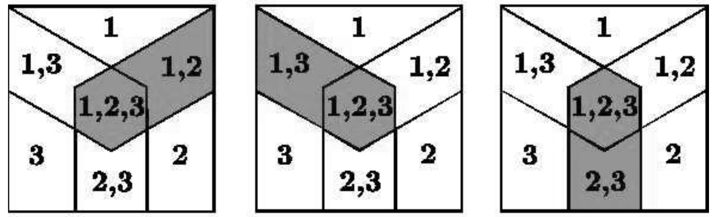
MCP2002 Bethesda



Demonstration of interlaboratory similarity of dose-response assays by modified interaction contrasts

Ludwig A. Hothorn
University of Hannover/ Germany
E-mail: hothorn@bioinf.uni-hannover.de

UNIVERSITÄT HANNOVER

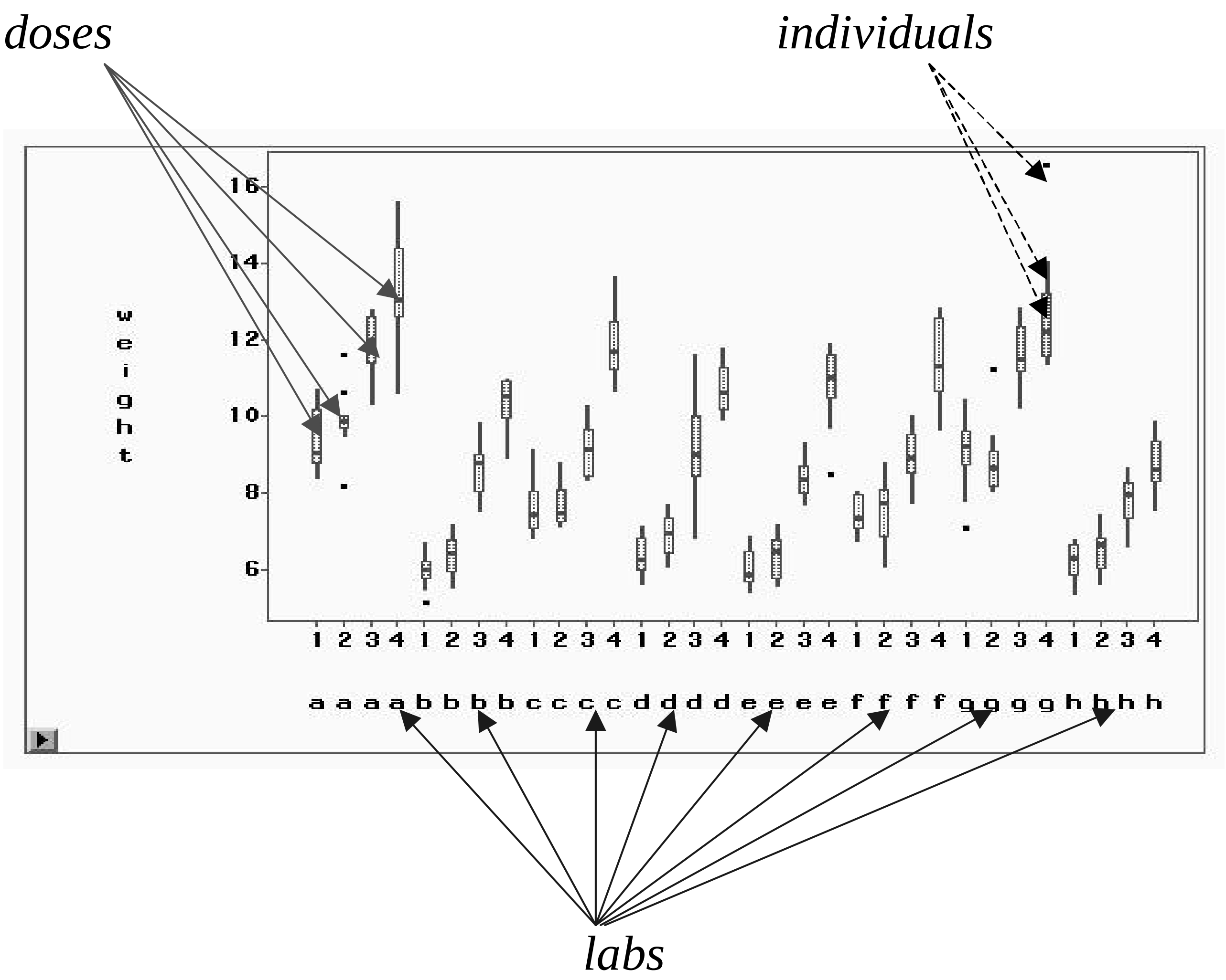
1. A toxicological example

- Validation process of new tox-assays:
  demonstrating similar sensitivity in several laboratories
  the majority of non-toxic chemicals ⇒ **-**
  the majority of toxic chemicals ⇒ **+**
  in all labs, i.e. no qualitative interaction

