3.4 Reliability

3.4.1 Inter-laboratory reproducibility for the common 10 chemicals

The final judgment for a chemical by the test is decided based on the comparisons of the IC50 value of test chemical with it of the concurrent relative control chemical (TEA) in experiments. As the inter-laboratory reproducibility, it was tabled whether the final judgment for each common chemical (P3-003, P3-005, P3-010, P3-012, P3-019, P3-020, P3-024, P3-028, P3-029 and P3-033) was concordant or not.

3.4.2 Distribution and summary statistics for the relative control chemical (TEA)

The IC50 values of the concurrent relative control chemical (TEA) are included in an experiment. To grasp the distribution of the IC50 values of TEA in each laboratory, the histograms were drawn and the summary statistics such as mean, standard deviation, minimum values, median and maximum value were tabled.

3.5 Relevance

3.5.1 Definition of the concordance, the sensitivity and the specificity

The proportions for the concordance, the sensitivity and the specificity were estimated as the indexes of relevance. These indexes are estimated by using frequency in the 2 by 2 contingency table. The definitions of these indexes are summarized in the following table 3.

Table 3 Definition of the concordance, sensitivity and specificity

Judgment from the SIRC-CVS test	Chemical category		Total
	Positive	Negative	
Positive	a	b	a+b
Negative	c	d	c+d
Total	a+c	b+d	N

Sensitivity = 100 x a/(a+c)
Specificity = 100 x d/(b+d)
Concordance = 100 x (a+d)/N

10 of 100 chemicals are commonly tested. Although the total number of the blinded chemicals is 120, the denominator of the proportion of the indexes should be 100 according to the design of this study. When this strategy is accepted, however, it is possible to calculate the different values for the proportion if the discordance is found in the common 10 chemicals. For the discordance between the common chemicals, we decided to calculate the following 3 cases,

Case A: Median judgment, which is decision to adopt majority judgment,

Case B: Judged as "positive" in case of the discordant judgment between laboratories,

Case C: Judged as "negative" in case of the discordance judgment between laboratories.

Although we presents the above 3 cases, the case A is primary one for the analysis and the both the case B and C are cases for a kind of the sensitivity analysis.

3.5.2 The 95% confidence interval for the concordance, the sensitivity and the specificity

The concordance, the sensitivity and the specificity are the estimated proportion. The precision of the proportion depends on the numbers of the chemical, that is, the denominator of the proportion. It is known that the precision also depends on the numerator of the proportion. The 95% confidence interval is widely used as the index showing the precision of the proportion, and there are many ways to estimate of it have been proposed.

This report describes the 95% confidence intervals with the proportion for the relevance in order to show the precision of the proportion. We used the method recommended by D. G. Altman et al. (2005). Although the method is not known widely as the authors are admitted, it has good statistical property compared with the traditional way and easy to calculate. The 95% confidence interval is calculated for the numerator, r, and the denominator, n, as

$100 \times (A-B)/C$ to $100 \times (A+B)/C$,

where $A=2r+z^2$, $B=z\sqrt{z^2+4r(1-r/n)}$, $C=2(n+z^2)$ and $z=1.96$.

3.5.3 Concordance, sensitivity and specificity for the GHS classification (Primary analysis)

In the protocol the GHS classification is dichotomized as the irritant for the 1, 2A and 2B, and the the non irritant (NI) for the not classified (Non). As the primary analyses, we estimate the concordance, the sensitivity and the specificity based on this dichotomization of the GHS classification on the case A described in 3.5.1.

3.5.4 Concordance, sensitivity and specificity for the EPA classification (Secondly analysis)

As the secondary analysis, the concordance, sensitivity and specificity based on the EPA classification are also estimated using 88 chemicals. In the protocol the

EPA classification is dichotomized as the irritant for the 1, 2 and 3, and the the non irritant (NI) for 4.

3.5.5 Analysis by each laboratory (other analysis)

Each laboratory conducted the tests for 40 chemicals. The indexes for the concordance, sensitivity and specificity in each laboratory were estimated.

4. Results

4.1 Submitted data files

We received 40 data files from the each laboratory, respectively. For the questionable items, we asked the person of the laboratory, and required the revised file if necessary.

The laboratory C conducted 3 runs for all 40 chemicals which is the way until the phase 2b study, but the 3rd run does not always required for the final judgment in the phase 3 study as described in 3.1. Therefore, the results of the 1st and 2nd runs take over the 3rd run for data of the laboratory C in our data analysis.

Basically, the judgment for a chemical into the constructed data set depends on the judgment on the developed data file. However, we gave opposite judgment only for the P3-066, because the IC50 value of the test chemical is “312.5” of the maxim concentration for the both 2 runs and the cell viability for the higher concentration than it did not obtained. Since the cell viabilities were not reach the 50%, the automatically data files calculated it as “negative”. Since the 312.5 is smaller concentration compared with the concurrent relative control, however, we gave the judgment as “positive” for the chemical and analyze the data set in this report. The chemical, P3-066, should have been eliminated from the data analysis. We will fix the treatment of the chemical for the later data analysis after the VMT meeting about discussion of the phase 3 study.

4.2 Data check

We also received the raw data which is obtained from the plate reader in each laboratory. Using these files, we verified values in the submitted datafiles with the values in the raw data, we asked the person of the laboratory if necessary.

4.3 Figure for the IC50 values in the each experiment

Figure 1(a)-(h) show the IC50 values of all the experiments. This figure can show the final judgment by seeing the position of the IC50 value of the test chemical and the relative control chemical in an experiment.

