

The correspondence section is a public forum and, as such, is not peer-reviewed. EHP is not responsible for the accuracy, currency, or reliability of personal opinion expressed herein; it is the sole responsibility of the authors. EHP neither endorses nor disputes their published commentary.

Instruments for Assessing Risk of Bias and Other Methodological Criteria of Animal Studies: Omission of Well-Established Methods

<http://dx.doi.org/10.1289/ehp.1307727>

In response to the systematic review by Krauth et al. (2013) of instruments for assessing animal toxicology studies for risk of bias and other aspects of quality, we propose the need for a broader perspective when appraising—and hopefully improving—such studies.

Krauth et al. (2013) reviewed 30 instruments, 4 of which were designed for environmental toxicology studies used to evaluate human and ecological health hazards. The authors noted that these instruments were derived from preclinical pharmaceutical research in animal models. Many of these instruments focus on efficacy and not toxicity, and—as acknowledged by the authors—they may have limited potential application in environmental health research because they often have criteria that are not relevant to hazard and risk assessments.

Based on these 30 instruments, Krauth et al. concluded that a limited number of risk of bias assessment criteria have been empirically tested for animal research, including randomization, concealment of allocation, blinding, and accounting for all animals. However, the authors did not discuss which elements of risk of bias criteria have been empirically tested, nor did they discuss how they were tested, leaving the reader with no information on their reliability or usefulness.

We would like to bring the readers' attention to several other important publications in environmental chemical health hazard assessment that are pertinent to this topic (Ågerstrand et al. 2011; Hulzebos et al. 2010; Schneider et al. 2009), along with a U.S. Environmental Protection Agency (EPA) approach developed under the High Production Volume Challenge (U.S. EPA 1999b) as well as relevant and potentially eligible guidance developed by the U.S. EPA (1999a) and the Food and Drug Administration (FDA 2003). In addition, the majority of the procedures specified in Good Laboratory Practices and regulatory *in vivo* toxicity test guidelines (e.g., U.S. EPA 2013; Organisation for Economic Co-operation and Development 1999) were specifically developed to minimize systematic errors, assure high quality data and produce scientifically reliable studies.

These additional publications describe design, conduct, and reporting criteria that form the basis of the methodologies employed globally to assure quality and reliability of *in vivo* toxicological investigations for regulatory assessment of human and ecological health hazards. Because the application of systematic review and related evidence-based approaches in toxicology is still in its infancy, it is especially important at this time to recognize the contributions of these publications.

The omission of these publications by Krauth et al. could have major science policy implications. The National Toxicology Program (NTP) (whose parent organization, the National Institute of Environmental Health Sciences, funded the research of Krauth et al.) has begun relying on Krauth et al. (2013) to identify elements of risk of bias in evaluating animal studies of environmental agents as part of its systematic reviews for assessing health effects (NTP 2013a, 2013b). The reliance on criteria that have not been transparently empirically tested instead of well-established methodological criteria developed by authoritative national and international organizations could result in biased systematic reviews that ultimately lead to regulations or classifications not supported by the science.

We suggest that further work is warranted in pulling together published perspectives on how to evaluate study quality in animal toxicology studies. Issues in appraising such studies for evaluating environmental hazards to humans and wildlife go well beyond those of human clinical trials, and would benefit from collaboration of experts in animal toxicology with experts in human clinical trials of medical interventions and human epidemiology.

The authors had complete control over the design, conduct, interpretation, and reporting of the analyses included in this letter. The contents are solely the responsibility of the authors and do not necessarily reflect the official opinions or policies of the authors' employers or clients.

None of the authors received specific financial support or honorarium as compensation for developing this letter. Several authors are members of the Evidence-Based Toxicology Collaboration (EBTC), and M.L. Stephens and S. Hoffmann serve as the secretariats for the North American and European EBTC Steering Committees, respectively, for which they are compensated for their time. The EBTC's overall aims are to improve toxicological decision making, facilitate the modernization of the toxicological toolbox, and reinvigorate the safety sciences (see <http://www.ebtox.com>).

