Predictive Models for Acute Oral Systemic Toxicity: A Workshop to Bridge the Gap from Research to Regulation

Nicole C Kleinstreuer[a], **Agnes Karmaus**[b], **Kamel Mansouri**[b], **David G Allen**[b], **Jeremy M Fitzpatrick**[c], and **Grace Patlewicz**[c,*]

[a]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA.

[b]Integrated Laboratory Systems, Inc., Research Triangle Park, North Carolina 27560, United States.

[c]National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA

Abstract

In early 2018, the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) published the "Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States" (ICCVAM 2018). Cross-agency federal workgroups have been established to implement this roadmap for various toxicological testing endpoints, with an initial focus on acute toxicity testing. The ICCVAM acute toxicity workgroup (ATWG) helped organize a global collaboration to build predictive in silico models for acute oral systemic toxicity, based on a large dataset of rodent studies and targeted towards regulatory needs identified across federal agencies. Thirty-two international groups across government, industry, and academia participated in the project, culminating in a workshop in April 2018 held at the National Institutes of Health (NIH). At the workshop, computational modelers and regulatory decision makers met to discuss the feasibility of using predictive model outputs for regulatory use in lieu of acute oral systemic toxicity testing. The models were combined to yield consensus predictions which demonstrated excellent performance when compared to the animal data, and workshop outcomes and follow-up activities to make these tools available and put them into practice are discussed here.

Abstract

Keywords

QSAR; read-across; acute oral toxicity; ICCVAM; workshop

Introduction

Background

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), consisting of representatives from 16 U.S. federal agencies, recently developed and published a national strategic roadmap with input from a broad range of stakeholder groups for incorporating new human-relevant approaches into safety testing of chemicals and medical products in the United States (ICCVAM 2018). The successful implementation of the roadmap depends upon coordinated efforts that address three strategic goals: 1) connecting end-users with developers of new approach methodologies; 2) fostering the use of efficient, flexible, and robust practices to establish confidence in new methods; and 3) encouraging the adoption and use of new methods and approaches by federal agencies and regulated industries. Towards these goals, ICCVAM establishes workgroups to perform specific tasks identified by the committee as being important for the development or validation of new approach methodologies (NAMs). ICCVAM workgroups develop detailed implementation plans to address roadmap goals, tailored to specific toxicological endpoints of concern. These implementation plans include four key elements: (1) definition of testing needs; (2) identification of any available alternative tests and computer models; (3) a plan to develop integrated approaches to testing and assessment and defined approaches for interpreting data; and (4) a plan to address both scientific and non-scientific challenges, including regulatory considerations such as international harmonization.

One of these workgroups is the Acute Toxicity Work Group (ATWG) which has developed an implementation plan for identifying, evaluating, and applying new approach methodologies that may serve as replacements for *in vivo* acute systemic toxicity testing studies (Lowit et al. 2017 [SOT Poster]). The ATWG comprises members from a number of different U.S. Agencies including the Department of Defense (DOD), U.S. Environmental Protection Agency (EPA), Occupational Safety and Health Administration (OSHA) and the US Consumer Product Safety Commission (CPSC) as well as International Cooperation on Alternative Test Method (ICATM) Liaison Members. The ATWG implementation plan covers the four key elements named above with respect to the area of acute systemic toxicity testing, and all activities are performed in coordination with a wide range of important

stakeholders. Current implementation activities include specifically: 1) draft a scoping document to identify US agency requirements, needs and decision contexts for acute toxicity data; 2) identify, acquire and curate high quality data from reference test methods; 3) identify, develop, and evaluate non-animal alternative approaches to acute toxicity testing; and 4) gain regulatory acceptance and facilitate application of non-animal approaches. Although the ATWG is focused on all three routes of exposure (dermal, inhalation, oral), the greatest progress has been made in realizing the implementation elements for acute oral toxicity. The remainder of this report is therefore focused on acute oral toxicity, specifically the rat oral LD50 (dose corresponding to 50% lethality) test.

