NORDIC WORKING PAPERS

SciRAP workshop report

Bridging the gap between academic research and chemicals regulation - the SciRAP tool for evaluating toxicity and ecotoxicity data for risk assessment of chemicals

Anna Beronius, Marlene Ågerstrand, Christina Rudén and Annika Hanberg

<http://dx.doi.org/10.6027/NA2017-905>
NA2017:905
ISSN 2311-0562

This working paper has been published with financial support from the Nordic Council of Ministers. However, the contents of this working paper do not necessarily reflect the views, policies or recommendations of the Nordic Council of Ministers.

Workshop report:

“Bridging the gap between academic research and chemicals regulation - the SciRAP tool for evaluating toxicity and ecotoxicity data for risk assessment of chemicals”

Authors: Anna Beronius¹, Marlene Ågerstrand², Christina Rudén² and Annika Hanberg¹

¹Institute of Environmental Medicine (IMM), Karolinska Institutet

²Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University

Date: 13 February 2017

Table of Contents

Preface.....	3
Summary	3
Background.....	3
Purpose and aim of the workshop	5
Participants.....	5
Pre-workshop exercise	6
Evaluation of a toxicity study	6
Evaluation of an ecotoxicity study	11
Participant presentations	15
Summary of breakout sessions	18
Breakout session 1	18
Breakout session 2	19
Overall conclusions from the workshop.....	20
Workshop evaluation	21
Next steps.....	21
Development of the SciRAP webpage and tools.....	21
Training and dissemination	22
Acknowledgements	22
References.....	23
Appendix I – Workshop agenda and discussion points.....	24
Appendix II – List of participants at the SciRAP workshop 15 – 16 November, 2016.	26
Appendix III – Evaluation of SciRAP workshop 15-16 November, 2016.....	28

Preface

The Science in Risk Assessment and Policy (SciRAP) project was initiated by researchers from the Department of Environmental Science and Analytical Chemistry (ACES) at Stockholm University and the Institute of Environmental Medicine (IMM) at Karolinska Institutet. In 2016 IMM received a grant from the Nordic Chemical Group (NKG) to organise a workshop in order to present and discuss the tools developed within SciRAP among intended end-users in the Nordic countries.

The workshop was organised together with the Nordic Risk Assessment Project (NORAP) and took place on November 15 – 16, 2016, at Karolinska Institutet in Stockholm. This report presents the results from the workshop, as well as a plan for how to continue the development and implementation of the SciRAP tools.

Summary

The Science in Risk Assessment and Policy (SciRAP) project is a research initiative with the purpose to increase and facilitate systematic use of academic (eco)toxicity studies in regulatory hazard and risk assessment. Within SciRAP, tools have been developed intended to promote structure and transparency in the process of evaluating the reliability and relevance of toxicity and ecotoxicity studies. The tools include sets of predefined criteria for evaluating reliability and relevance, as well as a colour-coding application, which are freely available online at www.scirap.org.

In November 2016 a workshop was organised to present and discuss the SciRAP tools and their application in hazard and risk assessment in the Nordic countries and the EU. Thirty-one experts in toxicology and ecotoxicology from authorities and academia in the five Nordic countries attended the workshop. The participants were encouraged to use the SciRAP tools and evaluate a specific ecotoxicity or toxicity study prior to the workshop. The results of the pre-workshop exercise, as well as the participants' general experience with the tools provided the basis for discussions during the workshop. In addition, five presentations were held where participants shared experiences from regulatory assessments of chemicals.

The overall conclusion from the workshop was that the SciRAP tools can be useful to Nordic authorities when evaluating the reliability and relevance of toxicity and ecotoxicity studies for hazard and risk assessment. The use of the SciRAP tools to increase transparency in evaluations and facilitate discussions to explain and resolve differences between evaluators were particularly highlighted. Future improvements were also suggested, for example that the website and online tools can be made more user-friendly and that guidance for how to interpret and use the output from the SciRAP tools is needed. This is especially important if the SciRAP approach is to be disseminated on the EU-level. The conclusions from the workshop will be useful in moving forward and making further developments to the SciRAP tools to meet end-users' needs.

Background

The Science in Risk Assessment and Policy (SciRAP) project was initiated by researchers from the Department of Environmental Science and Analytical Chemistry (ACES) at Stockholm University and the Institute of Environmental Medicine (IMM) at Karolinska Institutet. The purpose of SciRAP is to increase

and facilitate systematic use of academic research studies in regulatory hazard assessment (e.g. for hazard classification and setting of limit values, such as EQS values) and risk assessment and thereby bridge the gap between academic research and chemicals regulation (Molander et al. 2015). The aim of the project is primarily to provide tools that promote structure and transparency in the process of evaluating the reliability and relevance of toxicity and ecotoxicity data for hazard and risk assessment of chemicals.

Guidance for hazard and risk assessment of chemicals generally require or recommend that all relevant data should be considered in these processes (e.g. ECHA 2013; EFSA 2010; OECD 2005). Studies conducted in accordance with standardized test guidelines, such as the OECD test guidelines, and Good Laboratory Practices (GLP) are often preferred in regulatory hazard and risk assessment. But non-standard academic research studies can also contribute critical data and fill important information gaps (e.g. Beronius et al. 2013; Kortenkamp et al. 2012; Myers et al. 2009; Zoeller et al. 2012). However, current methods for evaluating the reliability and relevance of non-standard studies for hazard and risk assessment are often not very transparent or systematic (Beronius et al. 2010; SCENIHR 2012, Ågerstrand et al. 2011).

Within the SciRAP initiative frameworks that aim to promote systematic and transparent evaluation of (eco)toxicity research studies for the purpose of health and environmental hazard and risk assessment of chemicals have been developed. The proposed frameworks include detailed criteria for evaluating reliability and relevance of toxicity and ecotoxicity studies. The output can for example be used for ranking studies or for categorizing them as “reliable/relevant without restrictions”, “reliable/relevant with restrictions”, “not reliable/relevant” or “not assignable”. To facilitate the application of the evaluation criteria and illustrate the results of the evaluation, an online colour-coding tool has been developed. SciRAP also provides guidelines for researchers for how to report studies in order to meet the information requirements for regulatory hazard and risk assessment. The intended users of the SciRAP resources are risk assessors and researchers in the fields of ecotoxicology and toxicology. These tools and guidelines are publically available at the SciRAP website (www.scirap.org).

The overall objective of SciRAP has been to bridge the gap between academic research and chemicals regulation and to increase the usability of all studies that can contribute reliable and relevant data to health and environmental risk assessment. The purpose is to promote transparency, and to reduce scientific uncertainty in risk assessment conclusions, and in extension contribute to better-targeted policy decisions for health and environmental risk reduction. Ongoing work within the SciRAP project aims to investigate how the SciRAP approach can contribute to a wider “systematic review” or “weight-of-evidence” (WoE) approach. In such an approach available data is summarized and interpreted as one body of evidence, and evaluation of individual study quality, using predefined criteria and tools such as the ones provided in SciRAP, is one vital step.