- Common design **[C, D_1,...,D_k]**

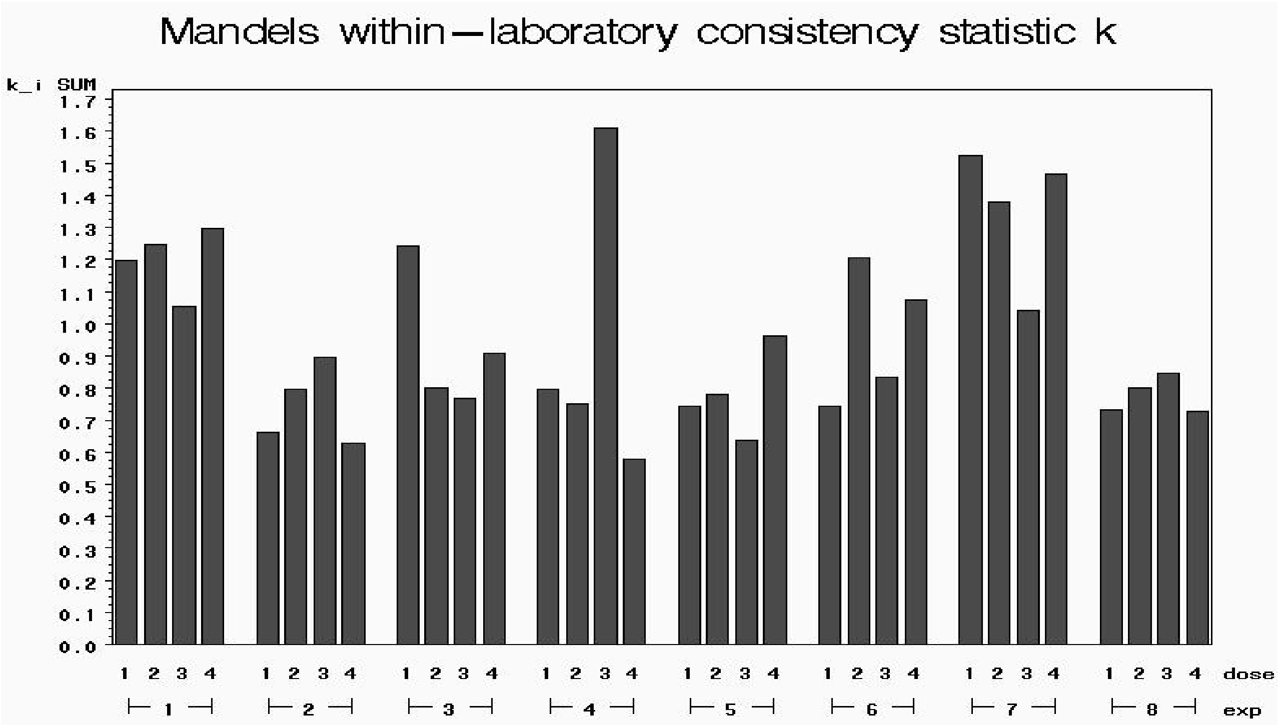
- Assay similarity= **similarity of the dose-response curves**, or at least similarity of ***increments versus C*** in all labs, i.e. **no quantitative interaction**