S. Hoffmann, J.R. Fowle III, and J. Goodman are consultants and have worked on a range of toxicity and risk assessment issues for a wide variety of clients. R.A. Becker and N.B. Beck are employed by the American Chemistry Council, a trade association of chemical manufacturers. A. Boobis, D. Fergusson, M. Lalu, and M. Leist are employed by institutes of higher education. In the past 3 years, A. Boobis and M. Leist have worked on a range of toxicity and risk assessment issues for a number of clients; this has included some consultancies.

All authors contributed equally and are listed in alphabetical order.

Nancy B. Beck,¹ Richard A. Becker,^{1*} Alan Boobis,^{2*} Dean Fergusson,^{3*} John R. Fowle III,^{4*} Julie Goodman,^{5*} Sebastian Hoffmann,^{6*} Manoj Lalu,^{7*} Marcel Leist,^{8*} and Martin L. Stephens^{9*}

¹Regulatory and Technical Affairs, American Chemistry Council, Washington, DC, USA; ²Department of Medicine, Imperial College, London, United Kingdom; ³Ottawa Hospital Research Institute, University of Ottawa, Ontario, Canada; ⁴Science to Inform, LLC, Pittsboro, North Carolina, USA; ⁵Gradient, Cambridge, Massachusetts, USA; ⁶seh consulting + services, Paderborn, Germany; ⁷The Ottawa Hospital, University of Ottawa, Ontario, Canada; ⁸University of Konstanz, Germany; ⁹Center for Alternatives to Animal Testing, John Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
E-mail: nancy_beck@americanchemistry.com

*Members of the Evidence-Based Toxicology Collaboration (EBTC), an initiative of scientists in academia, industry, and government who are interested in promoting evidence-based approaches to strengthen decision making in the safety sciences (see <http://www.ebtox.com>).

REFERENCES

- Ågerstrand M, Breitholtz M, Rudén C. 2011. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessments of pharmaceutical substances. *Environ Sci Eur* 23:17; doi:10.1186/2190-4715-23-17.
- FDA (Food and Drug Administration). 2003. General Guidelines for Designing and Conducting Toxicity Studies. In: *Guidance for Industry and Other Stakeholders, Toxicological Principles for the Safety Assessment of Food Ingredients*, Redbook 2000. Available: <http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078315.htm> [accessed 15 October 2013].
- Hulzebos E, Gunnarsdottir S, Rila JP, Dang Z, Rorije E. 2010. An Integrated Assessment Scheme for assessing the adequacy of (eco)toxicological data under REACH. *Toxicol Lett* 198(2):255–262.
- Krauth D, Woodruff TJ, Bero L. 2013. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. *Environ Health Perspect* 121:985–992; doi:10.1289/ehp.1206389.
- NTP (National Toxicology Program). 2013a. Appendix 2: Risk of Bias Guidance for BPA Exposure and Obesity Protocol. Available: http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/Appendix2BPA_Draft.pdf [accessed 13 February 2014].
- NTP (National Toxicology Program). 2013b. Draft OHAT Approach for Systematic Review and Evidence Integration for Literature-based Health Assessments. Available: http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/DraftOHATApproach_February2013.pdf [accessed 15 October 2013].
- Organisation for Economic Co-operation and Development. 1998. OECD Series on Principles of Good Laboratory

- Practice and Compliance Monitoring, No 1: OECD Principles on Good Laboratory Practice. ENV/MC/CHEM(98)17. Paris:OECD. Available: [http://search.oecd.org/officialdocuments/displaydocumentpdf?doclanguage=en&cote=env/mc/chem\(98\)17](http://search.oecd.org/officialdocuments/displaydocumentpdf?doclanguage=en&cote=env/mc/chem(98)17) [accessed 13 February 2014].
- Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, et al. 2009. "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol Lett* 189(2):138–144.
- U.S. EPA (U.S. Environmental Protection Agency). 1999a. Auditing General Toxicology Studies. Available: <http://www.epa.gov/compliance/resources/policies/monitoring/fifra/sop/glp-da-09.pdf> [accessed 15 October 2013].
- U.S. EPA (U.S. Environmental Protection Agency). 1999b. Determining the Adequacy of Existing Data. Available: <http://www.epa.gov/hpv/pubs/general/datadfin.htm> [accessed 15 October 2013].
- U.S. EPA (U.S. Environmental Protection Agency). 2013. OCSPP Harmonized Test Guidelines. Series 870: Health Effects Test Guidelines. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series870.htm [accessed 15 October 2013].

