The International Journal of Biostatistics

Volume 7, *Issue* 1 2011 *Article* 12

Consonance and the Closure Method in Multiple Testing

Joseph P. Romano, *Stanford University*
Azeem Shaikh, *University of Chicago*
Michael Wolf, *University of Zurich*

Recommended Citation:
Romano, Joseph P.; Shaikh, Azeem; and Wolf, Michael (2011) "Consonance and the Closure Method in Multiple Testing," *The International Journal of Biostatistics*: Vol. 7 : Iss. 1, Article 12.
Available at: http://www.bepress.com/ijb/vol7/iss1/12
DOI: 10.2202/1557-4679.1300

Consonance and the Closure Method in Multiple Testing

Joseph P. Romano, Azeem Shaikh, and Michael Wolf

Abstract

Consider the problem of testing s null hypotheses simultaneously. In order to deal with the multiplicity problem, the classical approach is to restrict attention to multiple testing procedures that control the familywise error rate (FWE). The closure method of Marcus et al. (1976) reduces the problem of constructing such procedures to one of constructing single tests that control the usual probability of a Type 1 error. It was shown by Sonnemann (1982, 2008) that any coherent multiple testing procedure can be constructed using the closure method. Moreover, it was shown by Sonnemann and Finner (1988) that any incoherent multiple testing procedure can be replaced by a coherent multiple testing procedure which is at least as good. In this paper, we first show an analogous result for dissonant and consonant multiple testing procedures. We show further that, in many cases, the improvement of the consonant multiple testing procedure over the dissonant multiple testing procedure may in fact be strict in the sense that it has strictly greater probability of detecting a false null hypothesis while still maintaining control of the FWE. Finally, we show how consonance can be used in the construction of some optimal maximin multiple testing procedures. This last result is especially of interest because there are very few results on optimality in the multiple testing literature.

KEYWORDS: multiple testing, closure method, coherence, consonance, familywise error rate

Author Notes: Research of the first author has been supported by NSF Grant DMS-0707085. Research of the second author has been supported by NSF Grant DMS-0820310. Research of third author has been supported by the Swiss National Science Foundation (NCCR FINRISK, Module A3).

1 Introduction

Consider the general problem of simultaneously testing s null hypotheses of interest $H_1, \ldots, H_s$. Data X with distribution $P \in \Omega$ are available, where the parameter space Ω may be a parametric, semiparametric or nonparametric model for P. In this setting, a general hypothesis H can be viewed as a subset ω of Ω. The problem is to test null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H_i' : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. Let $I(P)$ denote the indices of the set of true null hypotheses when P is the true probability distribution, that is, $i \in I(P)$ if and only if (iff) $P \in \omega_i$.

If tests for each of the s null hypotheses of interest are available, then one may simply disregard the multiplicity and test each hypothesis in the usual way at level α. However, with such a procedure, the probability of one or more false rejections generally increases with s and may be much greater than α. A classical approach to dealing with this problem is to restrict attention to multiple testing procedures that control the probability of one or more false rejections. This probability is called the *familywise error rate* (FWE). Here, the term "family" refers to the collection of hypotheses $H_1, \ldots, H_s$ which is being considered for simultaneous testing. Control of the FWE at level α requires that

$$\mathrm{FWE}_P \le \alpha \text{ for all } P \in \Omega ,$$

where

$$\mathrm{FWE}_P \equiv P\Big\{\text{reject any } H_i \text{ with } i \in I(P)\Big\} .$$

Note that we require $\mathrm{FWE}_P \le \alpha$ for all $P \in \Omega$. Control of the familywise error rate in this sense is called *strong* control of the FWE to distinguish it from *weak* control, where $\mathrm{FWE}_P \le \alpha$ is only required to hold when all null hypotheses are true, that is, when $P \in \bigcap_{1 \le i \le s} \omega_i$. Since weak control is only of limited use in multiple testing, control will always mean strong control for the remainder of this paper. A quite broad treatment of multiple testing procedures which control the FWE is presented in Hochberg and Tamhane (1987).

The most well-known multiple testing procedure is the Bonferroni procedure. If $\hat{p}_i$ denotes a p-value for hypothesis H_i, then the Bonferroni procedure rejects H_i iff $\hat{p}_i \le \alpha/s$. Other multiple testing procedures that only require the s p-values $\hat{p}_1, \ldots, \hat{p}_s$ as input are those of Holm (1979) and Simes (1986). Such procedures are very easy to apply, but they may be quite conservative, such as the procedures of Bonferroni and Holm (1979), or require strong restrictions on the dependence structure of the tests, such as the procedure of Simes (1986), which requires the p-values to be mutually independent.

A broad class of multiple testing procedures that control the FWE is provided by the closure method of Marcus et al. (1976), which reduces the problem of constructing such a multiple testing procedure to the problem of constructing tests of individual (or single) hypotheses that control the usual probability of a Type 1 error. Specifically, for a subset $K \subseteq \{1, \ldots, s\}$, define the intersection hypothesis

$$H_K : P \in \omega_K \ , \tag{1}$$

where

$$\omega_K \equiv \bigcap_{i \in K} \omega_i \ .$$

Of course, $H_i = H_{\{i\}}$. Suppose ϕ_K is a (possibly randomized) level α test of H_K, that is,

$$\sup_{P \in \omega_K} E_P[\phi_K] \leq \alpha \ . \tag{2}$$

The closure method, as a multiple testing procedure, rejects H_i iff H_K is rejected (based on ϕ_K) at level α for all subsets K for which $i \in K$. A thorough review of the topic of multiple testing, emphasizing, in particular, multiple testing procedures obtained using the closure method in a clinical trial setting, can be found in Bauer (1991). Unlike the procedures of Bonferroni, Holm (1979) and Simes (1986), as many as 2^s tests may in principle need to be carried out to apply the closure method. On the other hand, the resulting multiple testing procedure controls FWE without any restrictions on the dependence structure of the tests. To see why, we adapt the argument in Theorem 4.1 of Hochberg and Tamhane (1987). Let A be the event that at least one H_i with $i \in I(P)$ is rejected by the closure method, and let B be the event that the individual hypothesis $H_{I(P)}$ is rejected (based on $\phi_{I(P)}$) at level α. Since A implies B,

$$\text{FWE}_P = P\{A\} \leq P\{B\} \leq \alpha \ .$$

Moreover, as will be shown below, if the tests ϕ_K are designed properly, then the resulting multiple testing procedure may have desirable power properties.

By reducing the problem of controlling the FWE to that of constructing individual tests which control the usual probability of a Type 1 error, the closure method provides a very general means of constructing multiple testing procedures which control the FWE. Sonnemann (1982, 2008) in fact shows that any coherent multiple testing procedure can be constructed using the closure method. A multiple testing procedure is said to be *coherent* when the following condition holds: if a hypothesis $P \in \omega$ is not rejected, then any hypothesis $P \in \omega'$ with $\omega \subset \omega'$ must also not be rejected. (The symbol $\subset$ denotes a strict subset in contrast to the symbol $\subseteq$.) Sonnemann and Finner (1988)

DOI: 10.2202/1557-4679.1300

show further that any incoherent multiple testing procedure can be replaced by a coherent multiple testing procedure that rejects the same hypotheses and possibly more while still controlling the FWE. Hence, there is no loss in restricting attention to multiple testing procedures constructed using the closure method. These results are reviewed briefly in Section 2.

