Establishing Confidence in NAMs: Considering Variability in Reference Test Methods

David Allen[1*], John Rooney[1], Neepa Choksi[1], Patricia Ceger[1], Amber Daniel[1], Agnes Karmaus[1], Judy Strickland[1], Kim To[1], Jim Truax[1], Warren M. Casey[2], and Nicole Kleinstreuer[2]

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

*Presenting author

Many hazard classification and labeling systems are based on in vivo test method results. In vivo methods are also typically used as the benchmark against which new approach methodologies (NAMs) are compared. For many toxicity endpoints, there is no NAM accepted as a complete replacement of animal use due to a lack of concordance with the full spectrum of hazard categories. However, does discordance with in vivo results always indicate that the NAM is "wrong"? 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. It is critical to understand any variability inherent to the animal test, 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. This presentation will summarize efforts at NICEATM to characterize the variability of in vivo reference test methods for multiple endpoints, including skin and eye irritation, skin sensitization, and acute systemic toxicity. For example, we analyzed in vivo skin irritation data and found that chemicals classified as mild irritants were less than 50% likely to be repeated as such if retested. These efforts provide the basis for 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.