Nonclinical Statistics Conference Potsdam 2012

Statistical planning and analysis of the HET-MN Assay for genotoxicity testing

Ralph Pirow[1], Dagmar Fieblinger[1], Manfred Liebsch[1], Albrecht Poth[2], Kerstin Reisinger[3], Thorsten Wolf[4], Daniel Gerhard[5], Ludwig A. Hothorn[5]

[1] *Bundesinstitut für Risikobewertung,* [2] *Harlan CCR,* [3] *Henkel AG & Co KGaA,* [4] *Universität Osnabrück,* [5] *Leibniz Universität Hannover*

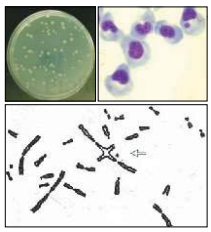
26. September 2012

Introduction

Alternative methods to animal testing play an important role in risk assessment of chemicals.

Legislative background

- EU Chemicals Regulation (REACH)
- EU Cosmetics Regulation 1223/2009 (→ complete ban on animal testing)



***in-vitro* Test battery for Genotoxicity**

- Bacterial mutation assay (Ames)
- Mammalian cell gene mutation assay
- Mammalian cell micronucleus[#] assay
  (# or chromosomal aberration) [Pfu10]

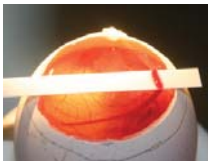
If positive, **follow-up testing** required!

The HET-MN: A new method for genotoxicity testing

The problem of *in-vitro* genotoxicity testing is the high number of **false positives** (75–95%) [Kir05].

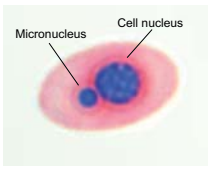
Optimization efforts (EU, OECD, COLIPA)

Development of new methods



The HET-MN

- = **H**en's **E**gg **T**est for **M**icro**N**uclei induction
- developed at the Univ. of Osnabrück [Wol08]
- detects clastogenic & aneugenic effects
- covers important toxicological processes (metabolic activation, elimination, excretion)



micronucleated erythrocyte

Pre-validation of the HET-MN in a ring trial

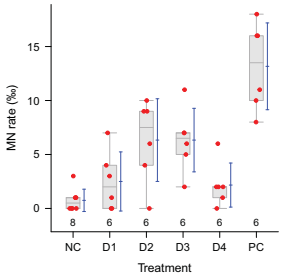
. . . to evaluate the transferability, reproducibility & predictivity.

Partner labs		Specific tasks
BfR	Federal Institute for Risk Assessment ZEBET	Statistical analysis Validation support
harlan	Harlan Cytotest Cell Research GmbH	
Henkel	Henkel AG & Co. KGaA	Coordination
Universität Osnabrück	University Osnabrück	Consulting

funded by the Federal Ministry of Education and Research (BMBF)

Experimental design of the HET-MN

- One-way layout, including 3–4 doses of the test compound, and a negative (NC) and positive control (PC)
- The egg is the randomized unit
- 6 eggs are used per group
- 1000 cells are scored per egg for the occurrence of micronuclei (MN)
- The number of MN can be considered as a count since the number of scored cells is constant and the incident rates are low



Previous approach to analyse the HET-MN

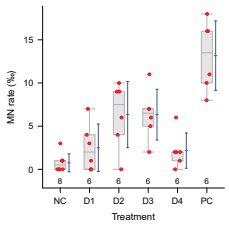
Proposal in the tox literature [Wol08]:
exact and asymptotic Wilcoxon tests against the concurrent and historical controls: $min(p^{exact.Wilcoxon}_{D_i vs.NC}, p^{asympt.Wilcoxon}_{D_i vs.histNC})$.

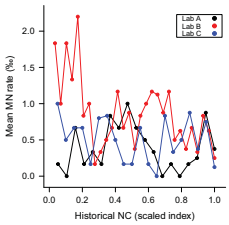
Problems:

- no control of the family-wise error rate (FWER)
- $p^{asympt.Wilcoxon}_{D_i vs.histNC}$ monotone in n_{histNC}
- no confidence intervals

Statistical challenges in the HET-MN data

1) **Small sample size**
(problem for asymptotic approaches)
2) **Between-egg heterogeneity**
(overdispersion leads to liberal decisions)
→ quasi-binomial/poisson methods
3) **Increased mean/variance in higher doses**
→ robustness against variance heterogeneity
4) **Downturn effects at high doses**
(problem for methods assuming monotonicity)
→ protection against downturn effects
5) **Near-to-zero counts in NC**
(can lead to unstable/biased results)
→ conditional use of historical NCs





