (Hayter and Hsu, 1994). Consequently, we search for IUT compatible confidence intervals which provide additional information to the IUT. It is not our aim to improve the IUT itself. Such improvements can be found in Berger (1989) and Sarkar, Snapinn and Wang (1995). A quite simple way of improving the IUT is to restrict the parameter space to $\tilde{\Theta} = \bigcap H_i \cup K$ (either both individual hypotheses are true or both are false) and reformulate the test problem as $\tilde{H} : H_1 \cap H_2$ against $\tilde{K} = \tilde{\Theta} \setminus \tilde{H}$. Applications of such restricted models could arise in gene expression settings (either all or none genes are active) or in repeated measures settings (there is a difference at either all or none time points). Here, $\tilde{H}$ is rejected iff $\min\{t_1, t_2\} > c_2$. Note that in contrast to c_1 the critical value c_2 depends on the correlation ρ. In fact, Slepians (1962) inequality yields that c_2 is non decreasing in $\rho \in (-1, 1)$. Since $c_2 < c_1$ for all possible values of ρ, the resulting procedure is more powerful than the IUT, but it should be preferred only if the restriction of the parameter space is feasible. Moreover it should be noted that, at least in the case of known σ, improving the IUT must cause some kind of incoherence. The rejection region of an improved procedure must be larger than that of the IUT. Consequently it can happen that, taken as one-sample problem, H_1 can not be rejected but taken as two-sample problem both, H_1 and H_2 can be rejected.

The paper is organized as follows. Section 2 introduces the Partitioning Lemma and a general method to construct confidence intervals which are compatible to a given test. Section 3 gives some examples of partitioning Θ. A detailed discussion of these partitions and the specification of the arising parameters is given. Extensions of the proposed techniques to test problems involving additional information about the ratio ϑ_2/ϑ_1 or the difference $\vartheta_2 - \vartheta_1$ are described. Section 4 applies and compares the confidence intervals using two clinical data examples. Concluding remarks are given in the final Section 5.

2. The Partitioning Lemma

Suppose that $P_{\vartheta,\sigma}$ is the probability measure inducing the distribution of a sampling statistic X. Let $\vartheta \in \Theta$ be the parameter of interest and let σ be a nuisance parameter. A test φ for a single hypothesis $\vartheta \in H$ against the alternative $\vartheta \in K = \Theta \setminus H$

does not provide much information about the unknown parameter ϑ. If we interpret the two possible decisions of the test in terms of confidence sets for ϑ we end up with $\vartheta \in \Theta$ if $\varphi(X) = 0$ or $\vartheta \in K$ if $\varphi(X) = 1$. The worst fist case could be a matter of greed. A more moderate test may have rejected $\vartheta \in H' \subset H$ and we could have get $\vartheta \in \Theta \setminus H'$ instead of no information. Also the decision for $\vartheta \in K$ is not satisfactory, because it may have been possible to reject $\vartheta \in H'' \supset H$ leading to the more informative confidence statement $\vartheta \in \Theta \setminus H''$.

Our approach to overcome this trouble follows the basic idea to break up H and K into small disjoint subhypotheses and to test each subhypothesis with an appropriate test. The union of all non rejected hypotheses then yields a confidence set for ϑ. The core piece of this idea is formalized in the following Lemma which is a direct consequence of the general partition principle introduced by Finner and Strassburger (2002). It can be viewed as an extension of the well known connection between tests and confidence sets (Lehmann, 1986, p. 90).

Lemma (Partitioning Lemma). *Let $\wp_\Theta = \{\Theta_i,\ i \in I\}$ be a partition of Θ for some index set I (i.e. $\Theta_i \cap \Theta_j = \emptyset$, $i \neq j$ and $\bigcup_{i\in I} \Theta_i = \Theta$) and let $\varphi = \{\varphi_i, i \in I\}$ be a family of local level α tests for $\wp_\Theta$, i.e.*

$$\inf_{\vartheta\in\Theta_i,\sigma} P_{\vartheta,\sigma}\big(\varphi_i(X) = 1\big) \leq \alpha, \quad i \in I,$$

then $C(X) = \bigcup_{i\in I:\varphi_i(X)=0} \Theta_i$ is an $(1-\alpha)-$confidence set for ϑ.

Note that the finest possible partition of Θ is given by $\wp_\Theta = \{\{\vartheta\},\ \vartheta \in \Theta\}$. In this case each point of Θ represent an element of the partition. Most of the classical (simultaneous) confidence intervals can be derived from the Partitioning Lemma using the finest partition and an appropriate family $\{\varphi_\vartheta, \vartheta \in \Theta\}$ of one or two sided tests. For example, under the normal model explained in the Introduction, simultaneous lower confidence bounds related to Dunnett's (1955) procedure can be obtained by setting $\varphi_\vartheta = 0$ iff $\max_{i=1,2}\{\hat{\vartheta}_i - \vartheta_i - d\lambda_i\hat{\sigma}\} \leq 0$, where d is the solution of $P(t_1 \leq d, t_2 \leq d|\vartheta = 0) = 1 - \alpha$. In this case the resulting confidence set $C = \{\vartheta : \vartheta_i \geq \hat{\vartheta}_i - d\lambda_i\hat{\sigma},\ i = 1, 2\}$ itself is a simultaneous confidence interval. This must not be the case in general, but for $\Theta \subseteq \mathbb{R}^k$, a confidence set $C(X)$ for ϑ can be used to construct simultaneous confidence intervals for the parameters ϑ_j, $j = 1, \ldots, k$, by simply projecting $C(X)$ on the coordinate axes. If one is interested in lower confidence bounds, the projection of a $(1-\alpha)-$confidence set leads to simultaneous lower $(1-\alpha)-$confidence bounds for ϑ_j of the form $L_j(X) := \min\{\eta_j : \eta \in C(X)\}$, $j = 1, \ldots, k$.

