Reliability Assessment of Guideline-Based Studies Using Systematic Review Critical Appraisal Tools

N. Y. Choksi, S. Fitch, M.A. Harris, C.M. Thompson, and D.S. Wikoff

Monday, March 20, 2023
9:00 AM – 10:45 AM
3182/P297
Regulation/Policy
Music City Center
Hall C

Abstract:

Systematic review methods and critical appraisal tools (CATs) have been implemented by authoritative and regulatory bodies in recent years to increase the transparency and data quality used in hazard and risk assessments. To assess (a) how well CATs capture critical elements of guideline-based toxicology studies and (b) how CATs that were developed for other disciplines are being modified to address study elements which are important to toxicology, we surveyed risk assessments conducted by authoritative and regulatory bodies, as well as those published in the literature, which used systematic review methods and included CATs to evaluate reliability of guideline (and similar) studies. Most assessments identified were issued by authoritative entities and involved assessment of environmental chemicals or food ingredients; CATs included Klimisch categorization and risk of bias-based assessments. Using a recent evaluation of hexavalent chromium (Cr(VI)) as one of the case studies, results demonstrates that several studies conducted according to or similar to guideline studies were determined to have low reliability. Rationale, however, was not due to limitations with methodological or reporting quality - notably, the internal validity was determined to have a very low risk of bias, which typically would support a high level of reliability. Rather, the lack of reliability was based on dose selection, despite being the dose determined via the guideline design. Rationale was not consistent with the CAT guidance, which describes parameters associated with the utility of the exposure design for the endpoint of interest; many of the studies determined to be unreliable were designed specifically to evaluate the mode of action associated with occurrence of a tumors at the given doses and were appropriate given the context of the scientific question under investigation. Such apparent disconnects between CAT criteria and scientific context highlights the importance of subject matter expertise in assessing the reliability of a study in context of the needs of the given risk assessment. In other case studies with food ingredients and pesticides, prompting questions from other CATs were added to risk of bias questions to provide a more granular and directed assessment of each domain. These examples demonstrate many instances where study elements relating to construct and external validity were included into risk of bias questions to fit the needs of the risk assessment or, in some cases, specific study types or subsets of studies. In addition to investigating the applications, a generic comparison of guideline study designs often mandated by authoritative agencies was compared to CAT criteria, demonstrating that several elements important to risk assessment inherent to guideline studies are not accounted for in standard risk of bias CATs - and vice a versa. This presents a clear paradox considering the mandate for the conduct of studies using GLP and guideline designs which remains unresolved. Collectively, this evaluation demonstrates the importance of continued refinement of critical appraisal tools designed for systematic review for the use in risk assessment- and in doing so, refining in a 'fit for purpose' manner to consider aspects of construct and external validity more directly, such as those inherent to guideline study design and conduct. Subsequently, expedient dissemination of such requirements to the research community is needed to ensure that any data generated using experimental animal models will be considered reliable.