Appendix B

Evaluating the Impact of Reducing the Sample Size from Five to Four Animals per Group on the Performance of the Ratio Rule of SI > 3 in LLNA Testing

[This Page Intentionally Left Blank]

1.0 Introduction

Test Guideline 429 issued by the Organisation for Economic Co-operation and Development (OECD; OECD 2002) states that "A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a negative control group treated only with the vehicle for the test substance, and a positive control, as appropriate. *In those cases in which individual animal data are to be collected, a minimum of five animals per dose group are used*." This analysis was undertaken to determine if the number of animals required for individual animal data collection could be harmonized with that required for pooled data without diminishing accuracy. This is important because most animal-use regulations require that the minimum number of animals be used in studies, which currently results in only pooled data being collected in many countries because it currently requires fewer animals.

Therefore, the issue under investigation in the evaluation that follows is the impact of modifying the murine local lymph node assay (LLNA) test method protocol by reducing the number of individual animals per group from 5 to 4. More specifically, the evaluation considers how often this reduction in animal usage would have an impact on the overall LLNA outcome when the decision criterion used to determine a sensitizer from a non-sensitizer is a stimulation index (SI) greater than or equal to 3 (i.e., the "Ratio Rule"). Since the "true" underlying sensitizer status for individual substances is generally not known, this investigation will focus on the degree of disagreement rather than on which observed outcome is the "correct" one. This evaluation focused primarily on the Ratio Rule, although the possible use of a formal statistical test will also be considered.

The results of the following analyses indicate that a reduction in the sample size from 5 to 4 animals per group is unlikely to have any significant impact on the results of the LLNA test when using the Ratio Rule. If using statistics, the power for detecting LLNA effects will be reduced slightly when using 4 animals per group relative to using 5 animals per group. However, the practical impact of this power difference may be minimal, in that the power difference appears to be small for detecting effects above the Ratio Rule cutoff point of SI = 3. Importantly, this analysis also indicates that a statistical test based on 4 animals per group will identify more sensitizers than using the Ratio Rule based on 5 animals per group.

2.0 Methods

The database evaluated includes three different strains of animals: CBA, BALB/c, and B6C3F1. This report evaluates in detail only the CBA database; the data from the other two strains are summarized (**Section 4.0** and **Table B-7**) and may be evaluated more definitively in due course. The CBA database consists of 83 individual studies, each with three or four dosed groups and a control group. There are not 83 distinct substances, because some substances are tested in multiple studies. The number of individual animals per group in these studies ranged from 2 to 9. There were a total of 277 dosed groups, two of which were excluded from the agreement-disagreement analysis since there were only 2 or 3 animals per group. Study results were evaluated on a dose-by-dose basis as well as on a study-by-study basis, recognizing that the doses within a study used a common control group. Also, for certain labs, a common control group was used for multiple substances.

For each study having 5 animals per group (i.e., N = 5), all possible random samples of size 4 (responses measured as disintegrations per minute [dpm] of a radiolabeled tracer compound) were taken from both the control and experimental groups (25 possible combinations), and the results of the Ratio Rule were compared for each of the samples with that of the full data set of 5 animals. The level of agreement was then determined.

For those studies having more than 5 animals per group, a similar procedure was applied, but in this case random samples were taken for both the N = 5 and N = 4 protocols, and there were far more combinations of samples to deal with (8100 rather than 25). Once again, the level of agreement between the N = 5 and N = 4 protocols were determined.

3.0 Results

Using the Ratio Rule criterion, the CBA mouse database consisted of a mix of sensitizers (49 studies) and non-sensitizers (33 studies), with one study (discussed in more detail below) producing a borderline effect. **Table B-1** shows the frequency of the various SI values in the 275 usable (for agreement-disagreement analysis) dosed groups, together with the average agreement seen between samples of N = 5 and N = 4. As can be seen in the table, the disagreement in study results is limited to SIs in the 2.1 to 4.7 range, with the disagreement increasing as the SI approaches 3. The overall average agreement between N = 4 and N = 5 studies is quite good: 97.5%. Moreover, as discussed in more detail below, the disagreement in outcome is due primarily to the inherent variability in the data (and the closeness of the SI to 3), not to the reduction in sample size.

Table B-1 Breakdown of Individual Dosed Group SIs: CBA Strain

SI	Frequency	Agreement between N = 5 and N = 4 samples
<2.1	154	100.00%
2.1 - 2.5	16	90.10%
2.6	2	85.00%
2.7	3	73.30%
2.8	2	64.00%
3.1	1	56.00%
3.2	2	55.50%
3.3	4	73.50%
3.4	1	88.00%
3.5	1	68.00%
3.6	1	84.00%
3.7	1	90.00%
3.8	1	100.00%
4.0 - 4.7	16	97.90%
>4.7	70	100.00%
Total	**275**	**97.50%**

Abbreviations: N = number of animals per dose group; SI = stimulation index

The individual study results for the CBA strain are summarized in **Annex I**.