In this paper, we consider the problem of how "best" to choose tests of H_K when constructing a multiple testing procedure using the closure method. Even in the case $s = 2$, little formal theory exists in the design of tests of H_K. In fact, even if the individual tests are constructed in some optimal manner, multiple testing procedures obtained by the closure method may in fact be inadmissible. This finding was previously obtained in Bittman et al. (2009) in a specific context. In our analysis, the notion of *consonance* becomes pertinent. A multiple testing procedure obtained using the closure method is said to be *consonant* when the rejection of an intersection hypothesis implies the rejection of at least one of its component hypotheses. Here, we mean that H_j is a *component* of H_i if $\omega_i \subset \omega_j$. For example, a hypothesis specifying $\theta_1 = \theta_2 = 0$ has component hypotheses $\theta_1 = 0$ and $\theta_2 = 0$, and a consonant procedure which rejects $\theta_1 = \theta_2 = 0$ must reject at least one of the two component hypotheses. A procedure which is not consonant is called *dissonant*. Both the notions of coherence and consonance were first introduced by Gabriel (1969).

We show that there is no need to consider dissonant multiple testing procedures when testing *elementary* hypotheses, defined formally in Section 3. Indeed, in such a setting, any dissonant multiple testing procedure can be replaced by a consonant multiple testing procedure that rejects the same hypotheses and possibly more. We show further that in many cases the improvement of the consonant multiple testing procedure over the dissonant multiple testing procedure may in fact be strict in the sense that it has strictly greater probability of detecting a false null hypothesis while still maintaining control of the FWE. Finally, in Section 4, we show how consonance can be used in the construction of some optimal maximin multiple testing procedures. This last result is especially of interest because there are very few results on optimality in the multiple testing literature so far. Proofs of all results can be found in the appendix.

We introduce at this point a classic, running example which will be revisited throughout the remainder of the paper.

Example 1.1 (Two-sided Normal Means) For $1 \leq i \leq 2$, let X_i be independent with $X_i \sim N(\theta_i, 1)$. The parameter space Ω for $\theta = (\theta_1, \theta_2)$ is the entire real plane. Let $s = 2$, so there are only two hypotheses, and null hypothesis H_i specifies $\theta_i = 0$. To apply the closure method, suppose the test of

H_i is the uniformly most powerful unbiased (UMPU) level α test which rejects H_i iff $|X_i| > z_{1-\frac{\alpha}{2}}$, where z_λ denotes the λ quantile of the $N(0,1)$ distribution. All that remains is to choose a test of the intersection hypothesis $H_{\{1,2\}}$. There are two well-known choices.

(i) *The uniformly most powerful (rotationally) invariant test.* Reject $H_{\{1,2\}}$ iff (X_1, X_2) falls in the rejection region $R_{1,2}(\alpha)$ given by

$$R_{1,2}(\alpha) \equiv \left\{(x_1, x_2) : \; x_1^2 + x_2^2 > c_2(1-\alpha)\right\} ,$$

where $c_d(1-\alpha)$ denotes the $1-\alpha$ quantile of the χ_d^2 distribution. This test is also maximin and most stringent; see Section 8.6 of Lehmann and Romano (2005).

(ii) *Stepdown test based on maximum.* Reject $H_{\{1,2\}}$ iff

$$\max\bigl(|X_1|, |X_2|\bigr) > m_2(1-\alpha) , \tag{3}$$

where $m_s(1-\alpha)$ is the $1-\alpha$ quantile of the distribution of

$$\max\bigl(|X_1|, \ldots, |X_s|\bigr)$$

when the X_i are i.i.d. $\sim N(0,1)$.

In both cases, the multiple testing procedure begins by testing $H_{\{1,2\}}$. If $H_{\{1,2\}}$ is retained, then there are no rejections by the multiple testing procedure. If, on the other hand, it is rejected, then H_i is rejected by the multiple testing procedure iff $|X_i| > z_{1-\frac{\alpha}{2}}$. It is easy to see that

$$z_{1-\frac{\alpha}{2}} < m_2(1-\alpha) < c_2^{1/2}(1-\alpha) . \tag{4}$$

The rejection region for test (i) is the outside of a disc centered at the origin of radius $c_2^{1/2}(1-\alpha)$, while the rejection region for test (ii) is the outside of a square centered at the origin and having side length $2 \cdot m_2(1-\alpha)$; see Figure 1. We refer to the multiple testing procedure which uses test (i) above for the test of $H_{\{1,2\}}$ as procedure (i), and analogously to the multiple testing procedure which uses test (ii) above for the test of $H_{\{1,2\}}$ as procedure (ii).

These procedures generalize easily when there are in general s hypotheses. Let $X_1, \ldots, X_s$ be independent with $X_i \sim N(\theta_i, 1)$, and H_i specifies $\theta_i = 0$. For an arbitrary subset $K \subseteq \{1, \ldots, s\}$, consider the intersection hypothesis

H_K that specifies $\theta_i = 0$ for all $i \in K$. In order to generalize the first construction, let ϕ_K be the test which rejects H_K iff

$$\sum_{i \in K} X_i^2 > c_{|K|}(1-\alpha) \ ,$$

where $|K|$ is the number of elements in K, so that $c_{|K|}(1-\alpha)$ is the $1-\alpha$ quantile of $\chi^2_{|K|}$, and then apply the closure method to this family of tests. In order to generalize the second construction, let ϕ_K be the test which rejects H_K iff

$$\max_{i \in K}(|X_i|) > m_{|K|}(1-\alpha) \ ,$$

and then apply the closure method to this family of tests. It is natural to ask which family should be used in the construction of the closure method. An optimality result is presented in Section 4.1 for a new procedure that is nearly the same as the procedure based on the maximum statistic. ■

2 Coherence

We first provide a lemma, which is a converse of sorts to the closure method. Indeed, the closure method starts with tests of H_K at level α, for any $K \subseteq \{1, \ldots, s\}$, to produce a multiple testing procedure concerning the hypotheses of $H_1, \ldots, H_s$ of interest. Conversely, given any multiple testing procedure (not necessarily obtained by the closure method) concerning the hypotheses $H_1, \ldots, H_K$, one can use it to obtain tests of H_K at level α for any $K \subseteq \{1, \ldots, s\}$.

Lemma 2.1 *Let $X \sim P \in \Omega$. Suppose a given multiple testing procedure controls the FWE at level α for testing null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H_i' : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. Define a test ϕ_K of the intersection hypothesis H_K in (1) as follows: If $K = \{i\}$ for some i, then test H_K by the test of that H_i; otherwise, reject H_K if the given multiple testing procedure rejects any H_i with $i \in K$. Then, ϕ_K controls the usual probability of a Type 1 error at level α for testing H_K, that is, it satisfies (2).*

Define a family of hypotheses $H_1, \ldots, H_s$ to be *closed* if each intersection hypothesis H_K is a member of the family. The *closure* of a given family is the family of all intersection hypotheses induced by the given family. In some cases, there really is nothing to lose by assuming the given family is closed (that is, by considering the closure of the given family in case the family is not closed to begin with). On the one hand, when applying the closure method, one gets a multiple testing procedure that controls the FWE, not just for the original family, but for the closure of the family. On the other hand, if one is concerned only with the original family of hypotheses, then we shall see that the notion of consonance may play a role in determining the tests of the additional intersection hypotheses.

By definition, the closure method always guarantees that the resulting multiple testing procedure is *coherent*, that is, if H_I implies H_K in the sense that $\omega_I \subset \omega_K$, and H_I is not rejected, then H_K is not rejected. So, if H_K is rejected and $\omega_I \subset \omega_K$, then the requirement of coherence means H_I must be rejected. The requirement of coherence is reasonable because if H_K is established as being false, then H_I is then necessarily false as well if $\omega_I \subset \omega_K$. As stated in Hochberg and Tamhane (1987), coherence "avoids the inconsistency of rejecting a hypothesis without also rejecting all hypotheses implying it." Note that even the simple Bonferroni procedure may not be consonant or coherent; see Remark 3.3.

