Nonclinical Statistics Conference Potsdam 2012

LOAEL Identification by Model Selection Procedures Under Order Restriction

Ludwig A. Hothorn
hothorn@biostat.uni-hannover.de
Institute of Biostatistics, Leibniz University Hannover, Germany
together with R.M. Kuiper, Utrecht Univ. and D. Gerhard LUH

September 2012

The problem I

- **LOAEL**: lowest observed adverse event level
  - identification of a specific dose, i.e. the lowest significant dose against negative control. Similar: MAXSD (maximal safe dose, NOAEL) or MED (minimal effective dose)
  - used in safety assessment, e.g. (environmental) toxicology or AEs in RC-DF-T
  - used in dose-response studies
  - using order restriction (not just a different effect)

- Today, let us focus on a simple one-way layout with $k > 2$ dose groups with $N(\mu_i, \sigma^2)$, such as laboratory endpoints or organ weights

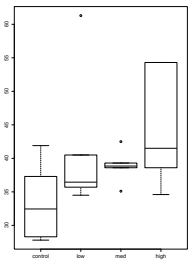
The problem II

- The aim of evaluation of these dose-response studies is manifold:
  - just any heterogeneity between the groups
  - a dose-related trend (global only)
  - the minimal effective dose for an efficacy endpoint
  - the LOAEL for a safety endpoint **today**

- Why order restriction?
  - Ordered alternative: $H_1 : \mu_0 \leq \mu_1 \leq ... \leq \mu_k | \mu_0 < \mu_k$,
  - Either increase the power and/or
  - achieve a specific claim, such as increasing monotone trend, or identification **LOAEL assuming a monotone increase** $\forall D_i \geq D_{LOAEL}$

A motivating example I

- Relative liver weight data of dogs in a chronic toxicity study on Mosapride Citrate (Fitzhugh et al. 1964)- used by [YK01]



- Question: $LOAEL = D_{min}$ *or* D_{med} *or* D_{max}
- Particularly limiting for testing approaches is the small sample size of $n_i = 6$

The crux of LOAEL identification by MCPs I

- [LR92] and [Kod09] criticized LOAEL approach. But be precise!

1. LOAEL is an experimental dose only; no inter/extrapolation possible and therefore $LOAEL > BMD$ commonly

* Yes, but not too serious for continuous data. Examples in [Kod09]:

Example	$BMD_{001}(10)$	LOAEL
HK	585	300 mg/kg
Hb	364	300 mg/kg
Bw	394	300 mg/kg
Neurotoxic	110	100 mg/kg
Hepatocellular carcinoma	260	600 mg/kg

I.e. do not believe in toooo precise estimates at all!
AND: what means $BMD_{001}(10)$ really- see later

The crux of LOAEL identification by MCPs II

2. LOAEL depends on sample size

* Yes, but today the **GORIC** (generalized order restricted information criterion) alternative approach is less depended on sample size. ONE advantage

3. LOAEL does not take monotone dose-response relationship into account

* Serious argument, see the next slides. But **GORIC** approach takes monotone dose-response relationship into account

4. Moreover [Kod09] confusion between LOAEL and NOAEL

* The proof of hazard (LOAEL) and proof of safety (NOAEL) can be perfectly formulated by MCPs [HH08]- not discussed today

5. Proposed alternative: benchmark dose (BMD) [Kod09], [WK05], but see their limitations below

LOAEL identification with order restricted MCPs? I

- Definition:

$$LOAEL = min(\xi \in 1, ..., k : H_1 : \mu_0 = ... = \mu_i < \mu_\xi \leq ... \leq \mu_k) | \mu_0 < \mu_k$$

- Naive approach is stepwise MCP: $k, (k-1), (k-2)..., 1$ trend test, each at level α (a-priori ordered IUT), stop with the first non-significant level i and $\xi = i + 1$
  E.g. using Helmert [Bau97] or reverse Helmert contrasts [Jan05]

- To test one-sided, monotone order restricted $H_1 : \mu_0 \leq \mu_1 \leq ... \leq \mu_k | \mu_0 < \mu_k$, at least two approaches exist:
  i MLE-test acc. to [Bar59] **quadratic test statistics**
  ii MCT **linear test statistics**

LOAEL identification with order restricted MCPs? II

- A **contrast** is a suitable linear combination of means:

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

- Notice, I use 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.

