The Annals of Statistics
2002, Vol. 30, No. 4, 1194–1213

THE PARTITIONING PRINCIPLE: A POWERFUL TOOL IN MULTIPLE DECISION THEORY[1]

BY H. FINNER AND K. STRASSBURGER

Deutsches Diabetes-Forschungsinstitut Düsseldorf

A first general principle and nowadays state of the art for the construction of powerful multiple test procedures controlling a multiple level α is the so-called closure principle. In this article we introduce another powerful tool for the construction of multiple decision procedures, especially for the construction of multiple test procedures and selection procedures. This tool is based on a partition of the parameter space and will be called partitioning principle (PP). In the first part of the paper we review basic concepts of multiple hypotheses testing and discuss a slight generalization of the current theory. In the second part we present various variants of the PP for the construction of multiple test procedures, these are a general PP (GPP), a weak PP (WPP) and a strong PP (SPP). It will be shown that, depending on the underlying decision problem, a PP may lead to more powerful test procedures than a formal application of the closure principle (FCP). Moreover, the more complex SPP may be more powerful than the WPP. Based on a duality between testing and selecting PPs can also be applied for the construction of more powerful selection procedures. In the third part of the paper FCP, WPP and SPP are applied and compared in some examples.

1. Introduction. A variety of multiple decision procedures has been proposed and examined over the past approximately fifty years. Besides selection and ranking approaches a convenient method to translate several questions concerning a statistical experiment into a tractable form is the formulation of a multiple hypotheses testing problem. Within this approach the most common requirement is the control of what we call the multiple level α, commonly known in the literature as strong control of the family-wise error rate at α, that is, the probability of any false rejection of a true (null) hypothesis should be bounded by a prespecified significance level α. This requirement causes serious problems in finding good and powerful multiple test procedures.

A helpful tool for the construction of multiple level-α tests is the so-called closure principle (CP for short). It was stated explicitly for the first time in Marcus, Peritz and Gabriel (1976) but had been in the air for a long time previously. Similar ideas closely related to the CP can be found, for example, in Fisher

---

Received December 2000; revised October 2001.

[1]Supported by the Deutsche Forschungsgemeinschaft.

AMS 2000 subject classifications. Primary 62J15, 62F07; secondary 62F03, 62C99.

Key words and phrases. Directional errors, formal closure principle, multiple comparisons, multiple level, multiple hypotheses testing, selection procedure, step-down procedure, step-up procedure, strong partitioning principle, weak partitioning principle.

(1935), Newman (1939), Keuls (1952), Tukey (1953), Hartley (1955), Miller (1966, 1981) and Naik (1975, 1977). A first more formal investigation of the CP is given in Sonnemann (1982). Application of the CP leads to the class of closed or stepwise test procedures. Although the CP is a very powerful method for constructing new and improving existing multiple test procedures, it generates a bunch of new questions and problems as, for example, the complexity of the resulting test procedures, the problem of directional errors in two-sided testing situations, problems in finding useful confidence sets being compatible with the results of the test procedure, or, difficulties in proving that the critical values of a stepwise procedure satisfy desirable monotonicity properties. Further and in general difficult problems are the comparison of different multiple test procedures and the related questions concerning admissibility.

Stefansson, Kim and Hsu (1988) first investigated methods based on partitioning the parameter space to derive confidence sets being compatible with the results of stepwise or closed multiple test procedures in specific situations; compare also Hayter and Hsu (1994) and the textbook of Hsu (1996). In Finner and Giani (1994, 1996) it is shown that the closure principle can also be applied for the construction of selection procedures. They developed a general duality between selection and multiple testing problems.

A question near at hand is whether the CP is the most powerful principle for the construction of a multiple test procedure for a prespecified family of hypotheses. The goal of this paper is to show that suitable modifications of the closure principle may lead to considerably more powerful decision procedures.

The paper is organized as follows. In Section 2 we first introduce some notation and basic definitions. Then we briefly discuss and review coherence, closure principle and interpretation of multiple decisions and the relationship to confidence sets. The closure principle is stated as Theorem 2.1, and Theorem 2.2 shows that only coherent tests should be considered. Moreover, we introduce strongly coherent multiple tests, which leads to a slight improvement of the CP (cf. Theorem 2.3) as outlined in Finner (1994a).

In Section 3 we first formulate a general partitioning principle (GPP) in Theorem 3.1 which is based on a suitable partition of the underlying family of hypotheses.

It will be seen in Lemma 3.1 that any family of hypotheses generates a natural partition of the parameter space. The relationship between a family of hypotheses being closed under intersections and the natural partition will be summarized in Remark 3.1.

Next we specify what we mean by a formal application of the closure principle (FCP).

Then we consider two variants of the GPP designed for application in practice; these are a weak partitioning principle (WPP) stated as Theorem 3.2 and a strong partitioning principle (SPP) stated as Theorem 3.3. Moreover, a first comparison of WPP, SPP and FCP is given. It will be shown that the WPP-related tests may be more powerful as FCP-related tests and that SPP-related tests may be more powerful than WPP-related tests. While Stefansson, Kim and Hsu (1988) used a partitioning principle to complement FCP-related multiple tests with confidence sets, the WPP and SPP, respectively, are designed to produce possibly more (and never less) powerful tests than FCP-related tests.

In Section 4 we apply and compare FCP, WPP and SPP in some selected examples. Some concluding remarks are given in Section 5.

2. Closure principle, coherence and strong coherence.

We first introduce some notation. Let $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be a family of probability measures defined on a common sample space $(\mathcal{X}, \mathcal{B})$ where $\mathcal{B}$ denotes a suitable σ-field over $\mathcal{X}$ and Θ is a parameter space with $|\Theta| \geq 2$ ($|A|$ denotes the cardinality of a set A). Let $\mathcal{H} = \{H_i : i \in I\}$ be a family of (null) hypotheses with $\varnothing \neq H_i \subset \Theta,\ H_i \neq H_j$ for $i \neq j,\ i, j \in I$, where I is any index set which may be finite or infinite (the notation $A \subset B$ will be used if $A \subseteq B$ and $A \neq B$). The alternative hypotheses are defined by $K_i = \Theta \setminus H_i$. If all hypotheses in $\mathcal{H} = \{H_i : i \in I\}$ are pairwise disjoint, then $\mathcal{H}$ will be called *disjoint*. For convenience, the union of disjoint sets or hypotheses $H_j,\ j \in J$, will be denoted by $\sum_{j \in J} H_j$. In the case that there exists at least one implication relation in $\mathcal{H}$, that is, $H_i \subset H_j$ for some $i, j \in I$, then various authors call $\mathcal{H}$ *hierarchical* (and *nonhierarchical* if there exists no such implication relation). Furthermore, $\mathcal{H}$ is said to be *closed* under (arbitrary) intersections (short: $\cap$-*closed*), if for all $\varnothing \neq J \subseteq I$ either $\bigcap_{j \in J} H_j \in \mathcal{H}$ or $\bigcap_{j \in J} H_j = \varnothing$. If $H_I = \bigcap_{i \in I} H_i \in \mathcal{H}$, then H_I is usually referred to as the *global* hypothesis.

Misleading is the definition of minimal and nonminimal hypotheses introduced by Gabriel (1969). He called a hypothesis $H_i \in \mathcal{H}$ minimal in $\mathcal{H}$, if there exists no $H_j \in \mathcal{H}$ with $H_j \supset H_i$. With respect to the natural order induced by the implication relation we reverse this definition. A hypothesis $H_i \in \mathcal{H}$ will be called *minimal* (*maximal*), if there exists no $H_j \in \mathcal{H}$ with $H_j \subset H_i$ ($H_j \supset H_i$). Sonnemann (1982) used the term *elementary* hypothesis for a maximal hypothesis. Often (but not always) the set of elementary or maximal hypotheses can be viewed as the hypotheses of main interest. If $\vartheta \in \Theta$ is the "true" parameter, then H_i is said to be *true* if $\vartheta \in H_i$. The index set $I(\vartheta) = \{i \in I : H_i \ni \vartheta\}$ will denote the set of all indices of true null hypotheses if ϑ is the true parameter. For convenience, throughout the paper we restrict attention to nonrandomized multiple tests, that is, the components of a multiple test $\psi = (\psi_i : i \in I)$ (say) take only values in $\{0, 1\}$ with the usual interpretation.