An application of the Partitioning Lemma yields confidence sets which are compatible to a given test ψ, if the following two conditions are satisfied:

(A) the partition $\wp_\Theta$ contains a subpartition $\wp_H = \{\Theta_i,\ i \in J\}$, $J \subseteq I$, with $H = \bigcup_{i\in J} \Theta_i$,

(B) the family $\{\varphi_i,\ i \in J\}$ of tests for $\wp_H$ fulfills $\psi = \min_{i\in J} \varphi_i$.

3. IUT compatible confidence intervals

Let $\Theta = \mathbb{R}^2$, $H = \{\eta : \min\{\eta_1, \eta_2\} \leq 0\}$ and $K = \Theta \setminus H$. In the following we apply the Partition Lemma to construct various confidence sets which are compatible with the IUT. Under the normal model described in the introduction we have $X = (\hat{\vartheta}_1, \hat{\vartheta}_2, \hat{\sigma})$. To define the pairs $(\wp_\Theta, \varphi)$ discussed in the following, let f_1 and f_2 be two functions defined on $\mathbb{R}$ satisfying $f_2(f_1(x)) \geq x$, $f_i(x) \leq (\geq)0$, for $x < (\geq)0$, $i = 1, 2$. Furthermore let

- $\Theta_{(1,r)} = \{\eta : \eta_1 = r, \eta_2 > f_1(r)\}$, $r \in \mathbb{R}$,
- $\Theta_{(2,r)} = \{\eta : \eta_2 = r, \eta_1 > f_2(r)\}$, $r \in \mathbb{R}$,
- $\Theta_{(3,r_1,r_2)} = \{(r_1, r_2)\}$, $(r_1, r_2) \in \Theta' = \{\eta : \eta_1 \leq f_2(\eta_2), \eta_2 \leq f_1(\eta_1)\}$.

Then the condition $f_2(f_1(x)) \geq x$ ensures that

$$\wp_\Theta = \{\Theta_{(1,r)},\ r \in \mathbb{R}\} \cup \{\Theta_{(2,r)},\ r \in \mathbb{R}\} \cup \left\{\Theta_{(3,r_1,r_2)} : (r_1, r_2) \in \Theta'\right\}$$

is a partition of Θ.

Let $B_i = \hat{\vartheta}_i - c_2\lambda_i\hat{\sigma}, i = 1, 2$. Then the family of local level-α tests φ for $\wp_\Theta$ is defined as follows:

- $\varphi_{(1,r)}$ rejects $\vartheta \in \Theta_{(1,r)}$, iff $A_1 \geq r$,
- $\varphi_{(2,r)}$ rejects $\vartheta \in \Theta_{(2,r)}$, iff $A_2 \geq r$,
- $\varphi_{(3,r_1,r_2)}$ rejects $\vartheta \in \Theta_{(3,r_1,r_2)}$ (i.e. $\vartheta = (r_1, r_2)$), iff $\min_{i=1,2}\{B_i - r_i\} \geq 0$.

With the index set $I = \{(j, r), j = 1, 2,\ r \in \mathbb{R}\} \cup \{(3, r_1, r_2), (r_1, r_2) \in \Theta'\}$ the Partition Lemma yields the $(1-\alpha)-$confidence set

$$\begin{aligned} C(X) &= \bigcup_{i \in I : \varphi_i(X) = 0} \Theta_i = \left\{\eta : \eta_1 > f_2(\eta_2), A_2 < \eta_2\right\} \\ &\cup \left\{\eta : \eta_2 > f_1(\eta_1), A_1 < \eta_1\right\} \\ &\cup \left\{\eta : \eta_1 \leq f_2(\eta_2), \eta_2 \leq f_1(\eta_1), \max\{\eta_1 - B_1, \eta_2 - B_2\} > 0\right\}. \end{aligned}$$

Projection of $C(X)$ on the coordinate axes yields the simultaneous lower $(1-\alpha)$-confidence bounds

$$L_i(X) = \min\{A_i, C_i, D_i\},\ i = 1, 2,$$

where $C_i = \inf\{r : f_i(r) > B_{3-i}\}$ and $D_i = \inf\{f_{3-i}(r) : r > A_{3-i}\}$, $i = 1, 2$, with the convention $\inf\{\emptyset\} = \infty$.

Note that the condition $f_i(x) \leq 0$, for $x < 0$, $i = 1, 2$, and the fact $B_i \geq A_i$, $i = 1, 2$, guarantees (A) and (B), so that the confidence sets and the confidence bounds are always compatible with the IUT.

In the following we give some examples for the choice of f_1, f_2 and the resulting confidence bounds. For convenience we drop the X in the notation of $L_i(X), i = 1, 2$. Note that by construction each resulting partition contains two subpartitions, one of H and one of K. Other partitions can be generated by combination of the resulting subpartitions.

(I) The first couple of examples deals with the case that f_1 is a non decreasing, piecewise continuous function with $f_1(x) \leq (\geq)0$, if $x < (\geq)0$. The function f_2 is the generalized inverse of f_1, i.e. $f_2(x) = \inf\{y : f_1(y) > x\}$. This choice of f_1 and f_2 results in $C_i \geq D_i$, $i = 1, 2$. The confidence bounds do not depend on the C_i's and are given by $L_i = \min\{A_i, f_{3-i}(A_{3-i})\}$, $i = 1, 2$. Note that there is no need to define the $\Theta_{(3,r_1,r_2)}$ explicitly for the partitions in (I), since the individual points could be included in either of the $\Theta_{(i,r)}, i = 1, 2$, without changing L_i. But to be consistent with the notation for the following partitions we decided to retain the $\Theta_{(3,r_1,r_2)}$.

