Vienna Section ROES

User-defined contrasts within multiple contrast tests- case studies using R

Ludwig A. Hothorn
hothorn@biostat.uni-hannover.de
Institute of Biostatistics, Leibniz University Hannover, Germany

October 2013

Differences Vienna- Hannover I

- *Weather, music, food, ...*
- sCI a MUST - see recent paper Phillips (2013):
  1. Because interpreting by decision makers, not statisticians
  2. Focusing on appropriate choice of **effect size** and their sCI
  3. (Adjusted) p-values are inappropriate from this perspective (although widely used)
  4. I.e. up to now: less focus on stepwise approaches or adaptive designs or gatekeeping (although 1991 ff papers)
- Nearly no FDR (just genome-wide Williams trend test using Benjamini-CI in package Isogene)
- Increasing power by:
  1. restricting the alternative
  2. taking the correlation(s) into account
  3. resulting in a **general non-product-moment structure**
  4. But after Ioannidis (2005) (Why most published research findings are false): **conservative is smart not painful- at least to some extend**

Differences Vienna- Hannover II

- Using R only, i.e. in papers and packages
- Gaussian distribution only a possibility, focus on GLMM
- Less focus on weighted procedures (choice of weights?)
- Not just RCT, but also genetics, toxicology, molecular biology
- Triple: superiority, non-inferiority, equivalence (by means of sCI)
- Non-parametric as well (co-working with Goettingen group in a joint DfG-project)
- Robustness, e.g. variance heterogeneity

The Problem I

- Charlies pioneering many21-procedure (Dunnett, 1955) is belonging to the most cited statistical papers. WebSci (04/2013): 4363 times
- Relevant up to now, e.g. for comparison of diversities in metagenomics (Pallmann et al., 2012)
- However, limited to **Gaussian errors with homogeneous variances**- and so in related software (SAS PROC MIXED, SPSS,...)
- But different endpoints occur commonly, e.g.:
  - i Proportions (Schaarschmidt and Biesheuvel, 2008)
  - ii Scores (count) data (Jaki et al., 2013)
  - iii Poly-3 estimates, i.e. mortality-adjusted tumor rates in carcinogenicity studies (Schaarschmidt et al., 2007)
  - iv Skewed-distributed endpoints:
    a) non-parametric (Konietschke and Hothorn, 2012),
    b) log-normal (Schaarschmidt, 2013),
  - v (Censored) time-to-event data (Herberich and Hothorn, 2012)

The Problem II

- Even when Gaussian errors can be assumed, a diversity of problems exist:

  i Inference for μ_i/μ_C instead of $\mu_i - \mu_C$ (Dilba et al., 2004, 2007)

  ii Correcting for heteroscedasticity in unbalanced designs (Hasler and Hothorn, 2008)

  iii Multiple endpoints:
  a) for superiority (Hasler and Hothorn, 2011),
  b) for non-inferiority (Hasler and Hothorn, 2013)

  iv Two-way layouts: claiming for or against qualitative interactions (Kitsche and Hothorn, 2013)

  v Mixed models (Kruppa, 2009)

  vi Using different contrasts, e.g. order restricted (Bretz, 2006), change-point (Hirotsu et al., 2011), vs. grand-mean (Djira and Hothorn, 2009)

  vii Replacing global ANOVA F-test by MCT vs. grand mean (Konietschke et al., 2013)

- Focusing on **simultaneous confidence intervals (sCI)**(*instead of adj. p-values*): interpretability *(single step procedures so far)*

MCP's formulated as MCT's I

- MCT: multiple contrast test
- A contrast is a suitable linear combination of estimators, e.g. means:

$$\sum_{i=0}^{k} c_i \bar{x}_i$$

- Here $i = 0...k$, focusing on comparisons vs. control
- A contrast test is standardized

$$t_{Contrast} = \sum_{i=0}^{k} c_i \bar{x}_i / S \sqrt{\sum_{i}^{k} c_i^2 / n_i}$$

where $\sum_{i=0}^{k} c_i = 0$ guaranteed a $t_{df,1-\alpha}$ distributed level-α-test