A multiple test $\psi = (\psi_i : i \in I)$ is said to control the multiple level α ($\alpha \in [0, 1]$) if

$$\forall\, \vartheta \in \Theta : P_\vartheta\left(\bigcup_{i \in I(\vartheta)} \{\psi_i = 1\} \right) \leq \alpha, \tag{2.1}$$

where the convention $\bigcup_{i \in \varnothing} A_i = \varnothing$ is used. The set of all multiple tests ψ for $\mathcal{H}$ satisfying (2.1) and some suitable measurability conditions is denoted by $\Phi_\alpha(\mathcal{H})$.

It has been shown in Sonnemann (1982) that condition (2.1) is equivalent to

$$\forall \varnothing \neq J \subseteq I : \forall \vartheta \in H_J = \bigcap_{j \in J} H_j : P_\vartheta\left(\bigcup_{j \in J} \{\psi_j = 1\}\right) \leq \alpha. \tag{2.2}$$

Sometimes the notion of local level-α tests is useful. A multiple test $\psi = (\psi_i : i \in I)$ is said to control the local level α ($\alpha \in [0, 1]$) if $\psi_i \in \Phi_\alpha(\{H_i\})$ for all $i \in I$. The set of all local level-α (multiple) tests for $\mathcal{H}$ is denoted by $\Phi_\alpha^{\text{loc}}(\mathcal{H})$. If $\mathcal{H}$ is disjoint, then $\Phi_\alpha(\mathcal{H}) = \Phi_\alpha^{\text{loc}}(\mathcal{H})$.

Several concepts to avoid contradictory results of multiple tests have been discussed in the literature; cf., for example, Lehmann (1957a, b) and Gabriel (1969). The following definition is due to Gabriel (1969).

A multiple test $\psi = (\psi_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ is said to be:

(a) *coherent*, if

$$\forall i, j \in I : [H_i \subset H_j \Longrightarrow \psi_i \geq \psi_j],$$

(b) *consonant*, if

$$\forall i \in I : \forall x \in \mathcal{X} : \left[\psi_i(x) = 1 \text{ and } \exists H_j \supset H_i \Longrightarrow \max_{r : H_r \supset H_i} \psi_r(x) = 1\right].$$

Conceptually, coherence is one of the most important terms in the theory of multiple hypotheses testing. If $H_i \subset H_j$ and H_j is rejected, it is more than logical that H_i should be rejected. A more formal reason for the requirement of coherence will be given below.

Comparisons of multiple tests are in general extremely difficult. However, it is often useful and possible to compare multiple tests pointwise and simultaneously in all components. To this end let $\psi^1, \psi^2 \in \Phi_\alpha(\mathcal{H})$. Then ψ^1 is said to be:

(a) *not less* than ψ^2 (short: $\psi^1 \geq \psi^2$ or $\psi^2 \leq \psi^1$), if $\forall i \in I : \forall x \in \mathcal{X} : \psi_i^1(x) \geq \psi_i^2(x)$,

(b) *greater* than ψ^2 (short: $\psi^1 > \psi^2$ or $\psi^2 < \psi^1$), if $\psi^1 \geq \psi^2$ and $\exists i \in I : \exists x \in \mathcal{X} : \psi_i^1(x) > \psi_i^2(x)$.

In this context terms like *not worse than* or *better than* are avoided because the previous definition does not distinguish whether a hypothesis is true or false. If $\psi^1, \psi^2 \in \Phi_\alpha(\mathcal{H})$ and $\psi^1 < \psi^2$, then ψ^1 is often referred to as a *conservative* test procedure.

The closure principle can be formally stated as follows.

THEOREM 2.1 (Closure principle). *Let $\mathcal{H}$ be $\cap$-closed, $\psi \in \Phi_\alpha^{\text{loc}}(\mathcal{H})$, and define*

$$\forall i \in I : \overline{\psi}_i = \min_{j : H_j \subseteq H_i} \psi_j. \tag{2.3}$$

Then $\overline{\psi} = (\overline{\psi}_i : i \in I) \in \Phi_\alpha(\mathcal{H})$, and $\overline{\psi}' = (\overline{\psi}_i : i \in I') \in \Phi_\alpha(\{H_i : i \in I'\})$ for all $\varnothing \neq I' \subset I$. Furthermore, $\overline{\psi}$ and $\overline{\psi}'$ are coherent.

It should be mentioned that if $\mathcal{H}$ is $\cap$-closed and if $\psi \in \Phi_\alpha^{\mathrm{loc}}(\mathcal{H})$ is coherent, then $\psi = \overline{\psi} \in \Phi_\alpha(\mathcal{H})$, where $\overline{\psi}$ is defined by (2.3) [cf. Sonnemann (1982)]. Hence, any coherent multiple test $\psi \in \Phi_\alpha(\mathcal{H})$ may be called a closed test or a closed multiple test procedure for $\Phi_\alpha(\mathcal{H})$. Theorem 2.1 shows that it is no restriction to consider a $\cap$-closed family $\overline{\mathcal{H}}$ generated by a given family $\mathcal{H} = \{H_i : i \in I\}$ by setting $\overline{\mathcal{H}} = \{H \subset \Theta : \exists\, J \subseteq I : H_J = H \neq \varnothing\}$, where H_J is defined as in (2.2). Subsequently, $\overline{\mathcal{H}}$ will be called the closure of $\mathcal{H}$. If $\psi \in \Phi_\alpha(\overline{\mathcal{H}})$ and one is only interested in a subfamily $\mathcal{H}' \subset \overline{\mathcal{H}}$ (e.g., the subfamily of all maximal hypotheses), the corresponding components of ψ yield a multiple level-α test ψ' for $\mathcal{H}'$. In most multiple hypotheses testing problems it is not difficult to construct a $\psi \in \Phi_\alpha^{\mathrm{loc}}(\overline{\mathcal{H}})$. The main practical difficulty will be the determination of $\overline{\psi}$ for $\overline{\mathcal{H}}$ defined in (2.3) or a subtest ψ' of $\overline{\psi}$ for a subfamily $\mathcal{H}' \subset \overline{\mathcal{H}}$, depending on the size of $\overline{\mathcal{H}}$ and the complexity of implication relations between the elements of $\overline{\mathcal{H}}$.

The levels of the components of the local level-α test are often called *nominal levels*, while the level of a component of $\overline{\psi}$, that is,

$$\alpha_i = \sup_{\vartheta \in H_i} P_\vartheta(\overline{\psi}_i = 1),$$

is sometimes called *true level* of $\overline{\psi}_i$ [cf., e.g., Hochberg and Tamhane (1987), page 67]. It is apparent that not the local levels but the true levels are the more interesting characteristic of a multiple test procedure (which may be closed or not). Notice that the acceptance region of $\overline{\psi}_i$ is given by

$$\{\overline{\psi}_i = 0\} = \bigcup_{j : H_j \subseteq H_i} \{\psi_j = 0\},$$

and this region may be very "large" compared with the acceptance regions $\{\psi_i = 0\}$ of the corresponding local level-α test ψ_i. In view of this fact it is obvious that a schematic application of the closure principle may lead to unnecessary conservative test procedures. On the other hand, this observation offers a chance to improve the closure principle.

We proceed with a result derived in Sonnemann and Finner (1988) which shows that there is no reason to use noncoherent multiple tests.

THEOREM 2.2. *Let* $\psi \in \Phi_\alpha(\mathcal{H})$ *and let* $\underline{\psi} = (\underline{\psi}_i : i \in I)$ *with* $\underline{\psi}_i = \max_{j : H_j \supseteq H_i} \psi_j$, $i \in I$. *Then* $\underline{\psi}$ *has the following properties*:

(a) $\underline{\psi} \in \Phi_\alpha(\mathcal{H})$,
(b) $\underline{\psi}$ *is coherent*,
(c) $\underline{\psi} \geq \psi$.

In other words, a noncoherent multiple level-α test can be replaced by a coherent multiple level-α test which is not less than the original test—and this can be done without additional costs, that is, without exceeding the multiple level α. Therefore, coherence is not only a natural but also a minimal requirement for multiple tests. Streitberg and Röhmel (1988) proposed to call $\underline{\psi}$ defined in Theorem 2.2 the co-closure of ψ.

