Options for data interpretation and analyses of avian reproduction endpoints

Manousos Foudoulakis[1], Timothy Fredricks[2], Tiffany Carro[3], Jonathan Maul[4], Ralf Barfknecht[5], Stephanie Plautz [6], Pablo Valverde[7], Adam Schapaugh[2], Xiaoyi Sopko[8], John W Green[9]

[1] Corteva Agriscience™ The Agriculture Division of DowDuPont™, 152 32 Halandri, Greece; [2] Bayer U.S. Crop Science, Chesterfield, Missouri, USA; [3] FMC Corporation, 2929 Walnut Street, Philadelphia, Pennsylvania, USA; [4] Syngenta Crop Protection LLC, Greensboro, NC; [5] Bayer AG, CropScience, 40789 Monheim, Germany; [6] BASF Corporation, 26 Davis Dr, Research Triangle Park, NC, USA; [7] Corteva Agriscience™ The Agriculture Division of DowDuPont™, Indianapolis, USA; [8] Corteva; Agriscience™ The Agriculture Division of DowDuPont™, Wilmington, USA; [9] John W Green Ecostatistical Consulting LLC, Newark, DE, USA

Background

Avian reproduction studies (OECD 206; OCSPP 850.2300) are required to assess the risks to wild birds from potential exposure to plant protection products in many regions (e.g. EU, USA). Numerous response variables are analysed to determine the no observed effect concentrations (NOECs) amongst treatment groups.

Due to the complexity of these studies and low statistical power of some tests and the large number of variables assessed, careful interpretation is required. Statistical differences between the concurrent control and treated groups can arise due to random sampling and can appear to be treatment-related.

A toolbox of different approaches (statistics, biological relevance, benchmark dose, historical control data (HCD) with a moving frame of reference) can improve interpretation of standard avian reproduction studies by providing further context in the data assessment, with the goal of avoiding unnecessary vertebrate testing.

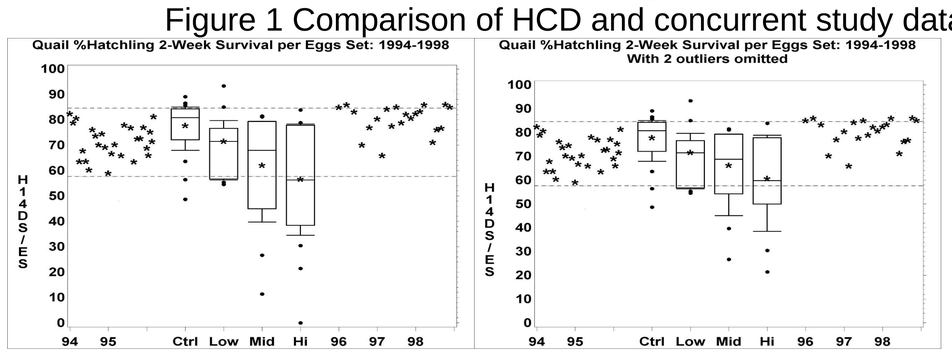
Case Study 1

14-day hatchling survivors (H14DS) /eggs set (ES), labeled H14DS_ES
Proportion data often have been modelled as continuous responses. Such proportions often are not normally distributed or have heterogenous variances. A standard normalizing, variance stabilizing transform is the arc-sine square-root (ARS). An alternative is the Freeman-Tukey transform, but this is discouraged because it is not order preserving and consequently can distort interpretation.

Transforms can be avoided by using generalized linear mixed models (GLMM) with a binomial distribution conditioned on replicate size, which treats data as actually observed (counts) rather than by approximate normality-based methods. H14DS_ES was not normal even after ARS transform so nonparametric analysis was done. Dunn test found NOEC=low dose, but stepdown Jonckheere-Terpstra (JT) test found NOEC=zero. Generally trend tests are preferred since they gain power by utilizing an anticipated dose-response monotonicity.
2 outliers were identified by the Tukey outlier test. With outliers omitted, data were normal and homogeneous, and Williams' and Dunnett NOEC=low dose.

GLMM analysis with H14DS binomially distributed conditioned on # at risk (ES) found Dunnett NOEC=low dose with all data, =mid-dose w/o outliers.
Only the high dose mean response is outside historical control range and with the 2 outliers omitted, it is also within the HCD range.

Figure 1 Comparison of HCD and concurrent study data



*=mean response, box from 25th to 75th percentile, dots=extremes
The 8% decrease at the low dose was not biologically meaningful and well within the HCD range. There was a clear downward trend, but the JT trend test was overly sensitive. Williams trend test is preferred when data meet its requirements. The impact of outliers should be examined.
GLMM analysis treats counts as counts w/o approximations and has good power properties.
NOEC ≥ mid-dose justified, effect at high dose questionable, given HCD.