Instruments for Assessing Risk of Bias and Other Methodological Criteria: Krauth et al. Respond

<http://dx.doi.org/10.1289/ehp.1307727R>

Beck et al. criticize our systematic review (Krauth et al. 2013) because we included instruments derived from preclinical animal research. Assessment instruments developed for preclinical animal models have criteria that are relevant to hazard and risk assessment because risk of bias in animal studies is not dependent on the data stream or the question being asked, but on the design of the study. Many instruments that have been developed (including those for evaluating animal toxicology studies) have criteria that have not been shown to bias research outcomes (see Supplemental Material, Table S1, of Krauth et al. 2013).

Furthermore, Table 1 of our paper (Krauth et al. 2013) lists the criteria found in most instruments we identified. In the "Discussion," we described the empirical evidence supporting the use of some of these criteria and cited the relevant references with the empirical data. By empirical evidence, we mean that a criterion (e.g., randomization) has been shown to be associated with overestimation or underestimation of effect (this could be an efficacy or harm outcome).

Beck et al. note several publications in environmental chemical health hazard assessment [Ågerstrand et al. 2011; Food and Drug Administration (FDA) 2003; Hulzebos et al. 2010; Organisation for Economic Co-operation and Development

(OECD) 1998; Schneider et al. 2009; U.S. Environmental Protection Agency (EPA) 1999a, 1999b, 2013]. All of these publications, except OECD (1998), were identified in our search; however, they did not meet the *a priori* inclusion criteria for our systematic review. As noted in our "Methods" (Krauth et al. 2013), we included the earliest publication of an instrument when it was used in subsequent reports. The article by Ågerstrand et al. (2011) was based on four earlier published papers (i.e., Durda and Preziosi 2000; Hobbs et al. 2005; Klimisch et al. 1997; Schneider et al. 2009). We cited three of these in our review, but excluded Schneider et al. (2009) because it appeared to be a description of software that could be used to operationalize the Klimisch criteria. After reviewing the criteria described by Schneider et al. (2009) in their supplemental file, we found no unique additional criteria that were not already included in our Table 1 and Supplemental Material, Table S1. The reports from the U.S. EPA (1999a, 1999b) and FDA (2003) were neither indexed in Medline nor found in screening of bibliographies. In addition, U.S. EPA (2013) was published after we ended our study. Because we did not find the OECD document (OECD 1998), we cannot conclude whether or not it should have been included in our study.

The comment by Beck et al. that the National Toxicology Program is relying on criteria that have not been "transparently empirically tested" is not correct. In our paper (Krauth et al. 2013), we recommended the use of empirically tested criteria and we pointed out criteria that have been shown to be a risk of bias.

We caution against gathering judgments on how to assess study quality and propose that evidence should guide such evaluations. We propose an empirically based approach—as opposed to consensus-based opinion of experts—as this would provide a more unbiased evaluation of the data.

The authors declare they have no actual or potential competing financial interests.