Implementation

Identify U.S. Agency Requirements, Needs and Decision Contexts—
Understanding agency requirements has a direct bearing on the types of information needed for different decision contexts and provides a framework for managing expectations for how new approach methodologies (NAMs) can be practically applied. A scoping document (Strickland et al. 2018) published in early 2018 revealed that multiple U.S. agencies use acute oral toxicity data in a variety of regulatory contexts. For instance, EPA, OSHA and Department of Transportation (DOT) utilize hazard categories based on ranges of LD50 values, although those numeric ranges vary between the respective agencies based on the EPA or the UN Globally Harmonized System of Classification and Labelling of Chemicals (GHS) schemas. On the other hand, CPSC and DOD rely on two hazard categories to discriminate highly toxic substances from everything else. These different approaches are driven by the differences in agency statutes and regulatory authorities.

Identify, Acquire and Curate High Quality Data—Collaboratively NICEATM and EPA's National Center for Computational Toxicology (NCCT) collected rat oral LD50 data on over 15,000 substances from a number of publicly available databases and resources (Karmaus et al., 2017 [ASCCT]; Karmaus et al., 2018 [SOT]). These included data from OECD's eChemPortal, National Library of Medicine Hazardous Substances Data Bank (NLM HSDB), ChemIDplus, and JRC AcutoxBase. A total of 15,688 chemicals (identified by their CAS registry numbers) were associated with 21,200 LD50 values. Structures were then identified for 11,992 chemicals (16,209 LD50 values) using the EPA Chemistry Dashboard (https://comptox.epa.gov/dashboard) and other public resources. This reference set of data provided the basis for evaluating the performance and coverage of new and existing models, as well as for understanding the inherent variability of the animal data (Fitzpatrick et al., 2017 [ASCCT]; Karmaus et al., 2017 [ASCCT]; Fitzpatrick et al., 2018 [SOT], Karmaus et al., 2018 [SOT]). A summary of the dataset compiled is available at https://ntp.niehs.nih.gov/go/tox-models. A separate manuscript describing the data curation and analyses to assess the variability of the data is currently in preparation.

Identify and Evaluate Non-Animal Alternative Approaches to Acute Toxicity Testing: Predictive Modelling Project—The two key implementation elements discussed above directly impact the ability to identify and evaluate existing approaches or develop new approaches. The scoping of agency requirements identified the endpoints that need to be modelled (Stickland et al., 2018), and the compilation of such a large dataset

enabled building models for broad chemistry space. Based on preliminary investigations within both NICEATM and EPA, *in silico* structure-based models were found to demonstrate the most promising performance when predicting acute oral toxicity endpoints (Fitzpatrick et al., 2017 [ASCCT]; Fitzpatrick et al., 2018 [SOT], manuscript in preparation). The ATWG therefore organized a project to leverage the expertise of the international modelling community to develop in silico models of acute oral systemic toxicity that would predict the specific endpoints required by US agencies (explained below). The timeline and resources are described in more detail on the NICTEAM website https://ntp.niehs.nih.gov/go/tox-models.

Facilitate Regulatory Acceptance and Application: International Workshop—In brief, the project was managed by a Workshop Organizing Committee who recused themselves from directly participating in the modelling exercise. A training set of chemicals (~9k), along with their associated LD50 information and hazard categories, was extracted from the large compiled dataset and made available to the modelling community in November 2017. A prediction set was then released in December 2017 which comprised a test set of chemicals (~3k) amongst a larger inventory of chemicals (~40k) of interest to different agencies. Modelers were asked to submit their model results and documentation by early February 2018 for consideration by the Organizing Committee. The committee evaluated each model qualitatively with respect to the OECD Validation Principles (2004, 2007) and quantitatively based on the predictive performance against the test set. Specific models were then selected for platform presentation at the workshop held in April 2018 (https://ntp.niehs.nih.gov/pubhealth/evalatm/3rs-meetings/past-meetings/tox-models-2018/index.html). Models meeting the qualitative and quantitative evaluation criteria were also included into consensus models for each endpoint to derive the Collaborative Acute Toxicity Modeling Suite (CATMoS), which leverages the strengths and compensate for the weaknesses of each individual approach. The format of the modelling project was largely based on other consensus modelling projects conducted previously (see Mansouri et al. 2016).