- To achieve compatible sCI $\sum sign^+(c_i) = 1, \sum sign^-(c_i) = 1$
- Notice, arbitrary c_i can be used in resampling tests- **one reason for their popularity?**

MCP's formulated as MCT's II

- A multiple contrast test is defined as maximum test:

$$t_{MCT} = max(t_1, ..., t_q)$$

  which follows jointly $(t_1, \ldots, t_q)'$ a q-variate t- distribution with degree of freedom df and correlation matrix R ($R = f(c_{ij}, n_i)$)

- R-library(mvtnorm): (non)-central multivariate t-distribution for any correlation matrix (Mi et al., 2009; Genz et al., 2012) **r-,d-,q-,p-**

- One-sided lower **simultaneous confidence limits**: $[\sum_{i=0}^{k} c_i \bar{x}_i - S \cdot t_{q,df,R,1-sided,1-\alpha} \sqrt{\sum_i^k c_i^2 / n_i}]$

MCP's formulated as MCT's III

- The choice of a particular contrast matrix defines the MCT
- Known examples (balanced design k=2 .... *just to keep it simple*)

Many-to-one, one-sided (Dunnett, 1955)

c_i	C	T_1	T_2
c_a	-1	0	1
c_b	-1	1	0

All pairs comparisons (Tukey1953)

c_i	C	T_1	T_2
c_a	-1	0	1
c_b	-1	1	0
c_c	0	-1	1
c_d	1	-1	0
c_e	-1	1	0
c_f	0	1	-1

Change-point comparisons (Hirotsu et al., 2011)

c_i	C	D_1	D_2
c_a	-1	0.5	0.5
c_b	-0.5	-0.5	1

Williams-type procedure (Bretz, 2006)

c_i	C	D_1	D_2
c_a	-1	0	1
c_b	-1	1/2	1/2

One- vs two-sided hypothesis I

- Simply using 2 different contrast matrices
- E.g. many21 One-sided:

c_i	C	T_1	T_2
c_a	-1	0	1
c_b	-1	1	0

Two-sided:

c_i	C	T_1	T_2
c_a	-1	0	1
c_b	-1	1	0
c_c	1	0	-1
c_d	1	-1	0

- Just two different related correlation matrices

sCI for ratios of μ_i I

- **Aim**: simultaneous confidence intervals for μ_i/μ_0

$$\omega_i = \mathbf{c}_i\boldsymbol{\mu}/\mathbf{d}_i\boldsymbol{\mu}$$

- $\mathbf{c}_i$ and $\mathbf{d}_i$ are the i^{th} row vector of **C** and **D** for numerator and denominator
- E.g. for Dunnett-type contrasts

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

sCI for ratios of μ_i II

- The simultaneous Fieller-type confidence intervals for ω_i are the solutions of the inequalities

$$T^2(\omega_i) = \frac{L^2(\omega_i)}{S^2_{L(\omega_i)}} \leq t^2_{q,\nu,R(\boldsymbol{\omega}),1-\alpha},$$

with the numerator

$$L(\omega_i) = \sum c_i \overline{Y}_i - d_i \omega_i \overline{Y}_0,$$

Notice, Sasabuchi's trick of a linear form

- $t_{q,\nu,R(\boldsymbol{\omega}_i),1-\alpha}$ is a central q-variate t-distribution with ν degrees of freedom and correlation matrix $R(\boldsymbol{\omega}_i) = [\rho_{ij}]$, where $\rho_{ii'}$ depend on c_{hi}, n_i **and on unknown ratios** ω_i: plug-in ML-estimators (Dilba et al., 2006) Trick no. 2
- The *mratios* R package (Dilba et al., 2007; Djira et al., 2012) can be used to make inferences about ratios of parameters in mixed models

sCI when variance heterogeneity occurs I

- Variance heterogeneity is quite common, i.e. $\varepsilon_{ij} \sim N(0, \sigma_i^2)$.
- Standard MCP do not control FWER, particularly for unbalanced n_i
- Modified test statistic $T^{2*}(\omega_i) = L(\omega_i)^2 / S^{2*}_{L(\omega_i)}$, where

$$S^{2*}{}_{L(\omega_i)} = \frac{\omega_i^2}{n_0} S_0^2 + \sum_{h=q+1-i}^{q} \frac{n_h}{\tilde{n}_i^2} S_h^2.$$