For sake of simplicity and some heuristic reasons discussed later we restrict ourselves to piecewise linear functions of the type $f_1(x) = \min\{0, (\gamma_1 + x)\tau_1\}$,

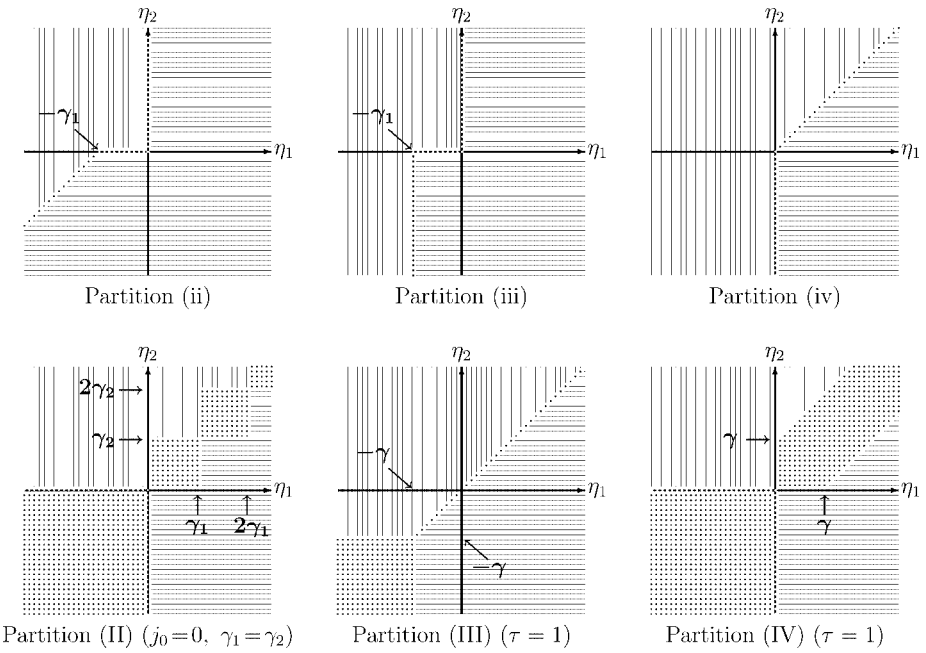

Figure 1: Graphical representation of some partitions leading to IUT compatible confidence intervals. The partition elements $\Theta_{(1,r)}$ and $\Theta_{(2,r)}$ are representetd by the vertical and horinzontal lines, respectively. The partition elements $\Theta_{(3,r_1,r_2)}$ are represented by dots.

if $x < 0$ and $f_1(x) = \max\{0, (\gamma_2 + x)\tau_2\}$ if $x \geq 0$, where $\tau_1, \tau_2 > 0$. This setting results in

$$L_1 = \begin{cases} \min\{0, A_1, A_2/\tau_1 - \gamma_1\} & \text{if } A_2 < 0 \\ \min\{A_1, \max\{0, A_2/\tau_2 - \gamma_2\}\} & \text{if } A_2 \geq 0 \end{cases},$$

$$L_2 = \begin{cases} \min\{0, A_2, (\gamma_1 + A_1)\tau_1\} & \text{if } A_1 < 0 \\ \min\{A_2, \max\{0, (\gamma_2 + A_1)\tau_2\}\} & \text{if } A_1 \geq 0 \end{cases}.$$

We get the following special cases:

(i) $\tau_1 = \tau_2, \gamma_1 = \gamma_2 = 0,$ i.e. $f_1(x) = \tau_1 x,$
(ii) $\tau_1 = 1, \tau_2 \to \infty,\ \gamma_1 > 0,\ \gamma_2 = 0,$
(iii) $\tau_1 \to \infty,\ \tau_2 \to \infty,\ \gamma_1 > 0, \gamma_2 = 0,$
(iv) $\tau_1 \to \infty,\ \tau_2 = 1, \gamma_1 = \gamma_2 = 0.$

Figure 1 visualizes the corresponding partitions for (ii)–(iv).

(II) Let $j_0 \in \mathbb{Z}$, $j_0 \leq 0$, and let γ_1, γ_2 be two positive real constants. For $m = 1, 2$ define

$$f_m(x) = \begin{cases} j_0\gamma_{3-m} & \text{if } x < j_0\gamma_m \\ j\gamma_{3-m} & \text{if } (j-1)\gamma_m \leq x < j\gamma_m,\ j \in \mathbb{Z}, j > j_0 \end{cases}.$$

This setting results in

$$L_i = \begin{cases} -\infty & \text{if } j_0 > B_{3-i}/\gamma_{3-i} \\ \min\{A_i, \gamma_i(\lfloor A_{3-i}/\gamma_{3-i} \rfloor + 1)\} & \\ \quad \text{if } \lfloor B_{3-i}/\gamma_{3-i} \rfloor \geq A_{3-i}/\gamma_{3-i} & \\ \min\{A_i, \gamma_i \lfloor A_{3-i}/\gamma_{3-i} \rfloor\} & \text{otherwise} \end{cases}, \quad i = 1, 2,$$

where $\lfloor x \rfloor = \max\{j \in \mathbb{Z} : j \leq x\}$ denotes the floor function.