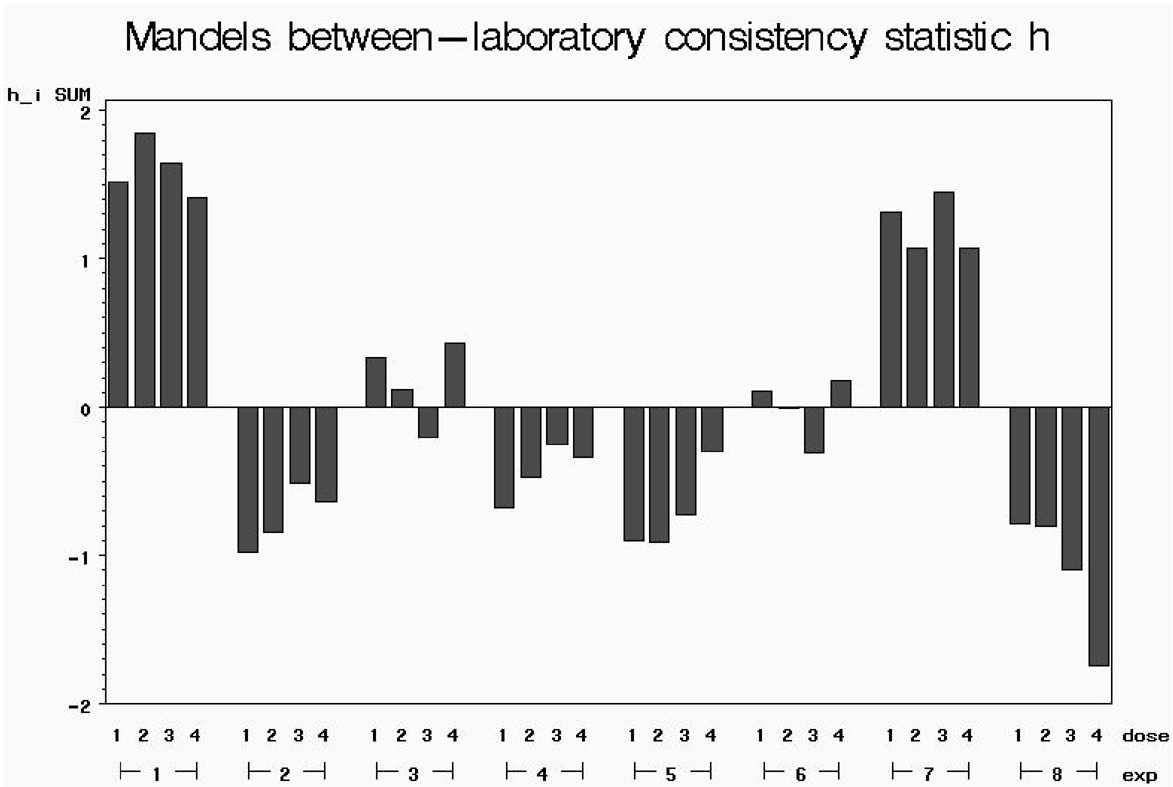
Example: HCB immunotox study (Schulte et al., 2002), liver weight



2. Common approaches of interlaboratory comparisons

Recommendation: Estimation of intra- and interlaboratory variance, typically as Mandel plots (ASTM Committee E-11 on Statistical Methods, 1993)

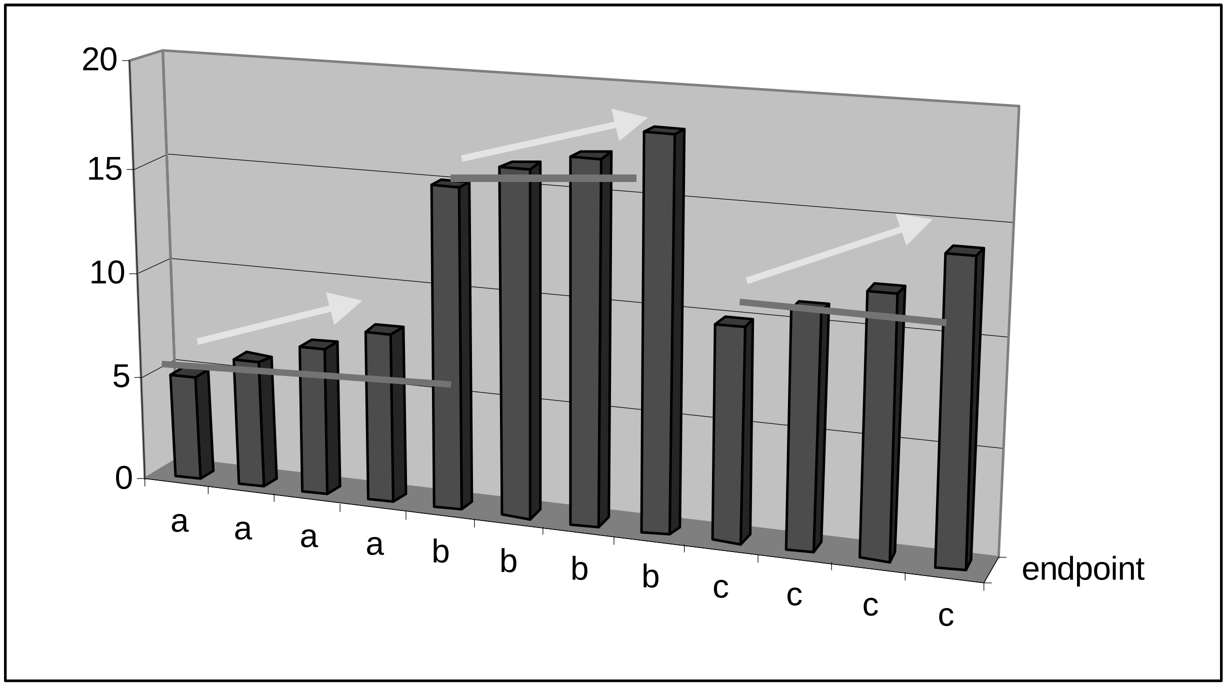




However: All toxicological studies would fail these criteria of reproducibility⇒ directed on all absolute values