Although the primary focus of this evaluation is on the Ratio Rule (i.e., SI > 3), it is possible that a formal statistical test may be used in addition to (or possibly even in place of) the Ratio Rule. For this reason, a simple Student's *t* test (based on the logged dpm data) was also used to compare each dosed group with its concurrent control. The results of this analysis are summarized in **Table B-2**. It is clear that using a formal statistical test will identify far more "positives" than the Ratio Rule, i.e., statistical significance ($p < 0.05$) was achieved for some dosed groups producing an SI well below 3. This matter is discussed in more detail below.

Table B-2 Distribution of Statistically Significant ($p < 0.05$) SIs: CBA Strain

SI	Frequency	Percentage of statistically significant ($p < 0.05$) SIs
<1.7	131	0.00%
1.7 - 1.9	23	52.20%
2.0 - 2.5	17	88.00%
2.6 - 3.0	7	85.70%
> 3.0	1	100.00%
Total	**277**	

Abbreviation: SI = stimulation index

4.0 Discussion

It was known in advance that the reduction in sample size from N = 5 to N = 4 would have essentially no impact on study results for "strong sensitizers" and for "clear non-sensitizers," and this is confirmed in **Table B-1**. What was not known was (1) how frequently such outcomes are seen in practice; (2) the specific range of SI values in which some impact on study outcome may be evident; (3) the magnitude of the impact for those studies having an SI close to 3; and (4) whether the disagreement in study outcome was due primarily to the reduction in sample size or to the inherent variability in the data (and the closeness of the SI to 3). The current investigation addresses all of these issues.

With regard to the first issue, for the CBA mouse database, only 34 of the 275 dosed groups (12%) had less than 100% agreement between N = 5 and N = 4 outcomes. Thus, for most dosed groups, the reduced sample size will not even be an issue when using the Ratio Rule.

Moreover, the reduced sample size becomes an issue only for a relatively narrow range of SI values. The range of SI values in this database producing less than 100% agreement was 2.1 to 4.7, but this may be somewhat misleading in that many studies in this range produced 100% agreement (see **Table B-1** and **Annex I**).

As the SI approaches 3, the disagreement between a sample of N = 5 and N = 4 increases notably (**Table B-1**). However, and this may be the single most important "take home" message of this entire analysis, the disagreement is far more a function of the animal-to-animal variability than it is to the reduction in sample size. That is, a second sample of 5 animals would show almost the same level of disagreement with the first sample of 5 animals, as would a sample of 4 animals. Thus, the reduction in sample size is a relatively small contributor to this difference. This important concept is illustrated below with two examples from the CBA mouse database, the first showing an SI of 2.8, just below the Ratio Rule threshold of SI = 3, the second showing an SI of 3.2, just above the Ratio Rule threshold.

The first example is the high dose of the third hexyl cinnamic aldehyde study, which had an SI of 2.8 for N = 6. This is the one study noted above with a borderline effect. Since N = 6, this required selection of samples of size 5 from both the control and dosed groups, and some of these samples did not give the same result as that seen for the full six animal sample. The results are summarized below and compared with the N = 4 strategy.

Table B-3 Example Showing Effect of Sample Size on Agreement of Results for a Test Substance with SI = 2.8

	Two N = 5 samples	**One N = 5 sample and one N = 4 sample**
Agreement (SI > 3)	7.7% (10/36) (10/36)	10.5% (10/36) (85/225)
Agreement (SI < 3)	52.2% (26/36) (26/36)	44.9% (26/36) (140/225)
Disagreement (one SI > 3; one SI < 3)	40.1% (by subtraction)	44.6% (by subtraction)

Abbreviations: N = number of animals per dose group; SI = stimulation index

As can be seen from these calculations (see also **Annex I**), the agreement between N = 5 and N = 4 strategies is "only" 55%. However, the disagreement is *not* due primarily to a reduction in sample size, since the agreement is very similar to that found for two N = 5 samples (60%). In other words, only 4.5% of the observed 45% disagreement is due to the reduction in sample size. The rest is due to the inherent variability among animals (and the closeness of the SI to 3) that would be evident even if a second sample of size 5 were used.

The second example is the mid-dose of the dipropylene triamine study, which had an SI of 3.2 also for N = 6. The results are summarized below and compared with the N = 4 strategy.

Table B-4 Example Showing Effect of Sample Size on Agreement of Results for a Test Substance with SI = 3.2

	Two N = 5 samples	**One N = 5 sample and one N = 4 sample**
Agreement (SI > 3)	56.25% (27/36) (27/36)	50.67% (27/36) (152/225)
Agreement (SI < 3)	6.25% (9/36) (9/36)	8.11% (9/36) (73/225)
Disagreement (one SI > 3; one SI < 3)	37.50% (by subtraction)	41.22% (by subtraction)

Abbreviations: N = number of animals per dose group; SI = stimulation index

The results are very similar to those of the first example, in that most of the 41% disagreement between the N = 4 sample and the N = 5 sample is due to the inherent variability of the data and the closeness of the SI to 3, not to the reduction in sample size.

Another point that should be noted: in the instances in which there is disagreement, the N = 4 strategy may actually have a higher likelihood of producing an SI > 3 result than using a sample of size 5. This occurs when the underlying SI is close to but below 3. For instance, consider the first example given above in which the observed SI = 2.8. A sample of size 4 would have a 38% chance (85/225) of producing an SI > 3 compared with only 28% (10/36) when using N = 5. In that sense, N = 4 could be regarded as having greater "power" than N = 5 for these data.