Sonnemann (1982, 2008) shows that any coherent multiple testing procedure can be constructed using the closure method. Sonnemann and Finner (1988) show further that any incoherent multiple testing procedure can be replaced by a coherent multiple testing procedure that rejects the same hypotheses and possibly more, while still controlling the FWE. Hence, there is no loss in restricting attention to multiple testing procedures constructed using the closure method. We restate these results below in Theorems 2.1 and 2.2 for completeness. A nice review can also be found in Finner and Strassburger (2002). Note, however, that we do not assume the family of hypotheses to be closed, and this feature will become important later when we discuss consonant multiple testing procedures.

Theorem 2.1 *Let $X \sim P \in \Omega$. Suppose a given multiple testing procedure controls the FWE at level α for testing null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H_i' : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. If the given multiple testing procedure is coherent, then it can be obtained by applying the closure method based on tests ϕ_K satisfying (2) for each $K \subseteq \{1, \ldots, s\}$. Thus, all coherent multiple testing procedures can be generated using the closure method.*

Remark 2.1 Note that the requirement of coherence does not restrict the multiple testing procedure unless any of the hypotheses imply any of the others, in the sense that there exist i and j with $\omega_i \subset \omega_j$. As a simple example,

DOI: 10.2202/1557-4679.1300

suppose $X = (X_1, \ldots, X_s)$ is multivariate normal with unknown mean vector $(\theta_1, \ldots, \theta_s)$ and known covariance matrix Σ. If H_i specifies $\theta_i = 0$, then no ω_i is contained in any other ω_j. Hence, in this case, the preceding theorem implies that *all* multiple testing procedures which control the FWE can be generated using the closure method. ■

The next theorem and remark show how an incoherent multiple testing procedure can be (weakly) improved upon by a coherent one in terms of the ability to reject false hypotheses in the sense that the new procedure is coherent, rejects the same hypotheses as the original one, and possibly rejects some additional hypotheses. In order to keep the statement of the theorem compact, we introduce the notion of *coherentization* of a general, incoherent multiple testing procedure. Specifically, assume without loss of generality (by including an auxiliary random variable in X, if necessary) that the incoherent multiple testing procedure rejects H_i when $X \in \tilde{R}_i$. Define the corresponding *coherentized* procedure as the one which rejects H_i when $X \in \tilde{R}'_i$ given by

$$\tilde{R}'_i \equiv \bigcup_{j : \omega_j \supseteq \omega_i} \tilde{R}_j \ . \tag{5}$$

Theorem 2.2 *Let $X \sim P \in \Omega$. Suppose a given multiple testing procedure controls the FWE at level α for testing null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H'_i : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. If the multiple testing procedure is incoherent, then it can be replaced by a coherent multiple testing procedure that rejects all the hypotheses rejected by the incoherent multiple testing procedure and possibly some further hypotheses while maintaining FWE control at level α. Specifically, the corresponding coherentized procedure based on* (5) *is coherent, rejects H_i whenever the incoherent multiple testing procedure rejects H_i, and maintains control of the FWE at level α.*

Remark 2.2 In Theorem 2.2, if we assume further that there exists i and some $P \in \omega_i^c$ such that

$$P\Big\{ \bigcup_{j : \omega_j \supset \omega_i} \tilde{R}_j \setminus \tilde{R}_i \Big\} > 0 \ ,$$

then the coherent multiple testing procedure is strictly better in the sense that for some $P \in \omega_i^c$, the probability of rejecting H_i is strictly greater under the coherentized procedure than under the given incoherent multiple testing procedure. Put differently, we must require that there exist i and j such that $\omega_i \subset \omega_j$ and $P\{\tilde{R}_j \setminus \tilde{R}_i\} > 0$ for some $P \in \omega_i^c$. ■

3 Consonance

It follows from Theorems 2.1 and 2.2 that we can restrict attention to multiple testing procedures obtained by the closure method. However, not all multiple testing procedures generated by applying the closure method are also consonant. Recall that a multiple testing procedure is consonant if at least one H_i with $i \in K$ is rejected by the multiple testing procedure whenever H_K is rejected at level α (based on ϕ_K). Hochberg and Tamhane (1987) write on page 46:

> Nonconsonance does not imply logical contradictions as noncoherence does. This is because the failure to reject a hypothesis is not usually interpreted as its acceptance. [...] Thus, whereas coherence is an essential requirement, consonance is only a desirable property.

A dissonant multiple testing procedure can, however, leave the statistician in a difficult situation when explaining the results of a study. Consider, for example, a randomized experiment for testing the efficiency of a drug versus a placebo with two primary endpoints: testing for reduction in headaches and testing for reduction in muscle pain. Suppose H_1 postulates that the drug is no more effective than the placebo for reduction of headaches and H_2 postulates that the drug is no more effective than the placebo for reduction of muscle pain. If the intersection hypothesis $H_{\{1,2\}}$ is rejected, but the statistician cannot reject either of the individual hypotheses, then compelling evidence has not been established to promote a particular drug indication. The net result is that neither hypothesis can be rejected, even though one might conclude that the drug has some beneficial effect. In this way, lack of consonance makes interpretation awkward.

More importantly, we will argue, not merely from an interpretive viewpoint, but from a mathematical statistics viewpoint, that dissonance is undesirable in that it results in decreased ability to reject false null hypotheses. (On the other hand, there may be applications where dissonance is informative. For example, it may be that a set of endpoints exhibit an overall effect in a common direction though not strongly enough where any individual endpoint is rejected. Such a diffuse overall effect may be of interest to warrant future experimentation.) For concreteness, let us revisit the running example.

Example 3.1 (Two-sided Normal Means, continued) It is easy to see that procedure (ii) is consonant, whereas procedure (i) is dissonant. In particular, it is possible in procedure (i) to reject the intersection hypothesis $H_{\{1,2\}}$,

DOI: 10.2202/1557-4679.1300

but to reject neither H_1 nor H_2 by the multiple testing procedure. For example, when $\alpha = 0.05$, then $c_2^{1/2}(0.95) = 2.448$; if $X_1 = X_2 = 1.83$, then $X_1^2 + X_2^2 = 6.698 = 2.588^2$, so $H_{\{1,2\}}$ is rejected, but neither X_i satisfies $|X_i| > 1.96$. For an illustration, see Figure 1.

Of course, it does not follow that procedure (ii) is preferred merely because it is consonant. More importantly, procedure (i) can be improved if the goal is to make correct decisions about H_1 and H_2. Moreover, this is true even though each of the tests of H_1, H_2 and $H_{\{1,2\}}$ possesses a strong optimality property. In particular, we see that the optimality of the individual tests of H_1, H_2 and $H_{\{1,2\}}$ does *not* translate into any optimality property for the overall multiple testing procedure.

In order to appreciate why, note that we may remove from procedure (i) points in the rejection region for testing the intersection hypothesis $H_{\{1,2\}}$ that do not allow for rejection of either H_1 or H_2. By doing so, we can instead include other points in the rejection region that satisfy the constraint that the overall multiple testing procedure be consonant, while still maintaining FWE control. In this example, this requires our test of $H_{\{1,2\}}$ to have a rejection region which lies entirely in

$$\left\{(x_1, x_2) : \ \max(|x_1|, |x_2|) > z_{1-\frac{\alpha}{2}}\right\} .$$

Any test of $H_{\{1,2\}}$ satisfying this constraint will result in a consonant multiple testing procedure when applying the closure method.