Problems in other statistical approaches

- U.S. NTP recommendation for proof-of-hazard analysis:
... continuous variables ... with the parametric multiple comparison procedures of Dunnett [Dun55] and Williams [Wil71]

- Dunnett/Williams-type MCP for overdispersed count data are asymptotically possible in the GLM [Hot08]:

```
library(multcomp)
f1<-glm(y ~ DOSE, data=HMN, family=quasipoisson(link="log"))
summary(glht(f1, linfct = mcp(DOSE = "Dunnett")))
```

- However, this approach
  i) has inadaquate asymptotic properties for small sample sizes (n=6) AND count data with overdispersion, and
  ii) is possibly intransparent for toxicologists

Problems solutions I

- **Transformation** of count data into pseudo-normally distributed data is common in toxicology
- For overdispersed near-to-zero counts, the **Freeman-Tukey root transformation** [FT50] can be recommended [Gua09]:

  $x_{ij} = \sqrt{x_{ij}} + \sqrt{x_{ij} + 1}$

  Further adjustments for heterogeneous variances, e. g. by using the sandwich estimator [Her10]), are not needed.
- The Dunnett/Williams-type MCP for the transformed endpoint can be formulated as **multiple contrast tests (MCT)**:

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

$$\text{where} \quad t_{MCT} = max(t_1, \ldots, t_q)$$

is jointly $(t_1, \ldots, t_q)'$ q-variate t-distributed with common df and correlation matrix R, with $R = f(c_i, n_i)$ only.

Problems solutions II

The contrast matrices for a balanced design with two doses are:

Dunnett procedure one-sided
[Dun55]

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

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

Williams procedure downturn-protected
[Hot04]

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

Williams-type procedure using historical controls

Toxicological endpoints such as MN induction represent the outcome of a specific pathological process. They are:

- counts or proportions,
- inherently increasing,
- and tend to be zero or near-to-zero (n-t-z) in NC.

The test sensitivity depends seriously on the number of zeros or n-t-z values. Whether 0 or 1 tumor occurs in 50 ctrl animals has an impact on the p-value.

Approaches using historical controls are available [TAR82, Din11, Kit12] but are rarely used in practice, because

- rather complex for toxicologists,
- unstable for $n_{HC} < 10$
- and do not follow US-FDA recommendation:

  The concurrent control group is always the most appropriate and important in testing drug related increases in tumor rates ... as long as the concurrent control data are within the range of historical control data [FDA01].

Conditional two-step approach using historical controls

- First, check whether the concurrent control data **are within** the range of historical controls, **or not**

  Naive 2σ intervals [Nel03] for FT-transformed variables can be recommended ($\rightarrow$ trade-off between simplicity and validity [AB11])
- When **within**, use the *common* Williams-type approach against the concurrent control
- When **outside**, use a *modified* Williams-type approach against the arithmetic mean of the control assays ϑ (not mean of all individual controls) (Jaki, Kitsche, Hothorn submitted)

$$t^{\text{vs. Standard, normal distr.}}_{\text{Contrast}} = (\sum_{i=1}^{k} c_{li} \bar{x}_i - \vartheta) / S_{i=1,\ldots,k} \sqrt{\sum_{i=1}^{k} c_{li}^2 / n_i}$$

- This approach can be easily realized by means of the parameter `rhs` in the function `glht` of the R package `multcomp`.

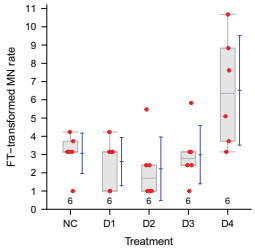
A real data example using historical controls

- $\bar{x}_{NC} = 3.07$ is outside the normal range of the historical controls $[1.05, 2.5]$
- Therefore, compare against the arithmetic mean of *historical* controls, not against the *concurrent* controls to avoid false negative decisions (i. e. too large *p*-values).

  Results for the Williams-type procedure:

Contrast	pValConcurrent	pValHist
C1	0.0036	0.0000
C2	0.0744	0.0000
C3	0.2678	0.0003
C4	0.3942	0.0005

- Signif. trend for contrast C1 (D_{max} *vs.* NC) for both comparisons against concurrent and historical controls, whereas the p-value for the comparison against the historical control is much smaller.