The criteria for evaluating reliability and relevance of ecotoxicity studies have been developed in the CRED project (Criteria for Evaluating and Reporting ecotoxicity Data). The CRED evaluation method was developed from available methods and OECD reporting recommendations for aquatic ecotoxicity studies. In order to assess if the CRED evaluation method would be a suitable replacer of the Klimisch method a ring test was performed in which the two methods were compared. The ring test concluded that the risk assessors preferred the CRED evaluation method since it was a more transparent and more detailed method. Specific guidance for evaluation of nanoecotoxicity studies has also been developed. The

criteria for evaluating reliability and relevance of *in vivo* toxicity studies developed in SciRAP were primarily based on recommendations and requirements set out in corresponding OECD test guidelines for such tests. The method was first published by Beronius et al. (2014) but has since been updated.

It is of critical importance to evaluate the feasibility and usefulness of the proposed evaluation criteria and reporting guidelines by involving the intended end-users, i.e. persons working with evaluating (eco)toxicity studies for hazard and risk assessment and researchers generating such studies. To that end, “ringtests” involving participants from authorities, industry and academia around the world have been performed with the CRED criteria for ecotoxicity studies (Kase et al. 2016) and the criteria for *in vivo* toxicity studies (not yet published). In these ringtests participants have been invited to apply and evaluate the proposed criteria and online tools. The intent of this Nordic workshop was to continue this evaluation process by inviting persons from academia and authorities with the relevant expertise to further discuss the SciRAP tools and how they can be applied in practice. The workshop was held in Stockholm over two half-days, from lunch November 15 to lunch November 16, 2016.

Purpose and aim of the workshop

This workshop was organized together with the Nordic Risk Assessment Project (NORAP), and funded by grants from the Nordic Chemical Group (NKG) of the Nordic Council of Ministers. The purpose was to present and discuss the SciRAP tools among intended end-users in the Nordic countries. The specific aims were to:

- discuss the practical use of the SciRAP criteria and tools and how they meet the needs of end-users
- identify and discuss how SciRAP can be further improved and developed
- discuss the benefits and possibilities of implementing SciRAP on the EU-level

A detailed agenda, with specific questions for the discussions, is attached in Annex I.

Participants

The workshop had 31 participants from the different Nordic countries (see list of participants in Annex II).

Table 1. Summary of workshop participants.

	Denmark	Finland	Iceland	Norway	Sweden	Total
Participants total	7 ^a	5	2	7	10 ^b	31
Authority	3	5	0	3	5	16
Academia	4	0	2	4	5	15
Expertise						
Ecotoxicology	3	1	1	3	4	12
Toxicology	4	4	1	4	6	19

^a Six participants funded by the NKG grant, one additional participant funded by other means.

^b Including 4 from the organizing group and one student from Stockholm University

Pre-workshop exercise

As preparation for the workshop, participants were sent one toxicity or ecotoxicity study (depending on expertise) one month beforehand and instructions on how to evaluate it according to the SciRAP approach. This exercise was intended to provide participants with the relevant background so as to be prepared for an interactive exchange of views and expertise during the workshop.

Evaluation of a toxicity study

The participants with primary expertise in toxicology and health risk assessment were provided the study “Obesogen effects after perinatal exposure of 4,40-sulfonyldiphenol (Bisphenol S) in C57BL/6 mice” by Ivry del Moral et al. (2016). The task was to evaluate the relevance and reliability of this study as if it were included in a health risk assessment of Bisphenol S, and specifically for the setting of a European tolerable daily intake (TDI) for this compound.

Results

Evaluations of the provided *in vivo* toxicity study were received from 12 participants. A comparison shows that there was significant variation between reliability evaluations, both in how the importance of the different criteria were weighed in this specific case and how they were evaluated (Figures 1-4). In terms of relevance, all evaluators considered the study relevant for the purpose of setting a TDI, but several evaluators emphasized that it would be used together with other evidence and not on its own. Evaluation of relevance of *in vivo* toxicity studies is currently not quantitative, and is restricted to judging whether the study is relevant or not. Therefore, there is no “score” for relevance.

Evaluation of the reliability of *in vivo* toxicity studies using the SciRAP tool is divided into evaluation of “reporting quality” and “methodological quality”. Scores for reporting quality of the provided study varied between 47.89 and 89.13 (Figure 1). Only 10 of the 12 evaluators completed the evaluation of reporting quality due to technical issues. Looking more in detail at the evaluations of reporting quality, Figure 2 shows how each criterion was weighed and judged (fulfilled, partially fulfilled, not fulfilled or not determined) by the 10 different evaluators. Scores for methodological quality varied between 40.0 and 84.38 (Figure 3). Figure 4 shows how each criterion for methodological quality was weighed and judged by the 12 different evaluators.

Figure 1. SciRAP scores for the reporting quality of the provided *in vivo* toxicity study, as reported by 10 different evaluators. The score represents the sum of % fulfilled criteria and % partially fulfilled criteria divided by 2, and also accounts for the individual weights attributed to each criterion (see www.scirap.org for more details). Out of the original 12 evaluators, two, ID# 10 and 12, did not complete the evaluation of reporting quality due to technical difficulties.

Evaluator ID	1	2	3	4	5	6	7	8	9	11
The chemical name, ID or CAS-number of the test compound was given.	3	3	1	2	2	2	2	2	3	3
The purity of the test compound was stated or is traceable...	3	2	2	2	3	2	2	2	3	3
The vehicle was described.	2	2	3	2	2	2	2	2	3	2
It was stated that a negative control group was included.	2	1	3	2	2	2	2	2	3	2
Any positive control, if used, was described.	3	2	1	2	2	2	N/A	N/A	3	N/A
The animal model (species, strain, age or life stage and sex) was described.	3	2	3	2	2	2	2	2	3	2
The method for individual identification of animals was described.	2	2	1	2	2	2	2	2	2	2
The housing temperature was stated.	2	2	2	2	2	2	N/A	2	2	2
The relative humidity was stated.	2	2	1	2	2	2	N/A	2	2	2
The light-dark cycle was described.	2	2	1	2	2	2	1	2	2	2
The number of animals per sex in each cage was stated.	3	2	2	2	2	2	2	2	3	2
The cage materials were described.	2	2	3	2	3	2	2	2	2	2
Any materials used for physical enrichment were described.	2	2	3	2	3	2	N/A	2	2	2
Water bottle materials were described.	3	2	3	2	3	2	2	2	2	2
The bedding material used was described.	2	2	3	2	2	2	N/A	2	2	2
The type and source of feed were reported.	3	2	3	2	2	2	2	2	2	2
The source of drinking water was reported.	3	2	3	2	3	2	1	2	2	2
The administered dose levels or concentrations were stated.	3	2	3	2	2	2	2	2	3	3
The method for allocating animals to different treatments was stated.	3	2	1	2	2	2	2	2	2	3
The total number of animals per dose group was stated.	3	2	3	2	2	2	2	2	3	3
The route of administration was stated.	3	2	3	2	2	2	2	2	3	3
The sex and age (or life stage) of the animals at start of dosing was stated...	3	2	3	2	2	2	2	2	2	3
The test and/or analytical methods used were sufficiently described to allow for evaluation of the reliability of results.	3	2	3	2	2	2	2	2	3	3
The method for allocating animals to different tests and measurements was stated.	2	2	3	2	2	2	2	2	2	2
The sex, age and number of animals per dose group subjected to separate tests and measurements was stated.	3	2	3	2	2	2	2	2	3	2
The statistical methods and software used were described.	3	2	3	2	2	2	2	2	3	3
The statistical unit, e.g. the individual or the litter, was stated.	3	2	3	2	2	2	2	2	2	3
All results for the investigated endpoints were reported...	3	2	3	2	2	2	2	2	3	3
The funding sources for the study were stated.	3	2	3	2	2	2	2	2	2	3
Any competing interests were disclosed...	2	2	1	2	2	2	N/A	2	2	3

Figure 2. Evaluation of reporting quality of the provided *in vivo* toxicity study from 10 different evaluators. Green, yellow, red and grey indicate that the criterion was reported, partially reported, not reported or not determined, respectively. The numbers indicate the weight attributed to each criterion where 2 is default, and 1 or 3 means the criterion was considered less important or specifically important, respectively, in this case. N/A = the criterion was considered not applicable in this case and has been removed from evaluation. Two evaluators, ID# 10 and 12, did not complete the evaluation of reporting quality due to technical difficulties.

