Instructions for evaluating the reliability and relevance of *in vivo* toxicity studies using the SciRAP tool.

Introduction:

The SciRAP tool for evaluating *in vivo* studies allows for evaluation of reliability (divided into reporting quality and methodological quality) and relevance. The evaluation often has to be endpoint-specific, meaning that the evaluation is carried out focusing on one of several endpoints/effects investigated in the study. Separate evaluations may thus be necessary for different endpoints in one study.

Download the excel file containing the assessment sheet available on the SciRAP website. The assessment sheet contains pre-defined criteria/items to be evaluated in 3 sections for **reporting quality**, **methodological quality**, and **relevance**. The reporting and methodological quality sections are divided in specific categories: Test compound and controls, Animal models and housing conditions, Dosing and administration of the test compound, Data collection and analysis, Funding and competing interests (only in the reporting quality section), and Other (**Fig. 1**).

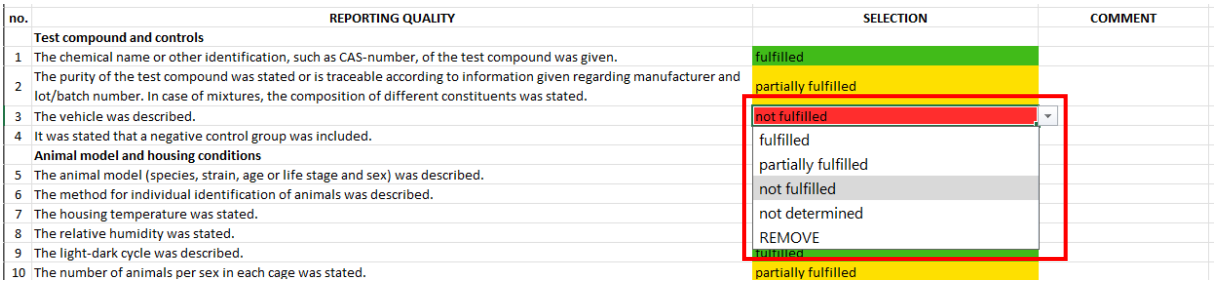
	METHODOLOGICAL QUALITY	SELECTION	COMMENT
	Test compound and controls		
1	The test compound or mixture was unlikely to contain any impurities that may significantly have affected its toxicity.		
2	An appropriate vehicle was used that is not expected to interfere with the absorption, distribution, metabolism, excretion or toxicity of the test compound.		
3	A concurrent negative control group was included.		
	Animal model and housing conditions		
4	A reliable and sensitive animal model was used for investigating the test compound and selected endpoints.		
5	Animals were individually identified.		
6	Housing conditions (temperature, relative humidity, light-dark cycle) were appropriate for the study type and animal model.		
7	The number of animals per sex in each cage were appropriate for the study type and animal model.		
8	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.		
	Dosing and administration of the test compound		
9	The allocation of animals to different treatments was randomized.		
10	The route of administration was appropriate and not likely to interfere with the study results.		
11	The timing and duration of administration were appropriate for investigating the included endpoints.		
12	The frequency of administration was appropriate for investigating the included endpoints.		
	Data collection and analysis		
13	The allocation of animals to different tests and measurements was randomized.		
14	Reliable and sensitive test methods were used for investigating the selected endpoints.		
15	Measurements were collected at suitable time points in order to generate sensitive, valid and reliable data.		
16	A sufficient number of animals per dose group were subjected to separate tests/data collection/measurements to generate reliable and valid results.		
17	The statistical methods have been clearly described and do not seem inappropriate, unusual or unfamiliar.		
	Other		
18	Are there any other aspects of study design, performance or reporting that influence reliability? (Comment in free text.)		

***Fig. 1** Categories of criteria in Methodological Quality section of the SciRAP tool.*

Evaluation may be conducted for either reporting quality, methodological quality, or relevance, or all three, depending on the purpose of evaluation. Although not required, evaluating reporting quality of the study before moving into the evaluation of methodological quality and relevance may in some cases save time and resources as it allows for identification of studies that have obvious deficiencies in reporting, hampering further evaluation.

Evaluation of the criteria:

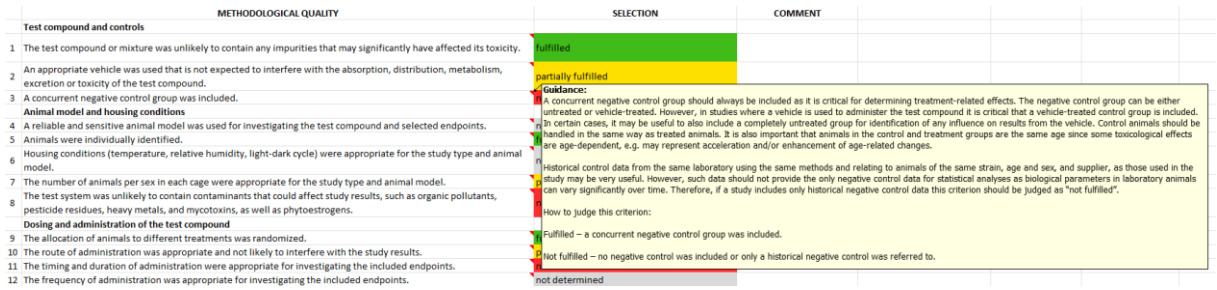
When you evaluate the criteria/items, choose one of the options from the drop-down menu in the "SELECTION" column (fulfilled, partially fulfilled, or not fulfilled for reporting and methodological quality, directly relevant, indirectly relevant, or not relevant in the relevance section, (**Fig. 2**). This drop-down menu is in almost every cell in the "SELECTION" column.



no.	REPORTING QUALITY	SELECTION	COMMENT
	Test compound and controls		
1	The chemical name or other identification, such as CAS-number, of the test compound was given.	fulfilled	
2	The purity of the test compound was stated or is traceable according to information given regarding manufacturer and lot/batch number. In case of mixtures, the composition of different constituents was stated.	partially fulfilled	
3	The vehicle was described.	not fulfilled	
4	It was stated that a negative control group was included.		
	Animal model and housing conditions		
5	The animal model (species, strain, age or life stage and sex) was described.		
6	The method for individual identification of animals was described.		
7	The housing temperature was stated.		
8	The relative humidity was stated.		
9	The light-dark cycle was described.		
10	The number of animals per sex in each cage was stated.	partially fulfilled	

Fig. 2 *Drop-down menu for the criteria in Reporting and Methodological Quality sections of the SciRAP tool.*

Guidance for evaluating individual methodological quality criteria and relevance items is available by pointing to the criterion with the cursor (the criterion containing the guidance has a red right corner, **Fig. 3**).

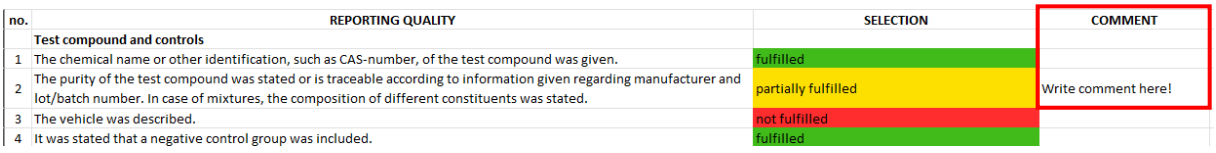


	METHODOLOGICAL QUALITY	SELECTION	COMMENT
	Test compound and controls		
1	The test compound or mixture was unlikely to contain any impurities that may significantly have affected its toxicity.	fulfilled	
2	An appropriate vehicle was used that is not expected to interfere with the absorption, distribution, metabolism, excretion or toxicity of the test compound.	partially fulfilled	
3	A concurrent negative control group was included.		
	Animal model and housing conditions		
4	A reliable and sensitive animal model was used for investigating the test compound and selected endpoints.		
5	Animals were individually identified.		
6	Housing conditions (temperature, relative humidity, light-dark cycle) were appropriate for the study type and animal model.		
7	The number of animals per sex in each cage were appropriate for the study type and animal model.		
8	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.		
	Dosing and administration of the test compound		
9	The allocation of animals to different treatments was randomized.		
10	The route of administration was appropriate and not likely to interfere with the study results.		
11	The timing and duration of administration were appropriate for investigating the included endpoints.		
12	The frequency of administration was appropriate for investigating the included endpoints.	not determined	

Fig. 3 *Guidance for evaluating each criterion in the SciRAP tool.*

Criterion no. 31 in the reporting quality section and criterion no. 18 in methodological quality section provide space for free text comments on additional aspects that affect study reliability. These criteria do not contain the drop-down menu with options.

You may use the "COMMENT" column to write free text comments, for example explaining your evaluation of a specific criterion (**Fig. 4**).



no.	REPORTING QUALITY	SELECTION	COMMENT
	Test compound and controls		
1	The chemical name or other identification, such as CAS-number, of the test compound was given.	fulfilled	
2	The purity of the test compound was stated or is traceable according to information given regarding manufacturer and lot/batch number. In case of mixtures, the composition of different constituents was stated.	partially fulfilled	Write comment here!
3	The vehicle was described.	not fulfilled	
4	It was stated that a negative control group was included.	fulfilled	

Fig. 4 *Writing a note in the "COMMENT" column.*

Judging criteria as "not determined"

If a criterion cannot be judged, you can select the option "not determined" in the drop-down menu (**Fig. 2**). This is primarily intended for methodological quality criteria when sufficient information is lacking to make a judgment regarding whether the criterion is fulfilled or not. Note that for reporting quality, if information is missing you should select "not fulfilled".