However, use of the Ratio Rule implicitly assumes that an SI less than 3 is biologically unimportant and thus should not be detected. Thus, the increased likelihood of exceeding the

Ratio Rule criterion using N = 4 in the example above could be regarded as an increase in the false positive rate, rather than an increase in power. Importantly, as N increases, the likelihood of detecting SI = 2.8 by the Ratio Rule approaches zero, with maximum "power" occurring for N = 1.

However, some investigators may regard an SI of 2.8 as biologically important, especially if seen at the top dose, as was the case in this study. Consequently, these investigators might actually prefer the performance of N = 4 rather than N = 5 in this example. Of course, if SI < 3 responses are considered important, it would make far more sense to carry out a formal statistical test to detect them rather than using the Ratio Rule, which will likely not detect them. Although not detected by the Ratio Rule, the SI = 2.8 effect noted above in the high dose hexyl cinnamic aldehyde study is highly significant ($p < 0.01$) by Student's *t* test.

Moreover, it is likely that this particular SI = 2.8 is a "real" effect, not only because it is highly significant statistically, but also because in four other studies with this compound, the SIs produced for this dose were 2.2, 4.1, 4.2, and 6.6, with higher doses producing even greater effects (see **Annex I**). Without these additional studies, it is possible that this effect would be "missed" since SI = 2.8 does not satisfy the Ratio Rule criterion of SI > 3, and without individual animal data, it would not be possible to determine whether or not this effect was statistically significant. This is another illustration of the value of individual animal data and also the value of using a formal statistical test. It also shows that in some cases a sample of N = 4 is actually more likely to produce the "correct" conclusion than N = 5 when using the Ratio Rule.

As can be seen in **Table B-2**, a formal statistical test will identify as statistically significant ($p < 0.05$) many responses that would not be detected by the Ratio Rule. In some cases, statistical significance is achieved for SI values as low as 1.7 (see **Annex I** and **Table B-2**). Normally, this "increased power" would be considered very desirable, but apparently it is possible that certain SIs in the 1.7 to 3.0 range, while truly different from controls, may be reflecting "irritation" rather than a true sensitizing effect, and thus may not be indicative of a meaningful human risk. Discussion of this matter is beyond the scope of this investigation, but it is logical to assume that since the Ratio Rule is widely used for LLNA data, while a formal statistical test is not, there must be concern that a formal statistical test will produce too many "significant effects" for SIs in the 2 to 3 range. That is, SIs below 3 may be statistically significant and reflect "real" dosed group effects, but responses in this range are considered biologically unimportant. As can be seen in **Table B-2**, most of the SIs in the 2 to 3 range are in fact statistically significant. Use of the Ratio Rule also implicitly assumes that false positives are more important than false negatives.

Any consideration of statistical power must take into account the variability in response among animals. To illustrate this, consider the 17 CBA mouse studies carried out at BASF (see **Table B-11** in **Annex I**). The mean control dpm response across these 17 studies was 552.3. The mean standard deviation (SD; based on the logged dpm responses) among the control animals was 0.4077. Based on this information, we can carry out a power calculation, which is summarized in **Table B-5**.

To explain further: Power is primarily a function of (1) the magnitude of the difference between the dosed and control groups, (2) the underlying variability among animals, and (3) the sample size. In the table below, "difference" is the size (on a log scale) of the "fold

increase" that is to be detected. The SD is the assumed underlying standard deviation among animals (on a log scale) as determined by the data from BASF (see **Table B-11** in **Annex I**). This SD is assumed to be the same in the dosed and control groups, an assumption consistent with the data from multiple labs obtained to date. Delta is the standardized (by SD) difference to be detected and is the key input variable into the power calculation program. The power calculations given below are based on a two-sided Student's *t* test, and assume an underlying normal distribution for the logged data. The specific power calculations were taken from http://www.danielsoper.com/statcalc/calc49.aspx. In this program "Cohen's *d*" is just the standardized difference, Delta. This is a very simple program to use, and alternative power calculations can easily be made.

Table B-5 Post-hoc Power Calculations Based on the BASF Control Data

	Dosed Group Increase Relative to Controls			
	3.5-fold	**3-fold**	**2.5-fold**	**2-fold**
Assumed control response	552.3	552.3	552.3	552.3
Log (Control response)	6.314	6.314	6.314	6.314
Dosed group response	1933.05	1656.90	1380.75	1104.60
Log (Dosed group response)	7.567	7.413	7.230	7.007
Difference (log scale)	1.253	1.099	0.916	0.693
Assumed SD (log scale)	0.4077	0.4077	0.4077	0.4077
Delta = Difference/SD	3.07	2.70	2.25	1.70
Power for N = 5	99.0%	96.4%	87.9%	65.8%
Power for N = 4	95.7%	89.8%	76.8%	53.0%

Abbreviations: N = number of animals per dose group; SD = standard deviation

From these calculations, the conclusion is that if the underlying variability among control animals is similar to that seen in an average BASF study, then there is an excellent chance that an underlying SI of 2.5 will be detected as statistically significant ($p < 0.05$), although this likelihood is higher for N = 5 (87.9%) than for N = 4 (76.8%). This power calculation is also consistent with the empirical results summarized in **Table B-2**. An underlying SI of 2.5 would almost certainly not be detected by the Ratio Rule, nor would one want it to be detected, since use of the Ratio Rule implicitly assumes that such an effect is of no consequence, as noted earlier.