Figure 3. SciRAP scores for the methodological quality of the provided *in vivo* toxicity study, as reported by 12 different evaluators. The score represents the sum of % fulfilled criteria and % partially fulfilled criteria divided by 2, and also accounts for the individual weights attributed to each criterion (see www.scirap.org for more details).

Evaluator ID	1	2	3	4	5	6	7	8	9	10	11	12
The test compound or mixture was unlikely to contain any impurities that may significantly have affected its toxicity.	3	3	2	2	2	2	2	2	3	3	3	3
An appropriate vehicle was used that is not expected to interfere with the absorption, distribution, metabolism, excretion or toxicity of the test compound.	2	2	3	2	2	2	2	2	3	2	3	2
A concurrent negative control group was included.	1	2	3	2	2	2	2	2	3	3	3	3
An appropriate positive control group was included, and the expected result was observed from this treatment	3	1	1	2	2	N/A	N/A	N/A	2	2	N/A	N/A
A reliable and sensitive animal model was used for investigating the test compound and selected endpoints.	3	2	3	2	2	2	2	2	2	2	3	2
Animals were individually identified.	3	2	2	2	2	2	N/A	2	2	2	2	2
Housing conditions (temperature, relative humidity, light-dark cycle) were appropriate for the study type and animal model.	2	2	2	2	2	2	N/A	2	2	2	2	1
The number of animals per sex in each cage were appropriate for the study type and animal model.	3	2	3	2	2	2	2	2	2	2	2	2
The test system was unlikely to contain contaminants that could affect study results, such as organic pollutants, pesticide residues, heavy metals, and mycotoxins, as well as phytoestrogens.	3	2	3	2	3	2	2	2	2	2	3	2
The allocation of animals to different treatments was randomized.	3	2	1	2	2	2	2	2	2	2	3	2
The route of administration was appropriate and not likely to interfere with the study results.	2	2	3	2	2	2	2	2	3	2	3	3
The timing and duration of administration were appropriate for investigating the included endpoints.	3	2	3	2	2	2	2	2	3	2	3	2
The frequency of administration is appropriate for investigating the included endpoints.	2	2	3	2	2	2	2	2	2	2	3	2
The allocation of animals to different tests and measurements was randomized.	3	2	1	2	2	2	2	2	3	2	3	2
Reliable, and sensitive test methods were used for investigating the selected endpoints.	3	2	3	2	2	2	2	2	2	2	3	2
Measurements were collected at suitable time points in order to generate sensitive, valid and reliable data.	2	2	3	2	2	2	2	2	2	2	3	2
A sufficient number of animals per dose group were subjected to separate tests/data collection/measurements to generate reliable and valid results.	3	2	3	2	2	2	2	2	2	2	3	2
The statistical methods have been clearly described and do not seem inappropriate, unusual or unfamiliar.	3	2	3	2	2	2	2	2	3	2	3	2

Figure 4. Evaluation of methodological quality of the provided *in vivo* toxicity study from 12 different evaluators. Green, yellow, red and grey indicate that the criterion was fulfilled, partially fulfilled, not fulfilled or not determined, respectively. The numbers indicate the weight attributed to each criterion where 2 is default, and 1 or 3 means the criterion is considered less important or specifically important in this case, respectively. N/A = the criterion is considered not applicable in this case and has been removed from evaluation.

Evaluation of an ecotoxicity study

The participants with primary expertise in ecotoxicology and environmental risk assessment were provided the study “Toxic effects of the non-steroidal anti-inflammatory drug diclofenac. Part I: histopathological alterations and bioaccumulation in rainbow trout”, by Schwaiger et al. (2004). The task was to evaluate the relevance and reliability of this study as if it were included in an assessment for an aquatic Environmental Quality Standard (EQS) for diclofenac.

Results

Evaluations of the provided ecotoxicity study were received from 7 participants. A comparison shows that there were variations in how many criteria were considered fulfilled, and how the different criteria were evaluated (Figures 5-8).

In terms of relevance, all evaluators considered the study “relevant with/without restrictions” for the purpose of setting an EQS-value. Aspects discussed were whether the endpoint was relevant for effects on population level, and whether the study represented the preferred life stage, exposure duration, and experimental conditions. The evaluators judge the study to be “reliable with restrictions”. Aspects discussed were validity criteria, acclimatization of test organisms, solvent concentrations, spacing between concentrations, biomass loading, number of replicates, and access to raw data. The conclusions regarding the reliability and relevance of the study are in line with the conclusion from the EQS dossier from the Sub-group on Review of the Priority Substances, as well as the European Commission’s SCHER Committee.

Figure 5. SciRAP scores for the reliability of the provided ecotoxicity study, as reported by 7 different evaluators. The score represents the sum of % fulfilled criteria and % partially fulfilled criteria divided by 2, and also accounts for the individual weights attributed to each criterion (see www.scirap.org for more details).

Figure 6. SciRAP scores for the relevance of the provided ecotoxicity study, as reported by 7 different evaluators. The score represents the sum of % fulfilled criteria and % partially fulfilled criteria divided by 2, and also accounts for the individual weights attributed to each criterion (see www.scirap.org for more details).

