Power and Type-1 Error of Global versus Local Tests

Dror Rom, Chen Chen, and Jaclyn McTague
Prosoft Clinical

Abstract

The Closure Principle of Marcus, Peritz, and Gabriel (1976) is re-visited, and a new implementation of this principle is devised which is shown to provide improvement of power over other closed testing procedures in many common settings. Such settings arise in medical studies when the hypotheses being tested correspond to treatment effects that are expected to trend similarly. As an example, a drug that is expected to quicken the time to improvement of a disease is usually also expected to prolong the improvement time.

In this paper, we first discuss the implementation of the proposed procedure in the case of two hypotheses where the test statistics are independent. We prove that the procedure maintains strong control of the familywise error rate (FWER) under mild regulatory assumptions on the distribution of the test statistics. We then devise an extension of Hochberg's step-up procedure using the same principle. We compare the extended procedure to the original Hochberg procedure under various non-null configurations. We also show how the procedure can be extended to correlated test statistics while maintaining strong control of the FWER. Finally, we propose an extension for more than two hypotheses.

1. Introduction

Closed testing procedures are based on the Closure Principle of Marcus, Peritz, and Gabriel (1976) which is stated below:

Let X be a random variable with distribution P_θ $(\theta \in \Omega)$. Let $= \{\omega_\beta\}$ be a set up null hypotheses, i.e. a set of subsets of Ω, closed under intersection: $\omega_i, \omega_j \in W$ implies $\omega_i \cap \omega_j \in W$. For each ω_β let $\phi_\beta(X)$ be a level α test, that is, $pr_\theta\{\phi_\beta(X) = 1\} \leq \alpha$ for all $\theta \in \omega_\beta$. Now consider the following procedure.

Any null hypothesis ω_β is tested by means of $\phi_\beta(X)$ if and only if all hypotheses ω that are included in $\omega_\beta(\omega \subset \omega_\beta)$ and belonging to $W(\omega \in W)$ have been tested and rejected. The probability of making no type I error with this procedure is at least $1 - \alpha$.

Loosely speaking, this principle states that with a closed family of hypotheses, in order to strongly control the familywise error rate, we use the following procedure: to reject any hypothesis in the family, we must test and reject it by an α-level test, and we also must test and reject any other hypothesis in the family that implies it.

The closure principle has been re-stated by many others. One such re-statement was done by Hochberg and Tamhane in the textbook *Multiple Comparison Procedures (1987)*, using slightly different notations but otherwise identical to the original.

Another common re-statement of the closure principle is as follows:

Suppose there are k hypotheses $H_1, \ldots, H_k$ to be tested and the overall type I error rate is α. The closed testing principle allows the rejection of any one of these elementary hypotheses, say H_i, if all possible intersection hypotheses involving H_i can be rejected by using valid **local level α tests**. It controls the familywise error rate for all the k hypotheses at level α in the strong sense."

Note that the tests are referred to as **local tests**, which can lead to loss of power (see section 2.1).

In this paper, a G-closed procedure, an extension of the closed testing procedure, is proposed. The extension allows to devise a family of tests that provides an increase of power under the alternative configuration where several of the tested hypotheses are false. An example of this generalization based on Hochberg's procedure, G-Hochberg, will be shown.

2. An Extended Class of Closed Testing Procedures

2.1 A closed testing procedure for two hypotheses

Suppose we are testing two means, μ_1 and μ_2, against one-sided alternatives. The global null is the intersection of H_1 and H_2, and it states that both means are zero. Let $H_1\colon \mu_1 = 0, H_2\colon \mu_2 = 0$, and $H_0 = H_1 \cap H_2$.

