A fallback test for three co-primary endpoints

Robin Ristl

Section for Medical Statistics, Center of Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna

joint work with
Florian Frommlet, Armin Koch and Martin Posch

Workshop Adaptive Designs and Multiple Testing Procedures
Basel, 5-6 June 2014

This work has been funded by the FP7-HEALTH-2013-INNOVATION-1 project Advances in Small Trials Design for Regulatory Innovation and Excellence (ASTERIX) Grant Agreement No. 603160.

Examples for co-primary endpoints

In complex diseases, more than one primary endpoint may be required for characterization of treatment effects.

- Alzheimer's disease:
  Cognitive functions and functions of daily living as co-primary, global assessment as secondary (EMA guideline)
- Duchenne and Becker muscular dystrophy:
  Motor functioning and muscle strength (EMA guideline)
- Lennox-Gastaut epilepsy syndrome (rare disease):
  Total seizure frequency, tonic/atonic seizure frequency and global improvement in seizure severity were used as three co-primary endpoints (Glauser et al., 2008)

Concepts in hypothesis testing

- Elementary null hypothesis H_i
  - H_i is true means "No effect in endpoint i"
- Level α Test for a single H_i
  - Reject H_i if the test statistic $T_i > c$
  - Choose c so that $P(T_i > c|H_i) = \alpha$
  - E.g. $T_i|H_i \sim N(0,1)$, one sided level $\alpha = 0.025$, $c = 1.96$
- Intersection null hypothesis $H_i \cap H_j$
  - H_i and H_j are true
  - Alternative: Effect in endpoint i or endpoint j or both
  - Multiple testing problem:
  - E.g. Uncorrelated endpoints i and j, $\alpha = 0.025$
  - Probability to individually reject H_i or H_j is $1 - (1 - 0.025)^2 = 0.049 \approx 2\alpha$

Regulatory position and reverse multiplicity problem

- Regulators: All co-primary endpoints must be significant at local level α (one-sided $\alpha = 0.025$)
- This means: Reject all H_i if all $T_i > c$. Othwerwise do not reject any null hypothesis.
- This may require increased sample sizes compared to single-endpoint-problems
- E.g.: Three uncorrelated co-primary endpoints with similar effects
- Power for each single-endpoint test is 80%
- Power to reject all three endpoints is $0.8^3 = 0.512$
- What to do in rare disease situation?

What to do if 2 of 3 co-primary endpoints are significant?

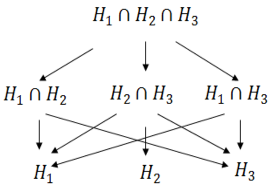
- Consider a trial with three co-primary endpoints.
- Can we perform some inference with level α control, in case that only two of three endpoints were signficant?
- All information is valuable, especially in rare diesease settings.
- Is there a "fallback" strategy to draw confirmative conclusions from a trial, that would otherwise be considered failed?

"A fallback test for three co-primary endpoints", R. Ristl, F. Frommlet, A. Koch, M. Posch, submitted

Fallback test for three co-primary endpoints

- Reject all H_i if all $T_i > c, i \in \{1, 2, 3\}$
- Reject $H_i \cap H_j$ if $T_i > c$ and $T_j > c$, for some $i \neq j$

Closed test scheme



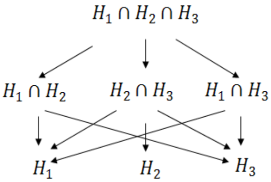
Intersection hypotheses	Reject if	Upper bound for local Level
$H_1 \cap H_2 \cap H_3$	At least 2 of 3 $T_i > c$	α, trivariate normal test statistics $3\alpha/2$, arbitrary distribution (Rüger)
$H_1 \cap H_2$, $H_2 \cap H_3$, $H_1 \cap H_3$	$T_i > c$ and $T_j > c$	α
H_1, H_2, H_3	$T_i > c$	α

Under normality assumption the family wise type I error rate (FWER) is bounded by α.

A liberal Fallback test for three co-primary endpoints

- Reject a pair H_i and H_j if both $T_i > c$ and $T_j > c$

Closed test scheme



Intersection hypotheses	Reject if	Upper bound for local Level
$H_1 \cap H_2 \cap H_3$	At least 2 of 3 $T_i > c$	α, trivariate normal test statistics 3α/2, arbitrary distribution (Rüger)
$H_1 \cap H_2$, $H_2 \cap H_3$, $H_1 \cap H_3$	$T_i > c$ or $T_j > c$	2α
H_1, H_2, H_3	$T_i > c$	α

Under global intersection null hypothesis $H = H_1 \cap H_2 \cap H_3$ the FWER is bounded by α.
Else, the FWER is bounded by 2α.