Which decision can be concluded from a multiple level-α test $\psi \in \Phi_\alpha(\mathcal{H})$? From a decision theoretic point of view a rejected null hypothesis H_i is commonly interpreted as a decision for the alternative $K_i = \Theta \setminus H_i$. If H_i is accepted, it is best to conclude nothing, that is, the decision for Θ; cf., for example, Lehmann (1957b). In a multiple hypotheses testing situation there are mainly two possibilities for interpretation of the results. First, each component ψ_i of $\psi = (\psi_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ may be considered separately in the aforementioned sense. In many cases this approach results in a bunch of inconsistent and contradictory decisions; cf., for example, Lehmann (1957a, b) or Sonnemann (1982). The second possibility is to summarize all partial decisions into a joint decision. For notational convenience we define the (random) sets

$$H_{\psi(x)} = \bigcup_{j\,:\,\psi_j(x)=1} H_j \quad \text{and} \quad K_{\psi(x)} = \Theta \setminus H_{\psi(x)}, \qquad x \in \mathcal{X}.$$

The set $H_{\psi(x)}$ can be interpreted as the set of all ϑ's rejected by one of the components of ψ given $x \in \mathcal{X}$. Hence, the joint decision induced by ψ should be $K_{\psi(x)} = \bigcap_{j\,:\,\psi_j(x)=1} K_j$, which is in line with Lehmann (1957b). It seems to the authors that this point of view is very helpful in understanding the nature of multiple hypotheses testing and the relationship to other decision procedures as, for instance, the confidence set approach. We note that $C = (C(x) : x \in \mathcal{X})$ with $C(x) = K_{\psi(x)}$ defines a $(1-\alpha)$ confidence set for $\vartheta \in \Theta$ [Finner (1994a)].

It suggests itself that a null hypothesis H_i with $H_i \subseteq H_{\psi(x)}$ should be rejected. It has been shown in Finner (1994a) that this requirement yields a natural generalization of the coherence principle and finally a generalization of Theorem 2.2. In Finner (1994a) a multiple test $\psi \in \Phi_\alpha(\mathcal{H})$ is said to be *strongly coherent* if

$$\forall x \in \mathcal{X} : \forall i \in I : [H_i \subseteq H_{\psi(x)} \Longrightarrow \psi_i(x) = 1]. \tag{2.4}$$

Equivalently, a multiple test $\psi \in \Phi_\alpha(\mathcal{H})$ is strongly coherent iff

$$\psi_i = \min_{\vartheta \in H_i} \max_{j \in I(\vartheta)} \psi_j \qquad \text{for all } i \in I.$$

We conclude this section with the announced generalization of Theorem 2.2 for strongly coherent multiple tests.

THEOREM 2.3 [Strong closure principle, Finner (1994a)]. *Let $\psi \in \Phi_\alpha(\mathcal{H})$ and let $\overline{\psi} = (\overline{\psi}_i : i \in I)$ with*

$$\overline{\psi}_i = \min_{\vartheta \in H_i} \max_{j \in I(\vartheta)} \psi_j \qquad \textit{for all } i \in I. \tag{2.5}$$

Then $\overline{\psi}$ has the following properties:

(a) $\overline{\psi} \in \Phi_\alpha(\mathcal{H})$,
(b) $\overline{\psi}$ *is strongly coherent*,
(c) $\overline{\psi} \geq \psi$,
(d) $\bigcup_{i\in I(\vartheta)}\{\psi_i = 1\} = \bigcup_{i\in I(\vartheta)}\{\overline{\psi}_i = 1\}$,
(e) $H_{\psi(x)} = H_{\overline{\psi}(x)}$ *for all* $x \in \mathcal{X}$.

3. The partitioning principle. It is well known and mentioned in Section 2 that in the case of a disjoint family of hypotheses $\mathcal{H}$ a local level-α test is a multiple level-α test, that is, $\Phi_\alpha(\mathcal{H}) = \Phi_\alpha^{\text{loc}}(\mathcal{H})$. This fact is the basis for the partitioning principle which can be stated in several variants. The main idea is to partition the union of all hypotheses under consideration into disjoint sets $\Theta_i \subset \Theta$ such that each hypothesis can be written as the sum (i.e., disjoint union) of some of the Θ_i's. Then we construct level-α tests for all Θ_i, $i \in I$, which are then used to construct a multiple level-α test for the hypotheses of interest. As we will see, this can be done in various ways. We start with the following theorem the proof of which is trivial.

THEOREM 3.1 [General partitioning principle (GPP for short)]. *Let* $\mathcal{H} = \{H_i : i \in I\}$ *be a family of hypotheses and let* $\Theta_J = \{\Theta_j : j \in J\}$ *denote a partition of* $\Theta' \supseteq \bigcup_{i\in I} H_i$ *with a suitable index set* J *and* $\Theta' \subseteq \Theta$ *such that*

$$\forall\, j \in J : \Theta_j \neq \varnothing \quad \textit{and} \quad \forall\, i \in I : \exists\, J(i) \subseteq J : H_i = \sum_{j\in J(i)} \Theta_j.$$

Moreover, let $\varphi = (\varphi_j : j \in J) \in \Phi_\alpha^{\text{loc}}(\Theta_J)$ *and define* $\psi = (\psi_i : i \in I)$ *by*

$$\psi_i = \min_{j\in J(i)} \varphi_j \qquad \forall\, i \in I.$$

Then $\psi \in \Phi_\alpha(\mathcal{H})$ *and* ψ *is strongly coherent.*

Finally, for the joint decisions induced by ψ *and* φ *we have*

$$K_{\psi(x)} = \Biggl(\bigcup_{i\in I\,:\,\psi_i(x)=1} H_i\Biggr)^c \supseteq \Biggl(\sum_{j\in J\,:\,\varphi_j(x)=1} \Theta_j\Biggr)^c = K_{\varphi(x)}.$$

Note that φ may yield a more precise decision $K_{\varphi(x)}$ than ψ. The finest partition is $\Theta_\Theta = \{\{\vartheta\} : \vartheta \in \Theta\}$. The approach to construct a multiple level-α test for a given family of hypotheses $\mathcal{H} = \{H_i : i \in I\}$ via level-α tests for each hypothesis $H_\vartheta = \{\vartheta\}$ is due to the pioneering work of Stefansson, Kim and Hsu (1988). Their aim was the construction of confidence sets being compatible with stepwise or closed multiple tests which is a serious issue and has been believed to be impossible for a long time. However, this approach has its difficulties, too. It is often not easy to find tests for each ϑ which yield a useful joint decision containing more desirable information than the multiple test itself. Moreover, converting the results of unconventional level-α tests φ_ϑ for testing $H_\vartheta = \{\vartheta\}$ versus $K_\vartheta = \Theta \setminus \{\vartheta\}$,

$\vartheta \in \Theta$, into a confidence set and into results for all hypotheses can be a hard job. Examples where this method works well can be found, for example, in Miwa and Hayter (1999) and Hayter, Miwa and Liu (2000). However, the advantage of formulating and testing a bunch of (composite) hypotheses instead of testing each ϑ is that they provide us with an idea in which direction the tests should be powerful.

In what follows we restrict attention to the case where we are mainly interested in the testing results for the hypotheses of interest. However, Theorem 3.1 tells us that we can test some disjoint sets outside $\bigcup_{i \in I} H_i$ all at level α without any additional costs. For example, an additional level-α test for $(\bigcup_{i \in I} H_i)^c$ versus $\bigcup_{i \in I} H_i$ may result in rejection which then indicates that at least one of the hypotheses H_i, $i \in I$, may be true. Hence, the saying "the more questions (hypotheses), the more you have to pay (α, power)" is not always true.

The following lemma shows that any family of hypotheses $\mathcal{H} = \{H_i : i \in I\}$ can be used to generate a natural partition of $\bigcup_{i \in I} H_i$.

LEMMA 3.1. *Let* $\Theta_J = \{\vartheta \in \Theta : I(\vartheta) = J\}$, $\varnothing \neq J \subseteq I$, *and* $\mathcal{J} = \{\varnothing \neq J \subseteq I : \Theta_J \neq \varnothing\}$. *Then*

$$\Theta_{\mathcal{J}} = \{\Theta_J : J \in \mathcal{J}\}$$

is a partition of $\bigcup_{i \in I} H_i$ *and* $H_i = \sum_{J \in \mathcal{J} : J \ni i} \Theta_J$ *for all* $i \in I$. *Moreover, the* $\cap$*-closed family of hypotheses induced by* $\mathcal{H}$ *generates the same partition as* $\mathcal{H}$ *and for an intersection hypothesis* $H_J = \bigcap_{j \in J} H_j \neq \varnothing$ *it is* $H_J = \sum_{R \in \mathcal{J} : R \supseteq J} \Theta_R$.