Evaluation criteria	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6	ID 7
Is a guideline method (e.g., OECD/ISO) or modified guideline used? (of minor importance for study reliability)	Grey	Red	Red	Yellow	Red	Grey	Yellow
Is the test performed under GLP conditions? (of minor importance for study reliability)	Grey	Red	Red	Red	Red	Yellow	Red
If applicable, are validity criteria fulfilled (e.g. control survival, growth)?	Grey	Yellow	Yellow	Yellow	Red	Grey	Yellow
Are appropriate controls performed (e.g. solvent control, negative and positive control)?	Green	Green	Green	Green	Green	Yellow	Green
Is the test substance identified clearly with name or CAS-number? Are test results reported for the appropriate compound?	Yellow	Green	Green	Green	Green	Green	Green
Is the purity of the test substance reported? Or, is the source of the test substance trustworthy?	Green	Green	Green	Green	Green	Green	Green
If a formulation is used or if impurities are present: Do other ingredients in the formulation exert an effect? Is the amount of test substance in the formulation known?	Green	Grey	Grey	Green	Green	Yellow	Grey
Are the organisms well described (e.g. scientific name, weight, length, growth, age/life stage, strain/clone, gender if appropriate)?	Green	Green	Green	Green	Green	Yellow	Green
conditions? Have the organisms not been pre-exposed to test compound or other unintended stressors?	Green	Green	Green	Green	Red	Yellow	Green
Is the experimental system appropriate for the test substance, taking into account its physico-chemical characteristics?	Green	Green	Green	Green	Green	Green	Green
medium or test water, feeding, water characteristics, temperature, light/dark conditions, pH, oxygen content)? Have conditions been stable during the test?	Green	Green	Yellow	Green	Green	Yellow	Green
Were exposure concentrations below the limit of water solubility (taking the use of a solvent into account)? If a solvent is used, is the solvent within the appropriate range and is a solvent control included?	Green	Green	Yellow	Green	Red	Yellow	Green
Is a correct spacing between exposure concentrations applied?	Yellow	Green	Yellow	Green	Green	Red	Red
Is the exposure duration defined?	Green	Green	Green	Green	Green	Green	Green
Are chemical analyses adequate to verify concentrations of the test substance over the duration of the study?	Green	Green	Green	Green	Green	Green	Green
Is the biomass loading of the organisms in the test system within the appropriate range (e.g. < 1 g/L)?	Green	Green	Red	Green	Red	Grey	Red
Is a sufficient number of replicates used? Is a sufficient number of organisms per replicate used for all controls and test concentrations?	Yellow	Yellow	Yellow	Yellow	Green	Red	Yellow
Are appropriate statistical methods used?	Green	Green	Green	Yellow	Green	Grey	Green
Is a concentration-response curve observed? Is the response statistically significant?	Grey	Yellow	Green	Green	Green	Yellow	Green
Are sufficient data available to check the calculation of endpoints and (if applicable) validity criteria (e.g., control data, concentration-response curves)?	Yellow	Red	Red	Green	Yellow	Yellow	Green

Figure 7. Evaluation of reliability of the provided ecotoxicity study, as reported by 7 different evaluators. Rows represent different evaluation criteria; columns represent the different evaluators. Green, yellow, red and grey indicate that the criterion was fulfilled, partially fulfilled, not fulfilled or not determined, respectively.

Evaluation criteria	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6	ID 7
Is the species tested relevant for the compartment under evaluation?	Green	Green	Green	Green	Green	Yellow	Green
Are the organisms tested relevant for the tested substance?	Green	Green	Green	Green	Green	Green	Green
Are the reported endpoints appropriate for the regulatory purpose?	Green	Yellow	Yellow	Green	Green	Yellow	Green
Are the reported endpoints appropriate for the investigated effects or the mode of action of the test	Green	Green	Green	Green	Green	Yellow	Green
Is the effect relevant on a population level?	Grey	Red	Yellow	Green	Red	Red	Green
Are appropriate life stages studied?	Green	Green	Green	Yellow	Red	Yellow	Yellow
Is the magnitude of effect statistically significant and biologically relevant for the regulatory purpose (e.g., EC10, EC50)?	Green	Green	Yellow	Green	Green	Yellow	Green
Are the experimental conditions relevant for the tested species?	Green	Green	Yellow	Yellow	Green	Yellow	Yellow
Is the exposure duration relevant and appropriate for the studied endpoints and species?	Green	Yellow	Green	Yellow	Yellow	Green	Yellow
If recovery is studied, is this relevant for the framework for which the study is evaluated?	Grey	Grey	Grey	Grey	Grey	Grey	Grey
In case of a formulation, other mixture, salts, or transformation products, is the substance tested representative and relevant for the substance being assessed?	Green	Grey	Grey	Grey	Grey	Green	Grey
Is the tested exposure scenario relevant for the substance?	Green	Green	Green	Green	Green	Green	Green
Is the tested exposure scenario relevant for the species?	Green	Green	Green	Green	Green	Yellow	Green

Figure 8. Evaluation of relevance the provided ecotoxicity study, as reported by 7 different evaluators. Rows represent different evaluation criteria; columns represent the different evaluators. Green, yellow, red and grey indicate that the criterion was fulfilled, partially fulfilled, not fulfilled or not determined, respectively.

Participant presentations

Participants were invited to share short presentations on specific cases or issues that have proven challenging from the hazard/risk assessment perspective and/or were relevant for discussing aspects of data evaluation for hazard and risk assessment. Five presentations were held on day 1 of the workshop and have been briefly summarized below.

Henning Clausen, Danish EPA: "Evaluating studies. Some general remarks"

Henning Clausen shared some general comments on the weaknesses in design of standard ecotoxicity and toxicity tests that limit conclusions that can be drawn from these tests. For example, you generally need to have more than 20% effect to achieve statistical significance.

He also discussed that differences in expert judgment will inevitably lead to differences in the evaluation of the "quality" of studies and their adequacy for use in hazard and risk assessment. Tools, such as the CRED criteria, may increase the harmonisation of how studies are evaluated but will not remove differences in conclusions from such evaluations. This is e.g. seen when comparing evaluations made by member state experts and experts from the involved industry.

Dag Eide, FHI: DecaBDE –neurotox/developmental neurotox a scirap case"

Dr. Eide presented an exercise conducted at FHI exploring the use of the SciRAP tools to evaluate the reliability and relevance of neurotoxicity and developmental neurotoxicity (DNT) studies on decaBDE. DecaBDE is a controversial case that has been discussed without reaching any consensus regarding risk for the past 14 years. Concern has been raised specifically regarding developmental toxicity and risk of this compound with a large body of non-guideline data from research studies reporting such effects. However, such studies show variations in results and generally have limitations that have hampered their use as evidence in risk assessment. A few guideline studies are available and they conclude that there are no DNT-effects of decaBDE. FHI attempted to conduct a systematic evaluation of the quality of the non-guideline neurotoxicity and DNT studies, using and adapting the SciRAP criteria and tool for evaluation of in vivo studies. The SciRAP spreadsheet was ported to a Filemaker Pro database that is also available as a standalone application. This app generates colour profiles for both reporting and methodological quality of the studies and also suggests a qualitative output for the study relevance (Figure 9). Another feature is that several studies can be evaluated graphically in the same panel. The results of the evaluation provided a transparent overview of the quality of the included studies. It also enabled identifying the specific limitations of the studies and their impact on study quality. As such, it provided a useful basis for further discussions about the use of these data in risk assessment of decaBDE. The exercise also generated insight into possible improvements to the SciRAP criteria and tool that could further enhance its application in study evaluation for risk assessment.

Figure 9. Colour profiles for reporting quality, methodological quality and relevance for 5 different *in vivo* toxicity studies generated by importing the evaluation conducted in the SciRAP tool to the Filemaker Pro application.

Hinni Papponen, Tukes: “DIURON – Substance evaluation under REACH”

Dr. Papponen presented the work of the Finnish Safety and Chemicals Agency (Tukes) to conduct a substance evaluation for Diuron. Diuron is regulated under REACH as a process regulator for polymerisation processes in production of resins, rubbers and polymers. It is also regulated under the Plant Protection Product Regulation as an herbicide and under the Biocidal Product Regulation as a preservative. It was brought up for substance evaluation based on concerns regarding its potential endocrine disrupting properties, as well as its wide dispersive use and pollution in ground and surface water. Dr. Papponen described the weight of evidence evaluation conducted at Tukes, evaluating both ecotoxicity and toxicity data to assess the hazards and endocrine disrupting potential of diuron. In addition to Klimisch method the SciRAP tool was used to evaluate reliability of the scientific papers. It was concluded that, based on ecotoxicity data, endocrine disrupting properties of diuron could not be

ruled out in terms of effects in the environment. However, available mammalian toxicity data did not allow to conclude that diuron is an endocrine disrupter for mammalian species.