For a concrete way to improve upon procedure (i), consider a rejection region $R'_{1,2}(\alpha)$ for the test of $H_{\{1,2\}}$ of the form

$$R'_{1,2}(\alpha) \equiv \left\{(x_1, x_2) : \ x_1^2 + x_2^2 > c'_2(1-\alpha) \ , \max\big(|x_1|, |x_2|\big) > z_{1-\frac{\alpha}{2}}\right\} , \qquad (6)$$

where the critical value $c'_2(1-\alpha)$ is chosen such that

$$P_{0,0}\left\{R'_{1,2}(\alpha)\right\} = \alpha .$$

Clearly, $c'_2(1-\alpha) < c_2(1-\alpha)$, and the resulting multiple testing procedure is consonant. For an illustration, see Figure 2. Moreover, for $1 \le i \le 2$,

$$P_{\theta_1,\theta_2}\left\{\text{reject } H_i \text{ using } R_{1,2}(\alpha)\right\} < P_{\theta_1,\theta_2}\left\{\text{reject } H_i \text{ using } R'_{1,2}(\alpha)\right\} .$$

In particular, the new consonant multiple testing procedure has uniformly greater power at detecting a false null hypothesis H_i. In this way, imposing consonance not only makes interpretation easier, but also improves our ability to detect false hypotheses as well. ■

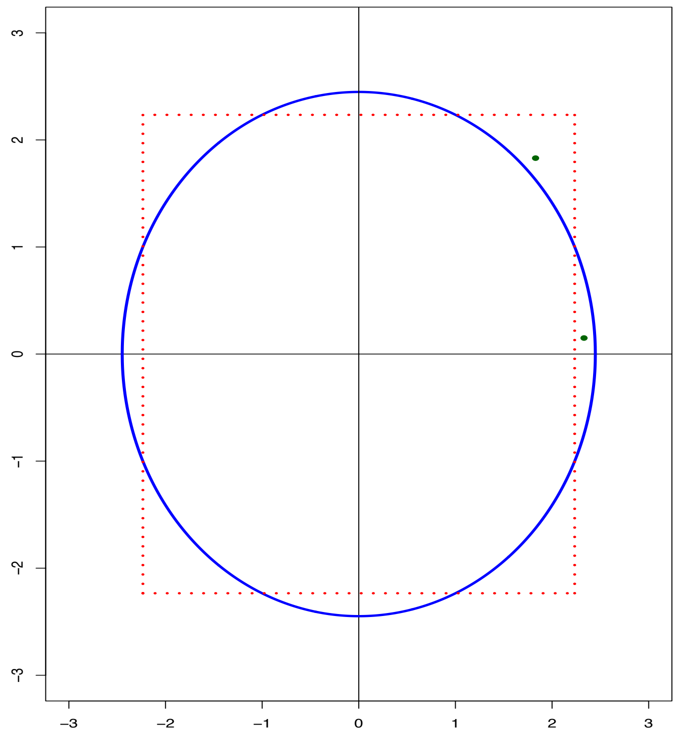

Figure 1: The rejection regions for the two intersection tests of Example 1.1 with nominal level $\alpha = 0.05$. Test (i) rejects for points that fall outside the solid circle with radius 2.448. Test (ii) rejects for points that fall outside the dashed square with length 2×2.234. For example, the point (1.83, 1.83) leads to rejection by test (i) but not by test (ii). On the other hand, the point (2.33, 0.15) leads to the rejection of H_1 by procedure (ii) but not by procedure (i).

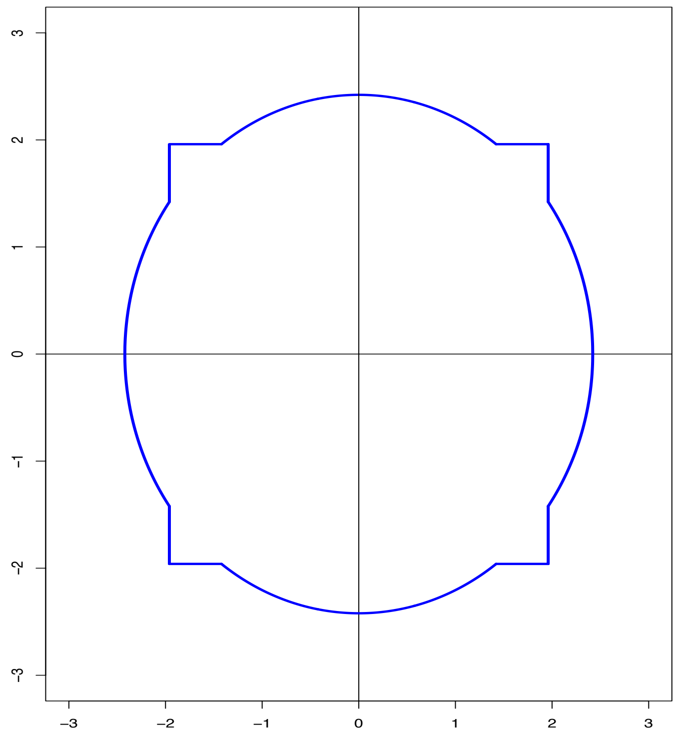

Figure 2: The rejection region $R'_{1,2}(\alpha)$ of the improved procedure (i) of Example 1.1 with nominal level $\alpha = 0.05$; see equation (6). This larger region is obtained as the intersection of the region outside of a circle with radius 2.421 and and the region outside a square with length 2×1.96.

In the previous example of testing independent normal means, note that if the original family of hypotheses had been H_1, H_2 and $H_{\{1,2\}}$, then we could not improve upon the original test in the same way. Such improvements are only possible when we use tests of intersection hypotheses as a device to apply the closure method without the intersection hypotheses themselves being of primary interest. For this reason, we consider the case where the family of hypotheses of interest $H_1, \ldots, H_s$ is the set of *elementary* hypotheses among the closure of the family of hypotheses of interest. Following Finner and Strassburger (2002), a hypothesis H_i is said to be *elementary* (or *maximal*) among a family of hypotheses if there exists no H_j in the family with $\omega_i \subset \omega_j$. So, in Example 1.1, H_1 and H_2 are the elementary hypotheses among the closure of the family of hypotheses of interest. In this setting, the following theorem shows that there is no need to consider dissonant multiple testing procedures when applying the closure method because any dissonant multiple testing procedure can be replaced by a consonant multiple testing procedure which reaches the same decisions about the hypotheses of interest. The main idea is that when applying the closure method, one should construct the tests of the intersection hypotheses in a consonant manner. In other words, the rejection region of the test of an intersection hypotheses should be chosen such that points in the rejection region lead to the rejection of at least one component hypothesis.

In order to keep the statement of the theorem compact, we introduce the notion of *consonantization* of a general, dissonant multiple testing procedure. Specifically, assume without loss of generality (by including an auxiliary random variable in X, if necessary) that the dissonant multiple testing procedure is obtained using the closure method based on tests of individual intersection hypotheses H_K with rejection region $X \in R_K$ for each $K \subseteq \{1, \ldots, s\}$. Define the corresponding *consonantized* procedure as the one based on the closure method with R_K replaced by R'_K given by

$$R'_K \equiv \bigcup_{i \in K} \bigcap_{J \subseteq S, i \in J} R_J \ . \tag{7}$$

Theorem 3.1 *Let $X \sim P \in \Omega$ and consider testing null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H'_i : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. Suppose a given multiple testing procedure is obtained using the closure method and controls the FWE at level α. Suppose further that each H_i is elementary. If the given multiple testing procedure is dissonant, then it can be replaced by a consonant multiple testing procedure that reaches the same decisions for the* *hypotheses of interest as the original procedure. Specifically, the corresponding consonantized procedure based on* (7) *is consonant, reaches the same decisions about the hypotheses of interest H_i as the original dissonant multiple testing procedure, and therefore controls the FWE.*