(III) $f_1(x) = \max\{-\gamma_0, \tau x\}, f_2(x) = \max\{-\gamma_0/\tau, x/\tau\}$, with $\gamma_0, \tau \geq 0$ results in

$$L_1 = \begin{cases} -\infty & \text{if } B_2 < -\gamma_0 \\ \min\{A_1, \max\{-\gamma_0, A_2\}/\tau\} & \text{if } B_2 \geq -\gamma_0 \end{cases},$$

$$L_2 = \begin{cases} -\infty & \text{if } B_1 < -\gamma_0/\tau \\ \min\{A_2, \max\{-\gamma_0/\tau, A_1\}\tau\} & \text{if } B_1 \geq -\gamma_0/\tau \end{cases}.$$

(IV) Let $f_1(x) = 0$, if $x < 0$, $f_1(x) = \gamma + \tau x$ if $x \geq 0$, $f_2(x) = 0$, if $x < 0$, $f_2(x) = (\gamma + x)/\tau$ if $x \geq 0$, with $\tau \geq 0$ and $\gamma > 0$. This setting results in

$$L_1 = \begin{cases} -\infty & \text{if } B_2 < 0 \\ \min\{0, A_1, (B_2 - \gamma)/\tau\} & \text{if } B_2 \geq 0,\ A_2 < 0 \\ \min\{A_1, \max\{(B_2 - \gamma)/\tau, 0\}, (A_2 + \gamma)/\tau\} & \text{if } A_2 \geq 0 \end{cases},$$

$$L_2 = \begin{cases} -\infty & \text{if } B_1 < 0 \\ \min\{0, A_2, \tau B_1 - \gamma\} & \text{if } B_1 \geq 0,\ A_1 < 0 \\ \min\{A_2, \max\{\tau B_1 - \gamma, 0\}, \tau A_1 + \gamma\} & \text{if } A_1 \geq 0 \end{cases}.$$

The performance of the proposed confidence bounds depends strongly on f_1 and f_2 (i.e. the γ's, τ's), the true values of ϑ_1, ϑ_2, σ and the design parameters λ_1, λ_2. Note that f_1 and f_2 must be defined without looking at the data. There is no uniformly best set of functions f_1 and f_2 available in the sense, that the resulting confidence bounds are always the best possible. In the following we will discuss some deterministic and some heuristic rules to choose the γ's and τ's in order tailor the tests to the experimenters goal.

Let us first consider the partitions in (I). If $\tau_1 = 1$ then (i) results in $L_i = \min\{A_1, A_2\}$. This seems to be a good choice if one is interested in a lower confidence bound only for $\min\{\vartheta_1, \vartheta_2\}$. Other values of τ_1 may lead to bounds that are sharper for one of the parameters (at the cost of the second parameter).

The setting (iv) provides the confidence bounds of Hsu and Berger (1999) if $H_1 = H_{(1)}$ is the priorized hypothesis. Choosing f_1 as in (ii) and (iii) yields other stepwise testing strategies, which are particularly useful in situations, where the hypotheses can be treated asymmetric and ϑ_2 is deemed to be more important than ϑ_1. Recall the gold standard trial from the Introduction where $\vartheta_1 = \mu_T - \mu_P - \delta_1$ and $\vartheta_2 = \mu_T - \mu_S + \delta_2$, $\delta_1, \delta_2 \geq 0$. Set $\gamma_1 = \delta_1$ and consider partition (iii) in more detail at first. Plugging in the values for γ_i and $\tau_i, i = 1, 2$, we obtain explicit

expressions of the lower confidence bounds for the ϑ_i's, $i = 1, 2$:

$$L_1(X) = \begin{cases} \min\{A_1, -\delta_1\} & \text{if } A_2 < 0 \\ \min\{A_1, 0\} & \text{if } A_2 \geq 0 \end{cases},$$

$$L_2(X) = \begin{cases} -\infty & \text{if } A_1 < -\delta_1 \\ \min\{A_2, 0\} & \text{if } -\delta_1 \leq A_1 < 0 \\ A_2 & \text{if } A_1 \geq 0 \end{cases}.$$

The logic behind this approach is that we are primarily interested in obtaining the sharpest confidence intervals possible for $\mu_T - \mu_S$, as long as $\mu_T - \mu_P$ has shown to be (relevantly) significant. The following stepwise confidence interval procedure illustrates this:

Step 1: If $A_1 < -\delta_1$, conclude $\mu_T - \mu_P > A_1 + \delta_1$ and stop; else go to step 2.

Step 2: If $A_2 < 0$, conclude $\mu_T - \mu_P > 0, \mu_T - \mu_S > A_2 - \delta_2$ and stop; else go to step 3.

Step 3: If $A_1 < 0$, conclude $\mu_T - \mu_P > A_1 + \delta_1, \mu_T - \mu_S > -\delta_2$ and stop ; else go to step 4.

Step 4: Conclude $\mu_T - \mu_P > \delta_1$ and $\mu_T - \mu_S > A_2 - \delta_2$.

The confidence bounds resulting from (ii) are the same at steps 3 and 4, but steps 1 and 2 are exchanged by

Step 1': If $A_1 < -\delta_1$, conclude $\mu_T - \mu_P > \min\{A_1 + \delta_1, A_2\}, \mu_T - \mu_S > \min\{A_1 + \delta_1, A_2\} - \delta_2$ and stop; else go to step 2'.

Step 2': If $A_2 < 0$, conclude $\mu_T - \mu_P > A_2, \mu_T - \mu_S > A_2 - \delta_2$ and stop ; else go to step 3.

Thus, in contrast to (iii) we get more information about $\mu_T - \mu_S$ but we have to pay the price of information loss for $\mu_T - \mu_P$. This sounds reasonable, if one assumes that the latter comparison serves for validation purposes only and the main interest lies in $\mu_T - \mu_S$.