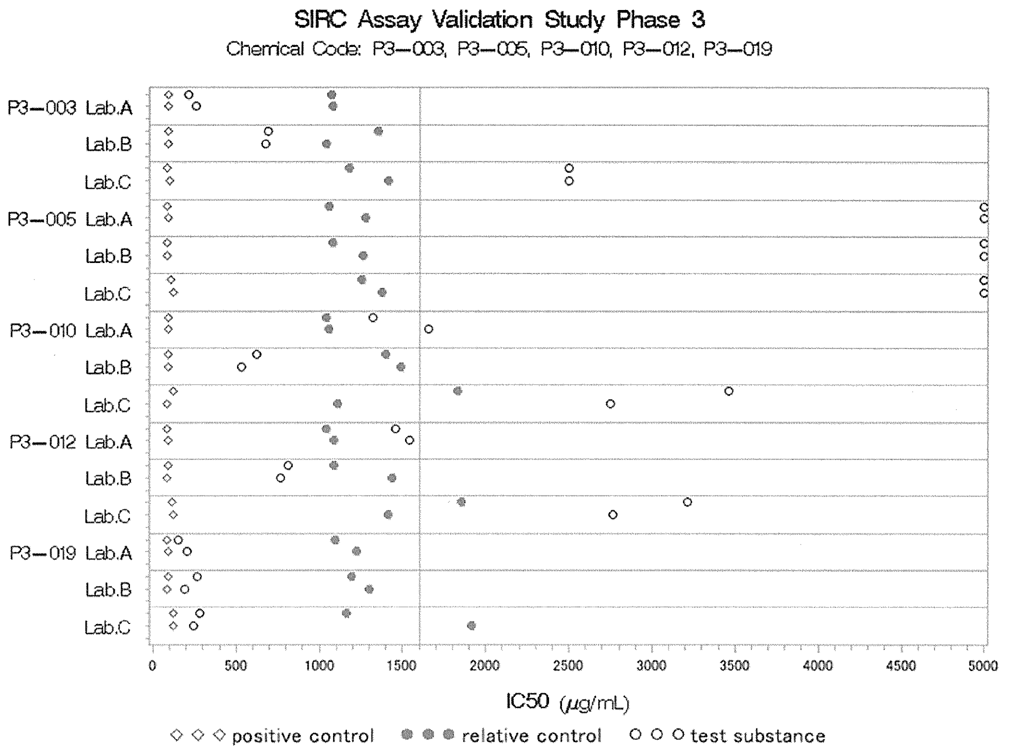


Figure 1 (a). IC50 values from the each run

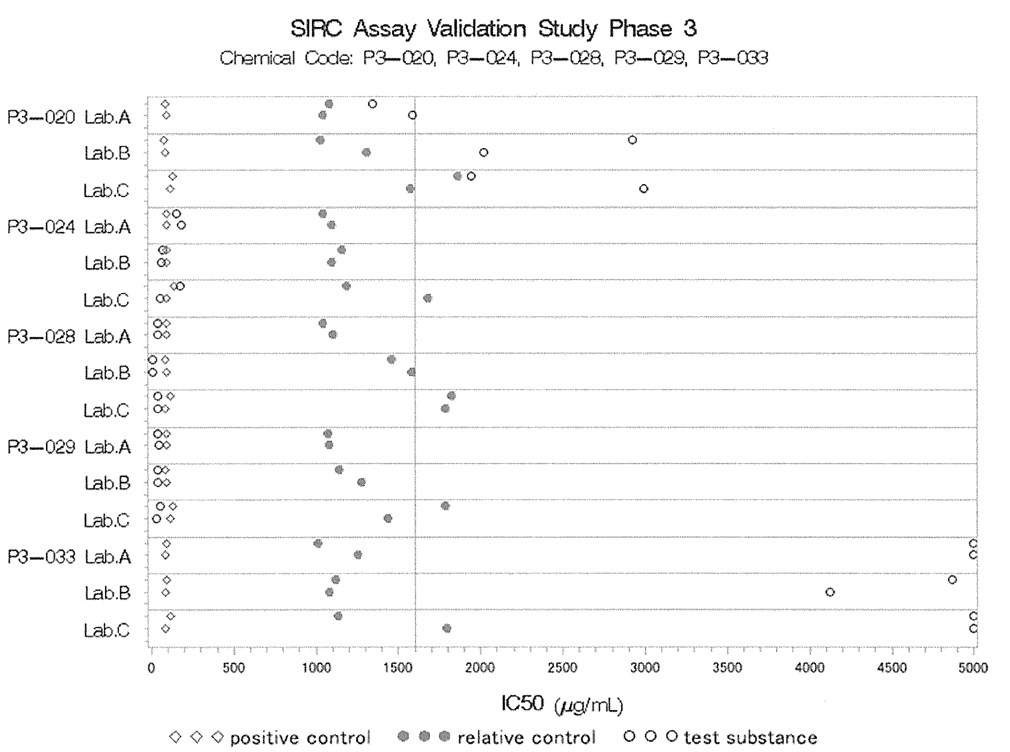


Figure 1 (b). IC50 values from the each run

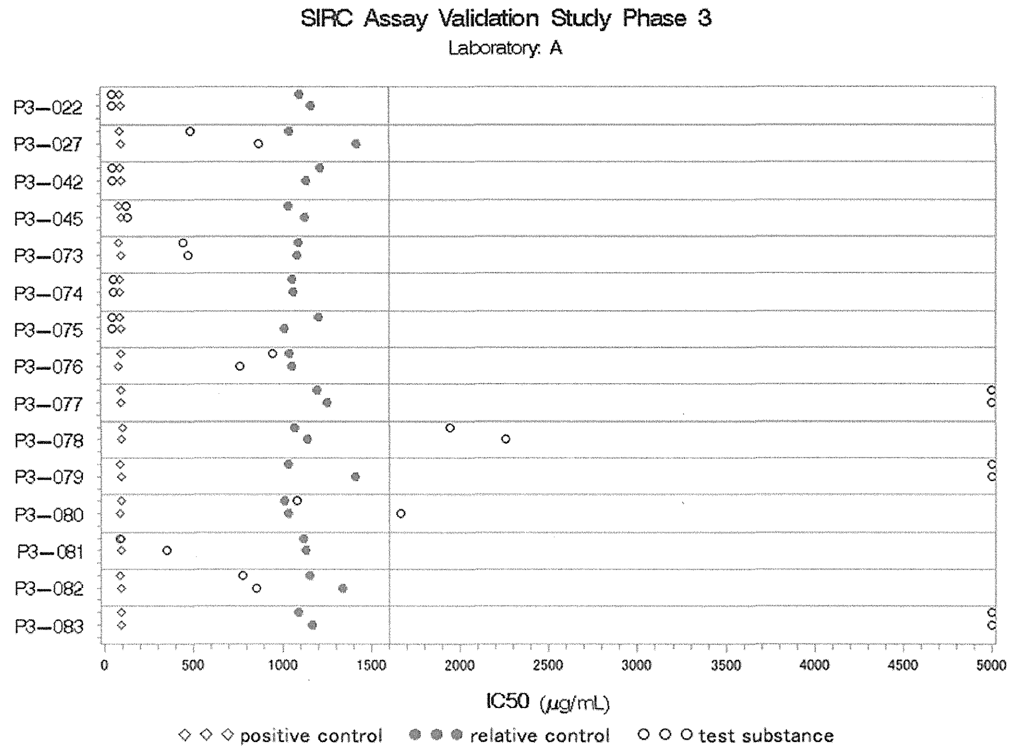


Figure 1 (c) . IC50 values from the each run

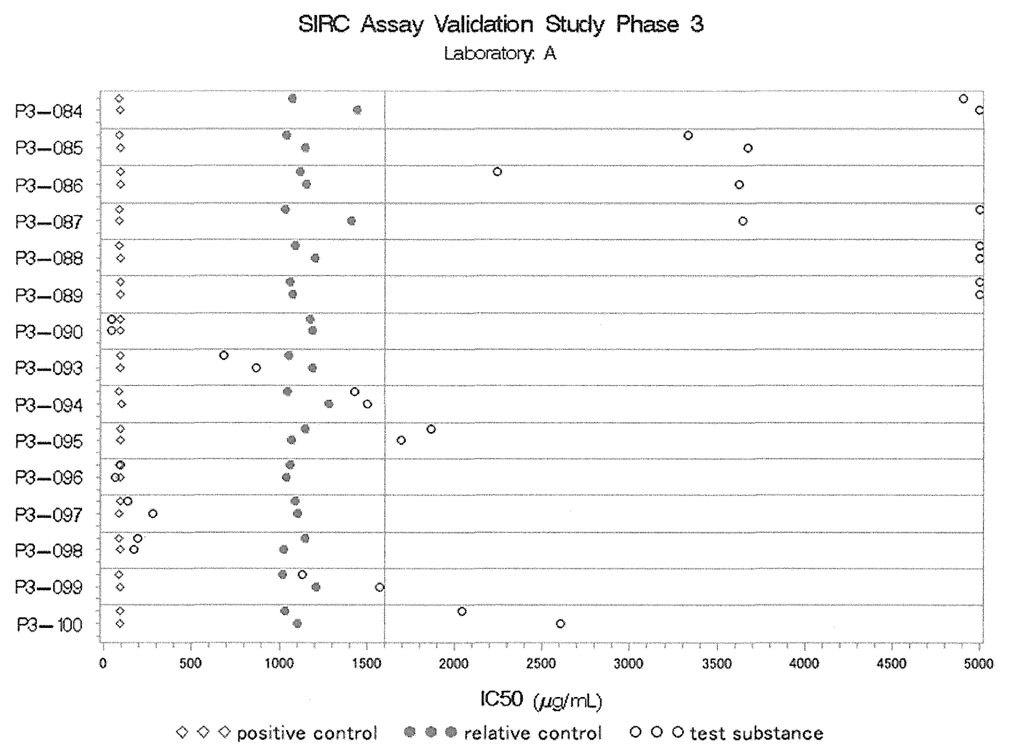


Figure 1 (d) . IC50 values from the each run

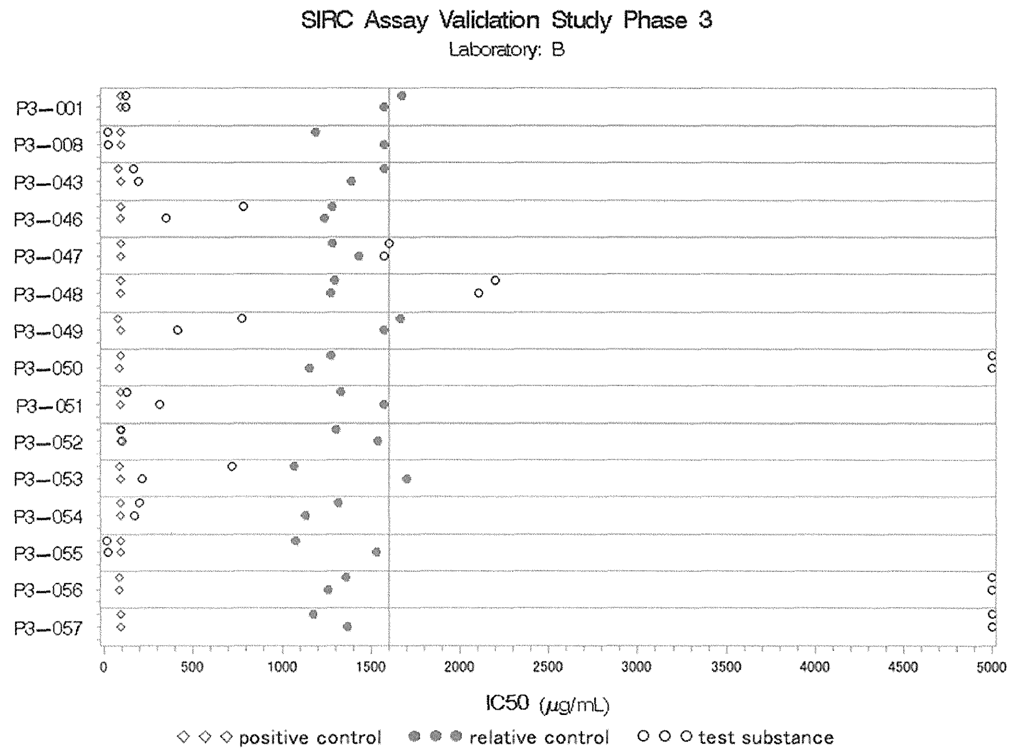