The number of elements of the natural partition may be much smaller than the number of elements of the underlying $\cap$-closed family of hypotheses. For example, consider the one-sided pair hypotheses $H_{ij} : \vartheta_i - \vartheta_j \leq \delta$ for $1 \leq i, j \leq 3$, $i \neq j$, $\delta > 0$ fixed. Then the induced $\cap$-closed family has 63 elements while the natural partition has only 19 elements.

REMARK 3.1. (a) The natural partition $\Theta_{\mathcal{J}}$ is the coarsest possible partition among all partitions of $\bigcup_{i \in I} H_i$ with the property that each H_i can be represented as a disjoint union of sets of the underlying partition.

(b) Suppose that $\mathcal{H} = \{H_i : i \in I\}$ is $\cap$-closed. Let $\Theta_i = H_i \cap (\bigcup_{j : H_j \subset H_i} H_j)^c$ for $i \in I$ and let $J_p = \{i \in I : \Theta_i \neq \varnothing\}$. Then the natural partition generated by $\mathcal{H}$ is given by

$$\Theta(J_p) = \{\Theta_i : i \in J_p\},$$

that is, $\Theta(J_p) = \Theta_{\mathcal{J}}$. If $J_p = I$, then each hypothesis H_i can be identified with Θ_i and vice versa. In general $I \neq J_p$.

(c) The (partial) order on $\mathcal{H}$ defined by the implication relation $\subseteq$ between elements of $\mathcal{H}$ can be utilized to define a (natural) order (denoted by $\to$) on the natural partition generated by $\mathcal{H}$. Without loss of generality, let $\mathcal{H}$ be $\cap$-closed. Then the (natural) (partial) order on Θ_{J_p} is defined by

$$\forall\, i, j \in J_p : [H_j \subseteq H_i \Longrightarrow \Theta_j \to \Theta_i].$$

Note that $\Theta_i \to \Theta_i$ for all $i \in J_p$. Keeping this in mind we get

$$\forall\, i \in J_p : H_i = \sum_{j \in J_p : \Theta_j \to \Theta_i} \Theta_j.$$

(d) Two families of hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$ may lead to the same (natural) partition $\Theta_{\mathcal{J}}$ (say) but the order on $\Theta_{\mathcal{J}}$ generated by $\mathcal{H}_1$ may differ from the order on $\Theta_{\mathcal{J}}$ generated by $\mathcal{H}_2$.

(e) It is consistent with the coherence requirement of a multiple test for a $\cap$-closed family $\mathcal{H}$ and often useful to require that $\varphi = (\varphi_i : i \in J_p) \in \Phi_\alpha^{\mathrm{loc}}(\Theta(J_p))$ should satisfy the consistency (coherence) condition

$$\forall\, i, j \in J_p : [\Theta_i \to \Theta_j \Longrightarrow \varphi_i \geq \varphi_j]. \tag{3.1}$$

Note that $\psi = (\psi_i : i \in J_p)$ with $\psi_i = \min_{j \in J_p : \Theta_j \to \Theta_i} \varphi_j$ for all $i \in J_p$ satisfies (3.1).

When we are faced with the problem of constructing a level-α test φ (say) for a (composite) hypothesis H (say) with $|H| > 1$, in general we try to find a least favorable parameter configuration (LFC) ϑ^* (say) such that $\alpha^* = P_{\vartheta^*}(\varphi = 1) = \sup_{\vartheta \in H} P_\vartheta(\varphi = 1)$ becomes as large as possible with the restriction $\alpha^* \leq \alpha$. It is often possible (depending on assumptions concerning the underlying distributions) to reduce the number of LFC-candidates to a finite number. In multiple testing problems with a large number of composite hypotheses LFC-problems occur repeatedly. It will be shown that the partitioning principle compared to a formal application of the closure principle has some advantages concerning the determination of LFCs and may result in more powerful testing procedures.

In the following we formulate two variants of the partitioning principle designed for the construction of multiple level-α tests. Both principles will be based on the natural partition generated by a $\cap$-closed family of hypotheses. However, we first recall the situation when the closure principle is applied in a more or less formal and schematic way. We refer to this method as *formal closure principle* (FCP).

(A) *The formal closure principle* (*FCP*). The first step is to construct a local level-α test for a $\cap$-closed family of hypotheses $\mathcal{H} = \{H_i : i \in I\}$. The easiest way to do this is to choose a powerful level-α test φ_i for H_i for each $i \in I$ independently of each other. Loosely speaking, it is not required that a test φ_i takes into account the structure of other tests φ_j, $j \neq i$. In general, we choose the φ_i's without further restrictions except that they should be as powerful as possible for the corresponding testing problem H_i versus K_i such that

$$\alpha_i^* = \sup_{\vartheta \in H_i} P_\vartheta(\varphi_i = 1)$$

is as large as possible with the restriction $\alpha_i^* \le \alpha$ for all $i \in I$. Note that we have to look for LFCs over H_i for all $i \in I$. Consider now the resulting closed test $\psi = (\psi_i : i \in I)$ with $\psi_i = \min_{j\,:\,H_j \subseteq H_i} \varphi_j$, $i \in I$. As mentioned in Section 2, the true local levels $\alpha_i = \sup_{\vartheta \in H_i} P_\vartheta(\psi_i = 1)$ of ψ may be much smaller than the α_i^*'s so that the closed test procedure may be unnecessarily conservative. However, the advantage of applying the FCP is that it is mostly much easier to choose the local level-α tests φ_j independently of each other than to choose a multiple level-α test $\psi = (\psi_i : i \in I)$ such that $\psi_i = \min_{j\,:\,H_j \subseteq H_i} \psi_j$ and $\alpha_i = \sup_{\vartheta \in H_i} P_\vartheta(\psi_i = 1)$ is as large as possible but less than or equal to α for all $i \in I$.

(B) *The weak partitioning principle* (*WPP*). Next we show that a mixture of partitioning principle and formal closure principle, respectively, may improve a test procedure derived by the FCP. The following theorem based on the natural partition generated by a $\cap$-closed family of hypotheses is a direct consequence of the GPP described in Theorem 3.1.

THEOREM 3.2 (Weak partitioning principle). *Let $\mathcal{H} = \{H_i : i \in I\}$ be $\cap$-closed and let $\Theta(J_p) = \{\Theta_i : i \in J_p\}$ denote the natural partition generated by $\mathcal{H}$* [*as described in Remark* 3.1(b)]. *For each $i \in J_p$ choose tests φ_i for testing H_i versus K_i such that*

$$\sup_{\vartheta \in \Theta_i} P_\vartheta(\varphi_i = 1) \le \alpha,$$

that is, $\varphi = (\varphi_i : i \in J_p) \in \Phi_\alpha^{\mathrm{loc}}(\Theta(J_p))$. Define $\psi = (\psi_i : i \in I)$ by

$$\psi_i = \min_{j \in J_p\,:\,\Theta_j \subseteq H_i} \varphi_j \qquad \forall\, i \in I.$$

Then $\psi \in \Phi_\alpha(\mathcal{H})$ and ψ is strongly coherent.

In contrast to the FCP we now choose in a first step level-α tests φ_i for testing H_i versus K_i for $i \in J_p$ (i.e., I is replaced by $J_p \subseteq I$) such that type I errors are only controlled over $\Theta_i \subseteq H_i$, that is, we now have to look for LFCs over Θ_i instead of H_i as one would do when applying the FCP. As a result this may lead to a test with a smaller acceptance region, hence a more powerful test for H_i versus K_i which is now defined by $\psi_i = \min_{j \in J_p\,:\,\Theta_j \subseteq H_i} \varphi_j$.

Whether the WPP indeed yields a more powerful test procedure than the FCP heavily depends on the structure of $\mathcal{H}$ and the structure of the acceptance regions. Often tests are of the type $\{\varphi_i = 0\} = \{T_i \le c_i\}$, where T_i denotes a suitable test statistic. Assume for a moment that T_i tends to larger values if ϑ moves away from H_i. Then in the case of the FCP $c_i = c_i(\mathrm{FCP})$ (say) is determined such that $\alpha_i = \sup_{\vartheta \in H_i} P_\vartheta(T_i > c_i)$ is as large as possible but less than or equal to α while in the case of the WPP $c_i = c_i(\mathrm{WPP})$ (say) is determined such that $\alpha_i = \sup_{\vartheta \in \Theta_i} P_\vartheta(T_i > c_i)$ is as large as possible but less than or equal to α (assuming that $\Theta_i \neq \varnothing$). Clearly, if the LFCs are different and $c_i(\mathrm{WPP}) < c_i(\mathrm{FCP})$, the WPP yields a more powerful test than the FCP.