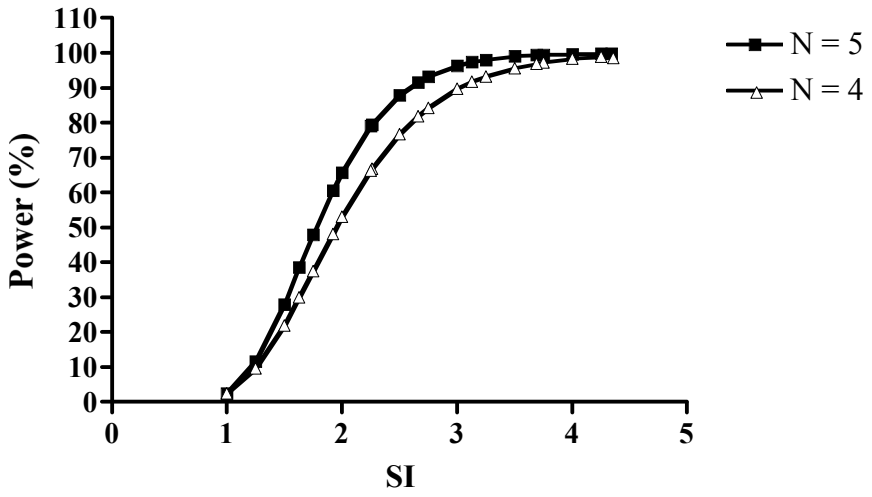
From the website given above, a general power curve can be constructed for N = 5 and N = 4 by specifying different values of Delta, which could reflect different "-fold increases (i.e., SI values)," different underlying variabilities, or a combination of these two factors. Such power comparisons are summarized below in **Table B-6** and **Figure B-1** and include the four from **Table B-5**.

Table B-6 Selected Power Comparisons for N = 5 and N = 4 Samples Based on BASF Control Data

SI	Delta	N = 5	N = 4
4.34	3.60	99.9%	99.1%
4.25	3.55	99.9%	98.9%
4.00	3.40	99.7%	98.3%
3.75	3.24	99.5%	97.2%
3.69	3.20	99.4%	96.9%
3.50	3.07	99.0%	95.7%
3.25	2.89	98.0%	93.3%
3.13	2.80	97.4%	91.8%
3.00	2.70	96.4%	89.8%
2.75	2.48	93.2%	84.3%
2.66	2.40	91.6%	81.9%
2.50	2.25	87.9%	76.8%
2.26	2.00	79.5%	66.8%
2.25	1.99	79.1%	66.3%
2.00	1.70	65.8%	53.0%
1.92	1.60	60.5%	48.2%
1.75	1.37	47.9%	37.4%
1.63	1.20	38.6%	30.0%
1.50	0.99	28.0%	21.9%
1.25	0.55	11.6%	9.7%
1.00	0.00	2.5%	2.5%

Abbreviations: N = number of animals per dose group; SI = stimulation index

Figure B-1 Power Curve for N = 5 and N = 4 Samples Based on BASF Control Data



Abbreviations: N = number of animals per dose group; SI = stimulation index

Although these particular "Deltas" could result from different combinations of -fold-increases and assumed variability, the power calculations for the BASF data indicate that the most notable differences in power between N = 5 and N = 4 occur for SIs below 3, a range for which detection of an effect is apparently viewed as a "false positive" as discussed earlier. That is, the Ratio Rule implicitly assumes that SIs less than 3 should not be detected, so the fact that samples of N = 5 are more likely than samples with N = 4 to detect significant effects for SIs below 3 could be viewed as a disadvantage rather than an advantage of a larger sample size. For SI = 3.5 (at least for the BASF data), the power is high and similar for N = 5 and N = 4 (99.0% vs. 95.7%).

Note also from **Table B-6** that there will be far more sensitizers identified by a statistical test based on 4 animals per group than would be identified by the Ratio Rule using 5 animals per group. For example, a formal statistical test with N = 4 would have approximately 90% power for detecting an SI = 3, compared with only 50% power by using the Ratio Rule (regardless of N).

Although this report focuses on the large CBA mouse database, there are two smaller LLNA databases involving BALB/c and B6C3F1 mice. Although these other databases were not evaluated in detail, the pattern of LLNA response seen in these two strains was very similar to that seen in the CBA database. This comparison is summarized in **Table B-7** below. In this table, the percentage of positive studies is the percentage of studies having SI > 3 in at least one dosed group. As can be seen in **Table B-7**, there is little evidence of a strain difference in the pattern of LLNA response, and thus there is very little likelihood that a detailed evaluation of these other two strains would change the conclusions of this report.

