Dose-response analysis in toxicology and epidemiology: considering dose both as qualitative factor and quantitative covariate - using R

Ludwig A. Hothorn
retired Leibniz University Hannover, Germany

Joint work with F. Schaarschmidt and Ch. Ritz

July 6, 2017
Aims for the next 60 mins I

- Why dose-response analysis in biostatistics at all?
- How dose is considered in selected real data examples?
- Multiple contrast tests vs. quasi-linear regression models
- Tukey trend test [TCH85] and its glmm modifications
- The package tukeytrend [F.17]
- Use for undesigned studies in epidemiology
- Four case studies
- Recommendations
Alternative R packages on CRAN

- **drc** nonlinear model fitting, prediction, averaging
  Remember Tuesday’s tutorial by Christian
- **IsoGene** order-restricted inference for multiple endpoints in microarrays by Ziv and colleagues (UseR2017 organizer)
- **DoseFinding, MCPMod, LRcontrast** combining multiple comparisons and nonlinear modeling by Bjoern and colleagues
- **drsmooth, drfit, dosresmeta, mixtox, bmd**
- **ORCME, cir, goric, morse, causaldrf**
- etc.
- Today: `library(tukeytrend)`

Different tools for different dose-response problems available!
Motivating examples I

Four randomly selected Nature papers and a recent epidemiological paper:

- left) strong log; many concentrations; no control; comparing 2
- right) log; zero dose control; displayed qualitatively
Motivating examples II

- left) both qualitative factor and quantitative dose levels
- right) reference drug and 2 quantitative dose levels mixed

![Graph showing relative p53 expression and normalized CC1 density](image-url)
Motivating examples III

- Recent epidemiological study: Risk of chronic kidney disease vs. categorized (tertiles) metabolic analytes [KYY^16]
Motivating examples IV

- Quantitative covariate and/or qualitative factor (even after categorization!)
- Assuming log transformation quite common
- Why log-transformation, and how log(C=0)?
- Both regression model I, i.e. grouped dose levels and model II: randomly chosen dose levels
- Multiple endpoints common, recently high-dim
- ...
Classification I

Basic model behind d-r is Paracelsius: *All things are poison and nothing is without poison. Solely the dose determines that a thing is not a poison.*

Two concepts: i) NOEC in toxicology, ii) (monotone) dose-response as Hill’s causation criterion

- I) Dose as **qualitative factor** ⇒ multiple contrast test (MCT)
- II) Dose as **quantitative covariable** ⇒ lin-log or nonlinear regression model

Choice (I,II) depends on number doses, aim, dose vs. concentration at target cells, ... *No a-priori best approach*

Tukey’s idea [TCH85]:

three dose metameters *arithmetic, ordinal, lin-log* covers most shapes of dose response, is simple and easy to interpret: just a single slope parameter
Classification II

- Main results of recent research:
  1. MCT and Tukey trend test
  2. Extensions within \texttt{glmm}
  3. Extension to regression model II, ie. (indiv.) random dose levels
MCT: Motivation I

- **Dose-response studies** manifold aims:
  - Just any heterogeneity between the groups
  - Comparison vs. control (0) for several effect sizes
    \[ \mu_i - \mu_0, \mu_i / \mu_0, \pi_i / \pi_0, \ldots \]
  - Claiming a global dose-related trend (only)
  - Estimate a particular dose: MED (efficacy), NOEC (safety)

- Why order restriction?
  - Ordered alternative: \( H_1 : \mu_0 \leq \mu_1 \leq \ldots \leq \mu_k \mid \mu_0 < \mu_k \),
  - Increase the power and/or
  - Achieve a **specific claim**, such as increasing monotone trend, or identification MED
  - Interpretation for causation: i) exposure-response relationship in epidemiology against just any significance, ii) **trend** e.g. CA-trend test in genetic association case-control studies

- Notice, trade-off between restriction and robustness
Order restricted tests and related simultaneous confidence intervals

- What means trend?
- Decomposition the monotone $H_1 : \mu_0 \leq \mu_1 \leq \ldots \leq \mu_k$
  into all linear-elementary alternatives; e.g. $k=3$

$H^a_1 : \mu_0 = \mu_1 = \mu_2 < \mu_3$
$H^b_1 : \mu_0 = \mu_1 < \mu_2 < \mu_3$
$H^c_1 : \mu_0 < \mu_1 = \mu_2 < \mu_3$
$H^d_1 : \mu_0 < \mu_1 = \mu_2 = \mu_3$
$H^e_1 : \mu_0 < \mu_1 < \mu_2 = \mu_3$
$H^f_1 : \mu_0 < \mu_1 = \mu_2 < \mu_3$
$H^g_1 : \mu_0 = \mu_1 < \mu_2 = \mu_3$
Order restricted tests and related simultaneous confidence intervals II