In general an optimal (but data dependent) choice of f_1 would be any function fulfilling $f_1(A_1) = A_2$. In such cases, $L_1(X) = A_1$ and $L_2(X) = A_2$, which are the best bounds we can get for this setting. To get free from the data dependence one could replace A_1 and A_2 by their expectations and search for functions f_1 with

$$f_1(\vartheta_1 - c_1\lambda_1\sigma) = \vartheta_2 - c_1\lambda_2\sigma. \tag{1}$$

After finding a solution for f_1, one has to redefine f_1 as $\min\{0, f_1(x)\}$ for $x < 0$ and $\max\{0, f_1(x)\}$ for $x \geq 0$, in order to produce IUT compatible confidence intervals. To determine f_1 from (1) we must have at least partial knowledge about the unknown parameters. We discuss two possible approaches, having in mind that other approaches can be devised if ϑ_1 and ϑ_2 were replaced by their parameters in the original testing problem. The first approach starts with fixing ϑ_1 and ϑ_2. Then, equation (1) is satisfied for all σ if $f_1(x) = \vartheta_2 - (\vartheta_1 - x)\lambda_2/\lambda_1$. This results in $\tau_1 = \tau_2 = \lambda_2/\lambda_1$. Assume that the experimenter has a guesstimate ϑ^* for ϑ, based on preceding studies or just based on intuition. Then the setting $\gamma_1 = \gamma_2 =$

$\vartheta_2^*\lambda_1/\lambda_2 - \vartheta_1^*$ would result in a reasonably practical choice for f_1. We can also choose two different values for ϑ^* in order to define $f_1(x)$ on $x < 0$ and $x \geq 0$ respectively. For $x < 0$ a careful choice would be $\vartheta^* = 0$ (i.e. $\gamma_1 = 0$), which is the limiting worst case if K is true. A second way to get candidates for f_1 is to fix the value of σ and set $\vartheta_2 = g(\vartheta_1)$ for some non decreasing function g. We obtain $f_1(x) = g(x + c_1\lambda_1\sigma) - c_1\lambda_2\sigma$ and we are left with the problem of finding a guesstimate for σ and a reasonable choice for g. A "linear" guess for g, i.e. $\vartheta_2 = a + b\vartheta_1$, results in piecewise linear functions f_1 as proposed for the partitions (I).

Let us consider the partitions in (II). To ensure $\lfloor B_{3-i}/\gamma_{3-i} \rfloor \geq A_{3-i}/\gamma_{3-i}$ in order to get the larger bound in (II), we require $B_{3-i}/\gamma_{3-i} - A_{3-i}/\gamma_{3-i} \geq 1$. On the other side, γ_i should be set as large as possible. This leaves $\gamma_i = B_i - A_i$ as a promising but data dependent candidate. Taking expectations yields $\gamma_i = (c_1 - c_2)\lambda_i\sigma$ and replacing σ with a guesstimate σ^* would yield a reasonable choice for γ_1 and γ_2. A good choice for j_0 seems to be a guesstimate for the expectation of $\min\{B_2/\gamma_2, B_1/\gamma_1, 0\}$. Using $\gamma_i = (c_1 - c_2)\lambda_i\sigma^*$ and the worst case setting for ϑ_1, ϑ_2 under K, i.e. $\vartheta_1 = \vartheta_2 = 0$, one obtains $j_0 = \lfloor -c_2/(c_1 - c_2) \rfloor$.

The partition resulting from (III) can be viewed as a limit of the partition (II) for values of γ_1 and γ_2 tending to zero, with fixed $\gamma_2/\gamma_1 = \tau$ and $j_0\gamma_2 = -\gamma_0$. Plugging in the estimates obtained above for partition (II) leads to $\tau = \lambda_2/\lambda_1$ and $\gamma_0 = c_2\lambda_2\sigma^*$. Under this setting partition (II) produces a larger (smaller) confidence bound L_i than (III) if $\lfloor B_{3-i}/\gamma_{3-i} \rfloor \geq (<) A_{3-i}/\gamma_{3-i}$.

Maximizing the lower confidence bounds in (IV) leads to $\gamma = \lambda_2\sigma^*(c_1 - c_2)/2$ and $\tau = \lambda_2/\lambda_1$. Partition (IV) does not provide information about ϑ_i if $B_{3-i} < 0$. To attenuate this disadvantage one can generalize partition (4) by setting $f_1(x) = \min\{0, \max\{-\gamma_0, \gamma + \tau x\}\}$, if $x < 0$, $f_1(x) = \gamma + \tau x$, if $x \geq 0$, and $f_2(x) = \min\{0, \max\{-\gamma_0, (\gamma + x)\}/\tau\}$, if $x < 0$, $f_2(x) = (\gamma + x)/\tau$, if $x \geq 0$, where γ_0 is an additional positive constant. In this case $L_1 = -\infty$ only if $B_2 < -\gamma_0$ and $L_2 = -\infty$ if $B_1 < -\gamma_0/\tau$. But since the formulae for the L_i's get more complicated, we dropped this case for the sake of simplicity.

It should be noted that the critical value c_2 depends on the correlation ρ as mentioned in the introduction. Consequently ρ is involved in the proposed choices for partitions of type (II)–(IV).

The proposed partitions and confidence intervals so far are designed to provide the IUT with lower confidence bounds for the individual parameters ϑ_i. To the end of this section we discuss some additional strategies, when the experimenter is additionally interested in the ratio ϑ_2/ϑ_1 or the difference $\vartheta_2 - \vartheta_1$. A clinical example for this type of question will be discussed in the next section.