Sofie Christiansen, DTU: "Evaluation of EFSA's new Scientific Opinion on Bisphenol A by DTU Food"

Dr. Sofie Christiansen shared work conducted at the National Food Institute (at the Technical University of Denmark, DTU) in relation to evaluation of effects of bisphenol A (BPA) on the developing mammary gland and their implications for the health risk assessment of BPA. Adverse effects on kidneys in adult animals and effects on the mammary gland development were in a recent EFSA opinion on BPA evaluated as "likely" and only these effects are taken forward to the EFSA hazard characterization. The EFSA hazard characterization was based on benchmark dose (BMD) response modelling, which is generally regarded as the best method if data allow this. This was only possible for kidney effects while effects on mammary gland development, a BMDL10 could not be derived due to very large confidence intervals on the BMDL's estimated from the models used. EFSA established a new temporary TDI for BPA at 4 µg/kg bw/day, based on the effects on the kidney in adult animals. In contrast, DTU decided to conduct a "classical" hazard characterisation of the effects on the mammary gland development, using observed NOAELs or LOAELs to estimate a sufficiently protective TDI for BPA. Based on this approach DTU concluded that the TDI for BPA should be 0.7 µg/kg bw/day or lower to be sufficiently protective with regards to endocrine disrupting effects of BPA. New data from a large study on BPA supports this TDI of 0.7 µg/kg or lower. Moreover DR. Christiansen mentioned that exposure to BPA can be 1.4-2 times higher than 0.7 µg/kg bw/day and DTU finds that this gives rise to concern with regards to risk for health effects of BPA for highly exposed persons.

Nanna Hartmann, DTU: "NanoCRED - a new tool for evaluating nanoecotoxicity studies"

Dr. Hartmann presented the new initiative to develop a tool, based on the CRED criteria and SciRAP online tools, for evaluating nanoecotoxicity studies. OECD test guidelines, that include specific considerations for testing nanomaterials, are currently lacking. Risk assessment therefore relies very much on the incorporation of data from academic research studies or tests performed according to modified test guidelines. Nanomaterials present additional challenges to risk assessment, in part because of their physico-chemical properties, the presence of coatings and ability to aggregate. Differences in these parameters can result in different (eco)toxicity and risks of two nanomaterials with the same chemical composition. It is therefore critical that, for example, studies clearly and unambiguously identify the nanomaterial that has been tested so that comparison to the material being risk assessed is possible. Consequently, in the NanoCRED framework, CRED reliability criteria, and accompanying guidance, have been modified to accommodate the properties and behaviour of nanomaterials in ecotoxicity test systems. Certain CRED guidance on relevance criteria have further been refined to include aspects specific to nanomaterials. A publication presenting this initiative has been submitted for publication and the NanoCRED tool will become available on the SciRAP website shortly.

Summary of breakout sessions

Breakout session 1

The purpose of breakout session 1 was to discuss the practical use of the SciRAP tools based on the pre-workshop exercise (see agenda, Annex I, for list of specific discussion points). The participants were divided into three groups (two groups with toxicity expertise and one group with ecotoxicity expertise) to discuss each item and then reported back in plenum. The participants' comments to each discussion point are summarized below.

General experience from the exercise in preparation for the workshop.

The participants were generally positive concerning the SciRAP tools for evaluating *in vivo* toxicity and ecotoxicity studies, particularly the increased transparency and harmonization of aspects to consider when evaluating study quality. Training with the tool may be an important feature to further increase harmonization and agreement between evaluators. The possibility to save the evaluation result and share with colleagues was considered useful. The evaluators also acknowledged the importance of the reporting recommendations since peer-reviewed studies often are poorly reported.

However, for *in vivo* toxicity studies some remarked that using the SciRAP tool was time consuming and that the stepwise process of first weighing criteria and then evaluating reporting quality and methodological quality was unnecessarily complicated and confusing. The purpose and process of weighing criteria seemed to have been especially unclear, and the usefulness of default weight of 2 was questioned. It was suggested to improve instructions and guidance to facilitate evaluation. After discussing the different criteria and tool some participants felt they would have evaluated the study in the pre-workshop exercise differently.

Suggestions for improvements of the tool were, specifically:

- Make less complex with fewer steps.
- Include a function that gives a warning if all criteria have not been addressed.
- Make it possible to save evaluations as you are working on them and facilitating coming back to evaluations to make changes.
- Make it possible to print out or save the results of the evaluation to facilitate comparing the results of several different studies.

Do the proposed criteria to evaluate (eco)toxicity studies for risk assessment meet the needs of the Nordic regulatory agencies? What works well? What can be improved?

Many of the same comments were made as for the first discussion point and the criteria and tool for evaluating *in vivo* toxicity and ecotoxicity studies were generally considered useful for the work conducted at the different Nordic agencies. However, some participants observed that this type of thorough evaluation of study reliability and relevance would only be feasible in certain cases, e.g. areas of controversy, or only for key studies. The flexibility of the tool, which allows it to be adjusted for different hazard and risk assessment situations and questions, was noted as a strength. The tool was also considered especially useful for identifying recurring weaknesses in study design and conduct for specific types of studies, for example developmental toxicity studies. It was discussed that some criteria could be combined, subdivided or rephrased. In addition, it was suggested that some criteria be added to better cover different types of studies (e.g. inhalation) and/or confounders or co-exposures (e.g. high-fat diet,

stress). Participants also discussed how uncertainty in the evaluation could be transparently expressed but no concrete suggestions were agreed upon. The ecotoxicity evaluation method is already used when deriving environmental quality standards within the Water Framework Directive. The experiences from that use are that it works satisfactorily.

Is the guidance provided adequate and sufficient? What can be improved?

The guidance was considered helpful but could be further developed and refined. Some aspects that could be useful guidance was missing, while other details that were included were not considered relevant. Guidance for when criteria are “partially fulfilled” is specifically critical as judging a criterion as “partially fulfilled” as opposed to “fulfilled” will have quite significant bearing on the evaluation outcome and SciRAP score. For ecotoxicity studies, additional guidance for the criteria concerning statistics was requested, as well as a clarification of certain criteria.

Is the colour-coding tool a useful tool in applying the evaluation criteria? What works well? What can be improved?

The colour profile for individual studies was considered very useful for visualisation of weaknesses in the study and disagreements between evaluators. It was also considered useful as a basis for discussions within expert groups, to focus on the most critical issues. However, it was noted that areas marked as yellow (i.e. “partially fulfilled” criteria) have to be interpreted with caution, given the subjectivity in judging criteria as “partially fulfilled”. It was also suggested to add a feature to enable “flagging” of criteria or issues that need follow-up, e.g. by contacting study authors. For *in vivo* toxicity studies, the tool only provides a colour profile for methodological quality and it was suggested to also add this feature for reporting quality as well as relevance.

How would you use the results from the SciRAP evaluation? E.g. to categorize studies into Klimisch categories, ranking of studies, etc?