- A trend test should be **sensitive against all (most) possible elementary alternatives.** Not against just one, e.g. the linear trend- as the wide-spread used Cochran-Armitage trend test [Arm55] for proportions or the Jonckheere trend test for pairwise ranks ⇒ **crazy**

- At least two approaches:
  i. MLE-test **quadratic test statistics** [Bar59] (not today)
  ii. MCT **linear test statistics**

- **A specific trend test, which compares vs. control:** **Williams trend test** [Wil71] typically in nonclin and clin dose finding studies
Order restricted tests and related simultaneous confidence intervals III

- **MCT I):** A contrast is a suitable linear combination of means (or other effect sizes, e.g. \( \pi_i \)): \( \sum_{i=0}^{k} c_i \bar{x}_i \)
- Here \( i = 0 \ldots k \), focusing on comparisons vs. control (placebo) (more general possible)
- **MCT II):** A contrast test is standardized

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

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

- To guarantee comparable simultaneous confidence intervals is needed: \( \sum \text{sign}^+(c_i) = 1, \sum \text{sign}^-(c_i) = 1 \)
Order restricted tests and related simultaneous confidence intervals IV

- **MCT III):** A multiple contrast test is defined as maximum test:

\[ t_{MCT} = \max(t_1, \ldots, t_q) \]

which follows jointly \((t_1, \ldots, t_q)'\) a \(q\)-variate \(t\)-distribution with degree of freedom \(df\) and the correlation matrix \(R\) depending on \(c_i, n_i\) but also \(s_i, \rho_i, \ldots\). May be complex.

- **MCT IV):** Just the choice of a particular contrast matrix defines the respective MCT.
Order restricted tests and related simultaneous confidence intervals V

Known examples (balanced design k=2)

- Dunnett one-sided [Dun55]

\[
\begin{array}{cccc}
  c_i & C & T_1 & T_2 \\
  c_a & -1 & 0 & 1 \\
  c_b & -1 & -1 & 0 \\
\end{array}
\]

- Tukey all pairs comparisons (two-sided) (Tukey1953)

\[
\begin{array}{cccc}
  c_i & C & T_1 & T_2 \\
  c_a & -1 & 0 & 1 \\
  c_b & -1 & 1 & 0 \\
  c_c & 0 & -1 & 1 \\
  c_d & 1 & -1 & 0 \\
  c_e & 1 & -1 & 0 \\
  c_f & 0 & 1 & -1 \\
\end{array}
\]
Order restricted tests and related simultaneous confidence intervals VI

- Change-point contrast [HYH11]

<table>
<thead>
<tr>
<th>$c_i$</th>
<th>$C$</th>
<th>$D_1$</th>
<th>$D_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_a$</td>
<td>-1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>$c_b$</td>
<td>-0.5</td>
<td>-0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

- Williams Procedure (as multiple contrast [Bre06])

<table>
<thead>
<tr>
<th>$c_i$</th>
<th>$C$</th>
<th>$D_1$</th>
<th>$D_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_a$</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$c_b$</td>
<td>-1</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>

- Much more.... (interesting GrandMean [PH16])

- **MCT V):** One-sided (lower) simultaneous confidence limits:

$$
\left[ \sum_{i=0}^{k} c_i \bar{x}_i - S * t_{q,df,R,2\text{-sided},1-\alpha} \sqrt{\sum_{i}^{k} c_i^2 / n_i} \right]
$$

- Multiplicity-adjusted p-values are available alternatively to simultaneous confidence intervals $\Rightarrow$ compatible!, when....
Modification: Effect sizes I

- **Different effect size**: sCI for $\omega_i = \mu_i/\mu_0$

- **Sasabuchi’s trick of a linear form**
  \[ L(\omega_i) = \sum c_i \bar{Y}_i - d_i \omega_i \bar{Y}_0 \]
  (nominator $c_i$, denominator $d_i$)

- Simultaneous Fieller-type confidence intervals for $\omega_i$
  - solutions of the inequalities
  \[
  T^2(\omega_i) = \frac{L^2(\omega_i)}{S^2_L(\omega_i)} \leq t^2_{q,\nu,R(1-\alpha)},
  \]
  - $t_{q,\nu,R(1-\alpha)}$ is a central $q$-variate $t$-distribution with $\nu$
    degrees of freedom and correlation matrix $R(1) = [\rho_{ij}]$, where
    $\rho_{ij}$ depend on $c_{hi}, n_i$ and on unknown ratios $\omega_i$:
  - plug-in ML-estimators [DBGH04] ⇒ second trick