Thank you for your attention

Take home message

- MN counts can be analysed after Freeman-Tukey root transformation: approx. normal and variance homogeneous
- Williams-type procedure against mean of historical controls proposed: simple, independent on n_{HC}, and stable for $n_{HC} < 10$; although it ignores between-assay-variability
- Easy-to-use by R code is available

References I

[AB11] AEBTARM, Surath ; BOUGUILA, Nizar:
An empirical evaluation of attribute control charts for monitoring defects.
In: *EXPERT SYSTEMS WITH APPLICATIONS* 38 (2011), JUN, Nr. 6, S. 7869–7880.

[Bre06] BRETZ, Frank:
An Extension of the Williams Trend Test to General Unbalanced Linear Models.
In: *Comput. Stat. Data An.* 50 (2006), Nr. 7, S. 1735–1748

[Din11] DINSE, Gregg E. ; PEDDADA, Shyamal D.:
Comparing Tumor Rates in Current and Historical Control Groups in Rodent Cancer Bioassays.
In: *STATISTICS IN BIOPHARMACEUTICAL RESEARCH* 3 (2011), FEB, Nr. 1, S. 97–105.

[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

[FDA01] ANONYMOUS:
Guidance for Industry: Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals / U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research.
2001. –
Forschungsbericht

[FT50] FREEMAN, MF ; TUKEY, JW:
TRANSFORMATIONS RELATED TO THE ANGULAR AND THE SQUARE ROOT.
In: *ANNALS OF MATHEMATICAL STATISTICS* 21 (1950), Nr. 4, S. 607–611.

[Gua09] GUAN, Yu:
Variance stabilizing transformations of Poisson, binomial and negative binomial distributions.
In: *STATISTICS & PROBABILITY LETTERS* 79 (2009), JUL 15, Nr. 14, S. 1621–1629.

[Her10] HERBERICH, E. ; SIKORSKI, J. ; HOTHORN, T.:
A Robust Procedure for Comparing Multiple Means under Heteroscedasticity in Unbalanced Designs.
In: *Plos One* 5 (2010), März, Nr. 3, S. e9788

References II

[Hof12] HOFFMANN, S et a.:
Two new approaches to improve the analysis of BALB/c 3T3 cell transformation assay data.
In: *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 744 (2012), S. 36–41

[Hot08] HOTHORN, Torsten ; BRETZ, Frank ; WESTFALL, Peter:
Simultaneous Inference in General Parametric Models.
In: *Biometrical J.* 50 (2008), Nr. 3, S. 346–363

[Hot04] HOTHORN, LA:
A robust statistical procedure for evaluating genotoxicity data.
In: *ENVIRONMETRICS* 15 (2004), SEP, Nr. 6, S. 635–641.

[Kir05] Kirkland D, Aardema M, Henderson L, Müller L, 2005a. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. *Mutat Res* 584, 1–256 (Erratum in: *Mutat Res* 588, 70).

[Kit12] KITSCHE:
The use of historical controls in estimation simultaneous confidence intervals for comparisons against a concurrent control.
In: *CSDA* 56 (2012), S. 3865–3875.

[Nel03] NELSON, LS:
When should the limits on a Shewhart control chart be other than a center line +/- 3-sigma?
In: *JOURNAL OF QUALITY TECHNOLOGY* 35 (2003), OCT, Nr. 4, S. 424–425. –

[Pfu10] Pfuhler S *et al* (2010). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: Genotoxicity. A COLIPA analysis. *Regul Toxicol Pharmacol* 57, 315-324.

[Reif12] REIFFERSCHEID, G. et al.:
International round-robin study on the Ames fluctuation test.
In: *ENVIRONMENTAL AND MOLECULAR MUTAGENESIS* 53 (2012), APR, Nr. 3, S. 185–197.

[TAR82] TARONE, RE:
THE USE OF HISTORICAL CONTROL INFORMATION IN TESTING FOR A TREND IN POISSON MEANS.
In: *BIOMETRICS* 38 (1982), Nr. 2, S. 457–462.

References III

[Wil71] WILLIAMS, D A.:
A Test for Differences Between Treatment Means When Several Dose Levels are Compared with a Zero Dose Control.
In: *Biometrics* 27 (1971), Nr. 1, S. 103–117

[Wol08] Wolf T, Niehaus-Rolf C, Banduhn N, Eschrich D, Scheel J, Luepke NP (2008). The hen's egg test for micronucleus induction (HET-MN): Novel analyses with a series of well-characterized substances support the further evaluation of the test system. *Mutat Res* 650, 150–164.