



TITLE: Is the Dunnett procedure, the standard method for analyzing chronic toxicity studies, truly appropriate for the purpose?

SPEAKER AND COAUTHOR: Ludwig A. Hothorn and Mario Hasler
Leibniz University Hannover and Christian-Albrechts-University zu Kiel, Germany

ABSTRACT: On the one hand, the Dunnett procedure is one of the most commonly used statistical tests; on the other hand, it is used primarily in toxicology and ecotoxicology. This involves simultaneously comparing treatment or dose groups against the negative control, under the condition that $FWER < \alpha$. In other words, the false-positive rate is directly and strictly controlled, resulting in high false-negative error rates (e.g., any-pairs power). This represents an inappropriate weighting between f_+ and f_- for the required proof of safety. Using equivalence or noninferiority hypotheses, one could account for a higher weighting of the f_- rate: this alone requires a definition of endpoint-specific thresholds—which is illusory in reality. Therefore, the proof-of-hazard using the Dunnett test is current practice, also in light of the methodological shortcoming that “absence of evidence is not evidence of absence.”

Kobayashi’s (2024) proposal to use independent t-tests instead, each with a comparison-wise f_+ control only, may seem counterintuitive, since the reduction in degrees of freedom (especially in the small-sample-size designs common in toxicology) can lead to an increase in f_- rates (even higher than with the Dunnett test). Our proposal employs a multiple comparison test with joint degrees of freedom and per-comparison f_+ control, modified with a sandwich variance estimator—since heteroscedasticity can be a source of bias—and takes into account the correlations among the simultaneous comparisons. This approach is illustrated using a real-world data example, with the use of CRAN R libraries (Hothorn and Hasler 2026).

This method discusses the different f_-/f_+ error rates balance between for demonstrating efficacy (clinical trials) or safety (toxicity studies) in multiple testing scenarios—under the condition that equivalence thresholds are not actually available.

References

Kobayashi, K and Pillai, K.S. Repeated dose administration toxicity studies – use of t-tests in multiplicity data analysis. *Fundamental Toxicological Sciences*, 11(1):1–10, 2024.

Hothorn, L.A. and M. Hasler Comparisons of several treatments or dose groups against a negative control in toxicological bio-assays: an alternative to the common-used Dunnett test. Submitted to *Biom J*.