3. Interlaboratory study in toxicology

Same dose-response   **Different absolute values**



Show similarity of dose-response curves over labs = similarity of <u>effect sizes</u>, but not similarity of the absolute values

4. Interaction contrasts

I) Testing interaction in two-way layout: DOSE x LAB common by F-test. Global test only; no confidence intervals

II) Demonstrating similarity between labs

Similarity $\Rightarrow$ equivalence of selected interaction contrasts $\Rightarrow$ TOST; but a-priori no equivalence margins known $\Rightarrow$ estimation of **(1-2α) confidence intervals** (Schuirmann, 1987).

III) Tetrad pairwise interaction contrasts

$\frac{L(L-1)}{2}\frac{(k+1)k}{2}$ **tetrad pairwise interaction contrasts**

$$\max\left\{|(\mu_{lj}-\mu_{lj'})-(\mu_{l'j}-\mu_{l'j'})|\right\}$$

l=1,..., L-1; l'=2,..., L, j=0,..., k-1; j'=1,..., k

Simultaneous confidence intervals:

$$\sum_{l=1}^{L}\sum_{i=1}^{k}c_{li}\gamma_{li}\in\left[\sum_{l=1}^{L}\sum_{i=1}^{k}c_{li}\hat{\gamma}_{li}\pm q_{1-\alpha}\sqrt{\frac{MSE}{n}}\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{k}|c_{li}|\right]$$

$\gamma_{li}=\overline{Y}_{il.}-\overline{Y}_{i..}-\overline{Y}_{.l.}+\overline{Y}_{...}$ tetrad difference,

$q_{1-\alpha}$ the quantile:

- Bradu and Gabriel (1974) Bonferroni approach
- Hochberg and Tamhane (1986) range statistics (balanced)
- Westfall et al. (1999) simulation-based (MC-sim)
- Bretz et al. (2002) any design based on the multivariate t-distribution with a related correlation structure

But all-pairs comparisons between the doses are not the experimental question $\Rightarrow$ reduce the number of comparison to **many-to-one** between the doses only

IV) Many-to-one interaction contrasts

$\frac{L(L-1)}{2}k$ **many-to-one-by-lab-interaction contrasts**

$$\max\left\{\left| (\mu_{lj} - \mu_{l0}) - (\mu_{l'j} - \mu_{l'0}) \right|\right\}$$

(l=1,...,L-1; l’=2,...,L; j=1,...,k; j=0 ... control).
This UIT allows all elementary decisions, i.e. simultaneous confidence intervals (Gabriel, 1969)

$$H_0 : \bigcup_{l=1,l'=2}^{L} \bigcup_{j=1,j'=2}^{k} \left|(\mu_{lj} - \mu_{l0}) - (\mu_{l'j} - \mu_{l'0})\right| > \delta$$

(Notice in the example with 1+3 doses and 8 labs: 168 all-pairs vs. 84 many-to-one comparisons)

Interpretation:

- based on a-priori chosen δ, e.g. ± 3g liver weight difference
- [or a-posteriori chosen δ, e.g. ....]

Example:

Contrast	90% Conf. interval Lower	Upper
(a 2-a 1)-(b 2-b 1)	-1,56	1,73
(a 3-a 1)-(b 3-b 1)	-1,88	1,41
(a 4-a 1)-(b 4-b 1)	-2,08	1,21
(a 2-a 1)-(c 2-c 1)	-1,02	2,27
(a 3-a 1)-(c 3-c 1)	-0,55	2,74
(a 4-a 1)-(c 4-c 1)	-1,83	1,46
(a 2-a 1)-(d 2-d 1)	-1,65	1,69
(a 3-a 1)-(d 3-d 1)	-1,86	1,47
(a 4-a 1)-(d 4-d 1)	-2,12	1,22
(a 2-a 1)-(e 2-e 1)	-1,36	1,93
(a 3-a 1)-(e 3-e 1)	-1,45	1,83
(a 4-a 1)-(e 4-e 1)	-2,44	0,84
(a 2-a 1)-(f2-f1)	-1,17	2,12
(a 3-a 1)-(f3-f1)	-0,69	2,59
(a 4-a 1)-(f4-f1)	-1,86	1,60
(a 2-a 1)-(g 2-g 1)	-0,92	2,37
(a 3-a 1)-(g 3-g 1)	-1,63	1,66
(a 4-a 1)-(g 4-g 1)	-1,43	1,86
(a 2-a 1)-(h 2-h 1)	-1,34	1,95
(a 3-a 1)-(h 3-h 1)	-0,75	2,54
(a 4-a 1)-(h4-h1)	**-0,22**	**3,07**
(b 2-b 1)-(c 2-c 1)	-1,10	2,19
(b 3-b 1)-(c 3-c 1)	-0,31	2,97
(b 4-b 1)-(c 4-c 1)	-1,39	1,90
(b 2-b 1)-(d 2-d 1)	-1,73	1,61
(b 3-b 1)-(d 3-d 1)	-1,63	1,71
(b 4-b 1)-(d 4-d 1)	-1,68	1,65
(b 2-b 1)-(e 2-e 1)	-1,44	1,84
(b 3-b 1)-(e 3-e 1)	-1,22	2,07
(b 4-b 1)-(e 4-e 1)	-2,01	1,28
(b 2-b 1)-(f2-f1)	-1,25	2,04
(b 3-b 1)-(f3-f1)	-0,46	2,83
(b 4-b 1)-(f4-f1)	-1,42	2,04
(b 2-b 1)-(g 2-g 1)	-1,00	2,29
(b 3-b 1)-(g 3-g 1)	-1,40	1,89
(b 4-b 1)-(g 4-g 1)	-1,00	2,29
(b 2-b 1)-(h 2-h 1)	-1,42	1,87
(b 3-b 1)-(h 3-h 1)	-0,51	2,78
(b4-b1)-(h4-h1)	**0,21**	**3,50**
(c 2-c 1)-(d 2-d 1)	-2,27	1,07
(c 3-c 1)-(d 3-d 1)	-2,96	0,38
(c 4-c 1)-(d 4-d 1)	-1,93	1,40
(c 2-c 1)-(e 2-e 1)	-1,99	1,30
(c 3-c 1)-(e 3-e 1)	-2,55	0,74
(c 4-c 1)-(e 4-e 1)	-2,26	1,03
(c 2-c 1)-(f2-f1)	-1,79	1,50
(c 3-c 1)-(f3-f1)	-1,79	1,50
(c 4-c 1)-(f4-f1)	-1,67	1,79
(c 2-c 1)-(g 2-g 1)	-1,54	1,75
(c 3-c 1)-(g 3-g 1)	-2,73	0,56
(c 4-c 1)-(g 4-g 1)	-1,25	2,04
(c 2-c 1)-(h 2-h 1)	-1,96	1,33
(c 3-c 1)-(h 3-h 1)	-1,84	1,45
(c 4-c 1)-(h4-h1)	**-0,04**	**3,25**
(d 2-d 1)-(e 2-e 1)	-1,41	1,93
(d 3-d 1)-(e 3-e 1)	-1,28	2,05
(d 4-d 1)-(e 4-e 1)	-2,02	1,32
(d 2-d 1)-(f2-f1)	-1,21	2,12
(d 3-d 1)-(f3-f1)	-0,52	2,81
(d 4-d 1)-(f4-f1)	-1,43	2,08
(d 2-d 1)-(g 2-g 1)	-0,96	2,37
(d 3-d 1)-(g 3-g 1)	-1,46	1,88
(d 4-d 1)-(g 4-g 1)	-1,00	2,33
(d 2-d 1)-(h 2-h 1)	-1,38	1,95
(d 3-d 1)-(h 3-h 1)	-0,57	2,76
(d4-d1)-(h4-h1)	**0,21**	**3,54**
(e 2-e 1)-(f2-f1)	-1,45	1,84
(e 3-e 1)-(f3-f1)	-0,88	2,40
(e 4-e 1)-(f4-f1)	-1,06	2,40
(e 2-e 1)-(g 2-g 1)	-1,20	2,09
(e 3-e 1)-(g 3-g 1)	-1,82	1,47
(e 4-e 1)-(g 4-g 1)	-0,63	2,66
(e 2-e 1)-(h 2-h 1)	-1,62	1,67
(e 3-e 1)-(h 3-h 1)	-0,94	2,35
(e 4-e 1)-(h4-h1)	**0,58**	**3,87**
(f2-f1)-(g 2-g 1)	-1,40	1,89
(f3-f1)-(g 3-g 1)	-2,58	0,71
(f4-f1)-(g 4-g 1)	-1,39	2,07
(f2-f1)-(h 2-h 1)	-1,82	1,47
(f3-f1)-(h 3-h 1)	-1,70	1,59
(f4-f1)-(h4-h1)	**-0,18**	**3,28**
(g 2-g 1)-(h 2-h 1)	-2,07	1,22
(g 3-g 1)-(h 3-h 1)	-0,76	2,53
(g 4-g 1)-(h 4-h 1)	-0,43	2,85

lab		$\bar{x}_4 - \bar{x}_{Control}$
a	1	3.92
b	2	4.35
c	3	4.10
d	4	4.37
e	5	4.72
f	6	4.05
g	7	3.70
h	**8**	**2.49**

- ⇨ Comparison C- vs. high dose between labs e, d, b vs. h
- ⇨ Lab h reveals a lower effect size
- ⇨ Eliminating lab h: global F-test p>>0.10, no CI relevant increased ==> all labs are similar according to their comparisons vs. C- ... perfect!