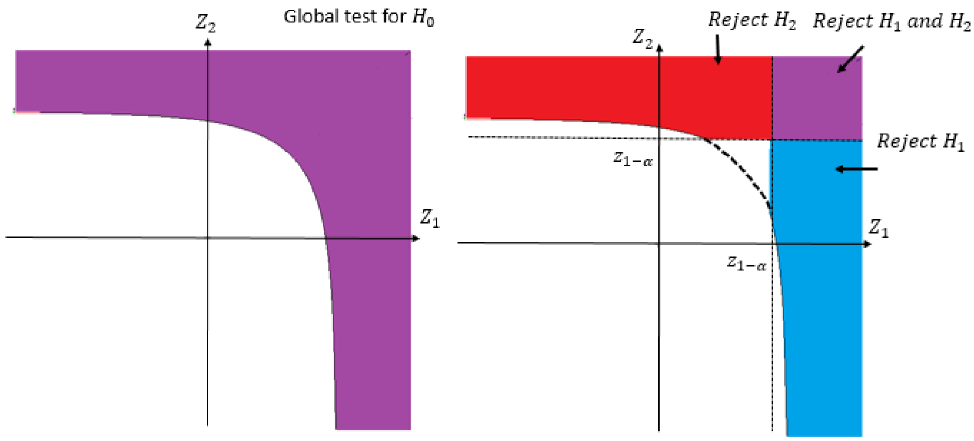
The closed family consists of the global null, H_0, and the two individual hypotheses, H_1 and H_2. Note that H_0 implies both H_1 and H_2, i.e., if H_0 is true, then it necessarily follows that both H_1 and H_2 are also true.

Suppose we have test statistics, Z_1 and Z_2, to test the individual hypotheses. The statistics are independent standard normal under their respective null hypotheses, have corresponding p-values P_1 and P_2, and ordered values $P_{(1)}$ and $P_{(2)}$. Suppose an α-level test $\varphi(Z_1, Z_2)$ is designed to test the global null hypothesis. A typical closed testing procedure for this setting is as follows:

Use the φ Test for the global null hypothesis, H_0; If H_0 is rejected, then test H_1 and H_2 each at level α .

A graphical display of the procedure is shown in Figure 1.

Figure 1: Typical Closed Testing Procedure



Notice that the area bounded by the dashed line is an area in which the global null H_0 is rejected, but the subsequent local tests do not reject either H_1 or H_2. This situation is termed "non-consonance," and it leads to some loss of power.

2.2 G-Closed testing procedure

Motivated by the example described above, we derive a different closed testing procedure for the same setting: 1. Let φ test for the global null hypothesis, H_0; 2. If H_0 is rejected, then reject $H_{(1)}$(the hypothesis corresponding to $P_{(1)}$); 3. reject $H_{(2)}$ (the other hypothesis) if $P_{(2)} \leq \alpha'$.