David Krauth,¹ Tracey J. Woodruff,^{2,3} and Lisa Bero^{1,4}

¹Department of Clinical Pharmacy, and ²Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, San Francisco, California, USA; ³Program on Reproductive Health and the Environment, Oakland, California, USA; ⁴Institute for Health Policy Studies, University of California, San Francisco, San Francisco, California, USA
E-mail: berol@pharmacy.ucsf.edu

REFERENCES

- Ågerstrand M, Breitholtz M, Ruden C. 2011. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessments of pharmaceutical substances. *Environ Sci Eur* 23:17; doi:10.1186/2190-4715-23-17.
- Durda JL, Preziosi DV. 2000. Data quality evaluation of toxicological studies used to derive ecotoxicological benchmarks. *Hum Ecol Risk Assess* 6(5):747–765.
- FDA (Food and Drug Administration). 2003. General Guidelines for Designing and Conducting Toxicity Studies. In: *Guidance for Industry and Other Stakeholders, Toxicological Principles for the Safety Assessment of Food Ingredients*, Redbook 2000. Available: <http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078315.htm> [accessed 15 October 2013].
- Hobbs DA, Warne MSJ, Markich SJ. 2005. Evaluation of criteria used to assess the quality of aquatic toxicity data. *Integr Environ Assess Manag* 1(3):174–180.
- Hulzebos E, Gunnarsdottir S, Rila JP, Dang Z, Rorije E. 2010. An Integrated Assessment Scheme for assessing the adequacy of (eco)toxicological data under REACH. *Toxicol Lett* 198(2):255–262.
- Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharmacol* 25(1):1–5.
- Krauth D, Woodruff TJ, Bero L. 2013. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. *Environ Health Perspect* 121:985–992; doi:10.1289/ehp.1206389.
- OECD (Organisation for Economic Co-operation and Development). 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No 1: OECD Principles on Good Laboratory Practice. ENV/MC/CHEM(98)17. Paris:OECD. Available: [http://search.oecd.org/officialdocuments/displaydocumentpdf?doclanguage=en&cote=env/mc/chem\(98\)17](http://search.oecd.org/officialdocuments/displaydocumentpdf?doclanguage=en&cote=env/mc/chem(98)17) [accessed 13 February 2014].
- Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, et al. 2009. "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol Lett* 189(2):138–144.
- U.S. EPA (U.S. Environmental Protection Agency). 1999a. Auditing General Toxicology Studies. Available: <http://www.epa.gov/compliance/resources/policies/monitoring/fifra/sop/glp-da-09.pdf> [accessed 15 October 2013].
- U.S. EPA (U.S. Environmental Protection Agency). 1999b. Determining the Adequacy of Existing Data. Available: <http://www.epa.gov/hpv/pubs/general/datadfin.htm> [accessed 15 October 2013].
- U.S. EPA (U.S. Environmental Protection Agency). 2013. OCSPP Harmonized Test Guidelines. Series 870: Health Effects Test Guidelines. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series870.htm [accessed 15 October 2013].

Erratum: “Instruments for Assessing Risk of Bias and Other Methodological Criteria of Animal Studies: Omission of Well-Established Methods”

In the letter “Instruments for Assessing Risk of Bias and Other Methodological Criteria of Animal Studies: Omission of Well-Established Methods” by Beck et al. [Environ Health Perspect 122:A66–A67 (2014); <http://dx.doi.org/10.1289/ehp.1307727>], information for Alan Boobis and Marcel Leist was omitted from the Competing Financial Interests declaration. The complete declaration is as follows:

None of the authors received specific financial support or honorarium as compensation for developing this letter. Several authors are members of the Evidence-Based Toxicology Collaboration (EBTC), and M.L. Stephens and S. Hoffmann serve as the secretariats for the North American and European EBTC Steering Committees, respectively, for which they are compensated for their time. The EBTC’s overall aims are to improve toxicological decision making, facilitate the modernization of the toxicological toolbox, and reinvigorate the safety sciences (see <http://www.ebtox.com>). S. Hoffmann, J.R. Fowle III, and J. Goodman are consultants and have worked on a range of toxicity and risk assessment issues for a wide variety of clients. R.A. Becker and N.B. Beck are employed by the American Chemistry Council, a trade association of chemical manufacturers. A. Boobis, D. Fergusson, M. Lalu, and M. Leist are employed by institutes of higher education. In the past 3 years, A. Boobis and M. Leist have worked on a range of toxicity and risk assessment issues for a number of clients; this has included some consultancies.

The authors regret the error.