Table B-7 Comparison of CBA, BALB/c, and B6C3F1 Databases

Strain	No. of Studies	No. of Doses	% Positive Studies	Distribution of SIs				
				<1.7	1.7 – 1.9	2.0 – 2.5	2.6 – 3.0	> 3.0
CBA	83	277	59 (49/83)	131 (47%)	23 (8%)	17 (6%)	7 (3%)	99 (36%)
BALB/c	41	133	63 (26/41)	67 (50%)	12 (9%)	8 (6%)	6 (5%)	40 (30%)
B6C3F1	10	28	70 (7/10)	15 (54%)	1 (4%)	1 (4%)	2 (7%)	9 (32%)

Abbreviation: No. = number; SI = stimulation index

There is one B6C3F1 mouse study that deserves special mention: the National Toxicology Program 2,4,5-trichlorophenoxyacetic acid study, which used a sample size of 6 animals per group. The top dose in this study produced a mean SI response of 3.03, which is the weakest "Ratio Rule positive" of any study in the three databases (control dpm responses were 63-69-75-90-119-133 compared with 213-229-244-249-325-405 in the top dosed group). The impact of reducing the sample size from 6 to 5 or 4 animals per group is summarized below.

Table B-8 Example Showing Effect of Sample Size on Agreement of Results for a Test Substance with SI = 3.03

	Two N = 5 samples	One N = 5 sample and one N = 4 sample
Agreement (SI > 3)	25.0% (18/36) (18/36)	26.4% (18/36) (119/225)
Agreement (SI < 3)	25.0% (18/36) (18/36)	23.6% (18/36) (106/225)
Disagreement (one SI > 3; one SI < 3)	50.0% (by subtraction)	50.0% (by subtraction)

Abbreviations: N = number of animals per dose group; SI = stimulation index

For these data, there is 50% disagreement between samples of size 4 and samples of size 5, but there is also 50% disagreement between two samples of size 5. This is a somewhat extreme example of the point made earlier, namely that most of the disagreement in Ratio Rule results observed between samples of size 5 and samples of size 4 shown in **Table B-1** is not due to the reduction in sample size, but rather due to the variability in response among animals and the closeness of the SI to the cutoff point of 3.

Finally, it is important to understand that **Table B-1** is not measuring accuracy; it is measuring agreement. That is, **Table B-1** assesses the reliability of N = 5 and N = 4 samples to produce the same classification outcome using the Ratio Rule; it does not assess the ability of N = 5 and N = 4 samples to produce the correct sensitizer classification (which for most substances is not known in any case). As illustrated in this report, as SI approaches 3, different samples may produce different classifications using the Ratio Rule, regardless of sample size, because of naturally occurring variability among animals. Importantly, most of the discordance between N = 5 and N = 4 samples shown in **Table B-1** is *not* due to the reduction in sample size.

With regard to accuracy of classification using the Ratio Rule, for 90% (75/83) of the CBA studies, there is no difference in accuracy using N = 5 and N = 4, based on the top dose group SI response. For eight studies, each with a top dose SI close to 3, there are slight differences in agreement, as shown in **Table B-9**.

Table B-9 Likelihood of SI > 3 for All CBA Studies Showing Less than Complete Agreement for the Top Dose Response Using N = 5 and N = 4 Samples

Substance	Top Dose SI	Likelihood of SI > 3 (%)	
		N = 5	N = 4
Formulation 54	2.3	0 (0/36)	7 (16/225)
Hexyl cinnamic aldehyde	2.8	28 (10/36)	38 (85/225)
Formulation 39	3.3	92 (33/36)	78 (175/225)
Bakelite EPR 161	3.5	83 (30/36)	77 (174/225)
Formulation 55	3.7	100 (36/36)	90 (202/225)
Potassium dichromate	4.1	100 (1/1)	92 (23/25)
Formulation 51	4.5[1]	100 (36/36)	96 (215/225)
1,6-(Bis(2-3-epoxypropoxy)hexane	4.7	100 (36/36)	94 (211/225)

Abbreviations: N = number of animals per dose group; SI = stimulation index
[1]Maximum response seen at mid-dose rather than top dose.

It is not known with certainty whether or not these eight substances are truly sensitizers. The one exception may be hexyl cinnamic aldehyde, which was confirmed in four other studies to be positive, with three showing SI > 4 at this dose. Thus, for this one compound the N = 4 sample may actually be more likely to be "accurate" than the N = 5 sample using the Ratio Rule.

If we assume that the Ratio Rule classifies all other substances correctly, and thus all six substances in **Table B-9** with SI > 3 are sensitizers, then there is a small loss in power by reducing the sample size per group from 5 to 4. However, this difference in power is small, and for all six substances, the likelihood is still quite high (77% - 96%) that the substance will be identified as a sensitizer using a sample of size 4. Recall also that these are "worst cases" and that for 90% of the CBA studies there is no difference in power at all between samples of N = 5 and N = 4. Thus, not only does the reduction in sample size from N = 5 to N = 4 have little impact on reliability using the Ratio Rule, it also appears to have little impact on the accuracy of classification.