(C) *The strong partitioning principle* (*SPP*). A more advanced and sometimes more powerful method than FCP and WPP, respectively, is what we call the *strong partitioning principle* (SPP).

THEOREM 3.3 (Strong partitioning principle). *Under the assumptions of Theorem* 3.2 *choose tests* φ_i *for testing* H_i (*or* Θ_i) *versus* K_i *for each* $i \in J_p$ *such that*

$$\forall\, i \in J_p : \varphi_i = \min_{j \in J_p : \Theta_j \subseteq H_i} \varphi_j \tag{3.2}$$

and

$$\forall\, i \in J_p : \sup_{\vartheta \in \Theta_i} P_\vartheta(\varphi_i = 1) \le \alpha.$$

Moreover, *set* $\varphi_i = \min_{j \in J_p : \Theta_j \subseteq H_i} \varphi_j$ *for* $i \in I \setminus J_p$. *Then* $\varphi = (\varphi_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ *and* φ *is strongly coherent*.

The difference between SPP and WPP is the additional requirement (3.2) for the (local) level-α tests φ_i for H_i (or Θ_i) versus K_i, $i \in J_p$, which ensures the consistency condition (3.1). Once all φ_i, $i \in J_p$, are constructed, the remaining tests for those H_i with $i \in I \setminus J_p$ are defined as in the case of the WPP.

Suppose we have the same testing problem with underlying test statistics T_i as described in connection with the WPP. Let $i \in J_p$ and suppose we already have tests φ_j [and critical values $c_j = c_j$(SPP) (say)] for all $\Theta_j \to \Theta_i$, $\Theta_j \neq \Theta_i$, such that $\{\varphi_j = 0\} = \bigcup_{r \in J_p : \Theta_r \subseteq H_j} \{T_r \le c_r\}$ and $\sup_{\vartheta \in \Theta_j} P_\vartheta(\varphi_j = 1) \le \alpha$. Then we choose a critical value $c_i = c_i$(SPP) as small as possible (which may be difficult) such that

$$\sup_{\vartheta \in \Theta_i} P_\vartheta\left(\bigcap_{j : \Theta_j \subseteq H_i} \{T_j > c_j\} \right) \le \alpha.$$

Choosing all critical values in this way results in c_i(SPP) $\le c_i$(WPP) for all $i \in J_p$, but there may be some strict inequalities. In the latter case the SPP yields a more powerful test procedure than the WPP.

To sum up, we now have three tools to construct a multiple level-α test, the FCP, the WPP and the SPP. Whether the WPP (SPP) leads to an improved multiple decision procedure compared with a FCP-related (WPP-related) decision procedure heavily depends on the underlying decision problem. Roughly speaking, in situations with many composite hypotheses (e.g., for $H_{ij} : \vartheta_i - \vartheta_j \le \delta$, $1 \le i, j \le k$, $i \neq j$) there is often a good chance to improve FCP-related procedures while in cases where only point hypotheses (e.g., $H_{ij} : |\vartheta_i - \vartheta_j| = 0$, $1 \le i < j \le k$) are considered a (meaningful) improvement of FCP-related procedures is often impossible.

A serious issue is optimality and admissibility of multiple decision procedures. For example, one may ask whether a consequent application of the SPP yields an admissible multiple test procedure which cannot be improved further. Clearly, there cannot be a general answer to this question. However, sometimes it is possible to show that a SPP-related multiple test $\psi \in \Phi_\alpha(\overline{\mathcal{H}})$ (say) satisfies

$$\forall \psi' \in \Phi_\alpha(\overline{\mathcal{H}}): [\psi' \geq \psi \Longrightarrow \psi' = \psi \text{ almost everywhere}]$$

which means admissibility of ψ in a weak sense. This may happen especially if all LFCs used for the determination of critical values are finite. But often some of the components of an LFC are $\pm\infty$ which causes serious problems in proving admissibility. A further discussion of this issue goes beyond the scope of this paper and is directed to future research.

4. Application and comparison of FCP, WPP and SPP: examples. In this section we apply the FCP, WPP and SPP, respectively, in multiple hypotheses testing problems and a subset selection problem. For convenience we consider a simple k-sample normal model, that is, let $X = (X_1, \dots, X_k)$ denote a vector of independent random normal variables with mean vector $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Theta = \mathbb{R}^k$ and common known variance $\sigma^2 = 1$. Moreover, let $I_n = \{1, \dots, n\}$ for $n \in \mathbb{N}$ and let $\alpha \in (0, 1)$. We start with an illustrative example with 2 hypotheses.

EXAMPLE 4.1. Assume $k = 2$ for the moment and suppose the hypotheses of interest are $H_1: \max\{\vartheta_1, \vartheta_2\} \leq \delta$ and $H_2: \min\{\vartheta_1, \vartheta_2\} \geq -\delta$ for some fixed $\delta > 0$, that is, $\mathcal{H} = \{H_1, H_2\}$. Then $H_{12}: \max\{|\vartheta_1|, |\vartheta_2|\} \leq \delta$ is the intersection of H_1 and H_2, hence $\overline{\mathcal{H}} = \{H_1, H_2, H_{12}\}$. Let c_1^{FCP} be the solution of $\inf_{\vartheta \in H_1} P_\vartheta(\max\{X_1, X_2\} \leq c) = P_{(\delta,\delta)}(\max\{X_1, X_2\} \leq c) = 1 - \alpha$, that is, $c_1^{\text{FCP}} = \delta + \Phi^{-1}(\sqrt{1-\alpha}) = \delta + u_{\sqrt{1-\alpha}}$, and let c_2^{FCP} be the solution of $\inf_{\vartheta \in H_{12}} P_\vartheta(\max\{|X_1|, |X_2|\} \leq c) = P_{(\delta,\delta)}(\max\{|X_1|, |X_2|\} \leq c) = 1 - \alpha$, that is, c_2^{FCP} is the solution of the equation $\Phi(c - \delta) - \Phi(-c - \delta) = \sqrt{1-\alpha}$. Hence, the definition of a level-α test for every hypothesis in $\overline{\mathcal{H}}$ is obvious. Application of the FCP leads to the following (step-down) procedure: reject H_{12} if $\max\{|X_1|, |X_2|\} > c_2^{\text{FCP}}$, reject H_1 (H_2) if H_{12} is rejected and $\max\{X_1, X_2\} \geq c_1^{\text{FCP}}$ ($\min\{X_1, X_2\} \leq -c_1^{\text{FCP}}$).

Next we apply the WPP. The natural partition induced by $\overline{\mathcal{H}}$ is given by $\Theta_{12} = H_{12}$, $\Theta_i = H_i \setminus \Theta_{12}$, $i = 1, 2$. This yields an improved critical value c_1^{WPP} (say) instead of c_1^{FCP}. Now we have to solve $\inf_{\vartheta \in \Theta_1} P_\vartheta(\max\{X_1, X_2\} \leq c) = 1 - \alpha$. A little reflection yields as LFC for this problem $\vartheta = (\delta, -\delta)$ [or $\vartheta = (-\delta, \delta)$], that is, c_1^{WPP} can be obtained as the solution of the equation $\Phi(c - \delta)\Phi(c + \delta) = 1 - \alpha$. Hence, the WPP leads to the same procedure as before but now with critical values $c_2^{\text{FCP}}, c_1^{\text{WPP}}, -c_1^{\text{WPP}}$ with $c_1^{\text{WPP}} < c_1^{\text{FCP}}$.