IV) Interaction contrasts based on order restriction

i) Incremental contrasts (not really order restricted)

$$\max\{| (\mu_{lj} - \mu_{lj-1}) - (\mu_{l'j} - \mu_{l'j-1}) |\}$$

ii) Isotonic contrasts

Order restriction: $H_1 : \mu_0 \leq \mu_1 \leq ... \leq \mu_k$ |<

<u>Confidence intervals:</u>

$\frac{L(L-1)}{2}(2^k - 1)$ Isotonic contrasts (Bretz 1999)

$$\max\left\{ | \sum_{j=0}^{k} c_j \mu_{lj} - \sum_{j=0}^{k} c_j \mu_{l'j} | \right\}$$

Still more contrasts (e.g. 196), but with very high correlated!

Alternative	**Contrast coefficients**
$H_1^1: \mu_C = \mu_1 = \mu_2 < \mu_3$	–0.09, –0.09, –0.09, 0.27
$H_1^2: \mu_C < \mu_1 = \mu_2 = \mu_3$	–0.27, 0.09, 0.09, 0.09
$H_1^3: \mu_C = \mu_1 < \mu_2 = \mu_3$	–0.16, –0.16, 0.16, 0.16
$H_1^4: \mu_C < \mu_1 = \mu_2 < \mu_3$	–0.27, 0, 0, 0.27
$H_1^5: \mu_C < \mu_1 < \mu_2 < \mu_3$	–0.28, –0.04, 0.04, 0.28
$H_1^6: \mu_C = \mu_1 < \mu_2 < \mu_3$	–0.16, –0.16, 0.04, 0.28
$H_1^7: \mu_C < \mu_1 < \mu_2 = \mu_3$	–0.28, –0.04, 0.16, 0.16

Isotonic contrasts for balanced design **[C, D_1, D_2, D_3]**

Margin for equivalence: ±30 g difference

Lab	Contrast	Lower 90% CL	Upper 90% CL
a vs b	$H_1^6 : \mu_C = \mu_1 < \mu_2 < \mu_3$	-17,9	9,0
a vs c		-16,6	10,2
a vs. d		-17,8	9,2
a vs. e		-21,5	5,3
a vs f		-16,5	12,1
a vs. g		-15,2	11,7
a vs h		-1,4	25,5
b vs c		-12,2	14,7
b vs. d		-13,3	13,7
b vs. e		-17,0	9,8
b vs f		-12,0	16,5
b vs. g		-10,7	16,2
b vs h		3,1	29,9
c vs. d		-14,5	12,5
c vs. e		-18,3	8,6
c vs f		-13,2	15,3
c vs. g		-11,9	14,9
c vs h		1,8	28,7
d vs. e		-17,3	9,7
d vs f		-12,3	16,4
d vs. g		-11,0	16,0
d vs h		2,8	29,8
e vs f		-8,4	20,1
e vs. g		-7,1	19,8
e vs h		6,7	33,6
f vs. g		-13,8	14,7
f vs h		0,0	28,5
g vs. h		0,3	27,2
a vs b	$H_1^1 : \mu_C = \mu_1 = \mu_2 < \mu_3$	-15,2	8,6
a vs c		-18,4	5,4
a vs. d		-15,3	8,6
a vs. e		-20,2	3,6
a vs f		-18,1	7,6
a vs. g		-12,2	11,7
a vs h		-3,0	20,8
b vs c		-15,2	8,7
b vs. d		-12,0	11,9
b vs. e		-16,9	6,9
b vs f		-14,8	10,9
b vs. g		-8,9	15,0
b vs h		0,3	24,1
c vs. d		-8,8	15,1
c vs. e		-13,7	10,2
c vs f		-11,5	14,2
c vs. g		-5,7	18,2
c vs h		3,5	27,3
e vs f		12,0	42,7
e vs. g		-3,9	20,0
e vs h		5,3	29,1
f vs. g		-7,9	17,8
f vs h		1,2	26,9
g vs. h		-2,8	21,1
a vs b	$H_1^3 : \mu_C = \mu_1 < \mu_2 = \mu_3$	-15,7	8,2
a vs c		-10,5	13,4
a vs. d		-15,3	8,7
a vs. e		-16,4	7,5
a vs f		-10,5	14,0
a vs. g		-14,4	9,4
a vs h		-1,8	22,0
b vs c		-6,7	17,1
b vs. d		-11,6	12,4
b vs. e		-12,6	11,2
b vs f		-6,8	17,7
b vs. g		-10,7	13,2
b vs h		1,9	25,8
c vs. d		-16,8	7,2
c vs. e		-17,8	6,0
c vs f		-12,0	12,5
c vs. g		-15,9	8,0
c vs h		-3,3	20,6
d vs. e		-13,1	10,9
d vs f		-7,3	17,4
d vs. g		-11,2	12,9
d vs h		1,4	25,5
e vs f		-6,0	18,4
e vs. g		-10,0	13,9
e vs h		2,6	26,5
f vs. g		-16,5	8,0
f vs h		-3,9	20,6
g vs. h		0,7	24,5
a vs b	$H_1^2 : \mu_C < \mu_1 = \mu_2 = \mu_3$	-13,6	10,2
a vs c		-7,5	16,4
a vs. d		-14,0	10,4
a vs. e		-12,9	11,0
a vs f		-8,3	15,8
a vs. g		-9,2	14,7
a vs h		-4,3	19,5
b vs c		-5,8	18,1
b vs. d		-12,3	12,1
b vs. e		-11,2	12,7
b vs f		-6,6	17,5
b vs. g		-7,5	16,4
b vs h		-2,6	21,2
c vs. d		-18,4	5,9