Following a call for participants, a total of thirty-five groups from the US, Europe, and Asia submitted 139 models for the different endpoints requested (covering five defined endpoint schema: EPA hazard categories, GHS hazard categories, very toxic (LD50 < 50 mg/kg), non-toxic (LD50 > 2,000 mg/kg), and LD50 point estimate predictions). Modelers represented various sectors including industry, academia, and government. The workshop entitled "Predictive Models for Acute Oral Systemic Toxicity" was convened on April 11 and 12, 2018 at the National Institutes of Health in Bethesda, Maryland. The workshop brought together representatives of both the regulatory and modeling communities, offering a unique forum to discuss the strengths and limitations of the different models developed and their implementation for regulatory use. Other participants included industry stakeholders and representatives of non-governmental organizations (NGOs). In total, over 70 attendees participated in person and another 30 contributed via webcast.

Workshop Overview—The workshop began with several introductory presentations to put into context the ICCVAM Roadmap, the ATWG implementation plan for acute toxicity,

and a summary of how ICCVAM member agencies currently use acute oral systemic toxicity (LD50) data. Next an overview of the compiled dataset was presented describing in particular the variability observed across independently replicated acute oral toxicity studies, the ways in which this had been assessed with respect to chemistry and chemical use categories, and deriving a quantitative measure of variability to aggregate and benchmark the data for modelling purposes. An overview of the preliminary results of the consensus modeling effort was then provided. Selected platform presentations and panel discussions from invited modelers followed. These were structured to summarize key learnings in the respective modeling approaches applied, areas of success, and challenges encountered during the project. The intent was to focus on the insights gained and the impact that the constraints of the project (e.g. timeline) might have had on the robustness and predictability of the models developed. Day 2 of the workshop began with perspectives from regulators and other end-users to explore both how predictive models might be used to replace in vivo acute systemic toxicity testing and the potential strengths and limitations of the models in various contexts. The remainder of the workshop was then organized in rotating breakout group sessions where participants discussed practical applications and explicit considerations for the interpretation, characterization, and extension of modeling data. These breakout groups provided an environment in which regulators and industry members were able to interact with modelers to discuss questions and learn from one another in order to better understand data needs, modeling approaches, use cases, and limitations. Further details are provided in the following sections, full webcasts of the presentations are available at https://ntp.niehs.nih.gov/pubhealth/evalatm/3rs-meetings/pastmeetings/tox-models-2018/index.html, and the complete agenda can be found at https://ntp.niehs.nih.gov/iccvam/meetings/at-models-2018/workshop-agenda-fd-508.pdf.

Introductory Presentations—The initial workshop presentations covered diverse perspectives on the use of acute oral systemic toxicity assay, curation and characterization of the existing animal data, predictive modeling approaches, and the adoption of modeling outputs in practice. The first presentation summarized the broad applications of acute oral LD50 data, e.g. for hazard classification and labeling of products, for determining acceptable human and ecological exposure limits, and for defining what personal protective equipment is required for handling products. With different hazard categorization schema and diverse regulatory needs across different agencies, various binary, categorical and continuous endpoints of interest were highlighted. These endpoints, as listed above, were the defined outputs for the modeling project.

Presentations by NICEATM scientists described the project details. NICEATM and the U.S. EPA's NCCT compiled a comprehensive inventory of acute oral systemic toxicity LD50 values from a multitude of sources and curated a subset of chemicals for which defined structures were available. There were 1,120 chemicals with three or more LD50 values, and this set was used to analyze and quantify the variability in the animal studies and to define a confidence interval for the in vivo LD50 values. The full structure-curated inventory comprised 11,992 chemicals, which were standardized in preparation for modeling then divided into training and test sets for the participants in the consortium to utilize in building predictive models for the mentioned acute oral toxicity endpoints.