- The *mratios* R package [DSH07] can be used to make
  inferences about ratios of parameters in mixed models

- Notice: usual log-transformation may be problematic
  [Sch13]
Modification: Variance heterogeneity I

- Variance heterogeneity is quite common, i.e. \( \varepsilon_{ij} \sim N(0, \sigma_i^2) \).
- Standard MCP do not control FWER, particularly when \( n_i \neq n_i' \).
- Modified test statistic \( T^{2*}(\omega_i) = \frac{L(\omega_i)^2}{S_{L(\omega_i)}^2} \), where

\[
S_{L(\omega_i)}^2 = \frac{\omega_i^2}{n_0} S_0^2 + \sum_{h=q+1-i}^{q} \frac{n_h}{\tilde{n}_i^2} S_h^2.
\]

- \( T^*(\omega_i) \) has an approximate \( t \)-distribution with approximate Satterthwaite-type \( \nu_i \).
  Under variance heterogeneity: both \( \nu_i \) and \( R() \) depend on the unknown ratios \( \omega_i \) and the unknown variances \( \sigma_i^2 \).
- Plug-in modification: \( sci.ratioVH \) function in the R package \( mratios \) [HH08]
Modification: Non-parametric 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 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
Modification: Non-parametric II

- Using relative effect size [BM00], [Ryu09]:

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

- \( p_{01} \) is a *win probability* in the sense of [Hay13]

- **sCl**: [KH12b] Let \( R_{ij}^{(0/l)} \) denote the rank of \( X_{ij} \) among all \( n_0 + n_l \) observations within the samples 0 and \( l \)

- The rank means can be used to estimate \( p_{0l} \)

\[ \hat{p}_{0l} = \frac{1}{n_0} \left( \frac{R_{i.}^{(0/l)}}{R_{i.}^{(0/l)}} - \frac{n_l + 1}{2} \right) \]

- Asymptotically \( \sqrt{N}(\hat{p}_1 - p_1, \ldots, \hat{p}_q - p_q)' \) follows a central multivariate normal distribution with expectation 0 and covariance matrix \( V_N \) [KH12a]
- Related approximate \((1 - \alpha)100\%\) one-sided lower simultaneous confidence limits are:

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

- R package \textit{nparcomp} [KPSH15]
Modification: Proportions I

- Three approaches

  1. Wald-type [SSH08a] ⇒ $n_i... \inf$
  2. Add1- adjusted [SBH09] ⇒ small $n_i$
  3. Profile likelihood [Ger16] ⇒ small $n_i$

- 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)
- [AC98] 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 ⇒ small \( n_i \) generalization for MCT [SSH08a]

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

- Choice of simultaneous confidence limits

<table>
<thead>
<tr>
<th>Notation</th>
<th>( \tilde{p}_i )</th>
<th>( \tilde{V}(p_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald</td>
<td>( Y_i/n_i )</td>
<td>( p_i (1 - p_i)/n_i )</td>
</tr>
<tr>
<td>add-1</td>
<td>( (Y_i + 0.5)/(n_i + 1) )</td>
<td>( \tilde{p}_i (1 - \tilde{p}_i)/(n_i + 1) )</td>
</tr>
<tr>
<td>add-2</td>
<td>( (Y_i + 1)/(n_i + 2) )</td>
<td>( \tilde{p}_i (1 - \tilde{p}_i)/(n_i + 2) )</td>
</tr>
</tbody>
</table>
- Comparing survival functions: i) Cox proport. hazards model or ii) the frailty Cox model to allow a joint analysis over sex and strains [HH12]
- Multiple endpoints: UIT(UIT) [HH13]: R package SimComp
- Poly-k estimates [SSH08b]
- ....
Tukey’s trend test- Intro I

- Up to now: dose as a qualitative factor

- Now: considering dose as quantitative covariate
Tukey’s trend test I

- Three decades ago: [TCH85] max-test on three regression models for the arithmetic, ordinal, and linear-log dose metameters of the covariate dose in a randomized one-way layout- without multiplicity adjustment.
- Joint distribution? Problem: \( \mathbb{R} \). Only SAS macro without max-t-test published.
- Tukey’s trend test based on \( \xi \) multiple linear regression models for the \( \xi \) dose transformation functions \( \psi^\xi(D_j) \) for a vector of response variables \( y_{ijk} \) with \( i = 1, \ldots, I \) multiple endpoints in \( j = 0, \ldots, J \) dose levels with \( k_j \) unbalanced replicates

\[
y_{ijk}^\xi = \alpha_i^\xi + \beta_i^\xi(\psi^\xi(D_{jk})) + \epsilon_{ijk}^\xi
\]
Tukey’s trend test II

- A maximum test on the slope parameters $\beta_{i\xi}$ from multiple marginal models for a global null hypothesis is performed

$$H_0 : \beta_{i\xi}(\psi^{\xi}(D_j)) = 0$$

representing a union-intersection test.

