Multiple testing approaches for evaluating the effectiveness of a drug combination in a multiple-dose factorial design.

Saswati Saha, Werner Brannath, *Competence Centre for Clinical Trials, University of Bremen.*
Bjoern Bornkamp, *Novartis Pharma AG, Basel, Switzerland.*
Correspondence email: saha@uni-bremen.de.

Abstract

Drug combination trials are often motivated from the fact that individual drugs target the same disease but via different routes. Hence by combining drugs one can achieve an overall better effect than conducting individual treatment. Several approaches have been explored for developing statistical methods that compare (single) fixed dose combination therapies to its component. But extension of these approaches to multiple dose combinations can be difficult and is not fully explored. We propose two approaches by which one can provide confirmatory assurance with familywise error rate control, that combination of two drugs is more effective than either component drug alone. These approaches involved multiple comparisons in multilevel factorial design where the type 1 error rate is controlled firstly, by bootstrap test, and secondly, by considering the least favorable null configurations under an union intersection test. In this poster we would like to demonstrate the implementation of these new approaches and compare their performance with a real data example from a blood pressure reduction trial and via extensive simulations.

Introduction

In the poster we focus on a multiple dose drug combination trial, where the objective is to show superiority of the combined drugs over the individual drugs while preserving the family wise error rate (FWER), and thereafter identifying the set of superior dose combinations that can be used in the drug development process. The multiple hypotheses testing for the above problem can be formulated as below:

$$\begin{aligned} H_0 &: \forall\, i,j;\ \mu_{ij} \leq \mu_{i0} \text{ or } \mu_{ij} \leq \mu_{0j} \\ H_1 &: \exists\, i,j;\ \mu_{ij} > \mu_{i0} \text{ and } \mu_{ij} > \mu_{0j} \end{aligned} \tag{1}$$

Testing the superiority in a single dose drug combination trial is not difficult but proposing a conservative test that controls the family wise error rate in a drug combination trial with multiple doses is not so trivial. We propose two single-step testing procedures using a MAX test statistics [1] which test if there exist atleast one combination in a multiple dose factorial drug combination study that performs better than the component drugs. Further the multiple testing procedures satisfy the subset pivotality condition [3] which ensures that FWER is controlled under both the approach.

Parametric set-up

Consider a $(r+1)\times(s+1)$ factorial design, where Drug A have dose levels $0,1,\ldots,r$ and Drug B have dose levels $0,1,\ldots,s$. We consider the following anova set up:

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{2}$$

where $i=0,\ldots,r$, $j=0,\ldots,s$, $k=1,\ldots,n_{ij}$. $Y=(((Y_{ijk})))$: observations for the drug combinations trial, μ_{i0} : mean effect of dose i of Drug A (monotherapy 1), μ_{0j} : mean effect of dose j of Drug B (monotherapy 2), μ_{ij} : mean effect of the combination: dose i of Drug A and dose j of Drug B and n_{ij}: number of subjects allocated in ij^{th} combination.

We propose the following max test statistics for testing H_0 against H_1:

$$T = \max_{i,j} T_{ij} = \max_{i,j}\{\min\{T^1_{ij}, T^2_{ij}\}\} \tag{3}$$

where T_{ij} denote the test statistic for testing $H_{0ij}: \mu_{ij} \leq \mu_{i0}$ or $\mu_{ij} \leq \mu_{0j}$ against $H_{1ij}: \mu_{ij} > \mu_{i0}$ and $\mu_{ij} > \mu_{0j}$. Here T^1_{ij} and T^2_{ij} are the contrast test statistics used for testing whether the $(i,j)^{th}$ drug combination is superior to the monotherapy 1 and 2 respectively. The raw p values (p_{ij}) corresponding to each H_{0ij} can be easily obtained using Min tests but for testing multiple combinations, these raw p-values need to be adjusted. Multiplicity adjustment is challenging here because computing the distribution of T is analytically quite difficult, specially for large r and s.

Methods

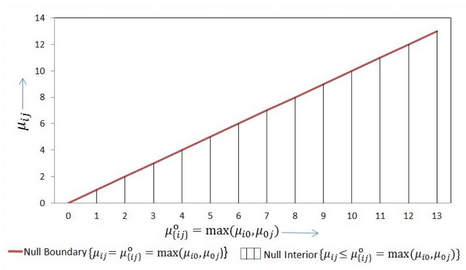
1. Bootstrap Approach

1. For the given data, compute the test statistics T (say t) and estimate the mean(μ) under

2a. NI constraint $\mu_{ij} \leq \max(\mu_{i0}, \mu_{0j})\ \forall\, i,j$

2b. NB constraint $\mu_{ij} = \max(\mu_{i0}, \mu_{0j})\ \forall\, i,j$

3. Simulate 5000 normal random variables with the estimated means. For each simulated data, compute the test statistics T (say T^k) and T_{ij} (say T^k_{ij}).



4. Compute $\sum_k I(T^k > t)/5000$ and $\sum_k I(T^k_{ij} > t)/5000$ (adjusted p-values), where I is the indicator function. Check if the adjusted p-values are within the significance level.

2. Least Favourable Null Configuration Approach (LFC)

The least favourable configurations (LFC) identifies the "worst case scenarios" that leads to the largest type 1 error rate.