Figure 1 (e) . IC50 values from the each run

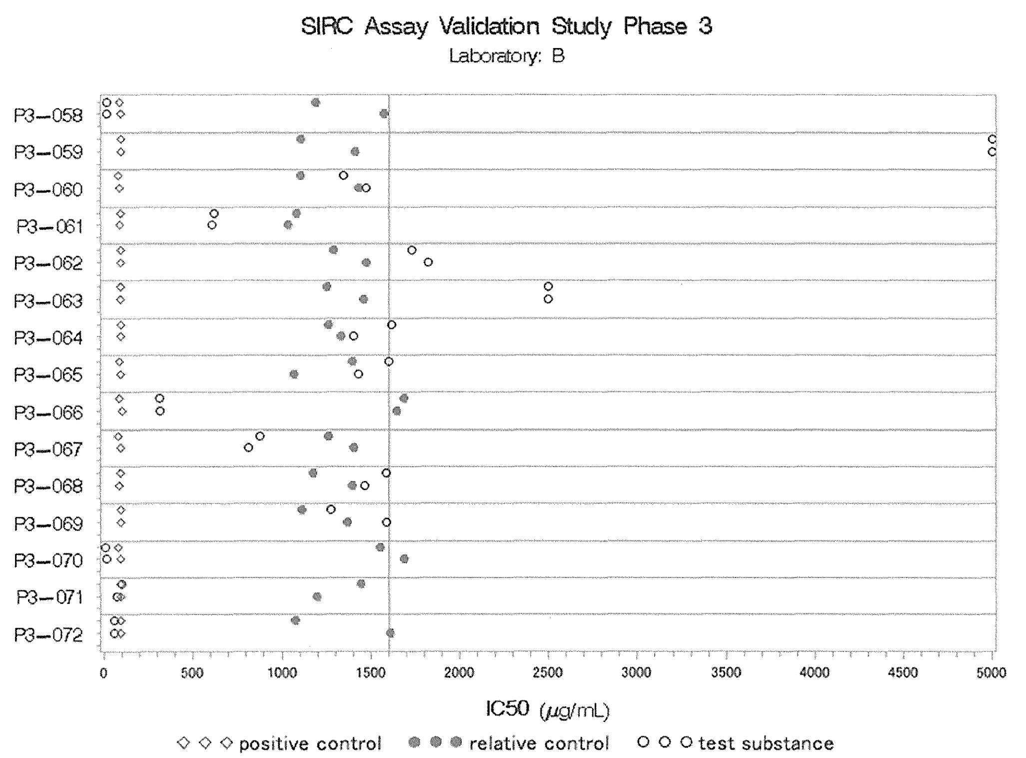


Figure 1 (f) . IC50 values from the each run

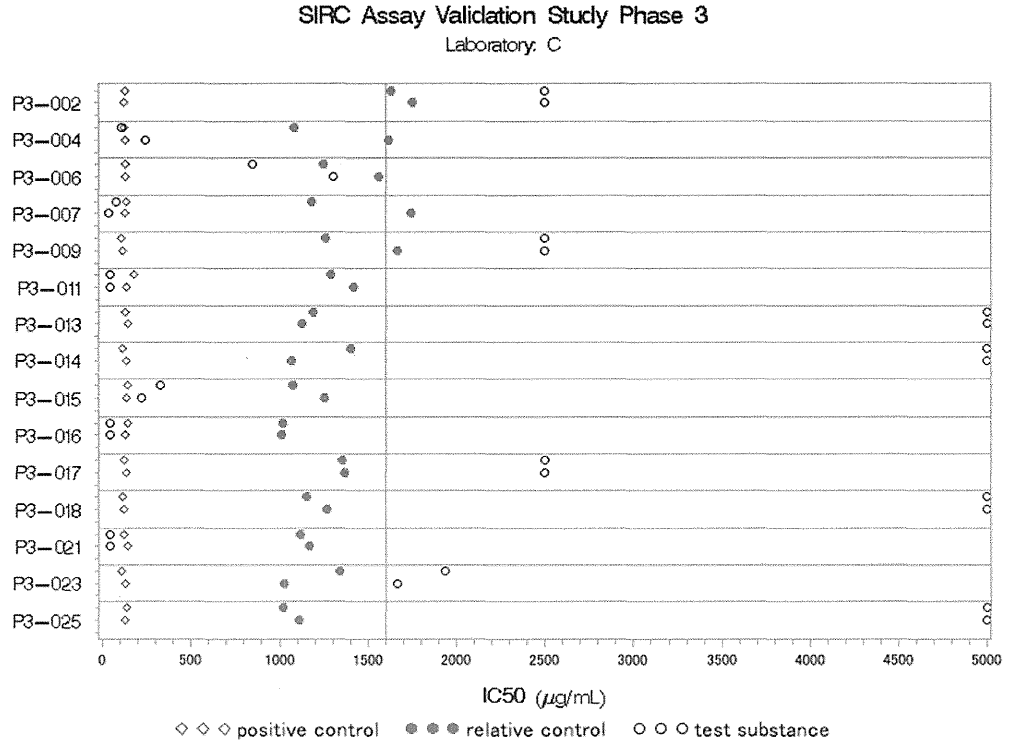


Figure 1 (g) . IC50 values from the each run

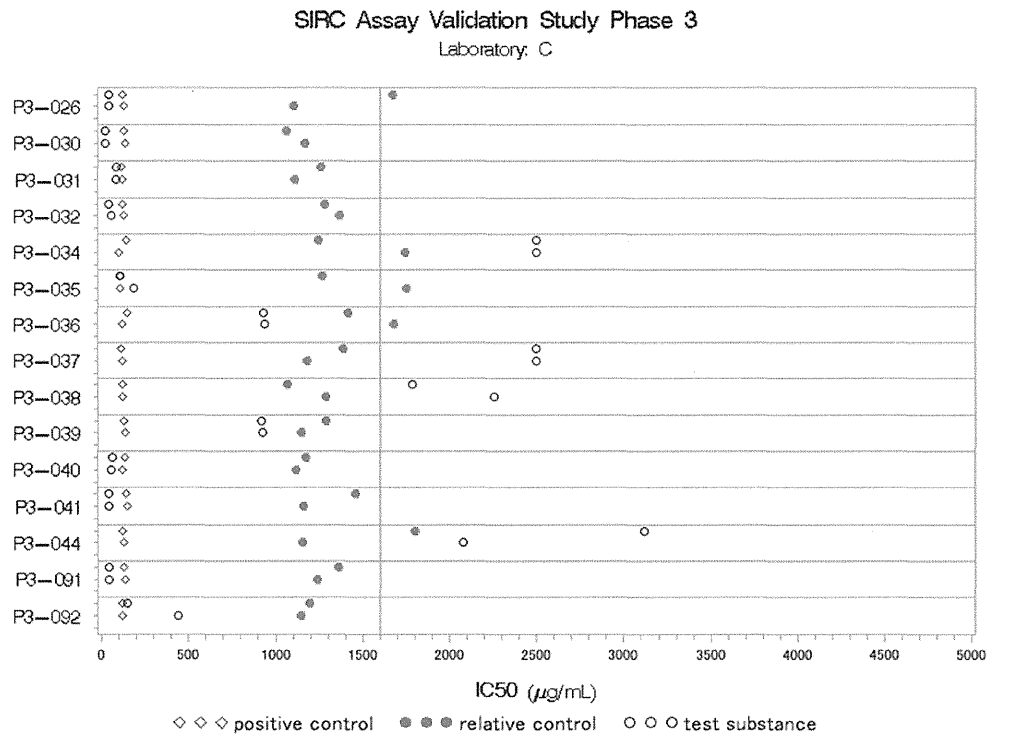


Figure 1 (h) . IC50 values from the each run

4.4 Reliability

4.4.1 Inter-laboratory reproducibility for the common 10 chemicals