- In the **multiple marginal models (mmm)-framework** [PRB12] $\xi$ marginal models for a *univariate* endpoint ($i = 1$) or ($\xi \ast I$) marginal models for $I$ *multiple endpoints* are included.

- From these parameter estimates the correlation matrix is estimated and the test is on the $\xi$ (respective ($\xi \ast I$)) slope parameters $\beta_{i\xi}$.

- Joint distribution of parameter estimates from **multiple marginal models** [PRB12]- without assuming a certain multivariate distribution for the data
Tukey’s trend test III

- Available as function `mms` in library(multcomp)
- Step 1: Derivation of asymptotic representation of parameter estimates from parametric or semi-parametric models using the derivative of the log likelihood (or other estimating functions). Such representations may be combined/stacked across model fits to provide a simultaneous asymptotic representation.
- Step 2: Calculation of the empirical variance-covariance matrix based on the components in the stacked asymptotic representation.
- I.e. empirical covariance based on functions of the data, not the data itself
Tukey’s trend test IV

- By plugging in parameter estimates a consistent sandwich estimator of the variance-covariance matrix is obtained.

- I.e. correlations between different parameter estimates obtained from different model fits to the same data.

- Allows **max-tests on multiple linear models** and estimation of adjusted p-values or simultaneous confidence intervals without the explicit formulation of the correlation matrix in \( \text{lm, glm, lmm} \)- asymptotically.

- For appropriate chosen df \( \nu \), finite versions works well (various simulations).

- Approach is abbreviated by \( \text{mmm} \) now.
Tukey’s trend test V

- Tox example: Bivariate normal endpoints in rats. Trend test for liver and body weight (to avoid relative organ weights)

<table>
<thead>
<tr>
<th>Dose</th>
<th>BodyWt</th>
<th>LiverWt</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>338</td>
<td>11</td>
</tr>
<tr>
<td>0</td>
<td>319</td>
<td>10</td>
</tr>
<tr>
<td>0</td>
<td>369</td>
<td>13</td>
</tr>
<tr>
<td>0</td>
<td>373</td>
<td>13</td>
</tr>
<tr>
<td>0</td>
<td>315</td>
<td>10</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1000</td>
<td>294</td>
<td>9</td>
</tr>
<tr>
<td>1000</td>
<td>294</td>
<td>9</td>
</tr>
<tr>
<td>1000</td>
<td>281</td>
<td>8</td>
</tr>
<tr>
<td>1000</td>
<td>317</td>
<td>9</td>
</tr>
<tr>
<td>1000</td>
<td>292</td>
<td>8</td>
</tr>
</tbody>
</table>

- mmm formulated for slope-to-zero test for three dose metameters and 2 endpoints: six highly correlated tests on slope parameters
Tukey’s trend test VI

▶ elementary code

```r
bN <- lm(LiverWt~DoseN, data=liv)  # arithm
bO <- lm(LiverWt~DoseO, data=liv)  # ordinal
bLL <- lm(LiverWt~DoseLL, data=liv)  # log-lin

lN <- lm(BodyWt~DoseN, data=liv)
lO <- lm(BodyWt~DoseO, data=liv)
lLL <- lm(BodyWt~DoseLL, data=liv)
library("multcomp")
BoLi <- glht(mmm(covarLiv=lN, ordinLiv=lO, linlogLiv=lLL,
                covarBody=bN, ordinBody=bO, linlogBody=bLL),
               mlf(covarLiv="DoseN=0", ordinLiv="DoseO=0", linlogLiv="DoseLL=0",
                   covarBody="DoseN=0", ordinBody="DoseO=0", linlogBody="DoseLL=0"))

▶

library(tukeytrend)

data("liv", package="SiTuR")
fitLl <- lm(LiverWt~Dose, data=liv)
fitLb <- lm(BodyWt~Dose, data=liv)
ttLl <- tukeytrendfit(fitLl, dose="Dose", scaling=c("ari", "ord", "arilog"))
ttLb <- tukeytrendfit(fitLb, dose="Dose", scaling=c("ari", "ord", "arilog"))
cttL <- combtt(ttLl, ttLb)
EXA11<-summary(glht(model=cttL$mmm, linfct=cttL$mlf))
```
Tukey’s trend test VII

- Interpret the adjusted p-values!