Modeler and End-User Perspectives—Presentations and panel discussions from modeling experts and end-users emphasizing case studies, challenges, and opportunities followed the introductory overview presentations. Submitted models had been reviewed by the Workshop Organizing Committee based on both quantitative and qualitative criteria, and 10 modelers were invited to give short platform presentations at the workshop and participate in panel discussions. All modeling participants were invited to present posters at the workshop. The platform presentations summarized a diversity of modeling approaches including random forest, QSAR, clustering-based methods, deep learning, and artificial intelligence methods. Key points made by the modelers included:

- Chemical descriptor selection is important and incorporating additional data inputs (e.g. other physicochemical properties, metabolism prediction, in vitro mechanistic data) could improve outcomes.

- Applicability domain assessment is necessary and can be accomplished in many ways.

- Data curation, and in particular access to a well-curated training set, is critical to modeling success.

On the second day of the workshop, presentations from end-users described case studies followed by a panel discussion addressing perspectives on how computational modeling outputs could be used in practice. Presentations represented industry and government viewpoints, highlighting different requirements, concerns, and information needs. For example, industry mainly utilizes alternatives to animal testing for product development and these presentations emphasized that any new models would have to be amenable to being run in-house for new compounds and protective of confidential business information (CBI). Representatives from regulatory agencies noted a need for training so that staff reviewing submissions could gain confidence in the predictions provided by the models and be able to defend their regulatory decisions. These diverse perspectives provided a foundation for breakout group discussions held at the end of the second day.

Workshop Breakout Sessions—Breakout groups were convened to facilitate smaller group interaction among stakeholders, and to allow participants to discuss applying computationally-derived predictive models to replace in vivo acute oral toxicity tests.

The first breakout group focused on practical applications. Participants in this group noted the importance for model transparency, and the ability to balance proprietary information and data security with the desire for open-source tools. Models must be sufficiently defined in order to be interpretable and defensible, as black-box methods with unclear or proprietary definitions will have limited regulatory applications (but may be suitable for industry use). The critical need to protect CBI was noted, as well as the desire for straightforward methods to assess domain of applicability and provide confidence estimates for predictions made on new chemicals. Finally, endusers urged modelers to demonstrate performance of new models using reference compound lists, relevant to specific regulatory programs, to clearly relay the usefulness and limitations of the models.

The second breakout group focused on model interpretation, characterization, and extension. Modelers in this session urged end-users to be realistic in their expectations of model outputs and their criteria for accepting model predictions as alternatives to the animal test. For example, expecting models to provide mechanistic interpretation may not be reasonable, given that the in vivo data as expressed by acute oral LD50 values do not describe the biological mechanism underlying the toxicity. However, it was acknowledged that predictive models have significantly more potential to yield mechanistic insights by providing associations between toxicity and chemical properties, structure, and mechanistic scaffolds. End-users requested clearly defined workflows, both for aiding in the interpretation of modeling approaches and for describing the curation of the input dataset. Participants discussed the types of approaches that could be used to account for variability in the data (and sources of such variability) or the uncertainties in the model predictions. Such information is critical to establishing confidence in these modeling approaches. Finally, participants all agreed that defining a uniform lexicon is needed to ensure the same language is used by everyone.

Conclusions

Workshop Outcomes

The consensus model predictions on the training and the test sets were equivalent in performance to the ability of the rat oral LD50 data to predict itself, i.e. the reproducibility of multiple independent studies on the same chemicals, for all the endpoints considered in the project. Various follow-up activities are therefore underway to assist in the implementation of these predictive models for acute oral toxicity. The project is currently being written up for publication, with one paper focusing on the compilation, characterization, and variability assessment of the reference animal data and another describing the international modeling effort and the resulting consensus predictions. Associated papers from individual groups will describe specific modeling efforts. The Collaborative Acute Toxicity Modeling Suite (CATMoS) will be incorporated into the OPERA package (Mansouri et al. 2018) and made available via the EPA Chemistry Dashboard (https://comptox.epa.gov/dashboard) and as standalone software. To help evaluate the model predictions within specific chemical domains and establish scientific confidence in their utility, ICCVAM agency representatives are compiling test chemical lists relevant to their respective regulatory applications, and the consensus model predictions will be generated and analyzed for these lists. Further prospective validation without additional testing could be accomplished via identification of proprietary datasets from industry, or recently generated data for regulatory submission, that could be used in evaluating the consensus model predictions on novel compounds. The development of training and outreach programs to help regulatory scientists and other endusers gain familiarity with the models is being discussed with NGO representatives and other stakeholder groups.