Participants discussed that SciRAP scores would mainly be used as a basis for screening and ranking studies in regard to their reliability, as evaluated by this tool. The colour profiles were considered particularly useful as basis for discussing strengths and weaknesses of key studies. It was also noted that the output from the SciRAP evaluation could be used for “sensitivity analysis”, investigating the consequence of different evaluations for the final conclusions, such as setting a tolerable daily intake (TDI). Importantly, more specific guidance for how to use the output from the SciRAP tool within EU authorities, such as EFSA or ECHA, should be developed. The evaluators also requested a function at the webpage where several studies can be compared. For ecotoxicity studies, the weighing of the criteria was hardly used but still the evaluators saw advantages of having that option.

Breakout session 2

The purpose of breakout session 2 was to discuss the use of the SciRAP tools for regulatory purposes at Nordic and European authorities (see agenda, Annex I, for list of specific discussion points).

The participants were divided into two groups (one with toxicity expertise and one with ecotoxicity expertise) to discuss each item and then reported back in plenum. The participants’ comments to each discussion point are summarized below.

Can the SciRAP tools be useful for regulatory risk assessment at Nordic authorities? At EU authorities?

In general, yes, although the needs and requirements will differ at different authorities and within different regulatory frameworks, as well as at different levels of hazard and risk assessment. It should be clearly communicated that the tool can be modified for use within different regulatory frameworks and for different problems or hazard and risk assessment questions.

How can the output from the SciRAP data evaluation be coupled to weight-of-evidence- or systematic review approaches used at different authorities?

Apart from the guidance on conducting systematic review published by EFSA, the concept of systematic review does not yet seem to be very well established at authorities responsible for hazard and risk assessment of chemicals. Further illustrations of systematic review in this context, as well as where in that process evaluation of single studies takes place, are needed. It was suggested that case studies would be a good method to illustrate and improve understanding of such issues.

How can the SciRAP tools be introduced at the EU-level, e.g. at ECHA, EFSA, etc? What channels/opportunities are available?

It was suggested to conduct case studies relevant for regulatory purposes on the EU level using the SciRAP tools. Such case studies could for example be based on substance evaluations under CoRAP and be carried out together with responsible member state authorities. These activities would illustrate the use of the tools in a real-life regulatory setting, which would also be relatively easy to communicate to relevant EU authorities.

Possibilities to address OECD and EU authorities and working groups involved in development of hazard and risk assessment guidance were also discussed. For example, it was suggested to approach the EFSA Working Group on the Application of systematic review methodology, as well as persons at ECHA responsible for development of guidance on weight of evidence evaluations and guidance for study evaluations in IUCLID.

Organising webinars were considered a useful and feasible approach to provide information about the SciRAP tools as well as more practical training. Webinars could easily be tailored to different groups of participants with different background, interests and needs.

It was also suggested to write letters to editors of scientific journals describing the problem with poorly reported *in vivo* toxicity and ecotoxicity studies, and how that hampers their use in hazard and risk assessment of chemicals.

Overall conclusions from the workshop

Overall, the following main conclusions were drawn based on discussions during the workshop and will provide the basis for moving forward in further developing and disseminating the SciRAP tools:

- SciRAP provides tools that can be useful for evaluating the reliability and relevance of toxicity and ecotoxicity studies for hazard and risk assessment in the Nordic countries.
- Future improvements to the website and online tools should include making the tool more user-friendly by reducing the number of steps. For example, by combining the weighing and evaluation of criteria so that it is done simultaneously in one step.

- Results of the reliability and relevance evaluations in the pre-workshop exercise varied between evaluators, especially the weights attributed to individual criteria in the evaluation of the toxicity study. After discussions, several participants noted that they would probably have adjusted their evaluations based on a better understanding of the criteria. It was discussed that some training with the tools before applying them would be useful and may result in smaller variations between evaluators, although this cannot be guaranteed.
- Guidance for how to interpret and use the output from the SciRAP tools, i.e. the SciRAP scores and colour profiles, is needed. This is especially important if the SciRAP approach is to be disseminated on the EU-level. Guidance should include how to express uncertainty in the evaluation in a structured manner.
- It was acknowledged that experts will never reach full agreement concerning the reliability and relevance of (eco)toxicity studies. The application of expert judgment is an integral part of evaluating studies for hazard and risk assessment and will inevitably lead to differences in study evaluation. Importantly, the aim of SciRAP is not to try to eradicate the need for expert judgement. However, by providing evaluators with a tool, such as SciRAP, evaluations become more transparent and discussions to explain and resolve differences are facilitated.

Workshop evaluation

After the workshop, participants were invited to evaluate the workshop in a short online evaluation form. Twenty-one of 27 (78%) external participants (i.e. excluding the organising group) responded in the workshop evaluation. A summary of the evaluation is provided in Appendix III to this report.

Evaluations were in general very positive; the workshop seems to have met the expectations of participants completely (35% of the answers) or almost completely (65% of the answers). All responded that they had gained new insights that were useful for their work. Forty-five percent of the participants answered that they had gained sufficient information at the workshop to be able to apply SciRAP in their work, while 50% answered that they had gained sufficient knowledge to look up information they need to apply SciRAP and 5 % would like to have had more information on some aspects. Commonly, participants would have liked to have practical training with the tools and more discussions on the pre-workshop exercises. Some suggestions for how to improve the practical aspects of the workshop, such as the venue, discussion sessions and accommodations, were also provided.

Next steps

Based on the discussions at the workshop, some future activities within the SciRAP initiative are planned. These include further development of the SciRAP webpage and tools, as well as activities for training and dissemination.

Development of the SciRAP webpage and tools

Up-dates and further development and refinement of the SciRAP on-line tools is currently underway. This work includes:

- Assessment and updates to increase user-friendliness of the webpage and existing online tools for evaluation of ecotoxicity and *in vivo* toxicity studies. For example, to limit the number of steps and simplify the evaluation process. Funding to carry out part of this work during January – April, 2017, has already been provided by the Swedish Research Council Formas.
- Development of tools for evaluation of *in vitro* studies following the same model as the tools for evaluating *in vivo* toxicity studies. This new tool is planned to be available online after summer, 2017.
- Publishing the NanoCRED tool for evaluating nanoecotoxicity studies on the SciRAP website.
- A scientific publication presenting the updated criteria for evaluating *in vivo* toxicity studies.

Training and dissemination

Webinars to provide forums for further discussions with potential end-users of the SciRAP tools on the national, Nordic, EU and international levels are planned. Persons from different sectors, e.g. authority, academia and industry, will be invited to the webinars. Different webinars will be organised with the aim to inform and discuss the SciRAP tools and to provide practical training with the tools, respectively. At least one training webinar is planned in April 2017 and for which preparation and organisation has been funded by the Swedish Research Council Formas. Additional webinars are planned to be organised later during 2017.

Different national, Nordic and EU authorities will be contacted during early 2017 and offered to host seminars for presenting the SciRAP tools to interested persons within their organisations. Suggestions were made at the workshop to contact, for example, relevant persons and working groups at ECHA, EFSA and the OECD. Participants at the workshop provided some names of specific contacts that could be useful. In addition, relevant persons within the contact networks of the SciRAP working group will be approached.

We also plan to, together with some of the workshop participants, write letters to editors of scientific journals describing the problem with poorly reported (eco)toxicity studies and consequences for the systematic inclusion of academic research in regulatory hazard and risk assessment of chemicals. In this context, tools and guidance that can be used to improve reporting of studies will be discussed.