<table>
<thead>
<tr>
<th>Model</th>
<th>Test stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>covarLiv: DoseN</td>
<td>-5.38</td>
<td>0.0000001</td>
</tr>
<tr>
<td>ordinLiv: DoseO</td>
<td>-3.98</td>
<td>0.0001253</td>
</tr>
<tr>
<td>linlogLiv: DoseLL</td>
<td>-3.98</td>
<td>0.0002106</td>
</tr>
<tr>
<td>covarBody: DoseN</td>
<td>-6.04</td>
<td>0.0000000</td>
</tr>
<tr>
<td>ordinBody: DoseO</td>
<td>-5.17</td>
<td>0.0000003</td>
</tr>
<tr>
<td>linlogBody: DoseLL</td>
<td>-5.17</td>
<td>0.0000004</td>
</tr>
</tbody>
</table>

Table: Tukey trend test for bivariate normal: body and liver weights

- Trends for both endpoints
- Alternatively, simultaneous confidence intervals for the single parameter slope available- more appropriate for interpretation!
- Remark: nonlinear models try an optimal fit, but need several parameters. Remember Box: All models are wrong, some are useful
Trend test using both a covariate and a factor I

- To assume dose as a **qualitative factor** or a **quantitative covariate** result in quite different- disjoint- approaches: trend tests or non-linear models.

- Common perception: trend test and (non)linear models are completely separate approaches - *not necessarily*.

- **Extension of the Tukey trend test:**
  i) three regression models for the arithmetic, ordinal, and logarithmic-linear dose metameasures \[\text{TCH85}\] **AND** ii) Williams multiple contrast.
Trend test using both a covariate and a factor II

- Example: litter weight data [Hot]

Directed: decreasing weights. No clear trend. A possible dose plateau?
Trend test using both a covariate and a factor III

- 4 marginal models for 6 hypotheses needed:
  - 3 regression models for arithmetic, ordinal and log-linear dose metameters **and** 3 Williams-type multiple contrasts

```
litter$dosen <- as.numeric(as.character(litter$dose))  # add a numeric dose variable
fitc <- lm(weight ~ dosen, data=litter)
dfn<-fitc$df.residual
ttw <- tukeytrendfit(fitc, dose="dosen",
scaling=c("ari", "ord", "arilog", "treat"),ctype="Williams")
exa1<-summary(glht(ttw$mmm, ttw$mlf), df=dfn)
```

<table>
<thead>
<tr>
<th>Dose metameter</th>
<th>Test statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dosenari: dosenari</td>
<td>-0.818</td>
<td>0.727</td>
</tr>
<tr>
<td>dosenord: dosenord</td>
<td>-1.703</td>
<td>0.212</td>
</tr>
<tr>
<td>dosenarilog: dosenarilog</td>
<td>-1.128</td>
<td>0.519</td>
</tr>
<tr>
<td>dosentreat: C 1</td>
<td>-1.863</td>
<td>0.156</td>
</tr>
<tr>
<td>dosentreat: C 2</td>
<td>-2.287</td>
<td>0.062</td>
</tr>
<tr>
<td>dosentreat: C 3</td>
<td>-2.759</td>
<td>0.018</td>
</tr>
</tbody>
</table>

- Look how insensitive any regression model for a plateau shape is!
Trend test using both a covariate and a factor IV

More general:

1. Power of Tukey trend test depends on dose metameters and design (unbalancedness, $k$) and ...

2. A tiny simulation study (log-scaled doses; rather new rank versions in a recent manuscript)

<table>
<thead>
<tr>
<th>shape</th>
<th>Williams</th>
<th>Tukey</th>
<th>TukeyWil</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose</td>
<td>quali</td>
<td>quanti</td>
<td>both</td>
</tr>
<tr>
<td>linear</td>
<td>0.85</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>plateau</td>
<td>0.95</td>
<td>0.76</td>
<td>0.87</td>
</tr>
<tr>
<td>sublinear</td>
<td>0.81</td>
<td>0.96</td>
<td>0.89</td>
</tr>
</tbody>
</table>

3. Serious power loss for plateau profiles when dose used as quantitative covariate

4. TukeyWilliams max-test: no serious power loss for any shape of dose response. Robustification!

5. TukeyWilliams max-test: interpreting covariate vs. factor (or pairwise comparison $Cvs.D_{max}$)

6. Can be formulated for non-monotonic trends (down-turn effects)
Generalizations of Tukey trend test I

1. Modification for variance heterogeneity using sandwich estimator of var-cov matrix ⇒ vignette
2. Extension to specific monotonic or non-monotonic dose-response relationships ⇒ vignette
3. Extension to multiple endpoints: normal, binary, multinomial, different-scaled ⇒ vignette
4. Modification for different arithmetic-logarithmic scores ⇒ vignette
5. Inclusion of the control vs. high dose comparison [AXL11]
Generalizations of Tukey trend test II