Remark 3.1 The preceding theorem only asserts that any dissonant multiple testing procedure can be replaced by a consonant multiple testing procedure that leads to the same decisions as the dissonant multiple testing procedure. In most cases, however, one can strictly improve upon a dissonant multiple testing procedure, as was done in Example 3.1, by removing points of dissonance from the rejection regions of the tests of the intersection hypotheses and adding to these rejection regions points that satisfy the constraint that the overall multiple testing procedure is consonant. ■

Remark 3.2 It is possible to generalize Theorem 3.1 to situations where the family of hypotheses of interest is a strict subset of their closure. For example, let $H_i : \theta_i = 0$ for $1 \leq i \leq 3$ and consider testing all null hypotheses in the closure of H_1, H_2 and H_3 except for $H_{\{1,2,3\}}$. Theorem 3.1 does not apply in this case since not all hypotheses are elementary. Even so, the idea of consonance can be applied as in the proof of the theorem when applying the closure method and choosing how to construct the rejection region for the test of $H_{\{1,2,3\}}$. As before, one should simply choose the rejection region for the test of $H_{\{1,2,3\}}$ such that points in the rejection region lead to the rejection of at least one of the other hypotheses. Any multiple testing procedure obtained using the closure method and based on a rejection region for $H_{\{1,2,3\}}$ that does not have this feature can be replaced by one that is at least as good in the sense that it rejects the same hypotheses and possibly more. ■

Remark 3.3 In the introduction, we mentioned several multiple testing procedures which only require the s p-values $\hat{p}_1, \ldots, \hat{p}_s$ as input. In general, these procedures are neither coherent nor consonant. We illustrate these two points for the Bonferroni procedure.

In order to illustrate that the Bonferroni procedure need not be coherent, consider setting of Example 1.1. Suppose that all three hypotheses are in fact of interest, that is, $s = 3$ and $H_3 \equiv H_{\{1,2\}}$. Let U_1, U_2, U_3 be three mutually independent uniform random random variables on [0,1]. For $1 \leq i \leq 3$, let $\phi_i \equiv 1_{\{U_i \leq \alpha/3\}}$, where $1_{\{\cdot\}}$ denotes the indicator function. The Bonferroni procedure rejects H_i iff $U_i \leq \alpha/3$. This multiple testing procedure controls the FWE at level α, but we may have that H_1 is rejected while H_3 is not rejected. (To make the construction a little more relevant, one might imagine dividing the data into three disjoint subsets so that the tests of H_1, H_2 and H_3 are mutually independent. Of course, this would be inefficient.) Hence, in this case, the Bonferroni procedure is not coherent.

In order to illustrate that the Bonferroni procedure need not be consonant, consider again the setting of Example 1.1. For $1 \leq i \leq 3$, let ϕ_i be defined as in procedure (i) of Example 1.1. The Bonferroni procedure rejects H_i iff H_i is rejected based on ϕ_i at level $\alpha/3$. Again, this multiple testing procedure controls the FWE at level α, but we may have that H_3 is rejected while neither H_1 nor H_2 is rejected. To see this, suppose $\alpha = 0.15$ so that $\alpha/3 = 0.05$. If $X_1 = X_2 = 1.83$, then $X_1^2 + X_2^2 = 6.698 = 2.588^2$, so $H_3 \equiv H_{\{1,2\}}$ is rejected at level 0.05, but neither X_i satisfies $|X_i| > 1.96$. Hence, in this case, the Bonferroni procedure is not consonant. ■

4 Optimality

4.1 Optimality Using Consonance

We now examine the role of consonance in optimal multiple testing procedures. We begin with the following general result. The main point of this result is that we can use an optimality property of a test of a single intersection hypothesis in order to derive an optimality property of a multiple testing procedure, as long as the assumption of consonance holds when the optimal test of the intersection hypothesis is used in an application of the closure method. Note that here we do not require the hypotheses of interest to be elementary.

Theorem 4.1 *Let $X \sim P \in \Omega$ and consider testing null hypotheses $H_i : P \in \omega_i$ versus alternative hypotheses $H_i' : P \notin \omega_i$ simultaneously for $i = 1, \ldots, s$. Let $S \equiv \{1, \ldots, s\}$. Suppose, for testing the intersection hypothesis H_S at level α, the test with rejection region R_S maximizes the minimum power over $\gamma_S \subseteq \omega_S^c$ among level α tests. Suppose further that when applying the closure method and using R_S to test H_S that the overall multiple testing procedure is consonant. Then, the multiple testing procedure maximizes*

$$\inf_{P \in \gamma_S} P\{\text{reject at least one } H_i\} \tag{8}$$

among all multiple testing procedures that control the FWE at level α.

Thus, the overall multiple testing procedure inherits a maximin property from a maximin property of the test of the intersection hypothesis H_S as long as the overall multiple testing procedure is consonant. We illustrate this result with an example.

Example 4.1 (Two-sided Normal Means, continued) For $\epsilon > 0$, define

$$\gamma_{1,2} \equiv \gamma_{1,2}(\epsilon) \equiv \big\{(\theta_1, \theta_2) : \text{ at least one } \theta_i \text{ satisfies } |\theta_i| \geq \epsilon\big\} \ . \tag{9}$$

For testing $H_{\{1,2\}}$ at level α, it is then straightforward to derive the maximin test against $\gamma_{1,2}(\epsilon)$. To see how, apply Theorem 8.1.1 of Lehmann and Romano (2005) with the least favorable distribution uniform over the four points $(\epsilon, 0)$, $(0, \epsilon)$, $(-\epsilon, 0)$ and $(0, -\epsilon)$. The resulting likelihood ratio test rejects for large values of

$$T \equiv T_\epsilon(X_1, X_2) \equiv \cosh\big(\epsilon|X_1|\big) + \cosh\big(\epsilon|X_2|\big) \ , \tag{10}$$

where the hyperbolic cosine function $\cosh(\cdot)$ is given by $\cosh(t) \equiv 0.5 \cdot \big(\exp(t) + \exp(-t)\big)$. The test has rejection region

$$R_{1,2}(\epsilon, \alpha) \equiv \big\{(x_1, x_2) : T_\epsilon(x_1, x_2) > c(1-\alpha, \epsilon)\big\} \ ,$$

where $c(1-\alpha, \epsilon)$ is the $1-\alpha$ quantile of the distribution of $T_\epsilon(X_1, X_2)$ under $(\theta_1, \theta_2) = (0, 0)$. It follows from Lemma A.2 in the appendix that the test with rejection region $R_{1,2}(\epsilon, \alpha)$ maximizes

$$\inf_{\theta \in \gamma_{1,2}(\epsilon)} P_{\theta_1, \theta_2}\{\text{reject } H_{\{1,2\}}\}$$

among level α tests of $H_{\{1,2\}}$.