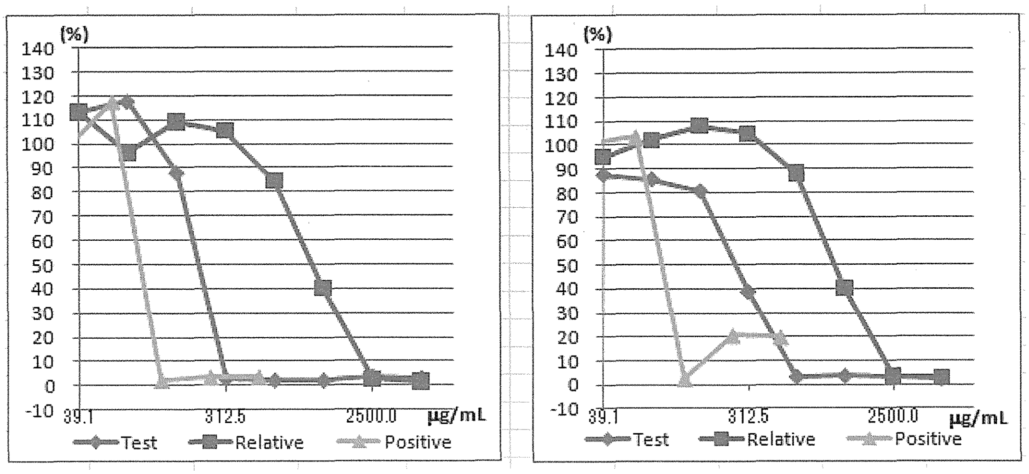
Table 4 shows each judgment for each common chemical.

Table 4

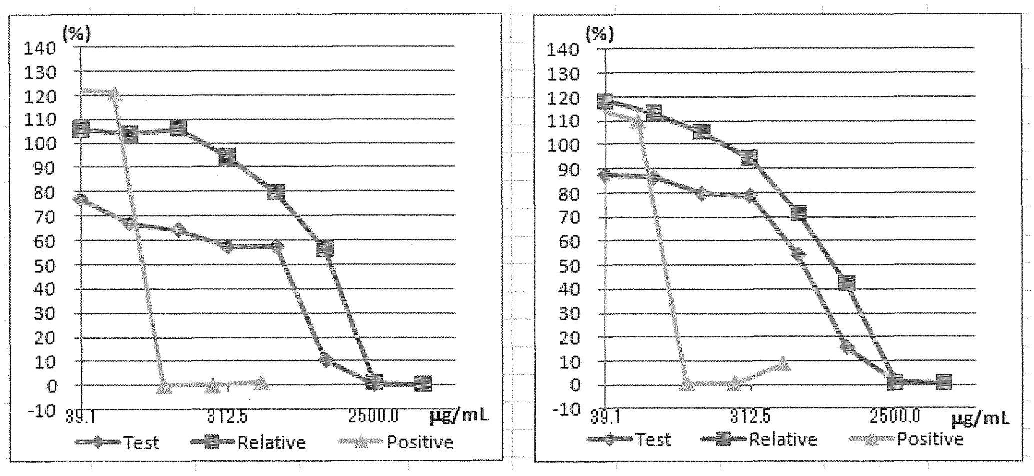
	Laboratory A	Laboratory B	Labotaroy C
P3-003	Pos.	Pos.	Neg.
P3-005	Neg.	Neg.	Neg.
P3-010	Neg.	Pos.	Neg.
P3-012	Neg.	Pos.	Neg.
P3-019	Pos.	Pos.	Pos.
P3-020	Neg.	Neg.	Neg.
P3-024	Pos.	Pos.	Pos.
P3-028	Pos.	Pos.	Pos.
P3-029	Pos.	Pos.	Pos.