- $T^*(\omega_i)$ has an approximate *t*-distribution with approximate Satterthwaite-type ν
  Under variance heterogeneity: both ν **and** $R(\omega)$ depend on the **unknown ratios** ω_i **and the unknown variances** σ_i^2
- Plug-in modification: *sci.ratioVH* function in the R package *mratios* (Hasler and Hothorn, 2008)

Non-parametric procedure I

- Commonly: $H_0^F : F_0 = ... = F_k$ formulated in terms of the distribution functions against simple tree $H_1^F : F_0 < F_i$

- But the distribution of the rank means is unknown under H_1, neither sCI are available for a general unbalanced design, nor power can be estimated

- AND: tied or ordered categorical data, such as severity counts, should be analyzed as well

- AND: variance heterogeneity occurs frequently; therefore a Behrens-Fisher (BF) version is needed

Non-parametric procedure II

- Using relative effect size (Brunner and Munzel, 2000), (Ryu and Agresti, 2008):

$$p_{01} = \int F_0 dF_1 = P(X_{01} < X_{11}) + 0.5P(X_{01} = X_{11}).$$

- p_{01} is a *win probability* in the sense of Hayter (2013)
- p_{01} can be interpreted for trials with subjects Browne (2010)
- **sCI:** Konietschke (2011) Let $R_{sj}^{(0s)}$ denote the rank of X_{sj} among all $n_0 + n_s$ observations within the samples 0 and s

- The rank means can be used to estimate p_{0s}

$$\widehat{p}_{0s} = \frac{1}{n_0}\left(\overline{R}_{s\cdot}^{(0s)} - \frac{n_s + 1}{2}\right)$$

Non-parametric procedure III

- Asymptotically $\sqrt{N}(\widehat{p}_1 - p_1, \ldots, \widehat{p}_q - p_q)'$ follows a central multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{V}_N$ (Konietschke, 2011)

- Related approximate $(1-\alpha)100\%$ one-sided lower simultaneous confidence limits are:

$$\left[\widehat{p}_\ell - t_{q,\nu,\mathbf{R},1-\alpha}\sqrt{S_\ell}; \quad \right], \; \ell = 1, \ldots, q, \tag{1}$$

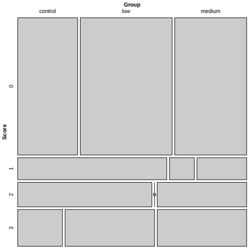
- E.g. relative Shirley-type effects for order restriction (Shirley, 1977)

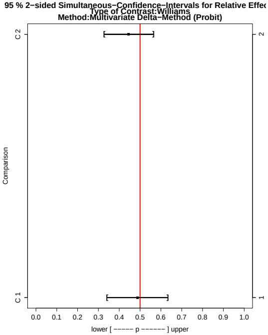
$$\begin{aligned} p_1 &= p_{0k} \\ p_2 &= \frac{n_{k-1}}{n_{k-1}+n_k} p_{0(k-1)} + \frac{n_k}{n_{k-1}+n_k} p_{0k} \\ &\vdots \\ p_q &= \frac{n_1}{n_1+\ldots+n_k} p_{01} + \ldots + \frac{n_k}{n_1+\ldots+n_k} p_{0k} \end{aligned}$$