1. Computing the distribution of T under the null parameter space is complicated so we focus on the LFC. For evaluating H_{0ij}, the LFC occur when $\mu \in LFC_{ij}$ where:

$$LFC_{ij} = \{\mu |\, (\mu_{ij} = \mu_{i0}, \mu_{ij} >> \mu_{0j}) \text{ or } (\mu_{ij} = \mu_{0j}, \mu_{ij} >> \mu_{i0})\}.$$

   The LFC for the global null H_0 occurs when $\mu \in LFC$ where $LFC = \bigcap_{i,j} LFC_{ij}$.

2. Consider the critical value C_α such that it satisfies: $\max_{\tau \in LFC}(1 - Pr_\tau(T \leq C_\alpha)) = \alpha$. We reject the global null in H_0 when the observed $T > C_\alpha$. Furthermore, all the component test statistics T_{ij} are tested against the critical value C_α and decisions are taken following a single-step testing procedure.

Implementation

A Case Study

A combination of a diuretic (drug B) and an ACE inhibitor (drug A) is tested for efficacy in decrease of sitting diastolic pressure (SiDBP) with a pooled standard deviation of $\sigma = 7.07$ ([2]). The response means and sample size allocation is summarized by the following table:

Table 1: Expected response structure of the drug combination study

	Drug A			
Drug B	0	1	2	3
0	0 (75)	1.4 (75)	2.7 (74)	4.6 (48)
1	1.8 (74)	2.8 (75)	5.7 (74)	8.2 (49)
2	2.8 (48)	4.5 (50)	7.2 (48)	10.9 (48)

Table 2: The unadjusted and adjusted p values under each component hypothesis for the different methods:

Dose Comb	TStat	UnadjP	BonfP	BootP	LFCP
(1,1)	0.863	0.194	1.000	0.481	0.709
(1,2)	1.190	0.117	0.703	0.284	0.512
(3,1)	2.507	0.006	0.037	0.006	0.036
(2,1)	2.581	0.005	0.030	0.004	0.029
(2,2)	3.049	0.001	0.007	0.001	0.007
(3,2)	4.365	0.000	0.000	0.000	0.000

Dose Comb: Different dose combinations
TStat : Test statistics T_{ij} testing for H_{0ij} .
UnadjP is the one-sided raw p-value testing H_{0ij} against H_{1ij}. BonfP, BootP and LFCP are the one sided Bonferroni adjusted, Bootstrap adjusted and LFC adjusted p-values respectively.

Simulations

- 3 dose groups (2 active and 1 placebo) for Drug A combined with 2 dose groups (1 active and 1 placebo) for Drug B.
- Balanced factorial design with $n_{ij} = n\ \forall\, i,j$ where $n = 10, 25, 50, 75, 100$
- 5000 simulations considered and standard normal errors assumed

Simulation Scenarios

Least Favourable Config	Drug B	Drug A		
		0	1	2
$H_0(LF_1)$	0	δ	δ	δ
	1	$\delta+a$	$\delta+a$	$\delta+a$
$H_0(LF_2)$	0	δ	δ	$\delta+2a$
	1	$\delta+a$	$\delta+a$	$\delta+2a$
$H_0(LF_3)$	0	δ	$\delta+2a$	δ
	1	$\delta+a$	$\delta+2a$	$\delta+a$
$H_0(LF_4)$	0	δ	$\delta+a$	$\delta+a$
	1	δ	$\delta+a$	$\delta+a$

Table 3: Scenario 1(Evaluating Type 1 error with $\delta = 2$, $a = 9999$)

	Drug A		
Drug B	0	1	2
0	2	2	2
1	2	2.5	2.5

Table 4: Scenario 2 (Evaluating power)

Results

Table 5: Table illustrating the empirical type 1 error rate for the different methods under Scenario 1:

Sample Size	Bonf	Hung	Boot	LFNull
10	0.045	0.057	0.047	0.045
25	0.049	0.055	0.053	0.051
50	0.046	0.049	0.048	0.047
75	0.042	0.045	0.045	0.044
100	0.048	0.050	0.048	0.048

Table 6: Table showing the empirical power of 5% level max test for the different method under Scenario 2:

Sample Size	Bonf	Hung	Boot	LFNull
10	0.149	0.172	0.220	0.153
25	0.433	0.443	0.508	0.438
50	0.789	0.794	0.835	0.790
75	0.929	0.931	0.945	0.930
100	0.980	0.979	0.983	0.981

- The type 1 error is controlled for all the methods under Scenario 1.
- The bonferroni method and the LF null approach is more conservative compared to the other approach.
- Bootstrap approach is showing better power performance across all scenarios.

Conclusions

- We observe from our simulations that both the bootstrap and LFC approach controls the FWER at nominal level α. While the LFC controls the FWER always below α bootstrap approach controls the FWER only asymptotically. But the bootstrap approach is more preferred because it gives better power as compared to the other approaches.
- The above methods provide a set of superior dose combinations but one cannot infer anything beyond the observed doses if the nature of dose response relationship is not known. Hence, it might be interesting to extend our bootstrap based multiple testing approach to a modelling framework and make inference on the dose-response relationship in drug combination studies.

References

[1] H. J. Hung. Evaluation of a combination drug with multiple doses in unbalanced factorial design clinical trials. *Statistics in medicine*, 19(16):2079–2087, 2000.

[2] H. J. Hung, G. Chi, and R. Lipicky. Testing for the existence of a desirable dose combination. *Biometrics*, pages 85–94, 1993.

[3] P. H. Westfall. Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, 92(437):299–306, 1997.

Acknowledgements

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567.