6. Adjustment against multiple covariates ⇒ see the case study here and the vignette
7. GLM: Proportions, Poisson (overdispersed) ⇒ vignette
8. Rank regression: both normal or rank ⇒ paper soon
9. Meta analysis in exposure epidemiology ⇒ vignette
10. Mixed effects model: repeated measures ⇒ see the case study here and the vignette
11. Multiple log-transformations for $D = 0$
12. ...
Mixed effects model: technical replicates

- The analysis of designs with **technical replicates** is performed often naive. But mixed model(s) are available
- Nonclin example: Per-litter data as natural replicates
- The jittered boxplots show rather different litter sizes
Mixed effects model: technical replicates II

![Graph showing pup weight distribution by sex and technical replicates.](image-url)
Using the `library(lme4)` and the Ritz function `lmer2lm` (in a recent Biometrics paper)

```r
library(lme4)
m4N <- lmer(weight ~ DoseN + sex+adjLitsize+(1|litter), data = Ratpup)
m4O <- lmer(weight ~ DoseO + sex+adjLitsize+(1|litter), data = Ratpup)
m4LL <- lmer(weight ~ DoseLL + sex+adjLitsize+(1|litter), data = Ratpup)

# De-correlating
d4N <- lmer2lm(m4N)
d4O <- lmer2lm(m4O)
d4LL <- lmer2lm(m4LL)
mixi <- glht(mmm(covariate=d4N, ordinal=d4O, linlog=d4LL),
             mlf(covariate="XDoseN=0", ordinal="XDoseO=0", linlog="XDoseLL=0"))
```

<table>
<thead>
<tr>
<th>Model</th>
<th>Test stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>covariate: XDoseN</td>
<td>-4.8388853</td>
<td>0.0000013</td>
</tr>
<tr>
<td>ordinal: XDoseO</td>
<td>-4.9169771</td>
<td>0.0000009</td>
</tr>
<tr>
<td>linlog: XDoseLL</td>
<td>-4.8388853</td>
<td>0.0000013</td>
</tr>
</tbody>
</table>

**Table:** Tukey trend test for mixed effects model
Considering of different covariance-adjusting models

- Interesting: simultaneous consideration of multiple models with different formulations for the adjusting covariate

- Nonclin example: Analysis of organ weights in toxicology. It is a priori not clear how to consider the body weight: i) as relative organ weight, ii) body weight as covariate or iii) ignoring body weight (quite different biological background!)

- Liver weights from a 13-week study on female F344 rats administered with sodium dichromate dihydrate [NTP]

```r
data("liv", package="SiTuR")
liv$relLiv <- liv$LiverWt/liv$BodyWt
LIVmod1 <- lm(LiverWt~Dose, data=liv)
LIVmod2 <- lm(relLiv~Dose, data=liv)
LIVmod3 <- lm(LiverWt~Dose+BodyWt, data=liv)

tt1 <- tukeytrendfit(LIVmod1, dose="Dose", scaling=c("ari", "ord"), tt2 <- tukeytrendfit(LIVmod2, dose="Dose", scaling=c("ari", "ord"), tt3 <- tukeytrendfit(LIVmod3, dose="Dose", scaling=c("ari", "ord"), cttC <- combtt(tt1, tt2, tt3)
LIVExa4 <- summary(glht(model=cttC$mmm, linfct=cttC$mlf))
```
<table>
<thead>
<tr>
<th>Model</th>
<th>Test stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>tt1.lm.LiverWt.Doseari: Doseari</td>
<td>-6.0358901</td>
<td>0.00000000</td>
</tr>
<tr>
<td>tt1.lm.LiverWt.Doseord: Doseord</td>
<td>-5.1678800</td>
<td>0.00000006</td>
</tr>
<tr>
<td>tt1.lm.LiverWt.Dosearilog: Dosearilog</td>
<td>-5.1678800</td>
<td>0.00000005</td>
</tr>
<tr>
<td>tt2.lm.relLiv.Doseari: Doseari</td>
<td>-4.7739259</td>
<td>0.00000049</td>
</tr>
<tr>
<td>tt2.lm.relLiv.Doseord: Doseord</td>
<td>-4.6579781</td>
<td>0.00000076</td>
</tr>
<tr>
<td>tt2.lm.relLiv.Dosearilog: Dosearilog</td>
<td>-4.6579781</td>
<td>0.00000091</td>
</tr>
<tr>
<td>tt3.lm.LiverWt.Doseari: Doseari</td>
<td>-2.4553875</td>
<td>0.0504257</td>
</tr>
<tr>
<td>tt3.lm.LiverWt.Doseord: Doseord</td>
<td>-2.9008284</td>
<td>0.0147290</td>
</tr>
<tr>
<td>tt3.lm.LiverWt.Dosearilog: Dosearilog</td>
<td>-2.9008284</td>
<td>0.0146386</td>
</tr>
</tbody>
</table>