Figure 2: G-Closed Testing Procedure

1

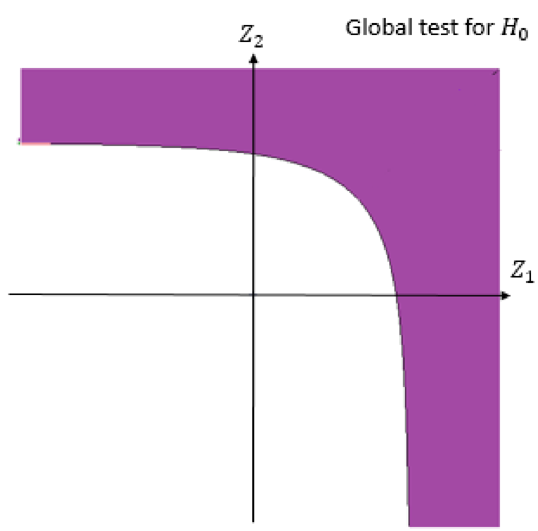
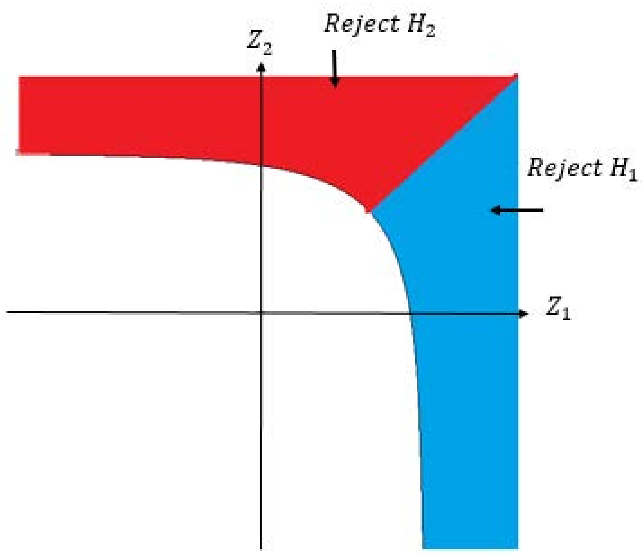
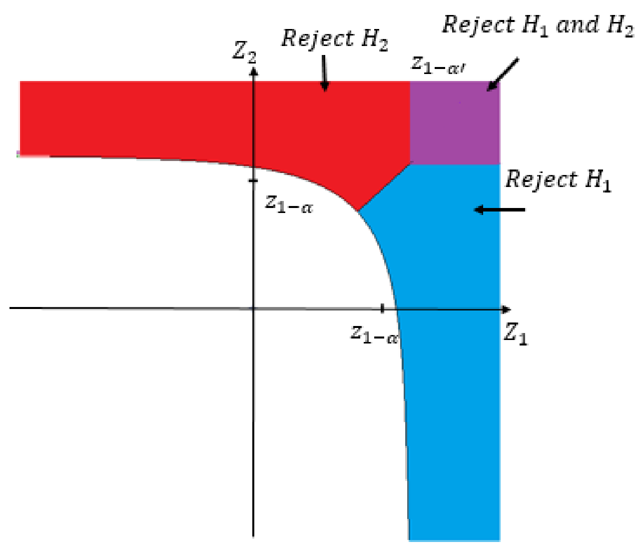


Figure 2 (cont.) : G-Closed Testing Procedure

2



3



Notice in this procedure, there is no loss of power since whenever H_0 is rejected, at least one of the individual hypotheses will be rejected. We will show that this procedure has strong control of the FWER for some choices of φ and α'. Such choices can be devised using an extension of Hochberg's step-up procedure.

3. An Extension of Hochberg's Step-up Procedure

3.1 G-Hochberg procedure

Figure 3 displays the Hochberg one-sided test for the setting described above. Hochberg's procedure rejects both hypotheses if both p-values $\leq \alpha$, or it rejects either hypothesis if its p-value $\leq \frac{\alpha}{2}$. When at least one hypothesis is rejected, the global null hypothesis, H_0, is rejected; thus, Hochberg's test is also a test of the global null.