Application to diagnostic trials

- Study design: Three readers, individually rating each of n patients as healthy or diseased
- Aim: Show that a prespecified sensitivity AND specificity can be reached.

Hypotheses and test statistics for reader i:

Hypothesis	Test statistic	Reject if
$H_{se,i} : sensitivity_i = sens_0$	$Z_{se,i}$	
$H_{sp,i} : specificity_i = spec_0$	$Z_{sp,i}$	
$H_i = H_{se,i} \cup H_{sp,i}$	$T_i = min(Z_{se,i}, Z_{sp,i})$	$T_i > z_{1-\alpha}$

- The fallback test can be applied to $T = (T_1, T_2, T_3)$
- FWER controlled, because there is an asymptotically multivariate normal vector $(Z_{s_1,1}, Z_{s_2,2}, Z_{s_3,3})$, $s_i \in \{se, sp\}$ so that $(T_1, T_2, T_3) \leq Z$

Power (%) to reject $H_1 \cap H_2 \cap H_3$ for standardized effects δ

δ_1	δ_2	δ_3	Correlation	Fallback	Bonferroni-Holm
3	0	0	0	4.2	73.2
			0.5	4.5	72.8
			0.85	3.7	72.8
3	3	0	0	73.0	92.6
			0.5	75.9	86.4
			0.85	80.0	80.1
3	3	3	0	94.0	98.0
			0.5	88.9	91.5
			0.85	86.1	83.3

Compare: Power for one primary endpoint (or hierarchical test) is 85.1 %.

Power (%) to reject all three H_i for standardized effects δ

δ_1	δ_2	δ_3	Correlation	Fallback	Bonferroni-Holm
3	0	0	0	0.1	0.1
			0.5	0.5	0.4
			0.85	1.3	1.1
3	3	0	0	1.8	1.6
			0.5	2.5	2.5
			0.85	2.5	2.5
3	3	3	0	61.6	59.8
			0.5	69.4	67.7
			0.85	77.0	73.7

Larger power for Fallback test.

Theorem

Assumptions:

- Trivariate normal random vector $Z \sim N_3(0, \Sigma)$
- $var(Z_i) = 1, i = 1, 2, 3$
- $\alpha \leq 1/2$
- $c = \Phi^{-1}(1 - \alpha)$

Theorem

Under the assupmtions, The probability π that at least two of the three random variables take values larger than c does not exceed α.

Remark

For $c \geq 0 : \pi \leq \alpha \Leftrightarrow$
$P(Z_1 > c, Z_2 < c, Z_3 < c) \geq P(Z_1 > c, Z_2 > c, Z_3 < c)$

Outline of proof

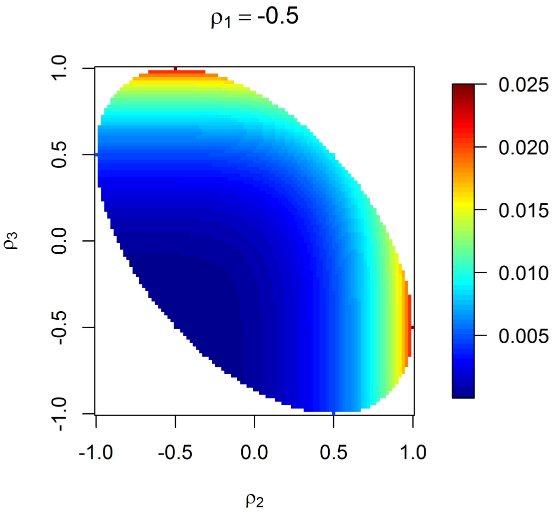
Special cases:

- Z uncorrelated, $\alpha \in [0; 0.5]$: $\pi = 3\alpha^2 - 2\alpha^3 < \alpha$
- Perfect correlation of any pair $(Z_i, Z_j), i \neq j$: $\pi = \alpha$

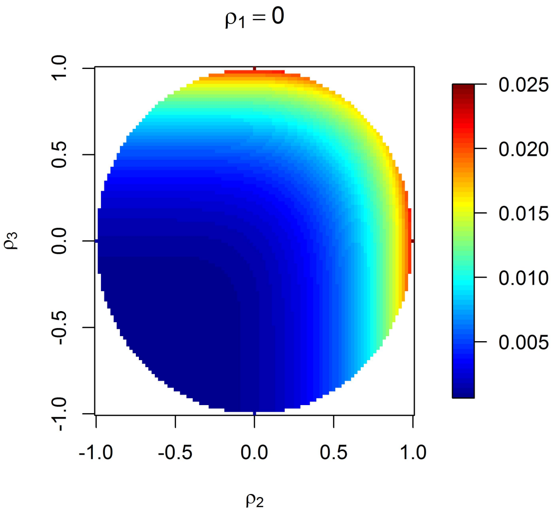
Arbitrary correlation structure:

- Study the gradient of π with respect to $\rho_{ij} = cor(Z_i, Z_j), i \neq j$
- Show that there is no local extreme value of π in the parameter space of $\{\rho_{ij}\}$, such that $det(\Sigma) > 0$
- At the boundary ($det(\Sigma) = 0$) the problem can be transformed to two dimensions.
- Geometric arguments show $\pi \leq \alpha$ on the boundary.