(V) Let $\Theta_{(1,r)} = \{\eta : \eta_1 = r\}$, $\Theta_{(2,r)} = \{\eta : \eta_2 = r\eta_1, \eta_1 > 0\}$ then $\wp_\Theta = \{\Theta_{(1,r)}, r \leq 0\} \cup \{\Theta_{(2,r)}, r \in \mathbb{R}\}$ is a partition of Θ fulfilling the compatibility requirement (A). The element $\Theta_{(1,r)}$ is rejected iff $A_1 \geq r$. The element $\Theta_{(2,r)}$ is rejected iff $\hat{\vartheta}_2 - r\hat{\vartheta}_1 \geq c_1\hat{\sigma}\lambda(r)$, where $\hat{\sigma}\lambda(r)$ is an estimate of the standard deviation of $\hat{\vartheta}_2 - r\hat{\vartheta}_1$, depending on r and the involved sample sizes. We get the confidence bounds

$$\begin{aligned} &\vartheta_1 > A_1 && \text{if } A_1 < 0, \\ &\vartheta_1 > 0, \vartheta_2 > r^*\vartheta_1 && \text{if } A_1 \geq 0, \end{aligned}$$

where $r^* = \min\{r \in \mathbb{R} : \hat{\vartheta}_2 - r\hat{\vartheta}_1 < c_1\hat{\sigma}\lambda(r)\}$. An explicit formula for r^* results in Fieller's (1954) confidence interval for ϑ_2/ϑ_1. We note that the confidence

intervals are compatible with the IUT only if compatibility requirement (B) is satisfied, i.e. if $\lambda_2 \geq \lambda(r)$ for all $r < 0$.

(VI) If one is only interested in confidence statements of the form $\vartheta_2 \geq r^*\vartheta_1$ when $r^* \geq 0$, one can use the partition $\wp_\Theta = \{\Theta_{(1,r)}, r \leq 0\} \cup \{\Theta_{(2,r)}, r > 0\} \cup \{\Theta_{(3,r)}, r \leq 0\}$, where $\Theta_{(3,r)} = \{\eta : \eta_1 > 0, \eta_2 = r\}$. An element $\Theta_{(3,r)}$ is rejected iff $A_2 \geq r$. This would result in confidence bounds

$$\begin{array}{ll} \vartheta_1 > A_1 & \text{if } A_1 < 0, \\ \vartheta_1 > 0,\ \vartheta_2 > A_2 & \text{if } A_1 \geq 0, A_2 < 0, \\ \vartheta_1 > 0, \vartheta_2 > \max\{0, r^*\vartheta_1\} & \text{if } A_1 \geq 0, A_2 \geq 0. \end{array}$$

Instead of getting a confidence statement $\vartheta_2 \geq r^*\vartheta_1$, for some $r^* < 0$ we now get lower bounds for ϑ_2. Moreover, the confidence intervals are compatible with the IUT.

(VII) If one is interested in confidence statements for the difference $\vartheta_2 - \vartheta_1$ instead of the ratio ϑ_2/ϑ_1 one can use the following additive analog of the last partition. Let $\wp_\Theta = \{\Theta_{(1,r)}, r \leq 0\} \cup \{\Theta_{(3,r)}, r \leq 0\} \cup \{\Theta_{(4,r)}, r \in \mathbb{R}\}$, where $\Theta_{(4,r)} = \{\eta : \eta_2 - \eta_1 = r, \eta_1 > 0, \eta_2 > 0\}$. An element $\Theta_{(4,r)}$ is rejected iff $A_3 > r$, where $A_3 = \hat{\vartheta}_2 - \hat{\vartheta}_1 - c_1\hat{\sigma}\lambda_{1,2}$ and $\hat{\sigma}\lambda_{12}$ is an estimate of the standard deviation of $\hat{\vartheta}_2 - \hat{\vartheta}_1$, depending on the involved sample sizes. This would result in confidence bounds

$$\begin{array}{ll} \vartheta_1 > A_1 & \text{if } A_1 < 0, \\ \vartheta_1 > 0,\ \vartheta_2 > A_2 & \text{if } A_1 \geq 0, A_2 < 0, \\ \vartheta_1 > 0, \vartheta_2 > \max\{0, \vartheta_1 + A_3\} & \text{if } A_1 \geq 0, A_2 \geq 0. \end{array}$$

Finally it should be mentioned, that there are situations where mixed confidence statements as $\vartheta_1 \geq \delta + r\vartheta_2$ are of interest. In an obvious way the above partition approaches can also be tailored to this goal.

4. Two gold standard trial examples

4.1. Example 1

Recall the gold standard trial described in the Introduction. We use this example data set to compare and discuss the different partitions of the previous section ($\alpha = 0.05$ throughout). Recall that in this example the IUT did not reject the union hypothesis H since one of the two individual comparisons failed to be significant. Thus, the conclusion drawn by Lange et al. (1998) was "$\vartheta \in \Theta$" (or, equivalently, $L_i(X) = -\infty, i = 1, 2$).

In contrast, computing the IUT compatible confidence intervals based on the partition (i) for $\tau_1 = 1$ (thus treating both hypotheses equally) yields $L_1 = L_2 = \min\{A_1, A_2\} = -25.39$, where $A_1 = -25.39$ and $A_2 = 23.92$. Therefore, $\vartheta_1 = \mu_T - \mu_P - \delta_1 > -25.39$ as well as $\vartheta_2 = \mu_T - \mu_S + \delta_2 > -25.39$ at the given confidence level. These lower bounds, however, can be improved substantially using any of the other partitions. If we follow the rules devised in the previous section, $\tau_1 = \lambda_2/\lambda_1 = 0.8$ should be used. We obtain $L_1 = -25.28$ and $L_2 = -20.31$, thus leading to slightly better results. Since $A_1 = -25.39 < 0$ in the present example and HCSE fails to be relevantly superior to placebo, $\tau_1 < 1$ leads always to better confidence intervals. In fact, $L_2 \nearrow 0$ as $\tau_1 \to 0$.