Figure 3: Hochberg Procedure

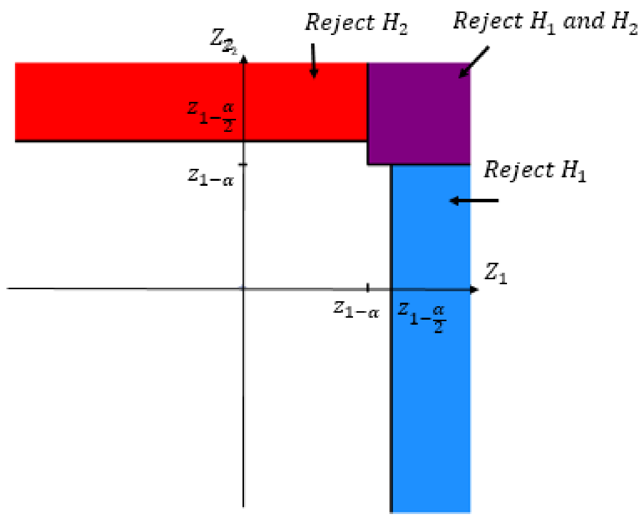
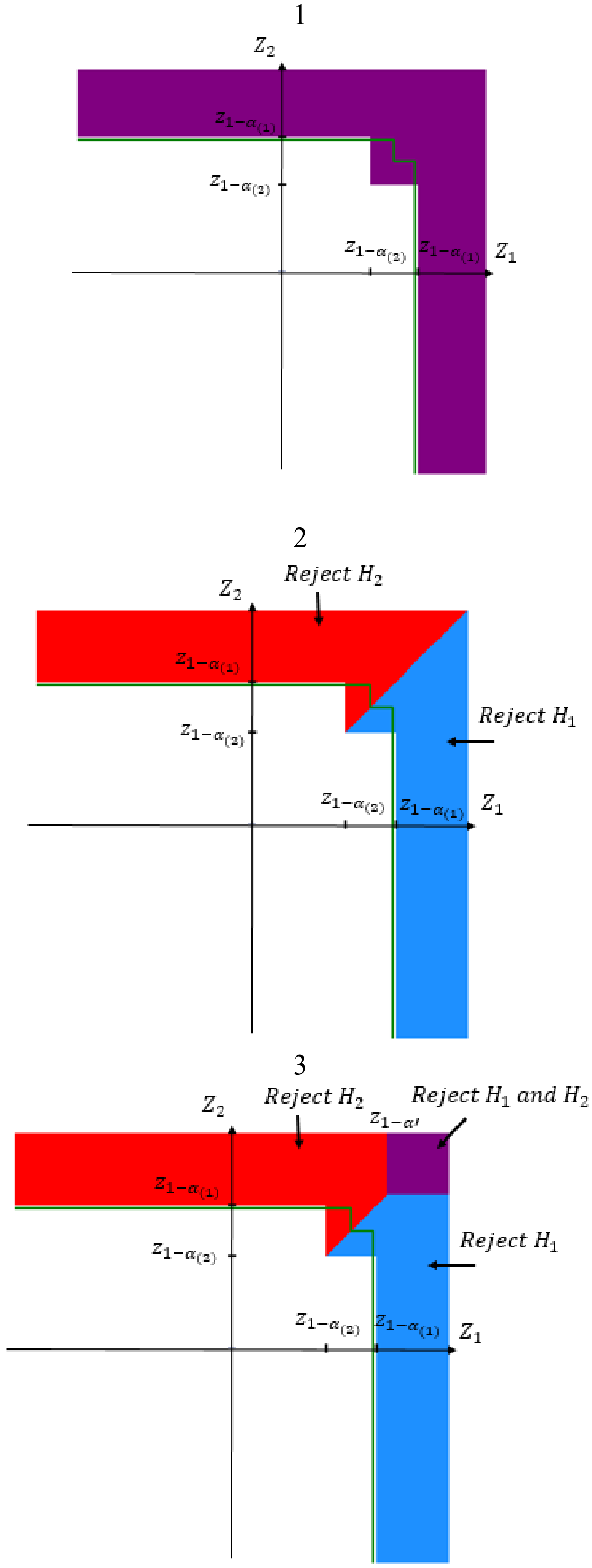


Figure 4 shows the G-Hochberg procedure. The green line represents the original Hochberg test for the global null, which has probability α. In the G-Hochberg procedure, the point where both p-values $\leq \alpha$ is re-positioned to $\alpha_{(2)}$, a point that is between α and 2α. The point corresponding to Hochberg's $\alpha/2$ is re-positioned to $\alpha_{(1)}$ so that the probability of the rejection region for the global null is hypothesis is maintained at α. Thus both the Hochberg and G-Hochberg procedures are α-level tests for the global null hypothesis.

The G-Hochberg procedure proceeds as follows and is graphically displayed in Figure 4:

1. Reject the global null hypothesis H_0 if $P_{(2)} \leq \alpha_{(2)} or\ P_{(1)} \leq \alpha_{(1)}$; 2. If H_0 is rejected, then reject $H_{(1)}$; and, 3. Reject $H_{(2)}$ if $P_{(2)} \leq \alpha'$.