Application of the SPP yields a further improvement. Now the structure of the local level-α tests for H_i, $i = 1, 2$, will change. We determine a suitable critical

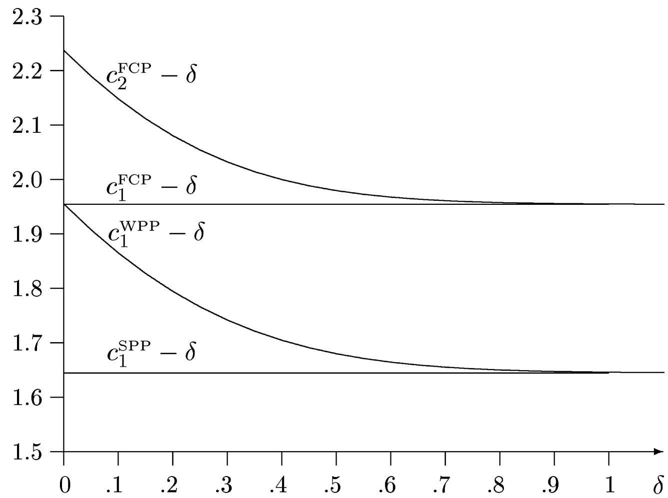

FIG. 1. *Values of* $c_2^{\mathrm{FCP}} - \delta$, $c_1^{\mathrm{FCP}} - \delta$, $c_1^{\mathrm{WPP}} - \delta$, $c_1^{\mathrm{SPP}} - \delta$ *in Example* 4.1 *for* $\alpha = 0.05$ *as functions of* δ.

value c_1^{SPP} such that

$$\inf_{\vartheta \in \Theta_1} P_\vartheta\big(\big\{\max\{|X_1|, |X_2|\} \le c_2^{\mathrm{FCP}}\big\} \cup \big\{\max\{X_1, X_2\} \le c_1^{\mathrm{SPP}}\big\}\big) = 1 - \alpha.$$

Although the determination of a LFC for this problem is less trivial than before, it can be shown that $\vartheta = (\delta, -\infty)$ [or $\vartheta = (-\infty, \delta)$] does the job, that is, $c_1^{\mathrm{SPP}} = \delta + \Phi^{-1}(1 - \alpha)$. Replacing c_1^{WPP} by c_1^{SPP} yields the announced improvement.

Figure 1 illustrates the improvement of SPP over WPP and WPP over FCP for $\alpha = 0.05$ and $0 \le \delta \le 1$. Finally we note that it is possible to add the hypotheses $H_3 : \max\{\vartheta_1, -\vartheta_2\} < -\delta$ and/or $H_4 : \min\{\vartheta_1, -\vartheta_2\} > \delta$ to, for example, $\overline{\mathcal{H}}$ without any additional costs, that is, one may test H_3 and H_4 each at full level α. For example, one may reject H_3 if $\max\{X_1, -X_2\} > c_3 = c_1^{\mathrm{FCP}} - 2\delta$.

The second example is devoted to the directional error problem in multiple hypotheses testing.

EXAMPLE 4.2. Let $\alpha \in (0, 1/2)$ for the moment and consider the family of hypotheses denoted by $\mathcal{H}$ containing $H_{2i-1} : \vartheta_i \ge -\delta$, $H_{2i} : \vartheta_i \le \delta$, $i \in I_k$. The $\cap$-closed family of hypotheses $\overline{\mathcal{H}} = \{H_J = \bigcap_{j \in J} H_j : \varnothing \ne J \subseteq I_{2k}\}$ generated by $\mathcal{H}$ has $2^{2k} - 1$ elements. The natural partition $\Theta_{\mathcal{J}}$ induced by $\overline{\mathcal{H}}$ has 3^k elements. First note that $H_J \setminus \bigcap_{R : J \subset R} H_R = \varnothing$ if there exists an $i \in I_k$ such that both $2i - 1$ and $2i$ are elements of J. Therefore,

$$\Theta_{\mathcal{J}} = \big\{\Gamma_1 \times \cdots \times \Gamma_k : \Gamma_i \in \{(-\infty, -\delta), [-\delta, \delta], (\delta, \infty)\} \text{ for all } i \in I_k\big\}.$$

A natural test statistic for testing a hypothesis $H_J \in \overline{\mathcal{H}}$ is $\max_{j \in J} T_j$ where $T_{2j-1} = -X_j$ and $T_{2j} = X_j$, $1 \le j \le k$. Application of the FCP requires level-α tests φ_J for each $H_J \in \overline{\mathcal{H}}$. Therefore, we define critical values c_J by

$$\inf_{\vartheta \in H_J} P_\vartheta\big(T_J(X) \le c_J\big) = 1 - \alpha, \qquad \varnothing \ne J \subseteq I_{2k}. \tag{4.1}$$

Since only one of the two hypotheses $H_{2i-1},\ H_{2i}$ $(1 \le i \le k)$ can be rejected, we can restrict attention to index sets $J \in \mathcal{M}$, where

$$\mathcal{M} = \big\{|J \cap \{2i-1, 2i\}| \ne 0 \text{ for all } 1 \le i \le k \text{ and } |J \cap \{2i-1, 2i\}| = 2| \text{ for at least one } i\big\},$$

that is, a hypothesis H_i is rejected if and only if $H_J \subseteq H_i$ is rejected at level α for all $J \in \mathcal{M}$. It turns out that the corresponding critical values depend only on $|J|$. Without loss of generality, we can restrict attention to J's of the type $J = I_{2r} \cup \{2r+2, 2r+4, \dots, 2k\}$. For $|J| = k + r$ the corresponding critical value will be denoted by c_r^{FCP}. It can easily be shown that a LFC under H_J is given by $\vartheta = (\delta, \dots, \delta)$. Hence, the corresponding critical value c_r^{FCP} is the solution of the equation

$$[\Phi(c - \delta) - \Phi(-c - \delta)]^r \Phi(c - \delta)^{k-r} = 1 - \alpha.$$

The critical values obviously satisfy $c_k^{\text{FCP}} > \cdots > c_1^{\text{FCP}}$. The resulting test procedure may be reformulated as a (short-cut) step-down procedure with directional decisions. Suppose w.l.o.g. that the realizations of $X_1, \dots, X_k$ satisfy $|x_k| > \cdots > |x_1|$. Then H_{2i-1} (H_{2i}) is rejected if $|x_j| > c_j^{\text{FCP}}$ for all $j \ge i$ and $x_i < -c_i^{\text{FCP}}$ $(x_i > c_i^{\text{FCP}})$. This procedure is slightly better than the Bonferroni–Holm procedure with critical values c'_j determined by $\Phi(c'_j) = 1 - \alpha/(k + j)$, $j = 1, \dots, k$. However, application of the WPP yields a first improvement upon the FCP-related procedure because LFCs change. The infimum in (4.1) is now taken over Θ_J instead of H_J. Consider again $J = I_{2r} \cup \{2r+2, 2r+4, \dots, 2k\}$. Then an LFC is given by $(\delta, \dots, \delta, -\delta, \dots, -\delta)$ with r entries δ. Hence, c_r^{WPP} is now the solution of

$$[\Phi(c - \delta) - \Phi(-c - \delta)]^r \Phi(c + \delta)^{k-r} = 1 - \alpha.$$

This results in $c_k^{\text{WPP}} > \cdots > c_1^{\text{WPP}}$, where $c_k^{\text{WPP}} = c_k^{\text{FCP}}$ and $c_i^{\text{WPP}} < c_i^{\text{FCP}}$ for $i = 1, \dots, k-1$. Finally, application of the SPP results in the classical directional error problem for stepwise multiple test procedures; cf., for example, Shaffer (1980) and Finner (1999). The results in, for example, Shaffer (1980) show that the critical values c_r^{SPP} can be chosen as solutions of

$$[\Phi(c - \delta) - \Phi(-c - \delta)]^r = 1 - \alpha, \qquad r = 1, \dots, k,$$

which results in $c_k^{\text{SPP}} > \cdots > c_1^{\text{SPP}}$, where $c_k^{\text{SPP}} = c_k^{\text{WPP}}$ and $c_i^{\text{SPP}} < c_i^{\text{WPP}}$ for $i = 1, \dots, k-1$. With the critical values $c_{|J|}^{\text{SPP}}$ we get in fact for each $J \in \mathcal{M}$ that

$$\inf_{\vartheta \in \Theta_J} P_\vartheta\Bigg(\bigcup_{R \in \mathcal{M}:\, J \subseteq R} \big\{T_R(X) \le c_{|R|}^{\text{SPP}}\big\}\Bigg) = 1 - \alpha.$$

For $J = I_{2r} \cup \{2r+2, 2r+4, \dots, 2k\} \in \mathcal{M}$ the infimum is attained for $\vartheta = (\delta, \dots, \delta, -\infty, \dots, -\infty)$ with r entries δ.