5.0 Conclusion

For strong sensitizers and for obvious non-sensitizers, the reduction in sample size from 5 to 4 will have essentially no impact on the observed study outcome using the Ratio Rule. For those substances having an SI between (approximately) 2 and 4, the outcomes may be different, especially as SI approaches 3, but any such differences reflect primarily the inherent variability among animals and the closeness of the SI to 3 rather than the impact of reducing the sample size. Empirical examination of data from 83 CBA LLNA studies confirms that it is very unlikely that a reduction in sample size from 5 to 4 animals per group would have any impact on the overall interpretation of study results using the Ratio Rule.

Although the BALB/c and B6C3F1 databases were not evaluated in detail, the pattern of LLNA response seen in these strains is very similar to that seen in the larger CBA database, so a more definitive analysis of these other two strains would almost certainly not change the conclusions of this report. We conclude that a reduction in the sample size from 5 to 4 animals per group is unlikely to significantly impact the results of the LLNA test when using the Ratio Rule.

If a formal statistical test is used rather than (or in addition to) the Ratio Rule, the effect of reducing the sample size from N = 5 to N = 4 is to decrease the power slightly. However, for SI > 3, the power differences between samples of N = 5 and N = 4 are minimal. Moreover, a statistical test based on 4 animals per group will identify more sensitizers than using the Ratio Rule based on 5 animals per group. Thus, even if a formal statistical test is used rather than (or in addition to) the Ratio Rule, the practical impact of reducing the sample size from 5 to 4 animals per group on the interpretation of experimental results appears to be minimal.

[This Page Intentionally Left Blank]

Annex I:
Summary of Study Results – CBA Mouse Database

Table B-10 Experiments Conducted at ECPA Laboratories

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Dincocap EC 0.8	5	175	50	5	471	198	2.7[3]	88 (22/25)
Dincocap EC 4.0	5	175	50	5	4007	1578	22.9[3]	100
Dincocap EC 10.0	5	175	50	4	7088	1863	40.5[3]	100[4]
Formaldehyde-1 1.0	5	163	59	5	125	12	0.8	100
Formaldehyde-1 5.0	5	163	59	5	208	147	1.3	100
Formaldehyde-1 20.0	5	163	59	5	781	439	4.8[3]	100
Formaldehyde-2 1.0	5	844	513	5	838	737	1.0	100
Formaldehyde-2 5.0	5	844	513	5	1824	1341	2.2	92 (23/25)
Formaldehyde-2 20.0	5	844	513	5	5188	2845	6.1[3]	100
HCA-1 3.0	5	430	154	5	571	153	1.3	100
HCA-1 10.0	5	430	154	5	955	368	2.2[3]	100
HCA-1 30.0	5	430	154	5	1870	376	4.3[3]	100
HCA-2 3.0	5	708	172	5	1353	649	1.9[3]	100
HCA-2 10.0	5	708	172	5	2981	1422	4.2[3]	100
HCA-2 30.0	5	708	172	5	6525	4014	9.2[3]	100
Oxyfluorfen EC 1	5	192	117	5	238	67	1.2	100
Oxyfluorfen EC 7	5	192	117	5	234	162	1.2	100
Oxyfluorfen EC 33	5	192	117	5	1043	311	5.4[3]	100
Potassium dichromate 0.02	5	153	84	5	260	139	1.7	100
Potassium dichromate 0.10	5	153	84	5	234	135	1.5	100
Potassium dichromate 0.50	5	153	84	5	626	390	4.1[3]	92 (23/25)
Quinoxyfen/ cyproconazole 7	5	226	86	5	283	102	1.3	100
Quinoxyfen/ cyproconazole 33	5	226	86	5	1470	276	6.5[3]	100
Quinoxyfen/ cyproconazole 100	5	226	86	5	3075	621	13.6[3]	100
Trifluralin EC 7	5	194	46	5	357	163	1.8[3]	100
Trifluralin EC 33	5	194	46	5	1585	349	8.2[3]	100
Trifluralin EC 100	5	194	46	5	3965	1456	20.5[3]	100

Abbreviations: EC = emulsion concentrate; ECPA = European Crop Protection Association; HCA = hexyl cinnamic aldehyde; N = number of animals per dose group; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] These SIs are significantly different ($p < 0.05$) from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

[4] Although N = 4 for the experimental group, the responses in this particular group clearly would have shown 100% concordance between the outcomes for N = 5 and N = 4.