Now construct a multiple testing procedure using the closure method as follows. Take the test with the rejection region $R_{1,2}(\epsilon, \alpha)$ as the test of $H_{\{1,2\}}$ and the usual UMPU test as the test of H_i, so that $R_i \equiv \big\{(x_1, x_2) : |x_i| > z_{1-\frac{\alpha}{2}}\big\}$. If we can show this multiple testing procedure is consonant, then it will maximize

$$\inf_{\theta \in \gamma_{1,2}(\epsilon)} P_{\theta_1, \theta_2}\{\text{reject at least one } H_i\} \tag{11}$$

among all multiple testing procedures controlling the FWE at level α. In fact, a consonant multiple testing procedure results for some values of ϵ. For large values of ϵ, the test statistic $T_\epsilon(X_1, X_2)$ is approximately equivalent to $\max(|X_1|, |X_2|)$, which does lead to a consonant multiple testing procedure. Indeed, see Figure 3 for an example with $\epsilon = 3$ and $\alpha = 0.05$. Thus, (11)

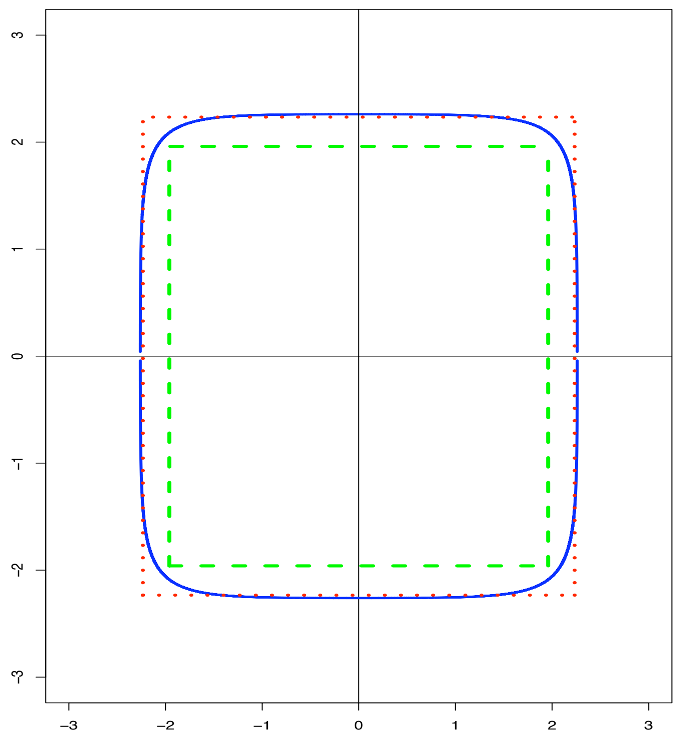

Figure 3: The test of Example 4.1 for $\epsilon = 3$ and nominal level $\alpha = 0.05$. The test rejects for points outside the solid curve. Points outside the inner square with length 2×1.96 lead to rejection of at least one H_i when the individual hypotheses are tested with the usual UMPU test. The outer square with length 2×2.234 is the rejection region of test (ii) of Example 1.1.

is maximized for this consonant multiple testing procedure when the tests of the hypotheses H_i are based on the rejection regions R_i. On the other hand, for small values of ϵ, rejecting for large values of the statistic $T_\epsilon(X_1, X_2)$ is approximately equivalent to rejecting for large values of $X_1^2 + X_2^2$, which we already showed in Example 3.1 does not lead to a consonant multiple testing procedure when the tests of the hypotheses H_i are based on the rejection regions R_i. So, we do not expect the theorem to apply for such values of ϵ. See Figure 4 for an example with $\epsilon = 0.25$ and $\alpha = 0.05$.

DOI: 10.2202/1557-4679.1300

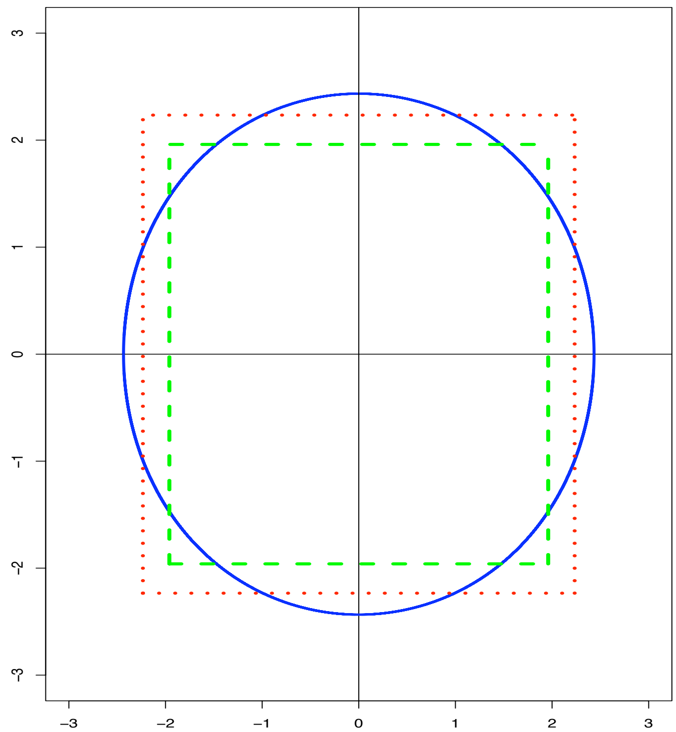

Figure 4: The test of Example 4.1 for $\epsilon = 0.25$ and nominal level $\alpha = 0.05$. The test rejects for points outside the solid curve. Points outside the inner square with length 2×1.96 lead to rejection of at least one H_i when the individual hypotheses are tested with the usual UMPU test. The outer square with length 2×2.234 is the rejection region of test (ii) of Example 1.1.

Next, consider

$$\bar{\gamma}_{1,2} \equiv \gamma_{1,2}(\epsilon) \cap \{(\theta_1, \theta_2) : \text{both } \theta_i \neq 0\} \ .$$

So, $\bar{\gamma}_{1,2}$ is the subset of the parameter space where both null hypotheses are false and at least one of the parameters is at least ϵ in absolute value. Then, since all power functions are continuous, (11) still holds if $\gamma_{1,2}(\epsilon)$ is replaced by $\bar{\gamma}_{1,2}(\epsilon)$. Moreover, because both H_1 and H_2 are false if $(\theta_1, \theta_2) \in \bar{\gamma}_{1,2}$, we can further claim that the multiple testing procedure maximizes

$$\inf_{\theta \in \bar{\gamma}_{1,2}(\epsilon)} P_{\theta_1,\theta_2}\{\text{reject at least one false } H_i\} \tag{12}$$

among all multiple testing procedures controlling the FWE at level α.

Finally, this example generalizes easily to higher dimensions where $X_1, \ldots, X_s$ are independent with $X_i \sim N(\theta_i, 1)$ and H_i specifies $\theta_i = 0$. For testing the intersection hypothesis H_K that specifies $\theta_i = 0$ for all $i \in K$, the maximin test statistic becomes

$$T \equiv T_\epsilon(X_1, \ldots, X_s) \equiv \sum_{i \in K} \cosh(\epsilon |X_i|) \ .$$

Again, for large enough ϵ, the closed testing method is consonant and an optimal maximin property can be claimed. For large ϵ, the test of H_K is approximately the test which rejects for large $\max(|X_i| : \ i \in K)$. ■

4.2 Weaker Optimality when Consonance Fails

When the multiple testing procedure obtained using the closure method and the maximin test of $H_{\{1,2\}}$ is not consonant, as above with $\epsilon = 0.25$, one can still derive an improved consonant multiple testing procedure, but we must then settle for a slightly more limited notion of optimality. In this case, not only must our multiple testing procedure satisfy the FWE level constraint, but we additionally restrict attention to multiple testing procedures based on the closure method where the individual tests of H_i have rejection region $R_i \equiv \big\{(x_1, x_2) : |x_i| > z_{1-\frac{\alpha}{2}}\big\}$. This constraint of forcing the individual rejection regions to be the UMPU tests does not appear unreasonable, though it is an additional assumption. Of course, because of Theorem 2.1, the restriction to multiple testing procedures obtained using the closure method is no restriction at all. (Indeed, the coherence condition is vacuous because the hypotheses are elementary.)

Therefore, rather than finding an overall maximin level α test of $H_{\{1,2\}}$, we must find the maximin level α test of $H_{\{1,2\}}$ subject to the additional constraint required by consonance that its rejection region $R_{1,2}$ satisfies

$$R_{1,2} \subseteq R_1 \cup R_2 \ .$$

Theorem A.1 in the appendix, a generalization of the usual approach, makes it possible. We illustrate its use with an example.