Figure 4: G-Hochberg Testing Procedure



The steps to determine the critical points $\alpha_{(1)}, \alpha_{(2)}$ and $\alpha'$are as follows:

1) Set $\alpha_{(2)}$, where $\frac{\alpha}{2} \leq \alpha_{(2)} \leq \alpha$.

2) Calculate $\alpha_{(1)}$ so that the test to reject the global null is an α-level test, i.e., $\alpha_{(1)} = \frac{(\alpha_{(2)})^2 - \alpha}{2\alpha_{(2)} - 2}$.

3) Calculate α' so that the probability to reject H_1 and H_2 is $\leq \alpha$, i.e., $\alpha' = \sqrt{\alpha^2 - (\alpha_{(2)} - \alpha)^2}$.

Note that for choice of $\alpha_{(2)} = \alpha$, the G-Hochberg procedure is equivalent to the original Hochberg procedure. Additionally, when selecting $\alpha_{(2)} > \alpha$, for example $\alpha_{(2)} = 2\alpha$, the G-Hochberg procedure leads to a rejection of at least one hypothesis whose p-value is $> \alpha$ (but $\leq 2\alpha$).

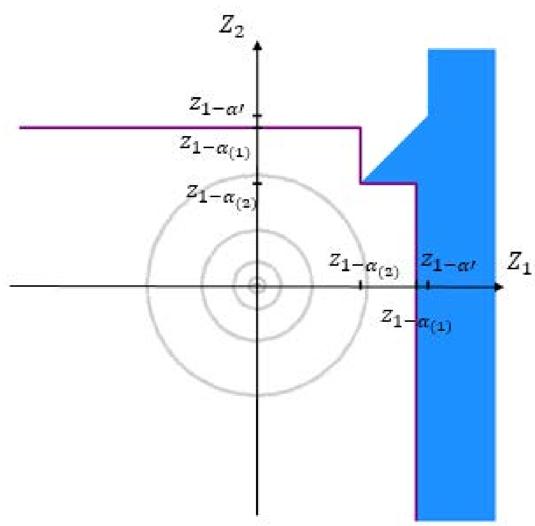
3.2 Strong Control of FWER

For strong control of the FWER, all hypotheses in the family must be protected at level α. Clearly, the global null is protected at level α because by design, the test for the global null is an α-level test.

To show that the individual hypotheses are protected at level α, we show that the probability to reject H_1 is no more than α, which by symmetry, also holds for H_2.

The blue area in Figure 5 represents the rejection region of H_1.

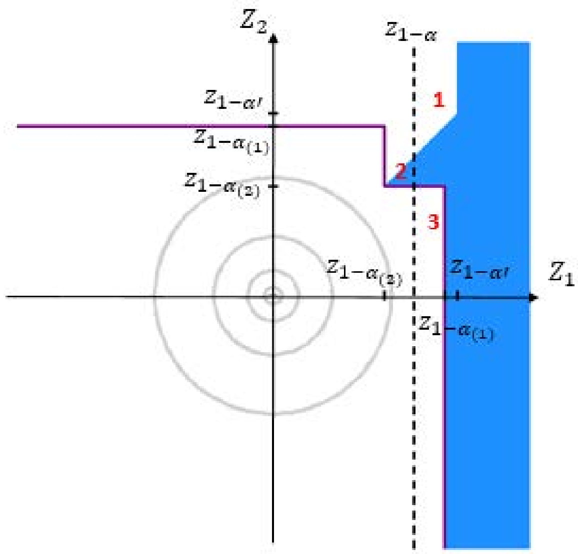
Figure 5: Rejection Region of H_1 Part 1



We assume that the two test statistics are independent and normally distributed and thus the joint density has a bell shape. This is represented by the concentric circles in Figure 5. Under the global null, the two means are zero. The probability of the blue area is clearly less than or equal to α since it is smaller than the probability of the rejection region of the global null, which is α by design.