d vs f		-6,7	17,8
d vs. g		-7,6	16,7
d vs h		-2,8	21,6
e vs f		-7,3	16,7
e vs. g		-8,2	15,6
e vs h		-3,4	20,5
f vs. g		-13,0	11,0
f vs h		-8,2	15,9
g vs. h		-7,1	16,8
a vs b	$H_1^4 : \mu_C < \mu_1 = \mu_2 < \mu_3$	-18,3	10,9
a vs c		-16,2	13,0
a vs. d		-18,7	10,9
a vs. e		-21,5	7,7
a vs f		-16,5	14,3
a vs. g		-12,7	16,5
a vs h		-2,2	27,0
b vs c		-12,4	16,8
b vs. d		-15,0	14,7
b vs. e		-17,8	11,4
b vs f		-12,7	18,0
b vs. g		-9,0	20,2
b vs h		1,5	30,7
c vs. d		-17,1	12,5
c vs. e		-20,0	9,2
c vs f		-14,9	15,9
c vs. g		-11,2	18,1
c vs h		-0,7	28,5
d vs. e		-17,9	11,8
d vs f		-12,8	18,4
d vs. g		-9,0	20,6
d vs h		1,4	31,1
e vs f		-9,5	21,2
e vs. g		-5,8	23,4
e vs h		4,7	33,9
f vs. g		-12,4	18,3
f vs h		-1,9	28,8
g vs. h		-4,1	25,1
a vs b	$H_1^7 : \mu_C < \mu_1 < \mu_2 = \mu_3$	-16,9	10,0
a vs c		-9,7	17,2
a vs. d		-16,9	10,4
a vs. e		-16,9	10,0
a vs f		-10,2	17,2
a vs. g		-13,3	13,6
a vs h		-2,2	24,6
b vs c		-6,3	20,6
b vs. d		-13,5	13,8
b vs. e		-13,4	13,5
b vs f		-6,8	20,6

c vs. g		-17,0	9,9
c vs h		-5,9	20,9
d vs. e		-13,8	13,5
d vs f		-7,2	20,7
d vs. g		-10,2	17,1
d vs h		0,8	28,1
e vs f		-6,8	20,6
e vs. g		-9,8	17,0
e vs h		1,2	28,1
f vs. g		-17,0	10,4
f vs h		-6,0	21,4
g vs. h		-2,4	24,5
a vs b	$H_1^5 : \mu_C < \mu_1 < \mu_2 < \mu_3$	-18,9	10,6
a vs c		-15,7	13,8
a vs. d		-19,2	10,8
a vs. e		-21,8	7,7
a vs f		-16,0	15,1
a vs. g		-13,9	15,7
a vs h		-1,6	27,9
b vs c		-11,6	18,0
b vs. d		-15,0	15,0
b vs. e		-17,7	11,9
b vs f		-11,8	19,3
b vs. g		-9,7	19,9
b vs h		2,5	32,1
c vs. d		-18,2	11,7
c vs. e		-20,9	8,7
c vs f		-15,0	16,0
c vs. g		-12,9	16,6
c vs h		-0,7	28,9
d vs. e		-17,9	12,1
d vs f		-12,0	19,5
d vs. g		-9,9	20,1
d vs h		2,4	32,3
e vs f		-8,9	22,1
e vs. g		-6,8	22,8
e vs h		5,4	35,0
f vs. g		-14,2	16,9
f vs h		-1,9	29,1
g vs. h		-2,5	27,0

Again lab h behaves not similar, but rejection of lab h shows similarity for all but one shape of dose-response.