Overall, participant feedback indicated that this workshop provided a highly productive forum for collaboration and discussion across sectors to facilitate progress toward the integration of alternative models to replace animal testing for acute oral systemic toxicity.

Acknowledgments:

References

Fitzpatrick J, Karmaus A, Patlewicz G, Development of an acute oral toxicity dataset to facilitate assessment of existing QSARs and development of new models. American Society for Cellular and Computational Toxicology Meeting, Gaithersburg, MD, September 21–22, 2017.

Fitzpatrick J, Pradeep P, Karmaus A, Patlewicz G, Using Chemical and Biological Descriptors to Develop Predictive Models for Rat Acute Oral Toxicity. Society of Toxicology, San Antonio, TX, 3 11–15, 2018.

ICCVAM 2018 A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. https://ntp.niehs.nih.gov/go/natl-strategy Accessed 22 June 2018

Karmaus AL, Allen DG, Kleinstreuer NC, Casey WM, Establishing a Rat Acute Oral Database and Characterizing Variability Across Studies: Implications for Alternative Model Development. American Society for Cellular and Computational Toxicology Meeting, Gaithersburg, MD, September 21–22, 2017.

Karmaus A, Fitzpatrick J, Allen D, Patlewicz G, Kleinstreuer N, Casey W, Variability of LD50 Values from Rat Oral Acute Toxicity Studies: Implications for Alternative Model Development. Society of Toxicology, San Antonio, TX, 3 11–15, 2018.

Lowit A, Schlosser C, Myska A, Patlewicz G, Paris M, Karmaus A, Strickland J, Allen D, Kleinstreuer N, Casey W, Replacing Animals for Acute Systemic Toxicity Testing: A U.S. Strategy Roadmap. Society of Toxicology, Baltimore, MD, 3 12–16, 2017.

Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, Zakharov A, Worth A, Richard AM, Grulke CM, Trisciuzzi D, Fourches D, Horvath D, Benfenati E, Muratov E, Wedebye EB, Grisoni F, Mangiatordi GF, Incisivo GM, Hong H, Ng HW, Tetko IV, Balabin I, Kancherla J, Shen J, Burton J, Nicklaus M, Cassotti M, Nikolov NG, Nicolotti O, Andersson PL, Zang Q, Politi R, Beger RD, Todeschini R, Huang R, Farag S, Rosenberg SA, Slavov S, Hu X, Judson RS, CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environ Health Perspect 124 (2016) 1023–1033. 10.1289/ehp.1510267 [PubMed: 26908244]

Mansouri K, Grulke CM, Judson RS, Williams AJ, OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform. 10 (2018):10 10.1186/ s13321-018-0263-1. [PubMed: 29520515]

OECD, OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) StructureActivity Relationship Models, 2004 https://www.oecd.org/chemicalsafety/riskassessment/ 37849783.pdf (Accessed 9 June 2018).

OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, 2007 http://www.oecd.org/env/guidance-document-on-the-validation-ofquantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm (Accessed 9 June 2018).

Strickland J, Clippinger AJ, Brown J, Allen D, Jacobs A, Matheson J, Lowit A, Reinke EN, Johnshon MS, Quinn MK, Jr, Mattie D, Fitzpatrick SC, Ahir S, Kleinstreuer N, Casey W, Status of acute toxicity testing requirements and data uses by U.S. regulatory agencies. Reg Toxicol Pharmacol 94 (2018) 183–196. 10.1016/j.yrtph.2018.01.022.

Highlights

- Towards implementation of the ICCVAM Strategic Roadmap, a global modeling project was organized to build predictive in silico models for acute oral systemic toxicity.

- An international workshop was held in April 2018 at the NIH to discuss the results of the modeling project, with a diverse group of scientists and stakeholders participating in 2 days of presentations and breakout group discussions.

- Relative strengths and weaknesses of the models for different regulatory purposes were discussed, recommendations and next steps are presented