P3-033	Neg.	Neg.	Neg.

3 of 10 chemicals were discordant. These were P3-003, P3-010 and P3-012. For these 3 chemicals, the dose-response curves were shown in Fig 2 (a)-(c).

[P3A82(Medium)]



[P3B79(Ethanol)]



[P3C61(DMSO) ＊Test substance conc.=0.5]

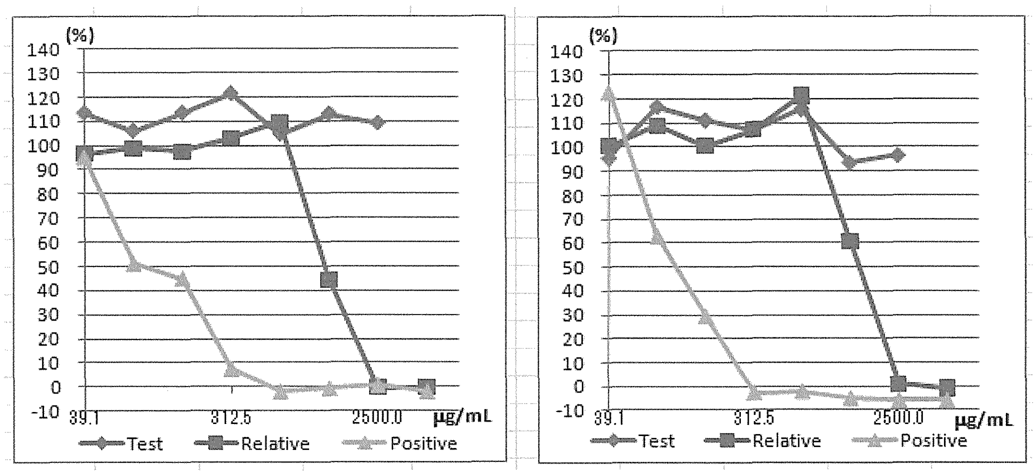
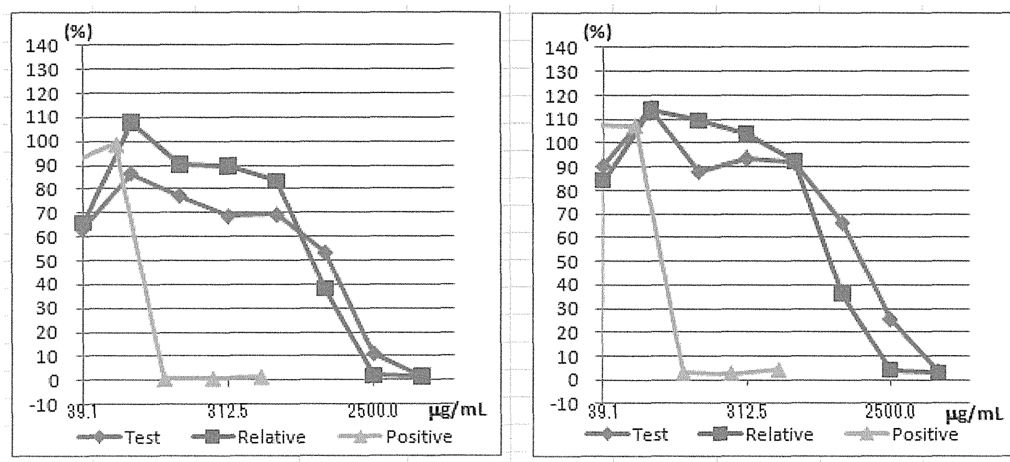
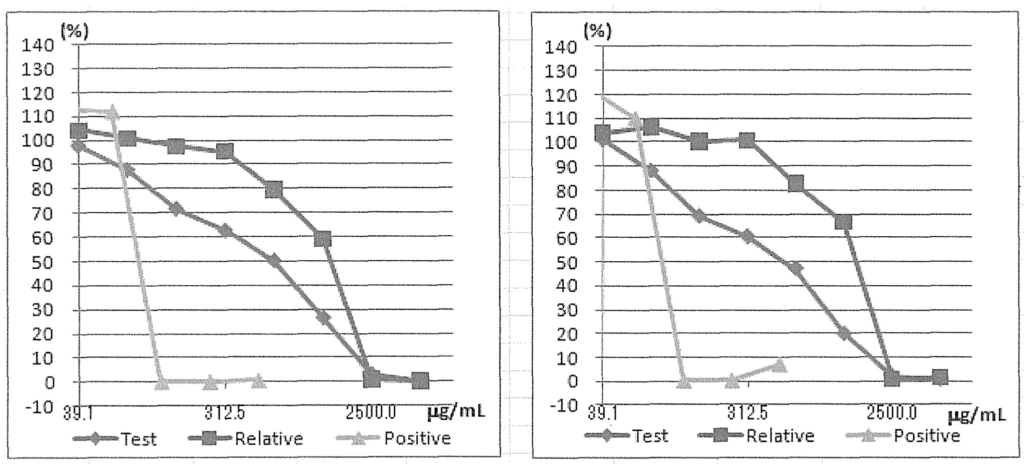