Table B-11 Experiments Conducted at BASF Laboratories

Study[1]	Control N	Control Mean	Control SD	Experi-mental N	Experi-mental Mean	Experi-mental SD	SI	Agreement (%)[2]
SC-1 3	6	626	216	6	511	124	0.8	100
SC-1 10	6	626	216	6	789	245	1.3	100
SC-1 30	6	626	216	6	1168	414	1.9[3]	100
HCA-3 2.5	6	1322	465	6	1479	161	1.1	100
HCA-3 5	6	1322	465	6	1571	921	1.2	100
HCA-3 10	6	1322	465	6	3749	1791	2.8[3]	55[4]
HCA-4 3	6	703	197	5	3209	1479	4.6[3]	100
HCA-4 10	6	703	197	6	4659	1409	6.6[3]	100
HCA-4 30	6	703	197	6	6929	1187	9.9[3]	100
HCA-5 10	5	176	26	5	711	240	4.1[3]	100
HCA-5 30	5	176	26	5	1362	611	7.8[3]	100
HCA-5 50	5	176	26	5	849	422	4.8[3]	100
1,6-Bis(2,3-epoxypropoxy)hexane 0.3	6	967	454	6	913	81	0.9	100
1,6-Bis(2,3-epoxypropoxy)hexane 1.0	6	967	454	6	1611	584	1.7	100
1,6-Bis(2,3-epoxypropoxy)hexane 3.0	6	967	454	6	4500	3061	4.7[3]	94 (211/225)
m-Phenylenebis (methylamine) 0.3	6	468	154	6	900	440	1.9[3]	100
m-Phenylenebis (methylamine) 1.0	6	468	154	6	4256	1298	9.1[3]	100
m-Phenylenebis (methylamine) 3.0	6	468	154	6	20691	6436	44.2[3]	100
Oxirane, mono((C12-14-alkyloxy) methyl) derivs 0.3	6	218	96	6	512	218	2.3[3]	92 (208/225)
Oxirane, mono((C12-14-alkyloxy) methyl) derivs 1.0	6	218	96	6	908	598	4.2[3]	92 (206/225)
Oxirane, mono((C12-14-alkyloxy) methyl) derivs 3.0	6	218	96	6	4963	1861	22.7[3]	100
1,2-Diaminocyclohexane 0.1	5	446	327	6	528	114	1.2	100
1,2-Diaminocyclohexane 0.3	5	446	327	6	810	290	1.8	100
1,2-Diaminocyclohexane 1.0	5	446	327	6	3736	1982	8.4[3]	100
Trimethylhexamine diamine 1.0	6	742	448	6	1599	400	2.2[3]	88[5]

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Trimethylhexamine diamine 3.0	6	742	448	6	2972	1191	4.0[3]	93 (209/225)
Trimethylhexamine diamine 10.0	6	742	448	6	6581	1250	8.9[3]	100
1-(2,3-epoxypropoxy)-2,2-bis[(2,3-epoxypropoxy) methylbutane 1.0	6	388	310	6	797	392	2.1[3]	81[6]
1-(2,3-epoxypropoxy)-2,2-bis[(2,3-epoxypropoxy) methylbutane 3.0	6	388	310	6	2531	1812	6.5[3]	100
1-(2,3-epoxypropoxy)-2,2-bis[(2,3-epoxypropoxy) methylbutane 10.0	6	388	310	6	4644	2150	12.0[3]	100
3-Aminomethyl-3,5,5-trimethylcyclohexylamine 0.3	6	309	85	6	384	134	1.2	100
3-Aminomethyl-3,5,5-trimethylcyclohexylamine 1.0	6	309	85	6	806	248	2.6[3]	86[7]
3-Aminomethyl-3,5,5-trimethylcyclohexylamine 3.0	6	309	85	6	6597	1867	21.4[3]	100
Dipropylene triamine 0.3	6	349	101	6	753	228	2.2[3]	100
Dipropylene triamine 1.0	6	349	101	6	1106	254	3.2[3]	59[8]
Dipropylene triamine 3.0	6	349	101	6	4344	1350	12.4[3]	100
N-(2-Hydroxyethyl)-ethylendiamine 3.0	6	445	179	6	891	277	2.0[3]	100
N-(2-Hydroxyethyl)-ethylendiamine 10.0	6	445	179	6	766	230	1.7[3]	100
N-(2-Hydroxyethyl)-ethylendiamine 30.0	6	445	179	6	2937	626	6.6[3]	100
p-tert-Butylphenyl 1-(2,3-epoxy)propyl ether 0.1	6	406	83	6	553	148	1.4	100
p-tert-Butylphenyl 1-(2,3-epoxy)propyl ether 0.3	6	406	83	6	681	230	1.7[3]	100
p-tert-Butylphenyl 1-(2,3-epoxy)propyl ether 1.0	6	406	83	6	5780	3279	14.2[3]	100
Bakelite EPR 161 0.1	6	770	189	6	789	108	1	100
Bakelite EPR 161 0.3	6	770	189	6	1825	733	2.4[3]	99 (222/225)
Bakelite EPR 161 1.0	6	770	189	6	2694	1652	3.5[3]	68[9]
Bakelite EPR 162 0.3	6	591	251	6	6225	3285	10.5[3]	100
Bakelite EPR 162 1.0	6	591	251	6	11790	4292	19.9[3]	100
Bakelite EPR 162 3.0	6	591	251	6	23583	3469	39.9[3]	100

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Bakelite EPR 164 0.3	6	463	208	6	2920	1049	6.3[3]	100
Bakelite EPR 164 1.0	6	463	208	6	8427	1833	18.2[3]	100
Bakelite EPR 164 3.0	6	463	208	6	10387	7000	22.4[3]	100

Abbreviations: EPR = epoxy resin; N = number of animals per dose group; SC = suspension concentrate; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses or footnoted indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] These SIs are significantly ($p < 0.05$) different from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

[4] 55% = (26/36 x 140/225) + (10/36 x 85/225)

[5] 88% = (35/36 x 204/225) + (1/36 x 21/225)

[6] 81% = (33/36 x 195/225) + (3/36 x 30/225)