Example 4.2 (Two-sided Normal Means, continued) Let $\gamma_{1,2}(\epsilon)$ be defined as in (9) and consider the problem of constructing the maximin test for $H_{\{1,2\}}$ over the region $\gamma_{1,2}(\epsilon)$ subject to the constraint that the rejection region is contained in the region where

$$\max\big(|x_1|, |x_2|\big) > z_{1-\frac{\alpha}{2}} \ .$$

We can apply Theorem A.1 to determine such a test. As before, the least favorable distribution is uniform over the four points $(\epsilon, 0)$, $(0, \epsilon)$, $(-\epsilon, 0)$, and $(0, -\epsilon)$ and large values of the likelihood ratio are equivalent to large values of the statistic $T_\epsilon(X_1, X_2)$ given in (10). The optimal rejection region for the intersection hypothesis $H_{\{1,2\}}$ is then

$$R'_{1,2}(\epsilon, \alpha) \equiv \left\{T_\epsilon(x_1, x_2) > t(1-\alpha, \epsilon),\ \ \max\bigl(|x_1|, |x_2|\bigr) > z_{1-\frac{\alpha}{2}}\right\} ,$$

where the constant $t(1-\alpha, \epsilon)$ is determined such that $P_{0,0}\{R'_{1,2}(\epsilon, \alpha)\} = \alpha$. ■

A Appendix

Proof of Lemma 2.1: If H_K is a member of the original family, the result is trivial. Otherwise, suppose H_K is true, that is, all H_i with $i \in K$ are true. Then, by construction, the probability that H_K is rejected is the probability any H_i with $i \in K$ is rejected using the given multiple testing procedure. Since the given multiple testing procedure is assumed to control the FWE at level α, the last probability is no bigger than α. ■

Proof of Theorem 2.1: Define tests of an arbitrary intersection hypothesis H_K as in the statement of Lemma 2.1. Applying the closure method with these tests for the tests of the intersection hypotheses, in fact, results in the same decisions for the original hypotheses. To see this, first note that any hypothesis that is not rejected by the original multiple testing procedure certainly cannot be rejected by the one obtained using the closure method in this way. Moreover, any hypothesis that is rejected by the original multiple testing procedure is also rejected by the one obtained using the closure method in this way. Indeed, if H_i is rejected by the original multiple testing procedure, then H_K must be rejected when $i \in K$. This occurs by construction if H_K is not a member of the original family and by coherence of the original multiple testing procedure otherwise. ■

Proof of Theorem 2.2: The new multiple testing procedure is coherent in the sense that if H_j is rejected, then so is any H_i for which $\omega_i \subset \omega_j$. To see this, simply note that (5) implies that $\tilde{R}'_j \subseteq \tilde{R}'_i$ whenever $\omega_i \subset \omega_j$.

Since $\tilde{R}_i \subseteq \tilde{R}'_i$, the new multiple testing procedure clearly rejects H_i whenever the incoherent multiple testing procedure rejects H_i.

Finally, the new multiple testing procedure also controls the FWE at level α. To see why, suppose a false rejection is made by the new multiple testing procedure, that is, $x \in \tilde{R}'_i$ for some i with $P \in \omega_i$. Then, it must be the case that $x \in \tilde{R}_j$ for some j such that $\omega_j \supseteq \omega_i$. Since $P \in \omega_i$, it follows that $x \in \tilde{R}_j$ for some j such that $P \in \omega_j$. In other words, the incoherent multiple testing procedure also made a false rejection. Control of the FWE therefore follows from the assumption that the incoherent multiple testing procedure controls the FWE at level α. ■

PROOF OF THEOREM 3.1: We claim that

$$R'_K = \bigcup_{i \in K} \bigcap_{J \subseteq S, i \in J} R'_J. \tag{13}$$

To prove (13), we first show that

$$R'_K \supseteq \bigcup_{i \in K} \bigcap_{J \subseteq S, i \in J} R'_J. \tag{14}$$

To see this, note that by intersecting over just the set K instead of many sets J in the inner intersection operation in the definition (7), one obtains

$$R'_K \subseteq R_K \ . \tag{15}$$

Replacing R_J with R'_J in the definition (7) by (15) establishes (14).

Next, we show that

$$R'_K \subseteq \bigcup_{i \in K} \bigcap_{J \subseteq S, i \in J} R'_J. \tag{16}$$

Suppose $x \in R'_K$. Then, there must exist $i^* \in K$ such that

$$x \in \bigcap_{J \subseteq S, i^* \in J} R_J \ . \tag{17}$$

It suffices to show that $x \in R'_L$ for any $L \subseteq S$ such that $i^* \in L$. But, for any such L, by only taking the union in the definition of R'_L in (7) over just i^* and not all $i \in L$, we have that

$$R'_L \supseteq \bigcap_{J \subseteq S, i^* \in J} R_J \ . \tag{18}$$

But, (17) and (18) immediately imply $x \in R'_L$, as required.

DOI: 10.2202/1557-4679.1300

The relationship (13) shows the new multiple testing procedure obtained is consonant. Indeed, (13) states that R'_K consists exactly of those x for which the multiple testing procedure obtained using the closure method with the rejection regions R'_J leads to rejection of some H_i with $i \in K$. Specifically, if $x \in R'_K$, then, for some $i^* \in K$,

$$x \in \bigcap_{J \subseteq S, i^* \in J} R'_J \ ,$$

so that H_{i^*} is rejected by the new multiple testing procedure.

Finally, we argue that both multiple testing procedures lead to the same decisions. By (15), the new multiple testing procedure certainly cannot reject any more hypotheses than the original multiple testing procedure. So, it suffices to show that if a hypothesis, say H_{i^*}, is rejected by the original multiple testing procedure when x is observed, that it is also rejected by the new multiple testing procedure. But, in order for the original multiple testing procedure to reject H_{i^*} when x is observed, it must be the case that

$$x \in \bigcap_{J \subseteq S, i^* \in J} R_J \ ,$$

which coupled with (18) shows that $x \in R'_L$ for any $L \subseteq S$ such that $i^* \in L$. The new multiple testing procedure then rejects H_{i^*} as well. ■

Proof of Theorem 4.1: If there were another multiple testing procedure with a larger value of (8), then we can assume without loss of generality (by including an auxiliary random variable in X, if necessary) that it is based on a multiple testing procedure obtained using the closure method with rejections regions R'_K. But for any such procedure

$$P\{\text{reject at least one } H_i \text{ based on the rejection regions } R'_K\} \le P\{R'_S\} \ ,$$

since H_S must be rejected in order for there to be any rejections at all. Therefore,

$$\begin{aligned}
& \inf_{P \in \gamma_S} P\{\text{reject at least one } H_i \text{ based on the rejection regions } R'_K\} \\
\le \ & \inf_{P \in \gamma_S} P\{R'_S\} \\
\le \ & \inf_{P \in \gamma_S} P\{R_S\} \\
= \ & \inf_{P \in \gamma_S} P\{\text{reject at least one } H_i \text{ based on the rejection regions } R_K\} \ ,
\end{aligned}$$

where the final equality follows because consonance implies that the rejection of H_S is equivalent to the rejection of at least one H_i by the multiple testing procedure. ■

We next state and prove two lemmas. Lemma A.2 is used in Example 4.1. In turn, Lemma A.1 is used in the proof of Lemma A.2.

