A Recommendation On Critical Effect Size (CES) For Quantitative Analyis Of Micronucleus In Vitro Concentration-Response Curves Based On Variability Of Solvent Controls

Andreas Zeller

* Pharmaceutical Sciences, Pharmaceutical Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Basel, CH

1 - Introduction

The benchmark-dose (BMD) approach is the preferred method to analyze dose-response relationships but data-based guidance on the appropriate choice of benchmark-response (CES in case of the most commonly used software, PROAST) is scarce. CES should reflect a small, yet measurable increase of effect that is associated with negligible risk [1]. As regulating something that cannot be measured appears difficult, it is conceivable to associate CES with the lower limit of detection (LLOD) of an experimental method. Analytical chemists often calculate LLOD as multiples of the background noise of a method, for example 3σ in chromatography applications when 3 standard deviations (SD) are added on top of the mean background to define the limit of detection. A more conservative approach would obviously be necessary for genotoxicity endpoints, and recommendations to use 1SD over mean as CES have been published and are default setting in EPA BMDS software. However, this approach would “reward” laboratories/studies with high variability in the control group and “penalize” laboratories performing meticulous work with tiny CES values. Therefore, an approach that removes 5% high outliers and averages over many *in vivo* laboratories has been published [2] as “trimmed historical control”. Based on [2] and a different approach coming to similar conclusions [3], a 50% increase over mean controls is now recommended by IWGT and has been used successfully for regulatory submissions.
No such recommendation currently exists for *in vitro* endpoints like the micronucleus assay with cell lines or cultured human lymphocytes.

2 – Material and Method

More than 15 laboratories kindly donated individual raw data of their historical solvent controls. Individual data and identity of laboratories are confidential. Based on a method described in [2], the uppermost 5%percentile of the data was removed to account for outliers, and arithmetic mean as well as standard deviation calculated. One SD increase over the mean was used to derive CES values, stated in decimal format (i.e. 0.5 which corresponds to 50%). Average data from up to 17 laboratories are given in Table 1, separated into microscopic (both manual and automated) and flow cytometric (Litron MicroFlow kits) evaluation. Most laboratories donated data on continuous treatment (24 – 29 h) in the absence of metabolic activation (rat liver S-9 mix), and pulse treatment (3 – 6 h, followed by recovery) in the presence of metabolic activation.

3 – Results

Both data and data annotation were heterogeneous, especially in terms of annotation/metadata granularity, database size, evaluation methods, and cell type. The mean variability of outlier-deprived background noise from MN vitro data 17 participating laboratories was between 0.13 and 1.0, calculated from a total of nearly 5000 individual studies. The mean CES was approx. 0.43 across methods and treatments. The presence of patterns that might be responsible for the heterogeneity of was investigated, but without clear result.

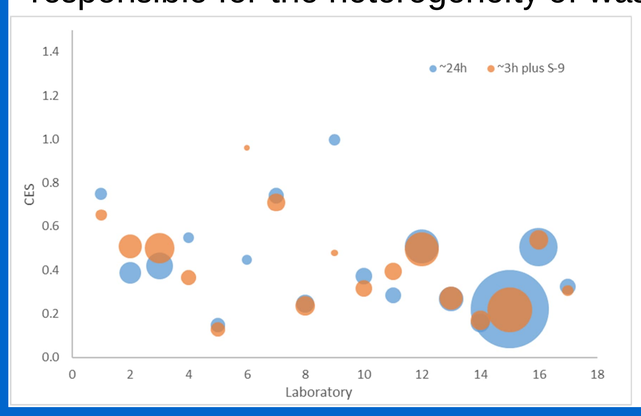

Figure: Individual lab data (coded). CES calculated for two protocol variants (bubble size indicates number of studies per lab)

Table: Summary statistics

Evaluation type	manual or automated microscopy		flow
Treatment	approx 24 h	approx 3h plus S-9	24 h
MIN	0.15	0.13	0.36
AVG	0.43	0.43	0.44
MAX	1.00	0.96	0.53
MEDIAN	0.39	0.39	0.43
SD	0.23	0.22	0.09
n (labs)	17	17	3
n (studies total)	3131	1276	481
MAX (studies per lab)	364	331	419

4 – Discussion

Data heterogeneity is likely related to the wide variety of laboratories from industry and CROs, located in 7 different countries, and using 5 different cell types evaluated with different stains/methods are represented in the dataset. No clear impact of any of those factors on CES has been identified. The consistency of calculated CES values suggests the results to be representative and robust, i.e. they can be expected to be applicable for the majority of other, non-participating laboratories as well.

5 – Conclusion

Based on a large number of data points from many experienced laboratories, a 50% increase over concurrent controls is recommended as CES for MN in vitro studies. Importantly, the 50% is not intended as evaluation criterion (POS vs NEG) but only to be used in PROAST to calculate BMDL values that possibly could be a starting point for risk assessment without animal experiments.

6 – References

(1) Edler, L.: (2014) in *Reichl and Schwenk: “Regulatory Toxicology”*, Springer, Berlin.
(2) Zeller et al., (2017) An appraisal of critical effect sizes for the benchmark dose approach to assess dose–response relationships in genetic toxicology. Arch Toxicol.
(3) White, P.A. et al: Benchmark response values for In Vivo Mutagencity Endpoints, (close to submission).

7 – Contributing Laboratories

Generous data donations from:

BASF, Ludwigshafen, Germany
Boehringer-Ingelheim, Biberach, Germany
Chugai, Gotemba, Japan
GenEvolution, Porcheville, France
Gentronix, Alderley Park, UK
GSK, Ware, UK
ICCR, Rossdorf, Germany
JTI, Kanagawa, Japan
LabCorp, Harrogate, UK
LSIM Safety Institute, Ibaraki, Japan
Merck KGaA, Darmstadt, Germany
Merck & Co, West Point PA, US
MTPC, Fujisawa, Japan
Nissan Chemical, Tokyo, Japan
Novartis, Basel, Switzerland
Roche, Basel, Switzerland
SNBL, Kagoshima, Japan
Takeda Pharma, Tokyo, Japan

THANK YOU! Your help made this present work possible and is greatly appreciated.