Non-parametric procedure IV

- Shirley-type test for graded histopathological findings using R package *nparcomp*
  Ordered categorical findings of non-neoplastic lesions in the P-Cresidine carcinogenicity study: hyperplasia in parotid gland

```
library(nparcomp)
nparcomp(Score~Group, data=parotid, asy.method = "probit",type="Willi
```





sCI for proportions I

- Three approaches
  1. *Wald-type (Hothorn et al., 2008)*
  2. Add1- adjusted (Schaarschmidt and Biesheuvel, 2008)
  3. *Profile likelihood (Gerhard, 2010)*
- For sample sizes of $n_i = 50...10$ there is no hope for valid $(1-\alpha)100\%$ Wald intervals. Therefore we need confidence intervals with coverage probability approximately 95% also for smaller samples
- And, for almost all proportions a one-sided alternative for an increase/decrease is appropriate
- As effect size the difference of proportions is common (alternatively RR, OR)

sCI for proportions II

- One-sided, lower $(1-\alpha)100\%$ Wald-type confidence limits for the difference of the proportions of treatments against C:

$$\left[\sum_{i=1}^{I} c_i p_i - z_{q,R,1-\alpha} \sqrt{\sum_{i=1}^{I} c_i^2 \hat{V}(p_i)} \, ; \right]$$

with $\hat{V}(p_i) = p_i (1 - p_i) / n_i$ and $z_{q,R,1-\alpha}$ denoting the $(1-\alpha)$ quantile of the q-variate normal distribution

- $\boldsymbol{R}$ depends not only on the known contrast coefficients c_{im} and sample sizes n_i **but also on the unknown** π_i **and** $V(p_i)$ where the plug-in of the ML-estimators $\hat{\pi}_i$ and $\hat{V}(\pi_i)$ works well.

sCI for proportions III

- Agresti and Coull (1998) showed that adding a total of four pseudo-observations to the observed successes and failures yields approximate confidence intervals for one binomial proportion with good small sample performance

- One-sided limits were investigated by Cai (2005) in the case of a single binomial proportion

$$\left[\sum_{i=1}^{I} c_i \tilde{p}_i - z_{q,R,1-\alpha} \sqrt{\sum_{i=1}^{I} c_i^2 \tilde{V}(\tilde{p}_i)}\right]$$

- Choice of simultaneous confidence limits

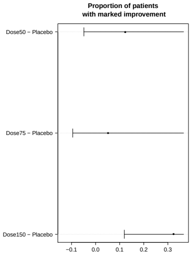
Notation	$\tilde{p}_i$	$\tilde{V}(p_i)$
Wald	Y_i / n_i	$p_i(1-p_i)/n_i$
add-1	$(Y_i + 0.5)/(n_i + 1)$	$\tilde{p}_i(1-\tilde{p}_i)/(n_i+1)$
add-2	$(Y_i + 1)/(n_i + 2)$	$\tilde{p}_i(1-\tilde{p}_i)/(n_i+2)$

sCI for proportions IV

- Simulation study (Schaarschmidt et al., 2008): use add1 approx. one-sided lower limits when n_i not too small
- Example: Simultaneous confidence limits for tubular epithelia hyaline droplet degeneration in male rats by means of MCPAN.

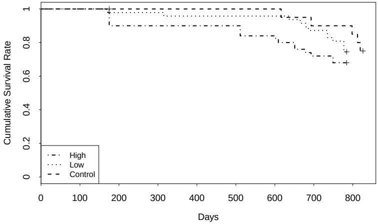
	Control	Dose50	Dose75	Dose150
with degeneration	2	6	4	13
n	32	27	32	21

```
library(MCPAN)
data(liarozole)
plot(binomRDci(tab, type="Dunnett", alternative="greater", method="ADD1"))
```



sCI for time-to-event data I

- Williams-type proc. comparing survival functions: i) Cox proport. hazards model or ii) the frailty Cox model to allow a joint analysis over sex and strains (Herberich and Hothorn, 2012)
- Example: Mortality in NTP-TR120 carcinogenicity



- Effect size: Hazard rate. Using Williams-type sCI

Comparison	Estimated HR	sim. 97.5%-Interval
C vs. D2	3.83	$[0.82, \infty)$
C vs. (D1, D2)	3.18	$[0.71, \infty)$