Table 2: IUT compatible confidence intervals for the HCSE clinical trial

Partition	L_1	L_2	Parameters
IUT	$-\infty$	$-\infty$	
(i)	-25.39	-25.39	$\tau_1 = 1$
(i)	-25.39	-20.31	$\tau_1 = 0.8$
(ii)	-25.39	0	$\gamma_1 = 50$
(iii)	-25.39	0	$\gamma_1 = 50$
(iv)	-25.39	$-\infty$	
(II)	-25.39	-17.59	$j_0 = -2, \gamma_1 = 11, \gamma_2 = 8.79$
(III)	-25.39	-14.92	$\tau = 0.8, \gamma_0 = 14.92$
(IV)	-25.39	$-\infty$	$\tau = 0.8$, $\gamma = 4.4$
Dunnett	-30.41	19.91	

For the partitions (ii) and (iii) a suitable choice is $\gamma_1 = \delta_1$. Thus, sharper bounds L_2 are obtained in dependence of the results for testing H_1. This corresponds roughly to a stepwise approach of first comparing HCSE with placebo and then trying to find a best confidence bound for the equivalence comparison of both treatments. As noted previously, the resulting bounds are the same at steps 3 and 4, which happens to be the case in our example ($L_1 = -25.28$ and $L_2 = 0$). In general, these two partitions strongly stress the importance of ϑ_2 at the cost of ϑ_1. Evaluation of other examples show that they frequently result in worse confidence bounds for ϑ_1 but much sharper ones for ϑ_2 in comparison to those of the competitors.

The final partition (iv) in this first set mimics a stepwise test as discussed by Hsu and Berger (1999): first test, whether μ_T is relevantly better than μ_P, and only if this is true get to the next step of comparing μ_T with μ_S. Since in our example HCSE fails to be relevantly superior to placebo, $L_1 = -25.28$ and $L_2 = -\infty$ and no information is obtained with regard to ϑ_2.

Consider partition (II) now. Following the advises of the previous section, we set $j_0 = \lfloor -c_2/(c_1 - c_2) \rfloor = -2$. For the remaining design parameters γ_1 and γ_2 we need a good estimate for σ. Lange et al. (1998) reported to have used $\sigma = 100$ in their power calculation, where the particular choice was based on previous experiences. Consequently, we use $\gamma_1 = 11$ and $\gamma_2 = 8.79$ for our calculations. Further on, $B_1 = -14.64$ and $B_2 = 32.52$. Correspondingly, we obtain $\tau = 0.8$ and $\gamma_0 = 14.92$ for partition (III). The resulting lower confidence bounds are given in Table 2.

The final partition (IV) will only be of use for the construction of confidence intervals for those ϑ_i, for which $B_{3-i} \geq 0$. In the present example this is not the case for B_1 and we fail to get any information about ϑ_2 and thus $L_2 = -\infty$. For the other parameter $L_1 = -25.28$. Note that all of the partitions considered here led to the same lower bound for ϑ_1, because $L_1(X) = A_1$ can not be improved upon in this example.

For reference purposes we also included a Dunnett (1955) type test, where HCSE treatment is the "control" to which the other two groups are compared with. The resulting confidence bounds $L_1 = -30.41$ and $L_2 = 19.91$ suggest that this procedure is particularly powerful in comparison to the IUT or its associated compatible confidence intervals, if one of the parameters is significant but the other not. However, if both parameters are effective (or ineffective) the IUT has the advantage of using the univariate quantile c_1, which is smaller than the bivariate quantile d used

by Dunnett (1955). Imagine that the present study would have been less ambitious and $\delta_1 = 20$ml would have been sufficient to show a relevant superiority of HCSE over the placebo group. In this scenario, the (single step) Dunnett (1955) procedure still fails to show that ϑ_1 is significant ($L_1 = -0.41, L_2 = 19.09$). But any of the partitions above provides sharper confidence bounds for ϑ_1 and possibly also for ϑ_2. Partition (II), for example, which treats both parameters equally, leads to $L_1 = 4.61$ and $L_2 = 8.79$. In contrast, partitions (ii) and (iii), which both emphasize inference on ϑ_2, yield $L_1 = 0$ and $L_2 = 23.92$.

4.2. Example 2

Pigeot et al. (2003) described a non-inferiority clinical trial on mild asthmatic patients consisting of a three armed gold standard design. The summary data of the trial are shown in Table 3. The primary endpoint was increase of forced vital capacity after 6 weeks. In order to be consistent with the results of Pigeot et al. (2003), we set $\alpha = 0.025$.

The hypotheses of interest in this trial are different to those discussed in the previous example. For the remaining section let $\vartheta_1 = \mu_S - \mu_P$ and $\vartheta_2 = \mu_T - \mu_P$. Pigeot at al. (2003) start considering the main hypothesis $H' : \mu_T - \mu_S \leq \delta$. The authors advocate that the relevance margin δ should not be chosen without taking the difference $\mu_S - \mu_P$ into account for "internal validation". Thus, above hypothesis is reformulated as $H_1 : \vartheta_2/\vartheta_1 \leq r$, where $r \in \mathbb{R}$ denotes the relevance fraction of the reference difference $\mu_S - \mu_P$. Pigeot et al. (2003) propose the use of Fieller's (1954) confidence intervals for testing the ratio in H_1 under the usual normality assumptions. In addition, it is proposed to test H_1 only if the gatekeeping hypothesis $H_{(1)} = H_2 : \vartheta_1 \leq 0$ is rejected before. This stepwise procedure, where the non-rejection of $H_{(1)}$ renders the testing of $H_{(2)} = H_1$ unnecessary, corresponds to the partition (V) in Section 3. Note that the stepwise testing approach of Pigeot et al. (2003) does not provide any confidence intervals if $H_{(1)}$ remains not rejected (i.e. $A_1 < 0$). For the present data set, $A_1 = 1.03 > 0$ and we assess that the standard treatment is indeed superior to placebo ($\vartheta_1 > 0$). For the second step, applying Fieller's (1954) formulae, one obtains $r^* = 0.39$. Thus, it has been shown that the difference $\mu_T - \mu_P$ is at least 39% of the reference value $\mu_S - \mu_P$. Since $r^* > 0$ and $\vartheta_1 > 0$ this also implies $\vartheta_2 > 0$.