LOAEL identification with order restricted MCPs? III

- 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 the correlation matrix R

- Known examples (balanced design k=2)
- Dunnett one-sided [Dun55]

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

LOAEL identification with order restricted MCPs? IV

- Williams Procedure (as multiple contrast [Bre06])

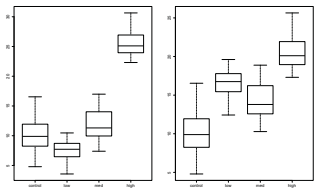
c_i	C	D_1	D_2
c_a	-1	0	1
c_b	-1	1/2	1/2

- Step contrasts [Bau97])

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

LOAEL identification with order restricted MCPs? V

- **All testing approaches using pooling means biased the LOAEL estimation**
- Two counter-examples with data-depended non-monotonicity:



TRUE	-	MED	-	MED	-	-
Helmert [Bau97]	0.99	0.003	0.0001	*0.0001*	**0.11**	0.0001
rev. Helmert[Jan05]	0.99	**0.81**	0.0001	0.0001	0.0026	0.0001
many-to-one [Dun55]	0.99	0.046	0.0001	0.0001	0.0001	0.0001

LOAEL identification with order restricted MCPs? VI

- **Therefore**, all testing approaches which used pooling means (e.g. contrasts or ML-estimations under order restrictions) biased the LOAEL estimation. Only the pairwise Dunnett approach does not:

$$H_1^{\xi=1} : \mu_0 < \mu_1$$

$$H_1^{\xi=2} : \mu_0 < \mu_2 | \ \mu_0 = \mu_1$$

$$H_1^{\xi=3} : \mu_0 < \mu_3 | \ \mu_0 = \mu_1, \mu_0 = \mu_2$$

- Analysis of the example by one-sided Dunnett procedure:

```
                   Estimate   Pr(>t)
low - control == 0    7.450   0.11
med - control == 0    5.500   0.21
high - control == 0  10.767   0.023 *
```

The LOAEL is dose $\xi = 3$ since the adjusted p-value is just below 0.05.

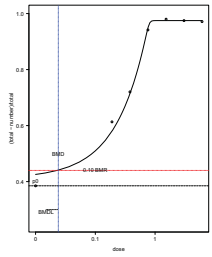
LOAEL identification with order restricted MCPs? VII

- An alternative concept is AIC-based model selection using **ORIC** (order restricted information criterion) [Anr99] and recently **GORIC**[KHS11].
  Without FWER-control but with optimal compromise between likelihood and model complexity

Problems with BMD approach for continous data I

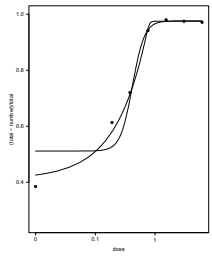
- GORIC in a minute- first a brief discussion: is BMD really THE alternative against LOEAL?
- BMD even for quantal response controversial defined:
  - The spontaneous rate p_0, the x% additional risk BMR (where x $0.01, ...0.1$ which one) (or extra risk definition)
  - The BMD or lower confidence level BMDL?
  - (Using earthworm data in library(drc))

Problems with BMD approach for continous data II



- BMD even for continuous data even more controversial:
  i) risk relative to the variation in the control [CRU94], ii) the risk relative to the control mean [SRE+04], iii) p_0 has to be estimated
- Depends on the underlying model

Problems with BMD approach for continous data III



- Model averaging (Ritz et al. 2012)

Model selection: GORIC as an alternative I

- Model selection via **generalized order-restricted information criterion** [KHS11]: to select the best out of a set of models- which represents alternative hypotheses- in which the population means (μ_i) may be restricted by a mixture of linear equality and inequality constraints.
- GORIC is similar to AIC [Aka73]: trade-off between the **fit of the hypothesis in the data** - the likelihood- and **complexity of the hypothesis** - number of distinct parameters
- GORIC is calculated by

$$GORIC = -2 \log L_m + 2\ PT_m, \tag{1}$$

with $\log L_m$ the (order-restricted) log likelihood and PT_m the penalty (complexity) term for H_m.