Acknowledgements

This workshop was made possible by grants from NKG. We would like to sincerely thank all the participants for contributing their time and expertise. We would also like to acknowledge and thank the members of the Nordic Risk Assessment Project (NORAP): Lena Höglund, Cecile Blom, Tor Fotland, Ísak Sigurjón Bragason, Lars Andersson, Daniel Borg and Tiina Suutari, for their valuable contributions to the planning and organization of the workshop.

References

- Beronius A, Johansson N, Rudén C, Hanberg A. 2013. The influence of study design and sex-differences on results from developmental neurotoxicity studies of bisphenol A, implications for toxicity testing. *Toxicology* 311:13-26.
- Beronius A, Rudén C, Håkansson H, Hanberg A. 2010. Risk to all or none? A comparative analysis of controversies in the health risk assessment of Bisphenol A. *Reprod Toxicol* 29(2): 132-146.
- ECHA. 2013. European Chemicals Agency. Evaluation under REACH, Progress Report 2012. Available on-line at: echa.europa.eu.
- EFSA. 2010. European Food Safety Authority Guidance: Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal* 8(6), 90 pp. Available on-line at: www.efsa.europa.eu.
- Kase R, Korkaric M, Werner I, Ågerstrand M. 2016. Criteria for Reporting and Evaluating ecotoxicity Data (CRED): comparison and perception of the Klimisch and CRED methods for evaluating reliability and relevance of ecotoxicity studies. *Environmental Sciences Europe* 28:7.
- Kortenkamp A, Martin O, Faust M, Evans R, McKinlay R, Orton F, Rosivatz E. 2012. State of the Art Assessment of Endocrine Disrupters - Final Report. EU Commission DG Environment.
- Molander L, Ågerstrand M, Beronius A, Hanberg A, Rudén C. 2014. "Science in Risk Assessment and Policy (SciRAP): An Online Resource for Evaluating and Reporting In Vivo (Eco) Toxicity Studies." *Human and Ecological Risk Assessment*. 21 (3), 753-762.
- Myers JP, vom Saal FS, Akingbemi BT, et al. 2009. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ Health Perspect* 117(3): 309-315.
- OECD. 2005. Manual for the assessment of chemicals. Chapter 3: Data evaluation. Available on-line at: <http://www.oecd.org>.
- SCENIHR. 2012. Scientific Committee on Emerging and Newly Identified Health Risks. Memorandum on the use of the scientific literature for human health risk assessment purposes - weighing of evidence and expression of uncertainty. Available on-line at: ec.europa.eu.
- Zoeller RT, Brown TR, Doan LL, et al. 2012. Endocrine-disrupting chemicals and public health protection: a statement of principles from The Endocrine Society. *Endocrinology* 153(9): 4097-4110.
- Ågerstrand M, Kuster A, Bachmann J, et al. 2011. Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. *Environ Pollut* 159(10): 2487-2492.

Appendix I – Workshop agenda and discussion points

Workshop: “Bridging the gap between academic research and chemicals regulation - the SciRAP tool for evaluating toxicity and ecotoxicity data for risk assessment of chemicals”

November 15 – 16, 2016

Karolinska Institutet, Stockholm

Venue: Room 216, Retzius väg 13A

Workshop Agenda

Day 1:

Time	Activity	Speaker
12:00 – 13:00	<i>Arrival and lunch</i> Room 216, Retzius väg 13A	
13:00	Welcome, purpose and outline of the workshop	Anna Beronius
13:10	Presentations of the SciRAP working group and workshop participants.	All
13:20	Background and presentation of the SciRAP initiative	Marlene Ågerstrand Anna Beronius
13:50	Participant presentations, 10-15 minutes each: Current approaches to data evaluation for risk assessment – benefits, challenges, practical examples	Henning Clausen, Danish EPA Dag Markus Eide and Trine Husøy, FHI
14:20	<i>Coffee break</i>	
14:45	Continued: Current approaches to data evaluation for risk assessment – benefits, challenges, practical examples	Hinni Papponen, Tukes Sofie Christiansen, DTU
15:15	”NanoCRED - a new tool for evaluating nanoecotoxicity studies”	Nanna Hartmann, DTU
15:30	Break-out session: Discussions on the practical use of the SciRAP tools based on the pre-workshop exercise. Group 1: Evaluation of ecotoxicity studies Groups 2 and 3: Evaluation of toxicity studies	One person takes notes from each group discussion
17:00	Groups report back main conclusions. Discussions in plenum.	
18:00	End of discussions	
18:30	<i>Joint workshop dinner at the hotel for those who want</i>	

Day 2:

Time	Activity	Speaker
9:00	Welcome and short introduction to next break-out session	Anna Beronius Marlene Ågerstrand
9:15	Break-out session: The use of the SciRAP tools for regulatory purposes at Nordic and European authorities. Group 1: Ecotoxicity/environmental focus Groups 2 and 3: Toxicity/human health focus	One person takes notes from each group discussion.
10:15	<i>Coffee break</i>	
10:45	Continue break-out session	One person takes notes from each group discussion.
11:45	Groups report back main conclusions. Discussions in plenum.	All
12:30	Sum-up of discussion points and conclusions from the workshop discussions.	Moderator Anna Beronius Marlene Ågerstrand
13:00	End of workshop	

Discussion points

Break-out session Day 1:

- General experience from the exercise in preparation for the workshop.
- Do the proposed criteria to evaluate (eco)toxicity studies for risk assessment meet the needs of the Nordic regulatory agencies? What works well? What can be improved?
- Is the guidance provided adequate and sufficient? What can be improved?
- Is the colour-coding tool a useful tool in applying the evaluation criteria? What works well? What can be improved?
- How would you use the results from the SciRAP evaluation? E.g. to categorize studies into Klimisch categories, ranking of studies, etc?

Break-out session Day 2:

- Can the SciRAP tools be useful for regulatory risk assessment at Nordic authorities? At EU authorities?
- How can the output from the SciRAP data evaluation be coupled to weight-of-evidence- or systematic review approaches used at different authorities?
- How can the SciRAP tools be introduced at the EU-level, e.g. at ECHA, EFSA, etc?
 - Is there a need for such a tool?
 - What channels/opportunities are available? E.g. opportunities for presentations/workshops on the EU level.

Appendix II – List of participants at the SciRAP workshop 15 – 16 November, 2016.

Names in italics indicate persons in the organising group.

