ANALYSIS TESTING HIERARCHIES WITH PRELIMINARY TESTS

The massive amount of data collected in toxicology and animal health studies are often processed via statistical analysis testing hierarchies or flowcharts. These analysis flowcharts are driven by introductory text books which state that two conditions must be present in order to correctly use parametric analyses (eg. ANOVA): 1) the data must come from a known distribution, which is usually a normal distribution, and 2) the variances among groups must be the same (homogeneity of variance). In these analysis flowcharts, formal statistical tests are often utilized to confirm these assumptions. Preliminary statistical tests such as Shapiro-Wilk are used to evaluate normality and Bartlett's or Levene's test is often used to examine homogeneity of variance. According to the flowchart, if both conditions cannot be confirmed, then a nonparametric test (eg. Kruskal-Wallis) should be used to test the data in question.

Problems with Preliminary Tests:

There are several problems with the preliminary formal testing of parametric assumptions. In fact, the usefulness of most preliminary tests is inversely related to where it is most powerful. For example, in small sample sizes, a nonparametric analysis may be more appropriate than its parametric counterpart. For small sample sizes, the preliminary tests are less likely to detect deviations from the assumptions. By contrast, preliminary tests for large sample sizes are more likely to detect significant deviations from normality and/or homogeneous variances. But with large, approximately equal sample sizes, the parametric analysis of variance is very robust to deviations from the assumptions. That is to say, with larger sample sizes, the loss of power as the result of deviations from the assumptions diminishes. So, a situation arises where for large sample sizes the preliminary testing works well but isn't as critical as compared to the small sample size situation.

As noted by the statistician G.E.P. Box,

"To make preliminary tests on variances is rather like putting to sea in a row boat to find out whether conditions are sufficiently calm for an ocean liner to leave port."

BioSTAT Perspective:

In toxicology, our preference is to utilize the knowledge of the vast historical data that is available to determine, a priori, the statistical analysis that should be used for each analysis endpoint. For example, historical body weight data meets the conditions of parametric analysis and should be analyzed accordingly. Similarly, it is known that sperm count data is subject to extreme values and skewed distributions, and a nonparametric approach is more appropriate. This approach of historical data driven analyses is supported by many statisticians in the toxicology industry [Hothorn and Hauschke (1998), Bailey (1998), Selwyn (1995), Holson et al (2008), Vidmar et al (2011, 2012).

Historical data also provides enough information about many toxicology endpoints to select a transformation of the data, achieving the required assumptions for a statistical test. For example, the viability index in a DART study (the percentage of pups that survive at least 4 days), has historically been transformed by use of the Freeman-Tukey arcsine to attain assumptions required by the analysis of variance.

References:

Selwyn, M. (1995), "The Use of trend Tests to Determine a No-Observable-Effect Level in Animal Safety Studies", J Am Coll Toxicol 14:158-168, 1995.

Developmental and Reproductive Toxicology: A Practical Approach, Third Edition Edited by Ali S. Faqi., Elsevier Inc., November 2012. [see section "Parametric vs. Nonparametric" in Chapter 30 (Biostatistics for Toxicologists)

*Design and Analysis of Animal Studies in Pharmaceutical Developmen*t, edited by Shein-Chung Chow and Jen-pei Liu. Marcel Dekker, 1998. [see section "Testing Hierarchies" in Chapter 4 (Principles in Statistical Testing in Randomized Toxicological Studies and "Parametric Statistical Methods" in Chapter 5 Subchronic Toxicity)

A Comprehensive Guide to Toxicology in Preclinical Drug Development, First Edition. Edited by Ron Hood., CRC Press, Inc., December 2011. [See section "Parametric vs. Nonparametric" in Chapter 21 (Understanding Statistics in Developmental and Reproductive Toxicology).

Holson, et al (2008) "Statistical issues and techniques appropriate for developmental neurotoxicity testing: A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints" Neurotoxicology and Teratology 30 (2008) 326–348.