A Dunnett-type approach for multiple endpoints I

- In RCT with several primary correlated endpoints **and** a multi-arm design, multiplicity adjustment should take both the endpoints and the treatment comparisons into account, i.e. global control of FWER
- Extension of the Dunnett procedure (Dunnett, 1955) for k multiple endpoints and q comparisons

$$\{X_{ipj} : p = 1, \ldots, k\} \sim \perp \mathrm{N}_k(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \quad (i = 0, \ldots, q, \; j = 1, \ldots, n_i).$$

- I.e. unknown covariance matrices $\boldsymbol{\Sigma}_l \in \mathbb{R}^{k \times k}$ with possibly different variances and covariances for the endpoints, but the same covariance matrices for all treatments
- Testing the hypotheses

$$H_0^{(ip)} : \eta_{ip} \leq \delta_p$$

A Dunnett-type approach for multiple endpoints II

- This is a union-intersection-test because the overall null hypothesis of interest can be expressed as an intersection of the local null hypotheses, i.e.,

$$H_0 = \bigcap_{i=1}^{q} \left\{ \bigcap_{p=1}^{k} H_0^{(ip)} \right\}.$$

- This means that the overall null hypothesis H_0 is rejected if and only if a local null hypothesis $H_0^{(ip)}$ is rejected for at least one treatment for at least one endpoint.
- The test of the above hypotheses based on (now for the difference!)

$$T_{ip} = \frac{\bar{X}_{ip} - \bar{X}_{0p} - \delta_p}{S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \quad (i = 1, \ldots, q,\ p = 1, \ldots, k).$$

A Dunnett-type approach for multiple endpoints III

- The distribution of the univariate $\boldsymbol{T}_i$ under $H_0^{(i)}$ is simply a k-variate t-distribution with ν degrees of freedom and the correlation matrix $\boldsymbol{R}$, i.e.,

$$\boldsymbol{T}_i \sim t_{k,\nu,\boldsymbol{R},1-\alpha}.$$

- Consequently, under H_0, the vector of **all** test statistics,

$$\boldsymbol{T} = (\boldsymbol{T}_1', \ldots, \boldsymbol{T}_q')' = (T_{11}, \ldots, T_{ip}, \ldots, T_{qk})',$$

  follows approximately a qk-variate t-distribution with ν degrees of freedom and a correlation matrix denoted by $\tilde{\boldsymbol{R}}$, i.e.,

$$\boldsymbol{T} \overset{appr.}{\sim} t_{qk,\nu,\tilde{\boldsymbol{R}},1-\alpha}$$

A Dunnett-type approach for multiple endpoints IV

The correlation matrix $\tilde{\boldsymbol{R}}$ is given by

$$\tilde{\boldsymbol{R}} = (\boldsymbol{R}_{ii'})_{i,ii'} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \dots & \boldsymbol{R}_{1q} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} & \dots & \boldsymbol{R}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{1q} & \boldsymbol{R}_{2q} & \dots & \boldsymbol{R}_{qq} \end{pmatrix}.$$

The submatrices $\boldsymbol{R}_{ii'} = (\rho_{ii',pp'})$ describe the correlations between the ith and the i'th comparison for all endpoints. Their elements are

$$\rho_{ii',pp'} = \begin{cases} \rho_{pp'}, & i = i' \\ \rho_{pp'} \dfrac{1}{\sqrt{\left(\frac{n_0}{n_i}+1\right)\left(\frac{n_0}{n_{i'}}+1\right)}}, & i \neq i' \end{cases}$$

A Dunnett-type approach for multiple endpoints V

- **sCI** The lower limits of the approximate $(1-\alpha)100\%$ sCIs for $(\eta_{11}, \ldots, \eta_{qk})'$ are given by

$$\hat{\eta}_{ip}^{low} = \bar{X}_{ip} - \bar{X}_{0p} - t_{qk,\nu,\hat{\tilde{\boldsymbol{R}}},1-\alpha}\, S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}$$

- The R package SimComp was developed

Some user-defined contrasts I

I: Concept: claim-wise error rate Phillips (2013)