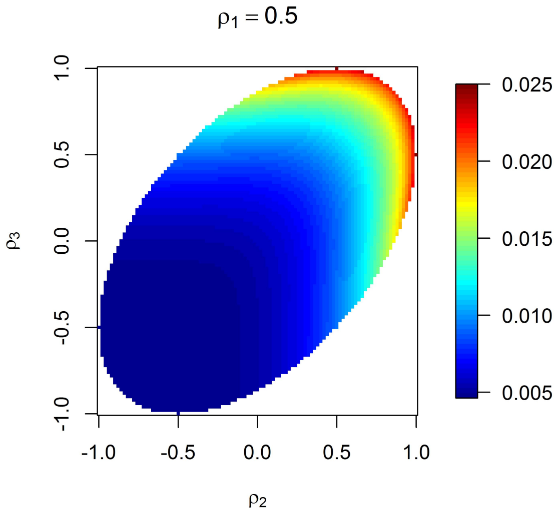
Numeric solution: Fallback test FWER, $\rho_1 = -0.5$



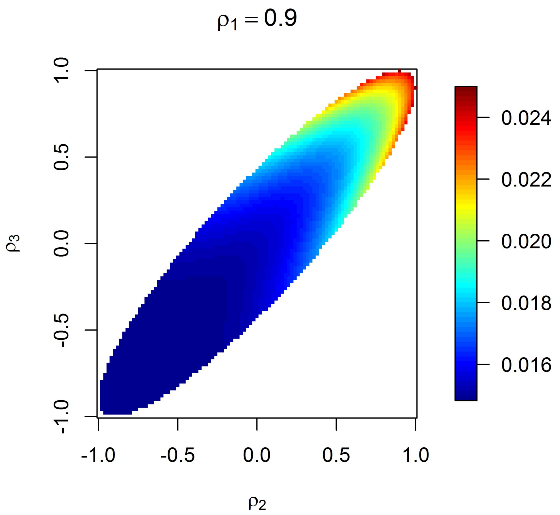
Numeric solution: Fallback test FWER, $\rho_1 = 0$



Numeric solution: Fallback test FWER, $\rho_1 = 0.5$



Numeric solution: Fallback test FWER, $\rho_1 = 0.9$



Summary fallback test

- Allows for proof of principle when two of three H_i are rejected at level α
- Reject $H_i \cap H_j$ with level α control
- Allows to reject significant elementary H_i and H_j with global level 2α
- Uniformly improvement of Rüger test under normality assumption
- Potentially useful in regulatory decision making
- Adds possibility for conclusion, which is especially desirable in the rare disease setting

Literature

CHMP, European Medicines Agency (2008), "Guideline on medicinal products for the treatment of Alzheimer's disease and other dementias."

CHMP, European Medicines Agency (2013), "Draft guideline on the clinical investigation of medicinal products for the treatment of Duchenne and Becker muscular dystrophy"

CHMP, European Medicines Agency (2002), "Points to consider on multiplicity issues in clinical trials"

Committee for medicinal products for human use, European Medicines Agency (2009), "Appendix 1 to the guideline on clinical evaluation of diagnostic agents (CPMP/EWP/1119/98 REV. 1) on imaging agents," Doc. Ref. EMEA/CHMP/EWP/321180/2008.

Genz, A. (2004), "Numerical computation of rectangular bivariate and trivariate normal and t probabilities", *Statistics and Computing*, 14, 251-260.

Glauser, T., G. Kluger, R. Sachdeo, G. Krauss, C. Perdomo and S. Arroyo (2008), "Rufinamide for generalized seizures associated with Lennox–Gastaut syndrome", *Neurology*, 70, 1950-1958

Marcus, R., E. Peritz and K.R. Gabriel (1976), "On closed testing procedures with special reference to ordered analysis of variance", *Biometrika*,63,655-660

Plackett, R. L. (1954), "A reduction formula for normal multivariate integrals", *Biometrika*, 41, 351-360.

Rüger, B. (1978), "Das maximale Signifikanzniveau des Tests: ,,Lehne H0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen", *Metrika*, 25, 171-178.

Thank you for your attention!
Any questions and discussion are welcome!