Figure 2 (a) dose-response curves for the P3-003.

[P3A90(Medium)]



[P3B71(Medium)]



[P3C63(Medium)]

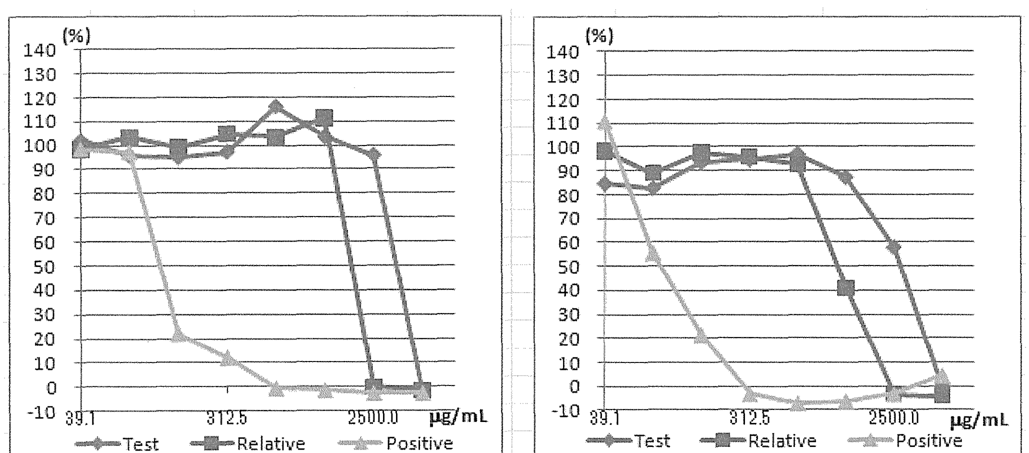
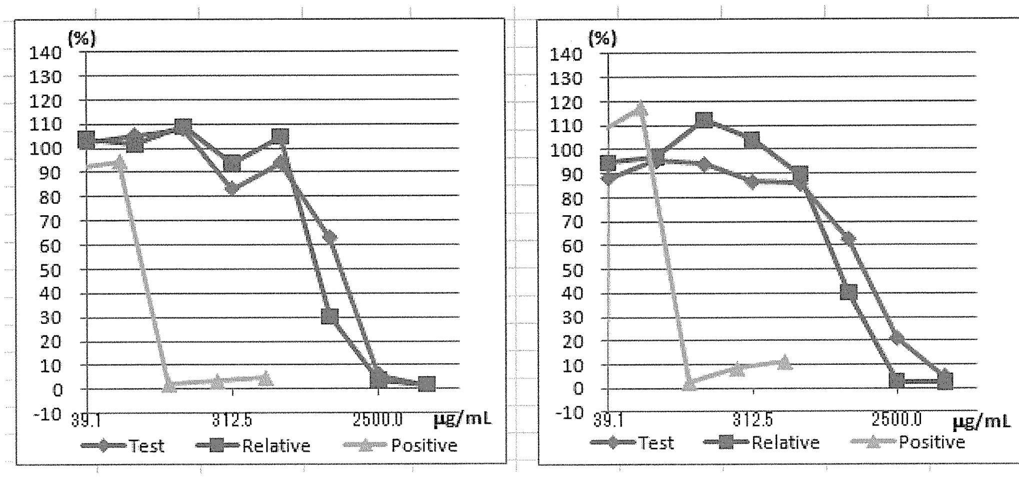
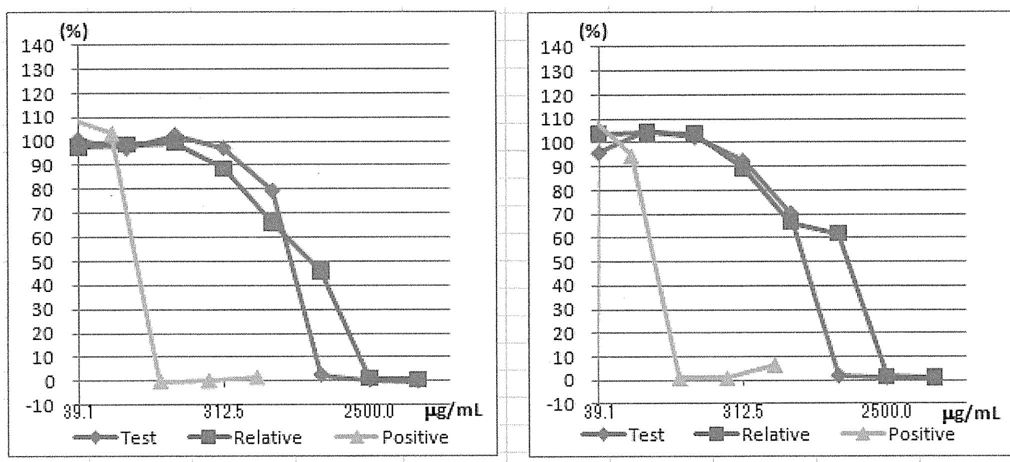


Figure 2 (b) dose-response curves for the P3-010.