II: Regulatory toxicology

- US-NTP recommends the use of Dunnett and Williams procedure. Which one really? Take both! (Jaki and Hothorn, 2013)

Dun

c_{qi}	NC	D_1	D_2	D_3
c_a	-1	0	0	1
c_b	-1	0	1	0
c_c	-1	1	0	0

Wil

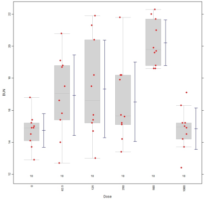
c_{qi}	NC	D_1	D_2	D_3
c_a	-1	0	0	1
c_b	-1	0	1/2	1/2
c_c	-1	1/3	1/3	1/3

UWil

c_{qi}	NC	D_1	D_2	D_3
c_a	-1	0	0	1
c_b	-1	0	1/2	1/2
c_c	-1	1/3	1/3	1/3
c_d	-1	0	1	0
c_e	-1	1/2	1/2	0
c_f	-1	1	0	0

Some user-defined contrasts II

- Blood urea nitrogen content after 13 weeks repeated administration of sodium dichromate dihydrate on male rats (NTP2012)



Comparison	Dun	Wil	DuWi	UWil
1000 − 0	0.80	0.60	0.80	0.80
500 − 0	6.8e-07	-	8.1e-07	8.4e-07
250 − 0	0.110	-	0.11	0.12
125 − 0	0.017	-	0.018	0.020
62.5 − 0	0.045	-	0.047	0.051
$(1000 + 500)/2 - 0$	-	0.0013	0.0030	0.0033
$(1000 + 500 + 250/3 - 0$	-	0.0029	0.0057	0.0063
$(1000 + 500 + 250 + 125)/4 - 0$	-	0.0021	0.0037	0.0042
$(1000 + 500 + 250 + 125 + 62.5)/5 - 0$	-	0.0022	0.0039	0.0043
$(500 + 250)/2 - 0$	-	-	-	< 0.001
$(500 + 250 + 125)/3 - 0$	-	-	-	< 0.001
$(500 + 250 + 125 + 62.5)/4 - 0$	-	-	-	< 0.001
$(250 + 125)/2 - 0$	-	-	-	0.023
$(250 + 125 + 62.5)/3 - 0$	-	-	-	0.015
$(125 + 62.5)/2 - 0$	-	-	-	0.015

Some user-defined contrasts III

III: Genetic association studies

Association between a di-allelic marker and a disease can be presented in a 2×3 contingency table, where *aa* is the high risk candidate allele and *AA* is any of the other alleles

	aa	aA	AA	Total
Cases	r_{aa}	r_{aA}	r_{AA}	r
Controls	s_{aa}	s_{aA}	s_{AA}	s
Total	n_{aa}	n_{aA}	n_{AA}	n

The global null hypothesis for the unknown proportions $\pi_j = E(r_j/n_j), j \in \{aa, aA, AA\}$

$$H_0 : \pi_{aa} = \pi_{aA} = \pi_{AA}$$

can be compared to either a global heterogeneity alternative

$$H_1^{heterogeneity} : \pi_j \neq \pi_{j'}, j \neq j' \in (aa, aA, AA)$$

e.g. by Pearson χ^2 test

Some user-defined contrasts IV

or to a global order restricted alternative

$$H_1^{ordered} : \pi_{aa} \leq \pi_{aA} \leq \pi_{AA}$$

$H_1^{ordered}$ can be decomposed in three elementary alternatives

$$H_1^{additiv} : \pi_{aa} < \pi_{aA} < \pi_{AA}$$

$$H_1^{dominant} : \pi_{aa} < \pi_{aA} = \pi_{AA}$$

$$H_1^{recessive} : \pi_{aa} = \pi_{aA} < \pi_{AA}$$

Some user-defined contrasts V

1. The quadratic form of the global heterogeneity test can be replaced by MCT against grand mean (Konietschke et al., 2013)

	π_{aa}	π_{aA}	π_{AA}
any wild type vs. risk	-1	0.5	0.5
heterocygotes vs. homocyg.	0.5	-1	0.5
any risk alle vs. homocyg. wild type	0.5	0.5	-1