Model selection: GORIC as an alternative II

- The order-restricted likelihood is based on order-restricted MLEs $\tilde{\mu}$.
- The penalty term is calculated by

$$PT_m = 1 + \sum_{l=1}^{k+1} LP(k+1, n_0, \ldots, n_k, H_m) \cdot l, \tag{2}$$

where $LP(.)$ is the level probability for hypothesis H_m, which depends on the number of dose levels (k), the number of observations per dose group (n_i), and the restrictions in hypothesis H_m.

- the hypothesis with the lowest GORIC value is the favored one of the set.

Model selection: GORIC as an alternative III

- The GORIC can be applied to hypotheses of the form

$$H_m : R_1\mu \geq 0, R_2\mu = 0, \text{ for } m \in \mathcal{M},$$

a **mixture** of:

- R_1 a $c_{m1} \times k+1$ matrix containing c_{m1} **inequality restrictions** on the $k+1$ means
  and
- R_2 a $c_{m2} \times k+1$ matrix with c_{m2} **equality restrictions** and the unconstrained hypothesis [KHS11].

- Instead of comparing with the model of H_0, here comparison against the unconstrained model (i.e., the hypothesis with no restrictions on the parameters, for example ANOVA-type heterogeneity alternative). I.e. to safeguard for selecting the best of a set of weak hypotheses, the unconstrained model is included

Model selection: GORIC as an alternative IV

- GORIC selects
  1. the correct hypothesis
  2. a similar one
  3. the unconstrained hypothesis
- To improve the interpretation, GORIC weights (w_m) -similar to AIC weights- can be used
- GORIC weight represents the relative likelihood, i.e. the relative support of one hypothesis of interest in comparison to the whole set
- Even better: the ratio of two weight gives the relative support of these two hypotheses, that is, H_m is $w_m/w_{m'}$ more likely than $H_{m'}$

Model selection: GORIC as an alternative V

- Increasing the number of observations does not affect the relative evidence (assuming that the data are still in agreement with the hypotheses)
- Properties:
  - $-2 \log L_m$ is max for unconstrained model
  - $2\ PT_m$ increases with increasing model complexity, $no_= \succ no_>$
  - impact of noncentrality is complex (see simulations)

LOAEL example I

- Decompensation of the global ordered alternative $H_1 : \mu_0 \leq .... \leq \mu_k, \mu_0 < \mu_k$ into all elementary alternatives
- E.g. $k = 3$ doses the global alternative can be decomposed in 7 elementary order restricted alternatives:

$$H_1^{\xi=1,a} : \mu_0 < \mu_1 = \mu_2 = \mu_3$$
$$H_1^{\xi=1,b} : \mu_0 < \mu_1 < \mu_2 = \mu_3$$
$$H_1^{\xi=1,c} : \mu_0 < \mu_1 = \mu_2 < \mu_3$$
$$H_1^{\xi=1,d} : \mu_0 < \mu_1 < \mu_2 < \mu_3$$
$$H_1^{\xi=2,a} : \mu_0 = \mu_1 < \mu_2 = \mu_3$$
$$H_1^{\xi=2,b} : \mu_0 = \mu_1 < \mu_2 < \mu_3$$
$$H_1^{\xi=3,a} : \mu_0 = \mu_1 = \mu_2 < \mu_3$$

- Similar to order restricted MCP [Hot06]

LOAEL example II

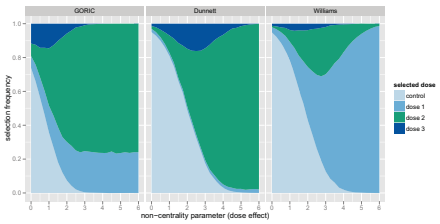
- Analysis of the example: GORIC estimates of the liver weight example