[P3A84(Medium)]



[P3B77(Medium)]



[P3C64(Medium)]

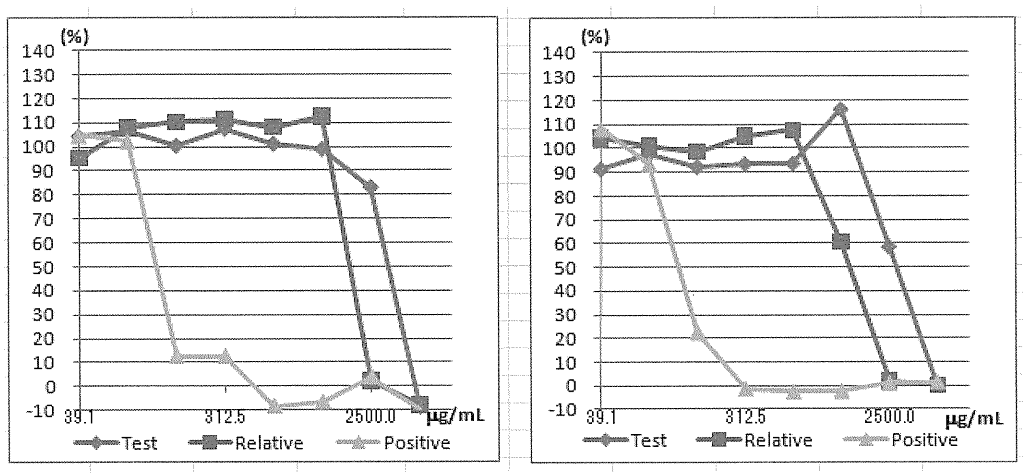


Figure 2 (c) dose-response curves for the P3-012.

4.4.2 Distribution and summary statistics for the relative control chemical (TEA)

Figure 3 and table 5 show the histograms and the summary statistics of IC50 values for the relative control chemical in each laboratory, respectively.

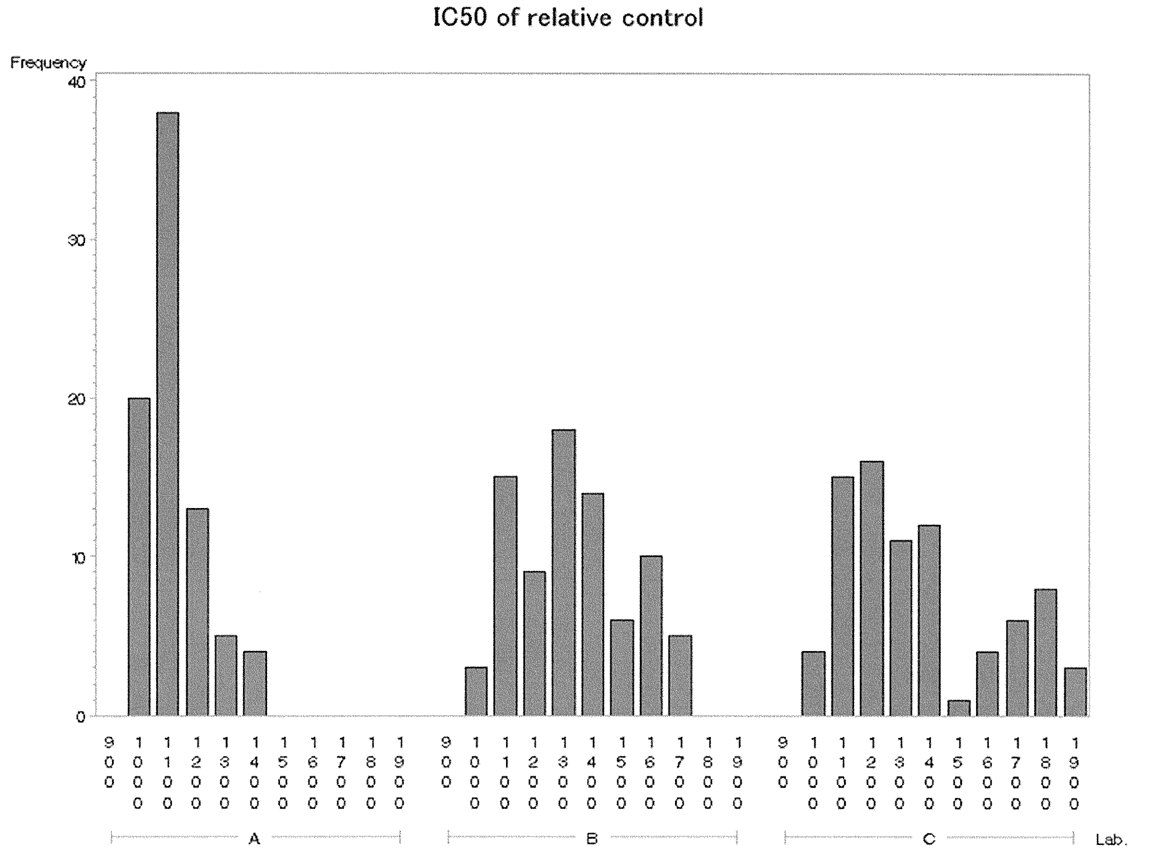


Figure 3. Histograms of the IC50 values for the relative control.

Table 5. Summary statistics of the the relative control

Lab.	N	Mean	Standard deviation	Min	Median	Max
A	80	1119.6	99.4	1010.2	1085.8	1446.4
B	80	1326.1	189.6	1026.6	1300.4	1704.3
C	80	1354.3	258.4	1013.8	1269.4	1913.3
Total	240	1266.7	219.6	1010.2	1190.85	1913.3

4.5 Relevance

4.5.1 Concordance, sensitivity and specificity for the GHS classification (Primary analysis)

Table 6 (a)-(c) show the 2 by 2 table for estimating the concordance, sensitivity and specificity. These indexes for the relevance are shown in the Table 7. Note that the case A is the result of the primary analysis.

