

Table I. Details of reliability assessment according to ToxRtool (21,25).

	Number of study	1	2	3	4
	Criteria	Bariani <i>et al</i> (8), 2017	Deng <i>et al</i> (11), 2016	Furuya <i>et al</i> (17), 2015	Novaković <i>et al</i> (31), 2015
No.	Criteria group I: Test substance identification				
1.	Was the test substance identified?	1	1	1	1
2.	Is the purity of the substance given?	1	0	1	1
3.	Is information on the source/origin of the substance given?	1	0	1	1
4.	Is all information on the nature and/or physico-chemical properties of the test item given, which you deem <u>indispensable</u> for judging the data?	1	1	1	1
	Subtotal Score I	4	2	4	4
	Criteria group II: Test organism characterisation				
5.	Are the species (for <i>in vivo</i> studies) or the test system provided (for <i>in vitro</i> studies)?	1	1	1	1
6.	Is the sex of the test organism (for <i>in vivo</i> studies) or the origin or source of the test system with the corresponding sex characteristic provided (for <i>in vitro</i> studies)?	1	1	1	1

7.	Is information provided on the strain of test animals (for <i>in vivo</i> studies) or the specification of cell/tissue culture (for <i>in vitro</i> studies)?	1	1	1	1
8.	Is the age or body weight of the test organisms at the start of the study provided (for <i>in vivo</i> studies) or the age or other relevant clinical characteristics of human donors for cell/tissue cultivation (for <i>in vitro</i> studies)?	0	0	0	0
9.	For repeated dose studies: is information provided on the housing or feeding conditions (for <i>in vivo</i> studies) or the conditioning and maintenance of the cultivation process (for <i>in vitro</i> studies)?	1	0	0	1
	Subtotal score II	4	3	3	4
	Criteria group III: Study design description				
10.	Is the method of administration given (see explanations for details)?	1	1	1	1
11.	Are doses administered or concentrations in application media given?	1	1	1	1
12.	Are frequency and duration of exposure as well as time-points of observations explained?	1	1	1	1
13.	Were negative and positive controls included? (it can not be only including negative controls)	0	0	0	0

14.	Is the number of animals (for <i>in vivo</i> studies) or the amount of cell/tissue culture (for <i>in vitro</i> studies) per group provided?	1	0	1	1
15.	Are sufficient details of the administration scheme given to judge the study?	1	1	1	1
16.	For repeated dose studies: Were achieved concentrations analytically verified or was stability of the test substance otherwise ensured or made plausible?	0	0	0	0
	Subtotal score III	5	4	5	5
	Criteria group IV: Study results documentation				
17.	Are the study endpoint(s) and their method(s) of determination clearly described?	1	1	1	1
18.	Is the description of the study results for all endpoints investigated transparent and complete?	1	1	1	1
19.	Are the statistical methods for data analysis given and applied in a transparent manner (give also point, if not necessary/applicable, see explanations)?	1	1	1	1
	Subtotal score IV	3	3	3	3
	Criteria group V: Plausibility of study design and data				
20.	Is the study design chosen appropriate for obtaining the substance-specific data aimed at (see explanations for details)?	1	0	1	1

21.	Are the quantitative study results reliable (see explanations for arguments)?	1	1	1	1
	Subtotal score V	2	1	2	2
	Total score I-V	18	13	17	18

Bariani *et al* (8), Deng *et al* (11) and Furuya *et al* (17) are *in vivo* studies, while Novakovic *et al* (31) was an *in vitro* study. For the references, please see the Reference list in the main manuscript.

Table II. Summary of scoring reliability assessment (21,25).

Quality of study assessment	Studies included			
	Bariani <i>et al</i> (8), 2017	Deng <i>et al</i> (11), 2016	Furuya <i>et al</i> (17), 2015	Novaković <i>et al</i> (31), 2015
Subtotal score of criteria I	4	2	4	4
Subtotal score of criteria II	4	3	3	4
Subtotal score of criteria III	5	4	5	5
Subtotal score of criteria IV	3	3	3	3
Subtotal score of criteria V	2	1	2	2
Total score	18	13	17	18
Reliability classification	Reliable without restrictions	Reliable with restrictions	Reliable with restrictions	Reliable without restrictions

Bariani *et al* (8), Deng *et al* (11) and Furuya *et al* (17) are *in vivo* studies, while Novakovic *et al* (31) was an *in vitro* study. For the references, please see the Reference list in the main manuscript.

Table III. Interpretation of reliability assessment (21,25).

Reliability categorisation				Consequence
	<i>In vivo</i>	<i>In vitro</i>	Reliability	
1	18-21	15-18	Reliable without restrictions	Useful, check relevance for intended purpose
2	13-17	11-14	Reliable with restrictions	Potentially useful, check relevance for intended purpose
3	<13	<11	Not reliable	Generally not to be used as key study, but depending on the shortcomings of the study it may still be useful in weight-of-evidence approaches or as supportive information