

Gottfried Wilhelm Leibniz Universität Hannover,
Institut für Biostatistik, Herrenhäuser Str. 2, 30419

Naturwissenschaftliche
Fakultät

Prof. Dr. L. A. Hothorn

Tel. +49 511 762 4998

Fax +49 511 762 4966

E-Mail:

hothorn@biostat.uni-
hannover.de

24.05.2010

**Answers on the questions from
May 20, 2010 webinar
Statistical Analysis of Short-
term Studies in Regulatory
Toxicology using R**

Q1: Often time, the obstacle we have when practice this is that our tox partners do not know what is are the acceptable range for the tox effect for some new tox studies. Your input?

A1: The a-priori endpoint-specific definition of acceptable ranges in regulatory tox (for either as relevance criteria in the proof of hazard or tolerable thresholds in the proof of safety) is important but hard to do in practice. Lab-specific reference values based on historical controls are useful for definition of such acceptable ranges (but a clear approach is still not really available)

Q2: Why would one not consider strictly controlling the false negative rate?

A2: Personally I argue for a proof of safety and hence a direct control of f- rate in regulatory toxicology. However 99.9% of decision making is based on the proof of hazard today and such a paradigm change takes time and is controversial. For example, a proof of safety for tumor-to-time relationships in a long-term carcinogenicity study with a maximum f- rate of say 10% would cause particularly large f+ rates >>50% for common tumors (i.e. spontaneous rates >5%) when tolerating less than tumor doubling rate.

Q3: Is issue 4 still under proof of hazard context? Should multiplicity be corrected on CI for proof of safety?

A3: Issue 4 discusses rather briefly the multiplicity adjustment for multiple endpoints in tox. In the proof of hazard, an UIT problem arises and a multiplicity adjustment would increase f- rates seriously. Therefore I argue NOT to adjust. In the proof of safety an IUT problem arises, i.e. global safety hold true if all endpoints are safe. This approach is extremely conservative for e.g. > 100 endpoints, and therefore I argue NOT for global safety (see a related discussion in Hothorn, LA; Oberdörfer, R: Statistical analysis used in the nutritional assessment of novel food using the proof of safety. Regul Toxicol. Pharmacol 44 (2006), 125-135)

Besucheradresse:
Herrenhäuser Str. 2
30419 Hannover
www. biostat.uni-
hannover.de

Q4: Is one-sided test evaluated at 0.025 rather than 0.05 here?

A4: No, classical definition: two-sided AND one-sided on ONE f+ rate of 0.05 (i.e. the both elementary directional hypotheses for two-sided hypotheses will be tested at 0.025 each)

Q5: Are these packages applicable for multivariate (repeated measure) analysis of variance and covariance?

A5: Yes for the package multcomp. But using a mixed model with several correlation structures for not necessarily complete repeated measures as a random factor when for the fixed factor DOSE simultaneous CI are estimated is now under research.

Q6: How do you evaluate effect size and its CI when non-parametric approach is used?

A6: At least two important effect sizes in non-parametric approach are important: i) difference of distribution functions (using by Kruskal-Wallis-, Steel-, Dunn-test) or ii) relative effect size (Brunner-Munzel-test). Only for ii) we are able to estimate confidence intervals in the k-sample one-way layout

Q7: Are stepwise MCP the same as gateway procedures?

A7: No. Gatekeeping procedures are conditional IUT stepwise procedures. I discussed stepwise MCP as stepwise UITs

Q8: Do you have a non-parametric version of repeated measure anova with covariate to recommend?

A8: No yet. But in May 2010 a three years project sponsored by the German Sci Foundation started together with the University of Goettingen started and this is ONE topic.

Q9: Can the methods be used on repeated measures data?

A9: See Answer 5

Q10: Can you give some more argument why you do use multiple comparison methods in the first place after saying that one option to avoid/reduce false negatives would simply be to not correct for multiplicity at all? Is it solely because of the NTP guideline?

A10: Yes. Only the U.S. NTP recommends details for stats evaluation, i.e. Dunnett/Williams respective Dunn/Shirley. We should be consequent: when multiplicity adjustment for dose group comparisons, than also multiplicity adjustment for multiple endpoints as well. But an extreme conservative approach results, i.e. we would control f+ for BOTH families (doses and endpoints) in a strict sense with the consequence of extreme high f- rates. Therefore my consequence is: no multiplicity adjustment at all: neither for dose comparisons nor for endpoints (in the proof of hazard).

Q11: Can SAS be used to perform the various tests?

A11: Yes, e.g. in PROC MIXED- but the various R libraries offer much higher functionality

Q12: For looking at proportions, are you not looking at overdispersion issues?

A12: If a certain histopathological finding occur in an animal or not, no overdispersion occur, just a 2 by k table. However, if within a female the individual pups are malformed or not, clearly a overdispersion problem occur. We are working on simultaneous CI when overdispersion (or extra-Poisson variability) occur

Yours sincerely
Mit freundlichen Grüßen

A handwritten signature in blue ink that reads "Rudolph Hothorn". The signature is written in a cursive, flowing style.

(Prof. Dr. L. Hothorn)