Table 6 (a). 2 by 2 tables for the case A

Case A			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	36	19	55
Neg.	27	18	45
Total	63	37	100

Table 6 (b) 2 by 2 tables for the case B

Case B			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	36	21	57
Neg.	27	16	43
Total	63	37	100

Table 6 (c) 2 by 2 tables for the case C

Case C			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	36	18	54
Neg.	27	19	46
Total	63	37	100

Table 7. Concordance, sensitivity and specificity for the dichotomized GHS classification

Case	Freq.	Concordance	Lower limit of Conc.	Upper limit of Conc.	Sensitivity	Lower limit of Sens.	Upper limit of Sens.	Specificity	Lower limit of Spec.	Upper limit of Spec.
A	**100**	**54.0**	**44.6**	**63.1**	**57.1**	**44.9**	**68.6**	**48.6**	**33.4**	**64.1**
B	100	52.0	42.8	61	57.1	44.9	68.6	43.2	28.7	59.1
C	100	55.0	45.5	64.2	57.1	44.9	68.6	51.4	35.9	66.6

4.5.2 Concordance, sensitivity and specificity for the EPA classification (Secondly analysis)

Table 8 (a)-(c) show the 2 by 2 table for estimating the concordance, sensitivity and specificity. These indexes for the relevance are shown in the Table 9.

Table 8 (a). 2 by 2 tables for the case A

Case A			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	37	12	49
Neg.	30	9	39
Total	67	21	88

Table 8 (b). 2 by 2 tables for the case B

Case B			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	38	13	51
Neg.	29	8	37
Total	67	21	88

Table 8 (c). 2 by 2 tables for the case C

Case C			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	37	11	48
Neg.	30	10	40
Total	67	21	88

Table 9. Concordance, sensitivity and specificity for the dichotomized EPA classification

Case	Freq.	Concordance	Lower limit of Conc.	Upper limit of Conc.	Sensitivity	Lower limit of Sens.	Upper limit of Sens.	Specificity	Lower limit of Spec.	Upper limit of Spec.
A	**88**	**52.3**	**42.3**	**62.1**	**55.2**	**43.4**	**66.5**	**42.9**	**24.5**	**63.5**
B	88	52.3	42.4	61.9	56.7	44.8	67.9	38.1	20.8	59.1
C	88	53.4	43.3	63.3	55.2	43.4	66.5	47.6	28.3	67.6

4.5.3 Analysis by each laboratory (other analysis)

Table 10 (a)-(c) show the 2 by 2 table for estimating the concordance, sensitivity and specificity for each laboratory for the GHS classification. These indexes for the relevance are shown in the Table 11.

Table 10 (a) Lab. A

Laboratory A			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	13	7	20
Neg.	12	8	20
Total	25	15	40

Table 10 (b). Lab B

Laboratory B			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	17	7	24
Neg.	11	5	16
Total	28	12	40

Table 10 (c). Lab C

Laboratory C			
JUDGE	GHS		
	Irritant	Non irritant	Total
Pos.	14	8	22
Neg.	8	10	18
Total	22	18	40

Table 11. Concordance, sensitivity and specificity for the dichotomized GHS classification in each laboratory

Lab.	Freq.	Concordance	Lower limit of Conc.	Upper limit of Conc.	Sensitivity	Lower limit of Sens.	Upper limit of Sens.	Specificity	Lower limit of Spec.	Upper limit of Spec.
A	40	52.5	37.4	67.1	52.0	33.5	70.0	53.3	30.1	75.2
B	40	55.0	40.7	68.4	60.7	42.4	76.4	41.7	19.3	68.0
C	40	60.0	45.2	73.0	63.6	43.0	80.3	55.6	33.7	75.4

Table 12 (a)-(c) show the 2 by 2 table for estimating the concordance, sensitivity and specificity for each laboratory for the EPA classification. These indexes for the relevance are shown in the Table 13.

Table 12 (a) Lab. A

Laboratory A			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	13	4	17
Neg.	12	3	15
Total	25	7	32

Table 12 (b). Lab B

Laboratory B			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	16	5	21
Neg.	13	2	15
Total	29	7	36

Table 12 (c). Lab C

Laboratory C			
JUDGE	EPA		
	Irritant	Non irritant	Total
Pos.	17	5	22
Neg.	12	6	18
Total	29	11	40

Table 13. Concordance, sensitivity and specificity for the dichotomized EPA classification in each laboratory

Lab.	Freq.	Concordance	Lower limit of Conc.	Upper limit of Conc.	Sensitivity	Lower limit of Sens.	Upper limit of Sens.	Specificity	Lower limit of Spec.	Upper limit of Spec.
A	32	50.0	33.9	66.1	52.0	33.5	70.0	42.9	15.8	75.0
B	36	50.0	35.2	64.8	55.2	37.5	71.6	28.6	8.2	64.1
C	40	57.5	42.4	71.3	58.6	40.7	74.5	54.5	28.0	78.7

References

1. Douglas Altman, David Machin, Trevor Bryant, Stephen Gardner. Statistics with Confidence: Confidence intervals and statistical guidelines, 2nd ed. BMJ, 2005.

Report on the selection of test substances for SIRC-CVS validation study

2013/11/20

SIRC-CVS Validation Management Team (VMT)

In this report, the selection process of test substances was described for the SIRC-CVS validation study.

The objective of this study in the study plan was to evaluate the within- and between-laboratory reproducibility and predictability of the SIRC-CVS assay on eye irritation (consistency with the two categories, Irritant and Non-irritant) in accordance with an initial step of bottom-up approach.

As a complementary study, the validation management team (VMT) evaluated the predictability on the United Nation Globally Harmonized System of Classification and Labeling of Chemicals (UN GHS) with three classifications (Category 1, Category 2, Non-irritant) and United States Environmental Protection Agency (EPA) with four classifications

For this purpose, phase II-A, phase II-B and phase III studies were conducted by three laboratories using test substances as shown in Table 1.

In addition, chemical categories or physical and chemical properties (molecular weight, solubility in the medium etc) were included in the list of these test substances to investigate applicable domain.

Table 1. Breakdown of the SIRC-CSV validation study

Phase	The number of the test substances	The number of the repetitions	Examination
II-A	5	3	Within- and between-laboratory reproducibility
II-B	15	3	
III	100 (Including common test substances)	1	Between- laboratory reproducibility and predictability

1. Phase II study

In the phase II study, twenty test substances were selected by VMT for within- and between-laboratory reproducibility as shown in Table 1. All substances were selected in consideration of valance of UN GHS or EPA labeling, solid: liquid from the following lists.

- The existing in individual animal data on test substances were available for classifying the eye irritating hazard under the GHS.
- Test substances had already been evaluated in other eye irritation tests in vitro.

Twenty test substances comprising 10 irritants and 10 non-irritants were listed in Table2. For within- and between-laboratory reproducibility, VMT distributed 3 coded bottles per each test substance to each laboratory. That is, VMT distributed 15 coded bottles (5 test substances) in the phase II-A study and 45 coded bottles (15 test substances) in the phase II-B study to each laboratory.