

Nonclinical Statistics Conference Padova 2026

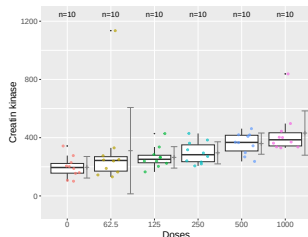
Comparisons of several treatments
against a control in regulatory toxicology:
**an alternative to the common-used
Dunnett test**

Ludwig A. Hothorn
retired Leibniz University Hannover, Germany

September 8, 2026

Is statistics in regulatory toxicology an old bone? I

- OECD -Guideline No. 211: *Multiple comparisons (e.g. Dunnett's test) ... can be used to calculate whether there are significant differences between the controls and the various test substance concentrations*
- Common data structure (chron tox biassays on rats)



Reg-Tox is safety assessment I

- The primary concern is a $\min(f\text{-rate})$: *be safe in negative results*
- Proof-of-safety is not used, because a-priori definition of equivalence thresholds is hopeless
- Therefore, proof-of-hazard is the common approach today
- But why control of FWER by Dunnett-test?: i) control f_+ rate is of secondary importance , ii) control of FWER increases f_- rate.

Constraintively

A recent proposal I

- Kobayashi, K and Pillai, K.S. Repeated dose administration toxicity studies - use of t-tests in multiplicity data analysis. Fundamental Toxicological Sciences 2024.
- Really independent t-tests (better Welch tests)?
- For common small n_i assays, this approach even comes out with higher f- rate.
- Why? Sacrificing global df. Can not be recommended

A solution I

- Dunnett-type multiple contrast tests without multiplicity adjustment.
- But global MQR estimator is non-robust in case of heteroscedasticity: use sandwich estimator instead

R-code I

- Litter weight data
- ```
library(multcomp); library(sandwich)
mod21<-lm(weight ~ dose, data=litter)
summary(glht(mod21, linfct = mcp(dose = "Dunnett"),
vcov=vcovHC), test=adjusted("none"))
```

# Take home I

- Control FWER is NOT a statistical issue. It depends on the claim.
- In safety studies another weight between  $f^-$  and  $f^+$  is appropriate.
- A related paper is accepted by Biometrical Journal
- Next available is the online report **Multiple Contrast Tests: Understanding By Running R Code Using Implemented Data Examples** - with 'no meaningful trend'