When H_1 is true but H_2 is not, the bivariate normal moves away from zero along the vertical axis, Z_2. Figure 6 illustrates that in this scenario, the rejection region is also less than α. A vertical line is drawn to create 3 areas: area 1, area 2 and area 3. If $P(Area\ 1) + P(Area\ 3) \geq P(Area\ 2)$ for every value of μ_2, then $P(Blue\ Area) \leq \alpha$. The proof follows below.

Figure 6: Rejection Region of H_1 Part 2



As defined in section 3.1, $\alpha' = \sqrt{\alpha^2 - (\alpha_{(2)} - \alpha)^2}$, i.e., α' is calculated so that the probability to reject H_1 and H_2 is less than or equal to α. Specifically, α' is calculated so that $P(Area\ 1) = P(Area\ 2)$ under the global null hypothesis.

Now, the bivariate normal distribution with non-negative correlation is totally positive of order 2 (TP2). The TP2 property implies that $\frac{P(Area\ 1)}{P(Area\ 2)}$ is an increasing function of μ_2. It follows that for all $\mu_2 \geq 0$, $P(Area\ 1) \geq P(Area\ 2)$ and $P(Area\ 1) + P(Area\ 3) \geq P(Area\ 2)$ is true for every value of μ_2. H_1 is therefore protected at level α. By symmetry, H_2 is also protected at level α.

3.3 Power Comparisons

Table 1 shows the power comparisons between the standard Hochberg and two different G-Hochberg procedures, one when $\alpha_{(2)} = 2\alpha$ and the other when $\alpha_{(2)} = 1.5\alpha$.

Table 1: Power (%) Comparison: Hochberg vs. G-Hochberg under independence

		Probability to Reject at least One Hypothesis (%)		
			G-Hochberg	G-Hochberg
μ_1	μ_2	Hochberg	$\alpha_{(2)} = 2\alpha,\ \alpha' = 0$	$\alpha_{(2)} = 1.5\ \alpha,\ \alpha' = 0.021651$
0	0	2.50	2.50	2.50
0	1	11.92	11.99	11.96
0	2	41.36	41.27	41.37
0	3	77.98	77.76	77.93
1	0	11.92	11.99	11.96
1	1	20.69	22.03	21.29
1	2	47.55	49.65	48.56
1	3	80.47	81.56	81.05
2	0	41.36	41.27	41.37
2	1	47.55	49.65	48.56
2	2	65.82	69.43	67.62
2	3	87.51	89.54	88.58
3	0	77.98	77.76	77.93
3	1	80.47	81.56	81.05
3	2	87.51	89.54	88.58
3	3	95.55	96.73	96.20

Power is calculated as the probability to reject at least one hypothesis. Note that when $\mu_1$and μ_2 are zero, the probability to reject at least one hypothesis is equal to the type 1 error for both procedures. When the distance from the null increases in only one direction, i.e. only one mean increases, Hochberg's procedure has slightly higher power. When the distance increases in both directions, i.e. both means increase, the power of the G-Hochberg increases relative to Hochberg, with a maximum difference of almost 4%.

4. G-Hochberg for Two Correlated Hypotheses

The bivariate normal with a positive correlation is TP2. Thus, the G-Hochberg procedure has strong control of the FWER for positively correlated normal statistics as long as the test for the global null hypothesis is of level α.

With a known correlation and by selecting $_{(1)}$, $\alpha_{(2)}$ and α', the G-Hochberg procedure can be constructed. In the following examples, we selected $\alpha_{(1)}$ to be the same as in the independence case, where $\alpha_{(2)} = 2\alpha$. For each correlation, $\alpha_{(2)}$ is calculated so that the type 1 error for testing the global null hypothesis is α.