In the following example we consider a selection goal for the k-sample normal model where the ϑ_i's are now interpreted as population means. Given $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Theta = \mathbb{R}^k$, the set of good populations is denoted by $G(\vartheta) = \{i : \vartheta_{k:k} - \vartheta_i \le \delta\}$ for some fixed $\delta > 0$, where $\vartheta_{1:k} \le \cdots \le \vartheta_{k:k}$ denote the ordered parameter values. In Finner (1994a) and Finner and Giani (1996) one can find a general theory concerning the duality between selecting and multiple testing based on the closure principle. By utilizing a special *structured gain function* a selection goal can be translated into a multiple testing problem and the resulting multiple test can be converted in a selection procedure satisfying the desired selection goal. The duality theory may be extended to multiple tests based on a partitioning principle and corresponding selection procedures. But this is not the aim of this paper. Instead, we give a brief description how the construction principles discussed in this paper apply.

EXAMPLE 4.3 (Selection of a subset that contains all good populations). The aim is the construction of a selection rule $S(X)$ taking values in $\{J : \varnothing \ne J \subseteq I\}$ such that

$$\forall \vartheta \in \Theta : P_\vartheta\big(G(\vartheta) \subseteq S(X)\big) \ge P^*, \tag{4.2}$$

where the selected subset should be as small as possible and $P^* \in (0, 1)$ is fixed. Based on the $\cap$-closed family of hypotheses $\mathcal{H} = \{H_J : \varnothing \ne J \subseteq I\}$ with $H_J : J \subseteq G(\vartheta)$, test statistics $T_J = X_{k:I} - X_{1:J}$, where $X_{1:J} \le \cdots \le X_{k:J}$ denote the order statistics of the X_j, $j \in J$ and critical values $c_1^{\mathrm{FCP}} \le \cdots \le c_k^{\mathrm{FCP}}$ satisfying

$$\inf_{\vartheta \in H_J} P_\vartheta\big(T_J(X) \le c_{|J|}^{\mathrm{FCP}}\big) = P^*, \tag{4.3}$$

the duality theory yields that $S(X) = \{i \in I : T_J(X) > c_{|J|}$ for all $J \ni i\}$ defines a (step-down) subset selection rule satisfying (4.2). Note that $\varphi = (\varphi_J : \varnothing \ne J \subseteq I)$ with $\varphi_J(X) = 1$ iff $T_J(X) > c_{|J|}^{\mathrm{FCP}}$ defines a local level-$(1 - P^*)$ test for $\mathcal{H}$. One of the main problems was the calculation of LFCs ϑ_J^* such that $\inf_{\vartheta \in H_J} P_\vartheta(T_J(X) \le c_{|J|}^{\mathrm{FCP}}) = P_{\vartheta_J^*}(T_J(X) \le c_{|J|}^{\mathrm{FCP}}) = P^*$. A partial solution of the LFC problem can be found in Finner and Giani (1996). But now the WPP says that we can take the infimum in (4.3) over $H_J \setminus \bigcup_{R : R \supset J} H_R = \Theta_J$ (say) instead of H_J. But $\Theta_J = \{\vartheta : G(\vartheta) = J\}$ is much smaller than H_J. And in fact, it turns out that the LFCs over Θ_J and H_J differ for certain J's which finally results in smaller critical values and a more powerful (step-down) subset selection procedure. Detailed results which can be proved with similar methods as in Finner and Giani (2001) will be reported elsewhere.

Probably, application of the SPP may yield a further improvement. In this case one may try to find a set of critical values such that

$$\forall \varnothing \neq J \subseteq I : \inf_{\vartheta \in \Theta_J} P_\vartheta \left(\bigcup_{R\,:\,R \supseteq J} \{ T_R(X) \leq c^{\mathrm{SPP}}_{|R|} \} \right) = P^*.$$

It was pointed out in Finner and Giani (1996) that this problem is closely related to directional error problems in multiple comparisons [cf. Finner (1999)]. It is conjectured that the (optimal) critical values are finally determined by $P_\vartheta(\max_{1\leq i\leq j} X_i - \min_{1\leq i\leq j} X_i \leq c_j^{\mathrm{SPP}}) = P^*$, $j = 2, \ldots, k$, with $\vartheta \in \mathbb{R}^j$ having m entries equal to 0 and $j - m$ entries equal to δ if $j = 2m$ or $j = 2m + 1$.

Construction of a corresponding so-called step-up procedure seems to be an extremely difficult job and will not be considered here; for a proposal, cf. Finner and Giani (1996).

Closely related to the subset selection approach in Example 4.3 are MCB-methods (multiple comparisons with the best). Stefansson, Kim and Hsu (1988) discuss simultaneous confidence intervals (depending on a single critical value) for $\vartheta_i - \max_{j\neq i} \vartheta_j$ based on partitioning the parameter space and pivoting level-α tests on the resulting partitions (which corresponds to the case $\delta = 0$ in the subset selection formulation). These confidence intervals reproduce the testing results of a single-step procedure. Moreover, Hsu (1992) derived confidence intervals for $\vartheta_i - \max_{j\neq i} \vartheta_j$ (by partitioning the parameter space) which reproduce the results of a FCP-related step-down test procedure for the hypotheses H_i : $\vartheta_i - \max_{j\neq i} \vartheta_j = 0$.

The procedures described in Example 4.3 will be available shortly in the software package SEPARATE (SElecting, PARtitioning And TEsting, developed by the authors and others) including the case of unknown variance σ^2. The FCP-related step-down procedure described in Example 4.3 is already available there. SEPARATE is designed for sample size determination and the calculation of critical values in selecting, partitioning and testing problems. A demo version can be found on the Web page www.ddfi.uni-duesseldorf.de/main/separate/index.htm.

5. Concluding remarks. The examples in the previous section illustrate that application of both the WPP and SPP can lead to more powerful decision procedures as a formal application of the closure principle. The most rigorous principle is the SPP. On the one hand, the number of hypotheses which have to be tested may be reduced by applying a partitioning principle, on the other hand LFC problems may become more difficult because unions of acceptance regions may no longer be convex. In the case of a family of point hypotheses there may be no improvement of a FCP-related procedure by applying WPP or SPP because LFCs do not change. However, although testing of point hypotheses has a long tradition, the authors' opinion in most practical problems it is more sensible to formulate hypotheses or selection goals in terms of some threshold values. Rejection of a point hypotheses $H:\vartheta_i = 0$ and a decision like $\vartheta_i \neq 0$ is not very informative. Testing for *material significance* [as introduced by Hodges and Lehmann (1954)] or relevant differences may yield much more informative results.

Finally, we look at multiple decision problems from a more general perspective. Often certain subsets of the parameter space can be associated with some verbal interpretation as, for example, "Π_i is the best population" or "the populations $\Pi_i,\ i = 1, 3$, are good," and so on. At a fixed level of significance α we would like to learn something about the true ϑ-value. Therefore, we try to reject as many ϑ's as possible and end up with some subset of non rejected ϑ's as the final decision which can be interpreted as a confidence set at confidence level $(1-\alpha)$.

Without referring to hypotheses a decision process which is closest to the SPP can be described as follows. In a first step, we collect ϑ's which should be rejected simultaneously in (pairwise disjoint) sets $\Theta_i \subset \Theta$ of ϑ's with $|\Theta_i| \geq 1$. This yields a partition $\Theta_I = \{\Theta_i : i \in I\}$ of Θ, where I is a suitable index set. In a second step, we fix a (partial) order "$\rightarrow$" on Θ_I (which determines the order of the rejection process) which is reflexive, antisymmetric, and transitive, that is, (i) $\forall\, i \in I : \Theta_i \rightarrow \Theta_i$, (ii) $\forall\, i, j \in I : [\Theta_i \rightarrow \Theta_j$ and $\Theta_j \rightarrow \Theta_i \Longrightarrow \Theta_i = \Theta_j]$, and, (iii) $\forall\, i, j, k \in I : [\Theta_i \rightarrow \Theta_j$ and $\Theta_j \rightarrow \Theta_k \Longrightarrow \Theta_i \rightarrow \Theta_k]$. In a third step, we construct a multiple level-α test for Θ_I satisfying the consistency condition (3.1). Finally, the multiple test procedure can be translated into a $(1-\alpha)$ confidence set.

The main problem is to fix a partial order on Θ_I to steer the rejection process. As we have seen in the examples in the previous section, application of the SPP often yields a natural order how to proceed through the thicket of partitions (hypotheses, questions) and how to incorporate previous acceptance or rejection regions into the next step.