Last name	First name	Primary expertise	Affiliation	Country
Andersen	Sjur	ecotox	Norwegian Environment Agency	Norway
Andersson	Lars	ecotox	Swedish Chemicals Agency	Sweden
Appelgren	Henrik	tox	Swedish Chemicals Agency	Sweden
<i>Beronius</i>	<i>Anna</i>		Institute of Environmental Medicine, KI	Sweden
Bjørge	Christine	tox	Norwegian Environment Agency	Norway
Christiansen	Sofie	tox	Technical University of Denmark	Denmark
Clausen	Henning	ecotox	Danish EPA	Denmark
Dahlgren	Elin	ecotox	Swedish EPA	Sweden
Eide	Dag Markus	tox	Norwegian Institute of Public Health	Norway
Glynn	Anders	tox	National Food Agency	Sweden
Halldórsson	Halldór Pálmar	ecotox	University of Iceland	Iceland
<i>Hanberg</i>	<i>Annika</i>		Institute of Environmental Medicine, KI	Sweden
Hartmann	Nanna	ecotox/nano	Technical University of Denmark	Denmark
Husøy	Trine	tox	Norwegian Institute of Public Health	Norway
Ingre-Khans	Ellen		ACES, Stockholm University	Sweden
Lillicrap	Adam	ecotox	Norwegian Institute for Water Research (NIVA)	Norway
Meister Beltoft	Vibe	tox	Technical University of Denmark	Denmark
Mörck	Anna-Karin	tox	Swedish Chemicals Agency	Sweden
Murtomaa-Hautala	Mari	ecotox	Regional State Administrative Agency for Northern Finland (AVI)	Finland
Nielsen	Elsa	tox	Technical University of Denmark	Denmark
Ólafsdóttir	Kristín	tox	University of Iceland	Iceland
Palomäki	Jaana	tox	Finnish Safety and Chemicals Agency (Tukes)	Finland

Papponen	Hinni	tox	Finnish Safety and Chemicals Agency (Tukes)	Finland
Petersen	Karina	ecotox	Norwegian Institute for Water Research (NIVA)	Norway
Reiler	Emilie	ecotox	Danish EPA	Denmark
Rudén	Christina		ACES, Stockholm University	Sweden
Stockmann-Juvala	Helene	tox	Finnish Institute of Occupational Health	Finland
Suutari	Tiina	tox	Finnish Safety and Chemicals Agency (Tukes)	Finland
Tyle	Henrik	tox	Danish EPA	Denmark
van der Hagen	Marianne	tox	Norwegian Environment Agency	Norway
Ågerstrand	Marlene		ACES, Stockholm University	Sweden

Appendix III – Evaluation of SciRAP workshop 15-16 November, 2016

1. How well did the workshop meet your expectations? (21 answers)

2. Did you gain any new insights at the workshop that are useful for your work? (21 answers)

Comments

- a better understanding of the scirap-tool and ways to use it
- As a head of group I would like to introduce the tool to my colleagues to see how much we would agree by scoring the same paper and then take a discussion in the group why/why not and learn from that
- Actually not so much. But nice to meet colleagues and exchange views / ideas especially with the SCIRAP folks...

- I think SciRap will be a valuable tool also for our agency's risk assessors that are not working on the international arena.
- Yes, I find the tool useful as a checklist when evaluating a study. It also makes the assessment more transparent. The problematic thing is the weighing. There will always be an element of subjectivity which is not possible (not even desirable) to remove. Expert judgement is in the end always needed.
- The use of SciRAP in REACH processes could be used and that this tool can be used internationally as well.
- How to properly use the tool to evaluate studies. Also good to know that the reporting checklist exists
- Mycket användbart verktyg!
- I gained new insight through the discussions on the practical use of the SciRAP tool, as well as the pre-workshop exercise, which were useful for my work. I found the comments and views of the working group and participants on the review process particularly useful.
- A good evaluation tool
- meeting more people with more experience in scientific reviews would have been even more rewarding.
- Better understanding of evaluation of toxicity data. Interesting talks.

3. Did you gain sufficient information at the workshop to be able to apply SciRAP in your work? (21 answers)

Comments

- I would have appreciated if we had gone through an example for the use of SciRAP.
- Hopefully we'll come back to this in a practical way if it based on a SCIRAP proposal will be agreed to do e.g. one of the SEV evaluations next year in each the Nordic countries , e.g. if each country could sign up to the idea to use the SCIRAP reliability rating for key studies on endpoints of concern...
- I enjoy the colour coding tool and the guidance given
- but scirap must be combined with other tools, I will keep you updated

4. Did the workshop provide you with the opportunity to network with your peers? (21 answers)

5. What would you have liked to discuss more at the workshop? (15 answers)

- How we evaluated the case study differently, and why
- Discussed the tool when going through an example
- The benefits of using the Tool..
- The audience was very mixed. Some are actually evaluating scientific articles and test guideline reports for regulatory use and have (some) knowledge about tox/ ecotox nasd regulatory use of such data - others not (so much)... Therefore also rather mixed discussion sometimes...
- Good balance
- Nothing
- How the Scope of SciRAP can be used also in US and OECD. It was mentioned that some working groups on pesticide have said that this tool is not relevant for their work. I would have liked to challenge this.
- More discussion about the challenges in the practical risk assessment in different agencies/organisations, and whether SciRAP could have been useful for some cases.
- development of the in-vitro tool
- Relevance of in vivo animal assays for predicting effects in humans
- Good as it was.
- The exercise
- Go through the practical examples how to use SciRAP
- klimisch, the ethical scores, ioannidis score sheets,
- How this tool could be implemented within regulatory frameworks

6. What would you have liked to discuss less at the workshop? (6 answers, two comments have been removed to protect the anonymity of the evaluation.)

- How to implement it - it is very hypothetical
- Good balance
- Nothing
- All the issues were relevant!
- Good as it was.
- Issues not directly relevant to the SciRAP Tool and exercise (although probably relevant for risk assessment of Chemicals in general)

7. Are there some practical aspects of the workshop that could have been improved (e.g. concerning travel, accommodation, venue, social program)? (15 answers)

- We could have used longer time to discuss the case to become more aware of different interpretations, and get better prepared for either coming discussion when using the tool later on for other studies. Break-out group on day 1 was a total mess in my group, thank you for organising it differently on day 2.
- all were fine.
- none - all worked well
- Maybe just one day meeting instead of two half days
- It would have been useful if in the invitation documents/email that it had been made clear that when signing up to the workshop it was only needed to book flights and arrange transport but that the hotel room was booked (or have I missed a written information about this ???). This would have avoided the situation that I first found out when arriving at the hotel - and then it was too late to cancel the room at the hotel next to the one where we had dinner. (When ordering flight and hotel I tried to ask the Travel Agent to book at the hotel but it was stated to be fully booked and so I had to ask the Agency to find a hotel nearby...)
- Practical arrangements worked well, thanks!
- No
- The meeting room was not optimal.
- No I liked the venue and accommodation
- Proper lunch for both days
- Anna Beronius and Marlene Ågarstrand did a great job in organizing the workshop and giving us sufficient and clear information on the practical aspects. Nothing to complain about.
- Venue - meeting room suited for students but not for me as I cannot handle papers etc. without a table
- Chairing of the discussion could have been more organised
- its fine
- All seemed to work smoothly

8. Please add any additional comments that can help us improve future events like this.

- In the last RIME meeting they seated groups of people at tables for discussions. This worked very well. Could have made more people participate more actively in the workshop.
- Thank you for doing this - I know that it is time consuming to get a meeting like this up and running.
Can't recall if this was discussed, but a repository with all the results from papers tested through this tool could be valuable for other peers. Also one could see how papers produced by a certain group are scored by others. It could - maybe - also be of help for you to see if there are certain questions that are always answered "partially" - that could indicate that the question might need to be changed (a bit)?
- Good luck :-)
- Thank's for your good work.
- Please consider using OECD as a platform and maybe also in collaboration with EFSA

- Please provide better instructions for the pre-exercise, also include all available guidances because only few had found them on the website. Could partly explain the huge variability in the results.
- Discuss relevance (and lack of relevance) of in vivo animal assays for predicting effects in humans
- one day is enough - start 1000 - end at 1700