References

- Cade BS 2015. Model averaging and muddled multimodel inferences. Ecology 96: 2370-2382. https://doi.org/10.1890/14-1639.1
- EFSA 2009. Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. *The EFSA Journal* (2009) 1150: 1-72.
- Green JW, Springer TA, Holbech H 2018. Statistical Analysis of Ecotoxicity Studies. Wiley. ISBN: 978-1-119-48881-1
- USEPA 2015. https://www.epa.gov/sites/production/files/2015-11/documents/support_material_for_model_averaging_workshop-11_06_2015-508.pdf
- Valverde-Garcia, P., Springer, T., Kramer, V., Foudoulakis M., Wheeler, J.R., 2018. An avian reproduction study historical control database: a tool for data interpretation.Regul. Toxicol. Pharmacol. 92, 295–302.

Normalization of data

An expected endpoint is **Hatchling survival at 14 days (HS14D) per maximum eggs set (MxES).** Problems include
(1) Once the study is complete and MxES is determined, this is equivalent to HS14D which is best analyzed by GLMM using a Poisson distribution.
(2) MxES is in fact a random variable and standard analysis of HS14D/ES ignore that and its effect on variance and test statistics. Resulting NOEC and ECx/BMD calculations can be grossly inaccurate.
(3) MxES is an extreme value statistic which can have high variance.
(4) The ratio depends on a single, possibly unusual, hen more than on the effect of the chemical generally.

Case Study 2

HSD14, 14-Day Survivors per Hatchling
Most avian repro studies have only 3 doses + control which is inadequate for regression modeling. This study had 4 doses + control, still poor for regression. Data were not normally distributed even after ARS transform.

JT and Dunn set NOEC=255ppm, where 23% decrease was observed. Preferred GLMM analysis with binomial distribution conditioned on Hatchlings set NOEC=85ppm.

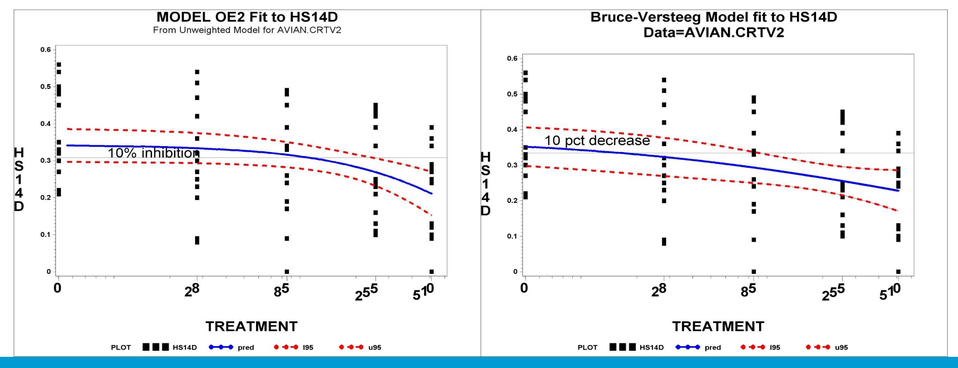
EFSA (2009) recommends regression with model averaging to obtain a benchmark dose (BMD) & its lower bound (BMDL). 5 models were fit: Bruce-Versteeg (BVP), log-logistic (LL3), simple exponential (OE2), exponential w/ shape parameter (OE3), exponential w/ shape & floor parameters (OE4).

Table 1. Model summary

Model	EC10	LB	UB
BVP	16.1	0.3	948
LL3	15.9	-53	84.6
OE3	15.9	-53	84.6
OE2	111	33.7	188
OE4	109	-92	310

There were 2 distinct groups of EC10 results. Model averaging means defining BMD & BMDL to be the minimum of the EC10 & LB values in the table, assuming each model fit is "adequate." Statistical model assessment tools are well developed. These models were not adequate by the usual measures, including underestimation of the control & one or more model parameters not different from 0. A negative BMDL is useless for risk assessment and BMDs differing by 7X are problematic. Figure 2 shows 2 models with very different BMD10. Visual inspection shows few clues to help choosing between them. BMD methods are best in principle, but avian data generally do not permit useful regression modeling due to too few doses & high within-dose variance.

Figure 2. Two models fit to HSD14



Conclusions

Avian endpoints present challenges for statistical analysis. Care must be taken in selecting the statistical tests to establish a NOEC and the influence and nature of outliers. A historical control database can be very useful in interpreting results by identifying unusual concurrent controls or identifying overly sensitive tests by which statistical significance and biological importance are in conflict. While regression methods to establish benchmark dose values (BMD and BMDL) are desirable in principle, these studies often do not support this approach due to few doses & high variability leading to wide confidence bounds.