A method described in Hayter and Hsu (1994) and Finner (1994b) in connection with two-sided tests and one-sided confidence intervals can be considered as an application of the SPP, too. Starting with, for example, a two-sided test for $H:\vartheta = 0$ we proceed after rejection of H in two directions simultaneously. A value $\vartheta > 0$ ($\vartheta < 0$) can only be rejected if all $\vartheta' \in [0, \vartheta)$ ($\vartheta' \in (\vartheta, 0]$) (or, alternatively, all hypotheses $H = [0, \vartheta']$ for $\vartheta' \in [0, \vartheta)$ ($H = [\vartheta', 0]$ for $\vartheta' \in (\vartheta, 0]$)) are rejected in previous steps. An appropriate choice of level-α tests for each ϑ then yields (in the case that $\vartheta = 0$ is rejected) a one-sided confidence interval which is close to the corresponding classical full level-α one-sided confidence interval.

Hayter and Hsu (1994) extended this method for two-dimensional problems. For higher dimensions it is mostly very difficult to fix a suitable order of testing for each $\vartheta \in \Theta$, at least, if all directions have the same weight and if there are no ticketed ones.

However, sometimes the hypotheses of interest are pre-ordered as for instance in dose-response and toxicity studies. Hsu and Berger (1999) derived stepwise confidence intervals related to pre-ordered hypotheses which can be considered as a result of a partitioning principle. Consider for example the hypotheses H_i : $\vartheta_i \le \vartheta_0 + \delta$, $i = 1, \dots, k$, $\delta > 0$ fixed, where ϑ_0 is the mean of a control group and $\vartheta_1, \dots, \vartheta_k$ are the mean responses corresponding to increasing doses of a test drug. One may test the hypotheses in a stepwise manner (H_1 first, then H_2 and so on) all at level α until the first acceptance occurs. This yields a multiple level-α test which is due to the fact that $\Theta_1 = H_1$ ($\equiv$ "is dose 1 not efficacious"), $\Theta_2 = H_1^c \cap H_2$ ($\equiv$ "is dose 1 efficacious but dose 2 not efficacious"), $\dots$, $\Theta_k = H_1^c \cap \cdots \cap H_{k-1}^c \cap H_k$ are disjoint. A more precise joint decision yields the following (partitioning) method. Let $\Theta_{i,\eta} = \{\vartheta : \vartheta_i = \vartheta_0 + \eta\} \cap H_1^c \cap \cdots \cap H_{i-1}^c$ for $i = 1, \dots, k$ and $\eta \le \delta$ with $H_0 = \Theta = \mathbb{R}^{k+1}$, and let $\Theta_{k+1,\eta} = \{\vartheta : \min_{i=1,\dots,k} \vartheta_i = \vartheta_0 + \eta\} \cap H_1^c \cap \cdots \cap H_k^c$ for $\eta > \delta$. This yields a partition of Θ. Define a partial order on this partition by $\Theta_{i,\eta_1} \to \Theta_{j,\eta_2}$ if $i < j$ and $\Theta_{i,\eta_1} \to \Theta_{i,\eta_2}$ if $\eta_1 \le \eta_2$, respectively [with $\eta_m \le (>) \delta$ according to i (or j) $\le$ (or $>$) k, $m = 1, 2$]. Testing all these $\Theta_{i,\eta}$ with appropriate one-sided tests all at level α with respect to the predetermined order leads to the stepwise confidence interval procedure described in Hsu and Berger (1999).

Undoubtedly, the method of testing each $\vartheta \in \Theta$ with a level-α test [which yields a family of $(1 - \alpha)$ confidence sets] is the most powerful method and yields the most precise decision. Therefore, one may argue that there is no necessity for something like a closure principle or partitioning principle. However, the authors' opinion such principles are a helpful tool for the construction of *meaningful* decisions. Moreover, one may consider the application of such principles as a first step in finding a decision. In a second step, one can try to improve the decision of a multiple hypotheses test or selection procedure with a method as described in Stefansson, Kim and Hsu (1988).

Acknowledgments. We are grateful to a referee and an Associate Editor for their valuable comments and suggestions.

REFERENCES

FINNER, H. (1994a). Testing multiple hypotheses: General theory, specific problems, and relationships to other multiple decision procedures. Habilitationsschrift, FB IV Mathematik, Univ. Trier.

FINNER, H. (1994b). Two-sided tests and one-sided confidence bounds. *Ann. Statist.* **22** 1502–1516.

FINNER, H. (1999). Stepwise multiple test procedures and control of directional errors. *Ann. Statist.* **27** 274–289.

FINNER, H. and GIANI, G. (1994). Closed subset selection procedures for selecting good populations. *J. Statist. Plann. Inference* **38** 179–199.

FINNER, H. and GIANI, G. (1996). Duality between multiple testing and selecting. *J. Statist. Plann. Inference* **54** 201–227.

FINNER, H. and GIANI, G. (2001). Least favorable parameter configurations for a step-down subset selection procedure. *Biom. J.* **43** 543–552.

FISHER, R. A. (1935). *The Design of Experiments.* Oliver and Boyd, London.

GABRIEL, K. R. (1969). Simultaneous test procedures—Some theory of multiple comparisons. *Ann. Math. Statist.* **40** 224–250.

HARTLEY, H. O. (1955). Some recent developments in analysis of variance. *Comm. Pure Appl. Math.* **8** 47–72.

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.

HAYTER, A. J., MIWA, T. and LIU, W. (2000). Combining the advantages of one-sided and two-sided procedures for comparing several treatments with a control. *J. Statist. Plann. Inference* **86** 81–99.

HOCHBERG, Y. and TAMHANE, A. C. (1987). *Multiple Comparison Procedures.* Wiley, New York.

HODGES, J. L., Jr. and LEHMANN, E. L. (1954). Testing the approximate validity of statistical hypotheses. *J. Roy. Statist. Soc. Ser. B* **16** 261–268.

HSU, J. C. (1992). Stepwise multiple comparisons with the best. *J. Statist. Plann. Inference* **33** 197–204.

HSU, J. C. (1996). *Multiple Comparisons: Theory and Methods.* Chapman and Hall, London.

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.

KEULS, M. (1952). The use of the "Studentized range" in connection with an analysis of variance. *Euphytica* **1** 112–122.

LEHMANN, E. L. (1957a). A theory of some multiple decision problems, I. *Ann. Math. Statist.* **28** 1–25.

LEHMANN, E. L. (1957b). A theory of some multiple decision problems, II. *Ann. Math. Statist.* **28** 547–572.

MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660.

MILLER, R. G., JR. (1966). *Simultaneous Statistical Inference.* McGraw-Hill, New York.

MILLER, R. G., JR. (1981). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York.

MIWA, T. and HAYTER, A. J. (1999). Combining the advantages of one-sided and two-sided test procedures for comparing several treatment effects. *J. Amer. Statist. Assoc.* **94** 302–307.

NAIK, U. D. (1975). Some selection rules for comparing p processes with a standard. *Comm. Statist.* **4** 519–535.

NAIK, U. D. (1977). Some subset selection problems. *Comm. Statist. Theory Methods* **A6** 955–966.

NEWMAN, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* **31** 20–30.

SHAFFER, J. P. (1980). Control of directional errors with stagewise multiple test procedures. *Ann. Statist.* **8** 1342–1347.

SONNEMANN, E. (1982). Allgemeine Lösungen multipler Testprobleme. *EDV in Medizin und Biologie* **13** 120–128.

SONNEMANN, E. and FINNER, H. (1988). Vollständigkeitssätze für multiple Testprobleme. In *Multiple Hypothesenprüfung* (P. Bauer et al., eds.) 121–135. Springer, Berlin.

STEFANSSON, G., KIM, W.-C. and HSU, J. C. (1988). On confidence sets in multiple comparisons. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **2** 89–104. Academic Press, New York.

STREITBERG, B. and RÖHMEL, J. (1988). Diskussion: Einige strukturelle Aspekte bei multiplen Testproblemen. In *Multiple Hypothesenprüfung* (P. Bauer et al., eds.) 136–143. Springer, Berlin.

TUKEY, J. W. (1953). The problem of multiple comparisons. Mimeographed monograph.

DEUTSCHES DIABETES-FORSCHUNGSINSTITUT AN DER
HEINRICH-HEINE UNIVERSITÄT DÜSSELDORF
ABTEILUNG BIOMETRIE UND EPIDEMIOLOGIE
AUF'M HENNEKAMP 65
D-40225 DÜSSELDORF
GERMANY
E-MAIL: finner@ddfi.uni-duesseldorf.de
strass@ddfi.uni-duesseldorf.de