As discussed previously, partition (V) is not the only way to approach the given test problem. In particular, if $r^* < 0$, the assessment of how much negative ϑ_2 will be compared to ϑ_1 might be replaced by a lower confidence bound for ϑ_2 itself. This leads to partitions as indicated in (VI). For the present data example, the conclusions remain the same as for partition (V), since $r^* > 0$.

Partition (VII) is designed for applications, where interest lies in the $\vartheta_i, i = 1, 2$, and the parameter of original interest $\vartheta_2 - \vartheta_1 = \mu_T - \mu_S$. Thus, instead of investigating $\mu_T - \mu_P$ on a relative scale (i.e. in comparison to $\mu_S - \mu_P$), we get

Table 3: Summary data for the comparative study in asthmatic patients

	Placebo	Treatment	Standard
Sample size	20	35	19
Mean	3.14	4.32	4.86
Std. dev.	0.97	1.16	1.03

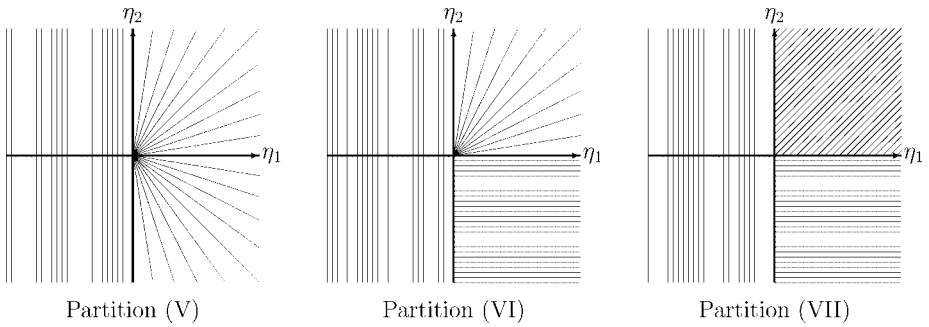

Figure 2: Graphical representation of the partitions (V)–(VII). The partition elements $\Theta_{(1,r)}$ and $\Theta_{(3,r)}$ are representetd by the vertical and horinzontal lines, respectively. The elements $\Theta_{(2,r)}$ and $\Theta_{(4,r)}$ are represented by angular and diagonal lines, respectively.

information about $\mu_T - \mu_S$ on the absolute scale. Since $A_3 = -1.15$, we conclude here that $\vartheta_i > 0, i = 1, 2$ and $\mu_T - \mu_S > -1.15$.

The last two partitions have shown that standard stepwise tests, as employed by Pigeot et al. (2003), for example, can be improved substantially and additional information is available without paying anything in terms of further multiplicity adjustment.

5. Conclusions

This paper addressed the problem of deriving simultaneous confidence intervals, which lead to the same hypothesis decisions as the associated tests. We focused on the evaluation of the IUT for two hypotheses. The number of partitions considered in this paper (and many other additional ones, which could have been included instead) indicate that research on selecting good partitions is still at its starting point. We tried to derive a few heuristic rules for a better understanding and to tailor this selection process. A suitable choice of such partitions and the concrete specification of their parameters seem to depend on the unknown constants and good estimates are required. Whether or not a satisfactory solution exists, remains to be seen. But the use of any IUT compatible confidence intervals substantially improves the standard IUT, as more information on the single parameters becomes available. In addition we have shown that in specific cases additional information on composite parameters are possible at no costs of further multiplicity adjustments. The two clinical examples have shown that the problems discussed in this paper are "real" and that better procedures than the current standard ones might be used in daily practice.

References

[1] Berger, R. L. (1982). Multi-parameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295–300. MR687187

[2] Berger, R. (1989). Uniformly more powerful tests for hypotheses concerning linear Inequalities and normal means. *J. Amer. Statist. Assoc.* **84**, 192–199. MR999678

[3] Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *J. Amer. Statist. Assoc.* **80**, 196–201.

[4] Fieller, E. C. (1954). Some problems in interval estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **16**, 175–185. MR93076

[5] Finner, H. and Strassburger, K. (2002). The partitioning principle: A powerful tool in multiple decision theory. *Ann. Statist.* **30**, 1194–1213. MR1926174

[6] Hayter, A. J. and Hsu, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *J. Amer. Statist. Assoc.* **89**, 128–136.

[7] Hsu, J. C. and Berger, R. L. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *J. Amer. Statist. Assoc.* **94**, 468–482.

[8] Lange, S., Freitag, G. and Trampisch, H. J. (1998). Practical experience with the design and analysis of a three-armed equivalence study. *Eur. J. Clin. Pharmacol.* **54**, 535–540.

[9] Laska, E. M. and Meisner, M. J. (1989). Testing whether an identified treatment is best. *Biometrics* **45**, 1139–1151. MR1040630

[10] Lehmann, E. (1986). *Testing statistical hypotheses* (2nd ed.). New York: Wiley. MR852406

[11] Pigeot, I., Schäfer, J., Röhmel, J. and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Stat. Med.* **22**, 883–899.

[12] Sarkar, S. K., Snapinn, S. and Wang, W. (1995). On improving the min test for the analysis of combination drug trials. *J. Statist. Comput. Simulation* **51**, 197–213. Correction: **60**, (1998) 180–181.

[13] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Sys. Tech. J.* **41**, 463–501. MR133183