Lemma A.1 *Suppose $Y_1, \ldots, Y_s$ are mutually independent. Further suppose the family of densities on the real line $p_i(\cdot, \eta_i)$ of Y_i have monotone likelihood ratio in Y_i. Let $\psi \equiv \psi(Y_1, \ldots, Y_s)$ be a nondecreasing function of each of its arguments. Then, $E_{\eta_1,\ldots,\eta_s}\big[\psi(Y_1, \ldots, Y_s)\big]$ is nondecreasing in each η_i.*

Proof: The function $\psi(Y_1, Y_2, \ldots, Y_s)$ is nondecreasing in Y_1 with $Y_2, \ldots, Y_s$ fixed. Therefore, by Lemma 3.4.2 of Lehmann and Romano (2005),

$$E_{\eta_1}\big[\psi(Y_1, \ldots, Y_s)|Y_2, \ldots, Y_s\big]$$

is nondecreasing in η_1. So, if $\eta_1 < \eta_1'$, then

$$E_{\eta_1}\big[\psi(Y_1, \ldots, Y_s)|Y_2, \ldots, Y_s\big] \leq E_{\eta_1'}\big[\psi(Y_1, \ldots, Y_s)|Y_2, \ldots, Y_s\big] \ .$$

Taking expectations of both sides shows the desired result for η_1. To show the result when $\eta_i \leq \eta_i'$ for $i = 1, \ldots, s$, one can apply the above reasoning successively to each component. ■

Lemma A.2 *In the setup of Example 4.1, the test with rejection region $R_{1,2}(\epsilon, \alpha)$ maximizes*

$$\inf_{\theta \in \gamma_{1,2}(\epsilon)} P_{\theta_1,\theta_2}\{\text{reject } H_{\{1,2\}}\}$$

among level α tests of $H_{\{1,2\}}$.

Proof: As is well known, the family of distributions of X_i has monotone likelihood ratio in $|X_i|$, and distribution depending only on $|\theta_i|$. Since T is increasing in each of $|X_i|$, it follows by Lemma A.1, with $Y_i \equiv |X_i|$ and $\eta_i \equiv |\theta_i|$, that the power function of this test is an increasing function of $|\theta_i|$, and therefore the power function is minimized over $\gamma_{1,2}(\epsilon)$ at the four points $(\epsilon, 0)$, $(0, \epsilon)$, $(-\epsilon, 0)$ and $(0, -\epsilon)$. By Theorem 8.1.1 of Lehmann and Romano (2005) the uniform distribution over these four points is least favorable and the test is maximin. ■

We finally consider the problem of constructing a maximin test where the test must satisfy the level constraint as well as the added constraint that the

DOI: 10.2202/1557-4679.1300

rejection region must lie in some fixed set R. Denote by ω the null hypothesis parameter space and by ω' the alternative hypothesis parameter space over which it is desired to maximize the minimum power. So, the goal now is to determine the test that maximizes

$$\inf_{\theta \in \omega'} E_\theta\big[\phi(X)\big]$$

subject to

$$\sup_{\theta \in \omega} E_\theta\big[\phi(X)\big] \le \alpha$$

and to the constraint that the rejection region must lie entirely in a fixed subset R. Let $\{P_\theta : \theta \in \omega \cup \omega'\}$ be a family of probability distributions over a sample space $(\mathcal{X}, \mathcal{A})$ with densities $p_\theta = dP_\theta/d\mu$ with respect to a σ-finite measure μ, and suppose that the densities $p_\theta(x)$ considered as functions of the two variables (x, θ) are measurable $(\mathcal{A} \times \mathcal{B})$ and $(\mathcal{A} \times \mathcal{B}')$, where $\mathcal{B}$ and $\mathcal{B}'$ are given σ-fields over ω and ω'. We have the following result.

Theorem A.1 *Let Λ, Λ' be probability distributions over $\mathcal{B}$ and $\mathcal{B}'$, respectively. Define*

$$\begin{aligned} h(x) &= \int_\omega p_\theta(x)\, d\Lambda(\theta) \\ h'(x) &= \int_{\omega'} p_\theta(x)\, d\Lambda'(\theta)\ . \end{aligned}$$

Let C and γ be constants such that

$$\varphi_{\Lambda,\Lambda'}(x) = \begin{cases} 1 & \textit{if} \quad h'(x) > Ch(x),\ x \in R \\ \gamma & \textit{if} \quad h'(x) = Ch(x),\ x \in R \\ 0 & \textit{if} \quad h'(x) < Ch(x),\ \text{or } x \in R^c \end{cases}$$

is a size-α test for testing the null hypothesis that the density of X is $h(x)$ versus the alternative that it is $h'(x)$ and such that

$$\Lambda(\omega_0) = \Lambda'(\omega_0') = 1,$$

where

$$\begin{aligned} \omega_0 &\equiv \Big\{\theta : \theta \in \omega \textit{ and } E_\theta \varphi_{\Lambda,\Lambda'}(X) = \sup_{\theta' \in \omega} E_{\theta'} \varphi_{\Lambda,\Lambda'}(X)\Big\} \\ \omega_0' &\equiv \Big\{\theta : \theta \in \omega' \textit{ and } E_\theta \varphi_{\Lambda,\Lambda'}(X) = \inf_{\theta' \in \omega'} E_{\theta'} \varphi_{\Lambda,\Lambda'}(X)\Big\}\ . \end{aligned}$$

Then, $\varphi_{\Lambda,\Lambda'}$ maximizes $\inf_{\theta \in \omega'} E_\theta \varphi(X)$ among all level-α tests $\phi(\cdot)$ of the hypothesis $H : \theta \in \omega$ which also satisfy $\phi(x) = 0$ if $x \in R^c$, and it is the unique test with this property if it is the unique most powerful level-α test among tests that accept on R^c for testing h against h'.

PROOF: It follows from Lemma 1 in Bittman et al. (2009) that $\varphi_{\Lambda,\Lambda'}$ is the most powerful test for testing h against h', among level α tests ϕ that also satisfy $\phi(x) = 0$ if $x \in R^c$. Let $\beta_{\Lambda,\Lambda'}$ be its power against the alternative h'. The assumptions imply that

$$\sup_{\theta \in \omega} E_\theta \varphi_{\Lambda,\Lambda'}(X) = \int_\omega E_\theta \varphi_{\Lambda,\Lambda'}(X)\, d\Lambda(\theta) = \alpha \ ,$$

and

$$\inf_{\theta \in \omega'} E_\theta \varphi_{\Lambda,\Lambda'}(X) = \int_{\omega'} E_\theta \varphi_{\Lambda,\Lambda'}(X)\, d\Lambda'(\theta) = \beta_{\Lambda,\Lambda'}.$$

Thus, the conditions of Theorem 1 in Bittman et al. (2009) hold, and the result follows. ■

References

Bauer, P. (1991). Multiple testing in clinical trials. *Statistics in Medicine*, 10:871–890.

Bittman, R. M., Romano, J. P., Vallarino, C., and Wolf, M. (2009). Testing multiple hypotheses with common effect. *Biometrika*, 96(2):399–410.

Finner, H. and Strassburger, K. (2002). The partitioning principle: a power tool in multiple decision theory. *Annals of Statistics*, 30:1194–1213.

Gabriel, K. R. (1969). Simultaneous test procedures — some theory of multiple comparisons. *The Annals of Mathematical Statistics*, 40(1):224–250.

Hochberg, Y. and Tamhane, A. (1987). *Multiple Comparison Procedures*. Wiley, New York.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, New York, third edition.

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.

DOI: 10.2202/1557-4679.1300

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73:751–754.

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

Sonnemann, E. (2008). General solutions to multiple testing problems. *Biometrical Journal*, 50(5):641–656. (English translation of Sonnemann (1982) with minor corrections, prepared by H. Finner).

Sonnemann, E. and Finner, H. (1988). Vollständigkeitssätze für multiple Testprobleme. In Bauer, P., Hommel, G., and Sonnemann, E., editors, *Multiple Hypothesenprüfung*, pages 121–135. Springer, Berlin.