- We pay a tiny penalty to consider several, similar models because they are highly correlated.
- We achieve: the best model. Here: just liver weight.
- Hard to solve this problem with model selection approaches (different models, but also different endpoints).
- Can be used e.g. in clinical Phase II dose findings studies with baseline values.
Multiple different-scaled endpoints I

- Rather common are multiple, but different-scaled endpoints, such as (normal and binomial), or (binomial and time-to-event).
- Adequate analysis of multiple endpoints is already a challenge, but for different-scaled endpoints to estimate the correlations may be problematic.
- Nonclin example. In the litter weight example (normal distributed), the number of litter mates can be considered as a second endpoint, Poisson distributed allowing extravariability between the litter.
Multiple different-scaled endpoints II

- Using function `combmep`

```r
data(litter, package="multcomp")
dl <- litter
dl$dosen <- as.numeric(as.character(dl$dose))
fitw <- lm(weight ~ dosen + gesttime, data=dl)  # lm-model
ttw <- tukeytrendfit(fitw, dose="dosen", scaling=c("ari", "ord", "arilog"))
fitqp <- glm(number~dosen + gesttime, data=dl, family=quasipoisson)
ttqp <- tukeytrendfit(fitqp, dose="dosen", scaling=c("ari", "ord", "arilog"))
cttqw <- combtt(ttqp, ttw)  # combine both models
Exa12 <- summary(glht(model=cttqw$mmm, linfct=cttqw$mlf))
```

<table>
<thead>
<tr>
<th>Model</th>
<th>Test stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ttqp.glm.number.dosenari: dosenari</td>
<td>-1.4378208</td>
<td>0.3707893</td>
</tr>
<tr>
<td>2 ttqp.glm.number.dosenord: dosenord</td>
<td>-0.3176185</td>
<td>0.9898792</td>
</tr>
<tr>
<td>3 ttqp.glm.number.dosenarilog: dosenarilog</td>
<td>-0.4540899</td>
<td>0.9604835</td>
</tr>
<tr>
<td>4 ttw.lm.weight.dosenari: dosenari</td>
<td>-1.1046626</td>
<td>0.5847639</td>
</tr>
<tr>
<td>5 ttw.lm.weight.dosenord: dosenord</td>
<td>-1.7556537</td>
<td>0.2136334</td>
</tr>
<tr>
<td>6 ttw.lm.weight.dosenarilog: dosenarilog</td>
<td>-1.1235982</td>
<td>0.5719765</td>
</tr>
</tbody>
</table>

- ...
Multinomial endpoint I

- Comparison of multinomial vectors represents a rather specific problem, such as differential blood count (Wald-type intervals [SGV16]).
- The problem is even more complicated when overdispersion may occur (recent research at LUH).
- Nonclin example: in a reproductive toxicity experiment $n_{ij}$ females are treated within the dose groups (and zero-dose control) and the health status of each pup within a single female is classified into unaffected, malformed or death.
- Up to now approaches for simultaneous inference for overdispersed multinomial vectors seems to be not available.
- **Approach I**: splitting into the 3 proportions $p_d = \frac{x_{death}}{x_{unaffected}}, \ldots, \ldots,$ followed by a trend test for correlated overdispersed binary endpoints.
Multinomial endpoint II

Figure: Multiple overdispersed proportions
Multinomial endpoint III

MmN <- glm(cbind(NResp.1, ClusterSize-NResp.1) ~ Dose, data=Dehp, family=quasibinomial(link="logit"))
DmN <- glm(cbind(NResp.2, ClusterSize-NResp.2) ~ Dose, data=Dehp, family=quasibinomial(link="logit"))
NmN <- glm(cbind(NResp.3, ClusterSize-NResp.3) ~ Dose, data=Dehp, family=quasibinomial(link="logit"))
tMmN <- tukeytrendfit(MmN, dose="Dose", scaling=c("ari", "ord", "arilog"))
tDmN <- tukeytrendfit(DmN, dose="Dose", scaling=c("ari", "ord", "arilog"))
tNmN <- tukeytrendfit(NmN, dose="Dose", scaling=c("ari", "ord", "arilog"))
ctMDN <- combtt(tMmN, tDmN, tNmN)
Exa18 <- summary(glht(model=ctMDN$mmm, linfct=ctMDN$mlf))
library("ggplot2")
EXA18 <- fortify(summary(Exa18))[, c(1, 5, 6)]
colnames(EXA18) <- c("Model", "Test stats", "p-value")
library("xtable")
print(xtable(EXA18, digits=7, caption="Combining splitted overdispersed multinomials into multiple binomials", label="tab:exa18"), include.rownames=FALSE)
Multinomial endpoint IV