Removing criteria

Individual criteria may be considered more or less critical in the specific case you are working on, and the SciRAP tool includes a function to remove criteria for reporting and methodological quality. In that case, choose "REMOVE" in the drop-down menu of the "SELECTION" column instead of fulfilled, partially fulfilled, not fulfilled (**Fig. 2**). Removed criteria will not be included in the colour profile or % fulfilled criteria calculation. Motivations for removing criteria can be provided in the "COMMENT" column (**Fig. 4**).

NOTE: removing criteria will have an impact on the colour profile and the % fulfilled criteria. It is therefore important that the same criteria are removed in evaluations that are going to be compared to each other. Items in the Relevance section cannot be removed.

<u>Interpreting the results of the SciRAP tool:</u>

Results of the study assessment are shown right below the relevance section of the SciRAP tool in the form of % fulfilled criteria, as well as a colour profile.

	% FULFILLED CRITERIA	
	REPORTING	METHOD
Study overall	50.00	43.75
Test compounds and controls	37.50	50.00
Animal model and housing conditions	50.00	37.50
Dosing and administration of the test compound	50.00	37.50
Data collection and analysis	50.00	50.00
Funding and competing interests	75.00	

Fig. 5 *Table with % fulfilled criteria.*

Percent fulfilled criteria

The results show % fulfilled criteria of for the study overall, as well as for the specific criteria categories (**Fig. 5**).

- The % fulfilled criteria is calculated as follows:
  $SciRAP\ score\ (\%) = \frac{F+(PF*0.5)}{T} * 100\%$ / $SciRAP\ score\ (\%) = \frac{DR+(IR*0.5)}{T} * 100\%$

where F is the number of fulfilled criteria, PF is the number of partially fulfilled criteria, and T is the total number of criteria. In other words, partially fulfilled criteria contribute half the value as fulfilled criteria. Criteria that have been removed are excluded from the calculation.

The % fulfilled criteria can have a value ranging from 0 (all criteria are judged as "not fulfilled"/"not determined") to 100 (all criteria are judged as "fulfilled").

NOTE:

- selecting “not determined” for a criterion will have the same impact as “not fulfilled” on the % fulfilled value. It is therefore advisable to leave as few criteria as possible as "not determined", and the user should take care to note the reason for leaving a criterion as "not determined".
- removing criteria will have an impact on the % fulfilled criteria, as well as the colour profile. It is therefore important that the same criteria are removed in evaluations that are going to be compared to each other.
- importantly, the % fulfilled criteria cannot be considered on its own but should be interpreted together with the colour profile when concluding on study reliability. The colour profile is crucial to identify where a study's strengths and weaknesses lie and is more informative than the % fulfilled criteria for this purpose.

Colour profile

In the colour profile, the evaluations of reliability and relevance are illustrated in bar charts (**Fig. 6**), showing green for fulfilled criteria, yellow for partially fulfilled and red for criteria that were not fulfilled. Criteria that were "not determined" will be shown as grey. Relevance items evaluated as relevant are shown as green, indirectly relevant items are shown as yellow, and if the item was evaluated as being not relevant for the risk assessment or problem formulation, it is shown as red. The bar charts do not include criteria that have been removed.

Fig. 6 *The evaluations of reliability and relevance are illustrated in bar charts.*

Categorisation of reliability and relevance

The SciRAP tool does not provide cut-off values or a pre-defined scheme for categorisation of the reliability and relevance of *in vivo* toxicity data. Principles for such categorisation needs to be established on a case-by-case basis and should be fit for purpose for the assessment at hand. Some examples of how the output of the SciRAP evaluation can be used in different contexts, including weight of evidence assessment, are provided in published articles. For example:

Holmer ML, Zilliacus J, Draskau MK, Hlisníková H, Beronius A, Svingen T. 2024. Methodology for developing data-rich Key Event Relationships for Adverse Outcome Pathways exemplified by linking decreased androgen receptor activity with decreased anogenital distance. Reprod Toxicol. 128:108662. doi: 10.1016/j.reprotox.2024.108662. Epub ahead of print. PMID: 38986849.

Röhl C, Batke M, Damm G, Freyberger A, Gebel T, Gundert-Remy U, Hengstler JG, Mangerich A, Matthiessen A, Partosch F, Schupp T, Wollin KM, Foth H. 2022. New aspects in deriving health-based guidance values for bromate in swimming pool water. Arch Toxicol. 96(6):1623-1659. doi: 10.1007/s00204-022-03255-9.PMID: 35386057; PMCID: PMC9095538.

Wiklund L and Beronius A. 2022. Systematic evaluation of the evidence for identification of endocrine disrupting properties of Bisphenol F. Toxicology. 476:153255. doi: https://doi.org/10.1016/j.tox.2022.153255

If you have any questions, please do not hesitate to contact us at anna.beronius@ki.se.