Interpretation:

- selecting 'best fitting' shape, here H^6

V) Special case: linear contrasts only

(Equivalence approach for parallel-line assay)

Comparison	90% lower CI	90% upper CI
a vs b	-9.4	4.2
a vs c	-8.5	5.2
a vs. d	-9.4	4.4
a vs. e	-11.5	2.2
a vs f	-8.5	6.1
a vs. g	-7.9	5.8
a vs h	0.3	14.0
b vs c	-5.9	7.8
b vs. d	-6.7	7.0
b vs. e	-8.9	4.8
b vs f	-5.7	8.8
b vs. g	-5.3	8.4
b vs h	2.9	16.6
c vs. d	-7.7	6.1
c vs. e	-9.8	3.9
c vs f	-6.7	7.8
c vs. g	-6.2	7.4
c vs h	1.9	15.6
d vs. e	-9.0	4.7
d vs f	-5.9	8.7
d vs. g	-5.4	8.3
d vs h	2.7	16.5
e vs f	-3.7	10.8
e vs. g	-3.3	10.4
e vs h	4.9	18.6
f vs. g	-7.2	7.3
f vs h	1.0	15.5
g vs. h	1.4	15.0

- I.e. homogeneity of 2 or more straight lines where doses are in groups (model I of linear regression), where for modell II see Bofinger (1999)

- This can be generalized to any contrast, e.g. similarity of a step contrast, or a concave contrast,...

5. TAKE HOME MESSAGE

- Similarity of multi-labs dose-response assays can be demonstrated by (1-2α) confidence intervals of specified interaction contrasts
- For unbalanced designs SAS and R software (Bretz et al., 2002) available: both resampling and parametric (can be downloaded from www.bioinf.uni-hannover.de)
- The a-priori or even post hoc definition of margins for equivalence for the differences in absolute terms or relative to SD remains difficult ==> confidence intervals for a ratio to control are more appropriate (% change) ==> further research

References

anonymous (1999). *International methods committee guidelines for validation of qualitative and quantitative microbiological offical methods of analysis*. AOAC

Bofinger E (1999) Homogeneity of two straight lines: Equivalence approach using fitted regressions. *Austral Nz J Statist* 41: 481-491

Bradu, D , Gabriel KR (1974) Simultaneous statistical inference on interactions in 2-way analysis of variance. *JASA* 69: 428-436.

Bretz, F. (1999). *Powerful modifications of William' test on trend*. PhD thesis. University of Hannover.

Bretz, F., Hothorn, T., Westfall, P.H. (2002).Multcomp: Multiple tests and simultaneous confidence intervals. The R-project (see cran.r-project.org)

Gabriel KR (1969) Simultaneous test procedures –some theory of multiple comparisons. *Ann. Math. Statist*. 40, 224-250.

Gabriel, K. R. , Putter, J. , and Wax, Y. (1973), Simultaneous confidence intervals for product-type interaction contrasts, *JRSS B*35 , 234-244

Hochberg Y Tamhane AC (1987) *Multiple comparison procedures*. Wiley NY.

Horwitz W, Britton P, Chirtel SJ. (1998) A simple method for evaluating data from an interlaboratory study. *J AOAC Int* 81,1257-65

Hothorn, L.A.: Statistics of interlaboratory in-vitro toxicological studies. *ATLA* 30 (2002) (in press)

Mandel, J. (1978). Planning, design, and analysis of interlaboratory studies. *Journal ASTM Standardization News* 6, 11-13.

Mukerjee, H., Robertson, T., Wright, F.T. (1986). Multiple contrast tests for testing against a simple tree ordering. In: Dykstra R. ed. *Advances in Order Restricted Statistical Inference*. Berlin: Springer, 203-230.

Schuirmann DJ. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal Pharmacokinetics Biopharmaceutics* 15, 657-680.

Schulte, A. et al. (2002) Two Immunotoxicity Ring Studies According to OECD TG 407 - Comparison of Data on Cyclosporin A and Hexachlorobenzene. (in press)

Westfall, P.H., Tobias, R.D., Rom, D. (1999). *Multiple comparisons and multiple tests*. SAS, Cary USA