<table>
<thead>
<tr>
<th>Model</th>
<th>Test stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>tMmN.glm.cbind(NResp.1, Doseari: Doseari)</td>
<td>5.0416539</td>
<td>0.0000009</td>
</tr>
<tr>
<td>tMmN.glm.cbind(NResp.1, Doseord: Doseord)</td>
<td>5.3329760</td>
<td>0.0000002</td>
</tr>
<tr>
<td>tMmN.glm.cbind(NResp.1, Dosearilog: Dosearilog)</td>
<td>5.4341268</td>
<td>0.0000001</td>
</tr>
<tr>
<td>tDmN.glm.cbind(NResp.2, Doseari: Doseari)</td>
<td>8.8977342</td>
<td>0.0000000</td>
</tr>
<tr>
<td>tDmN.glm.cbind(NResp.2, Doseord: Doseord)</td>
<td>7.6229984</td>
<td>0.0000000</td>
</tr>
<tr>
<td>tDmN.glm.cbind(NResp.2, Dosearilog: Dosearilog)</td>
<td>7.2209305</td>
<td>0.0000000</td>
</tr>
<tr>
<td>tNmN.glm.cbind(NResp.3, Doseari: Doseari)</td>
<td>-10.3417601</td>
<td>0.0000000</td>
</tr>
<tr>
<td>tNmN.glm.cbind(NResp.3, Doseord: Doseord)</td>
<td>-9.1127782</td>
<td>0.0000000</td>
</tr>
<tr>
<td>tNmN.glm.cbind(NResp.3, Dosearilog: Dosearilog)</td>
<td>-8.7953491</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

Table: Combining splitted overdispersed multinomials into multiple binomials

- Approach II Using VGAM
- Soon more general approaches with R functions (Schaarschmidt and Vogel, 2017)
Trends in epidemiology: individual random dose levels

Cross-sectional epidemiological example: association between age and metabolomic analytes on almost 300 volunteers [FBH^+16]- just a single analyte
A similar max-test alternative using nonlinear models is available in library(LRcontrast) [DTVB15, GB17]

Simulation snippet: $n_i = 10$, 1000 runs

**Table:** Some simulation results

<table>
<thead>
<tr>
<th>Hyp.</th>
<th>Shape</th>
<th>max(lin,loglin,ord)</th>
<th>max(lin,loglin,$E_{max}$,exp)</th>
<th>categor.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$-</td>
<td>-</td>
<td>0.058</td>
<td>0.057</td>
<td>0.060</td>
</tr>
<tr>
<td>$H_A$ lin</td>
<td>0.39</td>
<td>0.43</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>log-lin</td>
<td>0.78</td>
<td>0.77</td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td>$E_{max}$</td>
<td>0.69</td>
<td>0.72</td>
<td></td>
<td>0.58</td>
</tr>
</tbody>
</table>

Both approaches can be used (no umpt), whereas categorization is not a good idea. Tukey-test excels by its easy interpretability and generalizability.
Further general mmm applications I

- Multiple normal distributed endpoints in multi-arm trials: Dunnett-type sCI
- Multiple binary endpoints
- Multiple binary endpoints in multi-arm trials: Williams-type trend test
- Multiple regression models in genetic association test [ARH17]
- Composite binary endpoints [MRH16]
- Subgroup analysis with claim for total, targeted and complementary populations (soon Vogel et al.)
- Inference on dose (randomized) and time (dependent) [PPR15]
Take home message I

- Trend tests for several possible shapes are available with in glmm framework → rather relevant for nonclin and clin trials and epidemiological studies
- TukeyWilliams trend test can be recommended: R library available, simple interpretation
- Extension to model II of regression
- Flexibility is amazing: see vignette
- Use confidence intervals! Interpret the effect size first, for the simple slopes
- Comparison with nonlinear models (incl. model averaging resp. model selection) needed
- Basic property of mmm: no explicit formulation of $\mathbb{R}$
- The Tukey trend approach follows KISS


References II


[NTP] *National Toxicology Program. 13 Weeks gavage study on female F344 rats administered with Sodium dichromate dihydrate (VI) (CASRN: 7789-12-0, Study Number: C20114, TDMS Number:2011402.*
References III