Table 2: G-Hochberg for One-sided Positively Correlated Cases

	Critical values		
ρ	$\alpha_{(1)}$	$\alpha_{(2)}$	α'
0.1	0.011842	0.0458	0.0116
0.3	0.011842	0.0404	0.01569
0.5	0.011842	0.0383	0.01633
0.7	0.011842	0.0375	0.01617
0.9	0.011842	0.0349	0.01751

Note that as the correlation increases, $\alpha_{(2)}$ must approach α; Also, for this choice of $\alpha_{(1)}$, $\alpha_{(2)} > \alpha$, which means that one can reject a hypothesis even when its p-value > α.

Table 3 shows the power comparisons between Hochberg and G-Hochberg for the correlated cases demonstrated above.

Table 3: Power (%) Comparison: Hochberg vs. G-Hochberg when correlated

			Probability to Reject at least one hypothesis (%)	
ρ	μ_1	μ_2	Hochberg	G-Hochberg
0.1	0	0	2.49	2.50
	0	1	11.85	11.91
	0	2	41.18	40.99
	0	3	77.83	77.52
	1	1	20.33	21.47
	1	2	46.78	48.40
	1	3	79.86	80.57
	2	2	64.27	67.22
	2	3	86.33	88.01
	3	3	94.66	95.77
0.3	0	0	2.46	2.50
	0	1	11.69	11.71
	0	2	40.91	40.50
	0	3	77.67	77.18
	1	1	19.53	20.41
	1	2	45.29	46.24
	1	3	78.88	79.03
	2	2	61.23	63.33
	2	3	84.07	85.20
	3	3	92.75	93.72
0.5	0	0	2.40	2.50
	0	1	11.46	11.49
	0	2	40.66	40.11
	0	3	77.59	77.02
	1	1	18.50	19.46
	1	2	43.74	44.44
	1	3	78.13	77.95
	2	2	58.02	59.99
	2	3	81.86	82.75
	3	3	90.58	91.59

Table 3 (cont.): Power (%) Comparison: Hochberg vs. G-Hochberg when correlated

			Probability to Reject at least one hypothesis (%)	
ρ	μ_1	μ_2	Hochberg	G-Hochberg
0.7	0	0	2.28	2.50
	0	1	11.16	11.19
	0	2	40.48	39.79
	0	3	77.56	76.96
	1	1	17.19	18.49
	1	2	42.17	42.67
	1	3	77.68	77.22
	2	2	54.47	56.80
	2	3	79.74	80.41
	3	3	88.02	89.30
0.9	0	0	2.09	2.50
	0	1	10.79	10.59
	0	2	40.42	39.65
	0	3	77.56	76.96
	1	1	15.50	17.41
	1	2	40.71	40.40
	1	3	77.56	76.96
	2	2	50.24	53.37
	2	3	77.89	77.76
	3	3	84.80	86.60

As for independence, Hochberg's procedure has a slightly higher power when the deviation from the global null is in the direction of only one of the means. When the deviation is in the direction of both means, the G-Hochberg has higher power.

5. Conclusion

The G-Closed testing procedure proposed in this paper can be devised using global tests rather than local tests. It can make simultaneous inferences on individual hypotheses without going through the full closed testing procedure.

Inspired by the Hochberg procedure, an extended Hochberg procedure is proposed. In the independence case, the extended Hochberg procedure is shown to gain as much as 4% in power compared to Hochberg's procedure when both means are non-null. Examples of such configurations occur when testing an effect of a drug in several sub-populations, for example, Males/Females, Young/Old; Or when testing an effect of a drug at two dose levels. Extensions to more than two endpoints, as well as generalizations using other global tests are currently being pursued.

References

HOCHBERG, Y. (1998). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75 (4), 800–802.

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

MARCUS, R., PERITS, E. & GABRIEL, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63 (3), 655-660.

SARKAR, S.K. (1998). Some probability inequalities for ordered MTP random variables: a proof of the Simes conjecture. *The Annals of Statistics*, 26 no. 2, 494--504.

SIMES, R. J. (1986). Improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73 (3), 751-754.