2. For the order restricted alternatives the contrast coefficients c_{jq} are:

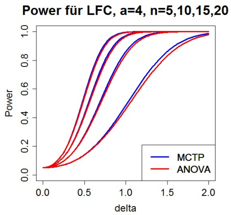
	π_{aa}	π_{aA}	π_{AA}
additive	-1	0	1
dominant	-1	0.5	0.5
recessive	-0.5	-0.5	1

3. Together:

	π_{aa}	π_{aA}	π_{AA}
additive mode	-1	0	1
dominant mode= any homocygotes vs GM	-1	0.5	0.5
recessive mode=risk homocygotes vs GM	-0.5	-0.5	1
over-dominance mode =heterocyg. vs homocyg	0.5	-1	0.5

Replacing ANOVA F-test by MCT vs. grand mean I

- Analyzing one-way layouts by F-test or Kruskal-Wallis test is common
- Quadratic F-test can be replaced by **max-test of linear contrasts vs. grand mean** Konietschke et al. (2013)



- Power:
  i) similar for least favorable configuration,
  ii) larger or smaller for any alternatives
- sCI available

R libraries - LUH and friends I

- multcomp
- mvtnorm
- mratio
- MCPAN
- SimComp
- goric
- mcprofile
- AND: pairwiseCI, BSagri, simboot, PropCIs, binMto

Take Home Message I

- (Single step) **sCI** are available for **most endpoints, designs and contrast formulations**
- Related R packages are available: UseR!
- (Not shown: power for the compatible tests can be calculated under some assumptions)
- I.e. unified analysis of all end-points in a trial/study is possible
- Alternative: hypothesis-restricted AIC-based model selection, e.g. for MED estimation (Kuiper et al., 2013)
- Focus now: mixed model applications

References I

Agresti, A. and Coull, B. A. (1998). Approximate is better than ëxactfor interval estimation of binomial proportions. *American Statistician*, 52(2):119–126.

Bretz, F. (2006). An extension of the williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis*, 50(7):1735–1748.

Browne, R. H. (2010). The t-test p value and its relationship to the effect size and p(x > y). *American Statistician*, 64(1):30–33.

Brunner, E. and Munzel, U. (2000). The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*, 42(1):17–25.

Cai, T. T. (2005). One-sided confidence intervals in discrete distributions. *Journal Of Statistical Planning And Inference*, 131(1):63–88.

Dilba, G., Bretz, E., Guiard, V., and Hothorn, L. A. (2004). Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods Of Information In Medicine*, 43(5):465–469.

Dilba, G., Bretz, F., Hothorn, L. A., and Guiard, V. (2006). Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. *Statistics In Medicine*, 25(7):1131–1147.

Dilba, G., Schaarschmidt, F., and Hothorn, L. (2007). Inferences for ratios of normal means. *R News*, 7:20–23.

Djira, G. and Hothorn, L. (2009). Detecting relative changes in multiple comparisons with an overall mean. *J Quality Control*, 41:60–65.

Djira, G. D., Hasler, M., Gerhard, D., and Schaarschmidt, F. (2012). *mratios: Inferences for ratios of coefficients in the general linear model*. R package version 1.3.17.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal Of The American Statistical Association*, 50(272):1096–1121.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2012). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9994.

Gerhard, D. (2010). Simultaneous small sample inference based on profile likelihood. Master's thesis, LUH.

Hasler, M. and Hothorn, L. (2008). Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 51:1.

Hasler, M. and Hothorn, L. (2013). Simultaneous confidence intervals on multivariate non-inferiority. *Statistics in Medicine*, DOI: 10.1002/sim.5633:DOI: 10.1002/sim.5633.

References II

Hasler, M. and Hothorn, L. A. (2011). A Dunnett-Type Procedure for Multiple Endpoints. *INTERNATIONAL JOURNAL OF BIOSTATISTICS*, 7(1).