Alternative	logLik	penalty	weight	LOAEL?
$H_1^{\xi=1,a} : \mu_0 < \mu_1 = \mu_2 = \mu_3$	-80.32	2.49	0.177	-
$H_1^{\xi=1,b} : \mu_0 < \mu_1 < \mu_2 = \mu_3$	-80.29	2.81	0.132	-
$H_1^{\xi=1,c} : \mu_0 < \mu_1 = \mu_2 < \mu_3$	-79.51	2.89	0.266	LOAEL
$H_1^{\xi=1,d} : \mu_0 < \mu_1 < \mu_2 < \mu_3$	-79.51	3.03	0.230	-
$H_1^{\xi=2,a} : \mu_0 = \mu_1 < \mu_2 = \mu_3$	-81.95	2.50	0.035	-
$H_1^{\xi=2,b} : \mu_0 = \mu_1 < \mu_2 < \mu_3$	-81.15	2.79	0.057	-
$H_1^{\xi=3,a} : \mu_0 = \mu_1 = \mu_2 < \mu_3$	-81.27	2.50	0.067	-
$H_1^{unconstrained} : \mu_0, \mu_1, \mu_2, \mu_3$	-79.38	5.00	0.037	-

- The ratio to the unconstrained model w_{fm1c}/w_{fmUnc} is 7.2, i.e. the relative support of LOAEL=1 is about 7fold with respect to any heterogeneity.
- Remember: Dunnett LOAEL=3
- GORIC differs from the first *non-testing model selection approach* by Yanagawa(2001) [YK01] by different penalty terms: they use simple AIC, i.e. ignore order restriction

LOAEL example III

- GORIC-LOAEL simulations. E.g. the true profile $0, 0, \delta, \delta$



- Williams procedure is biased (as other to-the-left pooling contrasts)
- Both model selection and Dunnett can correctly identify LOAEL
- But: Dunnett-type-LOAEL depends directly on non-centrality, whereas model selection represents an overlay of two effects which results in a first dependence phase followed by an independence phase
- Two serious advantages of model selection: i) consideration of order restriction, ii) less dependent on non-centrality

Take home message I

- MCT and **model selection** base on similar decomposed elementary alternatives, i.e. mixtures of $= ... =$ and $< ... <$
- MCT and **model selection** can be used for the same aim, e.g. LOAEL, but based on different principles
- Increasing number of models: MCT trade-off between multiplicity penalty and correlation. **Model selection**: free
- MCT: against H_0. **Model selection**: against unconstrained model (more precise: against all models/hypotheses in the set)
- MCT: p-value or sCI. **Model selection**: weight or ratio

References I

[Aka73] Akaike, H.: Budapest: Proc. 2nd Inter. Symposium on Information Theory. In: *Information theory and an extension of the maximum likelihood principle*, 1973, S. 267–281

[Anr99] Anraku, K.: An Information Criterion for Parameters under a Simple Order Restriction. In: *Biometrika* 86 (1999), S. 141–152

[Bar59] Bartholomew, D. J.: A Test Of Homogeneity For Ordered Alternatives .2. In: *Biometrika* 46 (1959), Nr. 3-4, S. 328–335

[Bau97] Bauer, P: A note on multiple testing procedures in dose finding. In: *BIOMETRICS* 53 (1997), SEP, Nr. 3, S. 1125–1128. http://dx.doi.org/{10.2307/2533569}. – DOI 10.2307/2533569. – ISSN 0006–341X

[Bre06] Bretz, F.: An extension of the Williams trend test to general unbalanced linear models. In: *Computational Statistics & Data Analysis* 50 (2006), April, Nr. 7, S. 1735–1748

[CRU94] CRUMP, KS: USE OF MECHANISTIC MODELS TO ESTIMATE LOW-DOSE CANCER RISKS. In: *RISK ANALYSIS* 14 (1994), DEC, Nr. 6, S. 1033–1038. http://dx.doi.org/{10.1111/j.1539-6924.1994.tb00073.x}. – DOI 10.1111/j.1539–6924.1994.tb00073.x. – ISSN 0272–4332

[Dun55] Dunnett, C. W.: A Multiple Comparison Procedure For Comparing Several Treatments With A Control. In: *Journal Of The American Statistical Association* 50 (1955), Nr. 272, S. 1096–1121

[HH08] Hothorn, Ludwig A. ; Hasler, Mario: Proof of hazard and proof of safety in toxicological studies using simultaneous confidence intervals for differences and ratios to control. In: *JOURNAL OF BIOPHARMACEUTICAL STATISTICS* 18 (2008), Nr. 5, S. 915–933. http://dx.doi.org/{10.1080/10543400802287511}. – DOI 10.1080/10543400802287511. – ISSN 1054–3406

