Variability in Reference Test Method Data and the Impact on NAM Evaluations

D. Allen[1*], J. Rooney[1], K. To[1], N. Choksi[1], P. Ceger[1], A. Daniel[1], A. Karmaus[1], J. Strickland[1], J. Truax[1], W. M. Casey[2], and N. Kleinstreuer[2]

[1]ILS, Research Triangle Park, NC, USA; [2]NIH/NIEHS/DNTP/NICEATM, Research Triangle Park, NC, USA

Historically, toxicity testing has been conducted using in vivo test methods. Confidence in data from these methods is such that regulatory hazard classification and labeling systems have been designed around their results and the methods are used as the benchmark against which new approach methodologies (NAMs) that replace or reduce animal use are compared. For many toxicity endpoints there is no NAM accepted as a complete replacement for animal use because hazard categorizations based on data from the NAM do not always agree with hazard categorizations based on in vivo data for the same chemical set. However, discordance with in vivo results may not always indicate that the NAM is generating an incorrect prediction. Variability of results from in vivo test methods could be an important contributor to such discordance and therefore should be carefully considered when comparing in vivo and NAM results. To establish confidence in NAMs, it is critical to understand any variability inherent to the in vivo test a NAM is intended to replace, as this variability will directly affect the expectations for performance of NAMs that seek to replace it. Sources of such variability might include both the inherent variability among animals and the subjective nature of observational in vivo endpoints. In this study, we characterized the variability of in vivo reference test methods for multiple endpoints, including skin and eye irritation, skin sensitization, and acute systemic toxicity. Our results indicate that in many cases in vivo test method variability is sufficiently high to warrant reassessing how NAMs are evaluated for these endpoints, particularly in the mild to moderate range of toxicity. For example, we found that chemicals classified as mild skin irritants by the in vivo test method were less than 50% likely to be classified as such if retested. These efforts provide the basis for redefining benchmarks against which to evaluate NAMs and thereby set appropriate expectations for NAM performance. This project was funded with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C.

Keywords: new approach methodologies, test method variability, establishing confidence