Hayter, A. J. (2013). Inferences on the difference between future observations for comparing two treatments. *JOURNAL OF APPLIED STATISTICS*, 40(4):887–900.

Herberich, E. and Hothorn, L. A. (2012). Statistical evaluation of mortality in long-term carcinogenicity bioassays using a williams-type procedure. *Regulatory Toxicology And Pharmacology*, yy:xx. In press.

Hirotsu, C., Yamamoto, S., and Hothorn, L. A. (2011). Estimating the Dose-Response Pattern by the Maximal Contrast Type Test Approach. *STATISTICS IN BIOPHARMACEUTICAL RESEARCH*, 3(1):40–53.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine*, 2(8):e124.

Jaki, T. and Hothorn (2013). Statistical evaluation of toxicological assays: Dunnett or williams test?- take both. *Arch Tox*, 00:00.

Jaki, T., Kitsche, A., and Hothorn, L. (2013). Statistical evaluation of toxicological assays with zero or near-to-zero proportions or counts in the concurrent negative control group: A tutorial. *Regul. Pharm . Tox.*, 00:000.

Kitsche, A. and Hothorn, L. (2013). Testing for qualitative interaction using ratios of treatment differences. *Statistics in M*.

Konietschke, F., Boesiger, S., Brunner, E., and Hothorn, L. (2013). Are multiple contrasts tests superior to the anova. *Int. J. Biostat*, 00:00.

Konietschke, F. and Hothorn, L. A. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. *ELECTRONIC JOURNAL OF STATISTICS*, 6:738–759.

Konietschke, F., H. L. (2011). Evaluation of toxicological studies using a non-parametric shirley-type trend test for comparing several dose levels with a control group. *Statistics in Biopharmaceutical Research*, xx:(in print).

Kruppa, J. (2009). Simultaneous confidence intervals for mixed effect parameters in a linear mixed model. Master's thesis, LUH.

Kuiper, R., Gerhard, D., and Hoth (2013). Identification of the minimum effective dose for normally distributed endpoints using a model selection approach. *tba*.

Mi, X., Miwa, T., and Hothorn, T. (2009). mvtnorm: New Numerical Algorithm for Multivariate Normal Probabilities. *R JOURNAL*, 1(1):37–39.

References III

Pallmann, P., Schaarschmidt, F., Hothorn, L. A., Fischer, C., Nacke, H., Priesnitz, K. U., and Schork, N. J. (2012). Assessing group differences in biodiversity by simultaneously testing a user-defined selection of diversity indices. *MOLECULAR ECOLOGY RESOURCES*, 12(6):1068–1078.

Phillips, A. (2013). Multiplicity: discussion points form the statstatistic in the pharmaceutical industry multiplicity expert group. *Pharmaceutical Stat*, 12:255–259.

Ryu, E. J. and Agresti, A. (2008). Modeling and inference for an ordinal effect size measure. *Statistics In Medicine*, 27(10):1703–1717.

Schaarschmidt, F. (2013). Simultaneous confidence intervals for multiple comparisons among expected values of log-normal variables. *COMPUTATIONAL STATISTICS & DATA ANALYSIS*, 58:265–275.

Schaarschmidt, F. and Biesheuvel, E. (2008). Asymptotic simultaneous confidence intervals for many-to-one comparisons of binary proportions in randomized clinical trials"]. *J. Biopharmaceutical Statistics*, 18:(accepted).

Schaarschmidt, F., Sill, M., and Hothorn, L. (2007). Poly-k-trend tests for survival adjusted analysis of tumor rates formulated as approximate multiple contrast test. Technical report, Leibniz University Hannover, Institute of Biostatistics.

Schaarschmidt, F., Sill, M., and Hothorn, L. A. (2008). Poly-k-trend tests for survival adjusted analysis of tumor rates formulated as approximate multiple contrast test. *Journal Of Biopharmaceutical Statistics*, 18(5):934–948.

Shirley, E. (1977). Nonparametric equivalent of WILLIAMS test for contrasting increasing dose levels of a treatment. *Biometrics*, 33(2):386–389.