[HL09] Hothorn LA, Hothorn T. Vaeth M M. Vaeth M: Trend tests for the evaluation of exposure-response relationships in epidemiological exposure studies. In: *Epidemiologic Perspectives and Innovations* 6 (2009), S. 1–10

[Hot06] Hothorn, Ludwig A.: Multiple comparisons and multiple contrasts in randomized dose-response trials-confidence interval oriented approaches. In: *JOURNAL OF BIOPHARMACEUTICAL STATISTICS* 16 (2006), Nr. 5, S. 711–731. http://dx.doi.org/{10.1080/10543400600860576}. – DOI 10.1080/10543400600860576. – ISSN 1054–3406

Take home message II

- My favorite properties of **model selection**:
  1. unbiased LOAEL estimation taken order restriction into account
  2. direct selection of the true elementary alternative
  3. takes order restriction into account
  4. less dependent on non-centrality
  5. suitable for small sample sizes
- Further examples: change point threshold in exposure epidemiology [HL09], mode of inheritance in genetic association studies, MED,...
- An R package *goric* exists (Gerhard/Kuiper, IBC Kobe 2012 talk)
- Question: is LOAEL- estimated by GORIC- now again an alternative to benchmark dose [LR92]?

References II

[Jan05] Jan, SL: Step contrasts for identifying the minimum effective dose. In: *COMMUNICATIONS IN STATISTICS-THEORY AND METHODS* 34 (2005), Nr. 1, S. 45–57. http://dx.doi.org/{10.1081/STA-200045872}. – DOI 10.1081/STA–200045872. – ISSN 0361–0926

[KHS11] Kuiper, R. M. ; Hoijtink, H. ; Silvapulle, M. J.: An Akaike-type Information Criterion for Model Selection under Inequality Constraints. In: *Biometrika* 98 (2) (2011), 495–501. doi:10.1093/biomet/asr002

[Kod09] Kodell, Ralph L.: Replace the NOAEL and LOAEL with the BMDL(01) and BMDL(10). In: *ENVIRONMENTAL AND ECOLOGICAL STATISTICS* 16 (2009), MAR, Nr. 1, S. 3–12. http://dx.doi.org/{10.1007/s10651-007-0075-3}. – DOI 10.1007/s10651–007–0075–3. – ISSN 1352–8505

[LR92] LEISENRING, W ; RYAN, L: STATISTICAL PROPERTIES OF THE NOAEL. In: *REGULATORY TOXICOLOGY AND PHARMACOLOGY* 15 (1992), APR, Nr. 2, Part 1, S. 161–171. http://dx.doi.org/{10.1016/0273-2300(92)90047-D}. – DOI 10.1016/0273–2300(92)90047–D. – ISSN 0273–2300

[SRE+04] Sand, S ; Rosen, D von ; Eriksson, P ; Fredriksson, A ; Viberg, H ; Victorin, K ; Filipsson, AF: Dose-response modeling and benchmark calculations from spontaneous behavior data on mice neonatally exposed to 2,2 ',4,4 ',5-pentabromodiphenyl ether. In: *TOXICOLOGICAL SCIENCES* 81 (2004), OCT, Nr. 2, S. 491–501. http://dx.doi.org/{10.1093/toxsci/kfh222}. – DOI 10.1093/toxsci/kfh222. – ISSN 1096–6080

[WK05] West, RW ; Kodell, RL: Changepoint alternatives to the NOAEL. In: *JOURNAL OF AGRICULTURAL BIOLOGICAL AND ENVIRONMENTAL STATISTICS* 10 (2005), JUN, Nr. 2, S. 197–211. http://dx.doi.org/{10.1198/108571105X46525}. – DOI 10.1198/108571105X46525. – ISSN 1085–7117

[YK01] Yanagawa, T ; Kikuchi, Y: Statistical issues on the determination of the no-observed-adverse-effect levels in toxicology. In: *ENVIRONMETRICS* 12 (2001), JUN, Nr. 4, S. 319–325. http://dx.doi.org/{10.1002/env.467}. – DOI 10.1002/env.467. – ISSN 1180–4009. – Conference on Environmetrics, VICTORIA FALL, ZIMBABWE, DEC 07-11, 1998