[7] 86% = (35/36 x 198/225) + (1/36 x 27/225)

[8] 59% = (27/36 x 152/225) + (9/36 x 73/225)

[9] 68% = (30/36 x 174/225) + (6/36 x 51/225)

Table B-12 Experiments Conducted at DuPont Laboratories

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
DU-1A 5	5	506	185	5	284	122	0.6	100
DU-1A 25	5	506	185	5	596	166	1.2	100
DU-1A 50	5	506	185	5	354	198	0.7	100
DU-1A 100	5	506	185	5	526	313	1.0	100
DU-1B 1	5	1067	301	5	635	202	0.6	100
DU-1B 5	5	1067	301	5	1165	386	1.1	100
DU-1B 10	5	1067	301	5	1413	1145	1.3	100
DU-1B 25	5	1067	301	5	1144	388	1.1	100
DU-1C 5	5	617	265	5	419	156	0.7	100
DU-1C 25	5	617	265	4	883	517	1.4	100[3]
DU-1C 50	5	617	265	5	1075	432	1.7	100
DU-1C 100	5	617	265	4	779	262	1.3	100[3]
DU-1D 5	5	1067	301	5	755	196	0.7	100
DU-1D 10	5	1067	301	5	1019	266	1.0	100
DU-1D 25	5	1067	301	5	1337	493	1.3	100
DU-1D 50	5	1067	301	4	1086	281	1.0	100[3]
DU-2A 5	5	992	446	5	4132	815	4.2[4]	100
DU-2A 25	5	992	446	5	5422	939	5.5[4]	100
DU-2A 50	5	992	446	5	6604	1282	6.7[4]	100
DU-2A 100	5	992	446	5	6482	724	6.5[4]	100
DU-2E 5	5	452	219	5	433	169	1.0	100
DU-2E 25	5	452	219	5	370	142	0.8	100
DU-2E 50	5	452	219	5	509	285	1.1	100
DU-2E 100	5	452	219	5	623	200	1.4	100
DU-3 5	5	917	533	5	531	231	0.6	100
DU-3 10	5	917	533	5	720	306	0.8	100
DU-3 25	5	917	533	5	699	174	0.8	100
DU-3 50	5	917	533	5	538	179	0.6	100
DU-4 5	5	516	114	5	439	203	0.9	100
DU-4 25	5	516	114	5	505	257	1.0	100
DU-4 50	5	516	114	5	500	200	1.0	100
DU-4 100	5	516	114	5	538	65	0.9	100
DU-5A 5	5	589	317	5	1576	504	2.7[4]	76 (19/25)
DU-5A 25	5	589	317	5	903	534	1.5	100
DU-5A 50	5	589	317	5	915	223	1.6	100

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
DU-5A 100	5	589	317	5	499	230	0.8	100
DU-5B 5	5	1057	256	5	835	406	0.8	100
DU-5B 25	5	1057	256	5	1168	352	1.1	100
DU-5B 50	5	1057	256	5	1087	200	1.0	100
DU-5B 100	5	1057	256	5	1200	394	1.1	100
DU-5C 1	5	354	140	5	491	136	1.4	100
DU-5C 5	5	354	140	5	692	313	2.0[4]	100
DU-5C 25	5	354	140	5	429	195	1.2	100
DU-5C 100	5	354	140	5	312	124	0.9	100
DU-6 5	4	468	290	5	503	300	1.1	100[3]
DU-6 25	4	468	290	5	381	106	0.8	100[3]
DU-6 50	4	468	290	5	400	176	0.9	100[3]
DU-6 80	4	468	290	5	440	211	0.9	100[3]
DU-7 5	5	721	191	5	1394	1154	1.9	100
DU-7 25	5	721	191	5	846	331	1.2	100
DU-7 50	5	721	191	5	817	286	1.1	100
DU-7 80	5	721	191	5	915	249	1.3	100
DU-8A 1	9	486	186	4	680	178	1.4	100[3]
DU-8A 10	9	486	186	5	658	261	1.4	100
DU-8A 50	9	486	186	4	391	184	0.8	100[3]
DU-8A 100	9	486	186	5	473	263	1.0	100
DU-8B 5	5	786	312	5	916	460	1.2	100
DU-8B 25	5	786	312	5	1515	621	1.9	100
DU-8B 50	5	786	312	5	1121	764	1.4	100
DU-8B 100	5	786	312	5	1422	921	1.8	100
DU-9A 5	5	677	307	5	2405	1569	3.6[4]	84 (21/25)
DU-9A 25	5	677	307	5	3354	1463	5.0[4]	100
DU-9A 50	5	677	307	5	5975	773	8.8[4]	100
DU-9A 100	5	677	307	5	9118	3211	13.5[4]	100
DU-9B 5	5	1049	285	5	809	362	0.8	100
DU-9B 25	5	1049	285	5	822	195	0.8	100
DU-9B 50	5	1049	285	5	622	242	0.6	100
DU-9B 100	5	1049	285	5	493	88	0.5	100
DU-10 0.5	5	177	67	5	174	25	1.0	100
DU-10 1.0	5	177	67	5	230	73	1.3	100
DU-10 2.5	5	177	67	5	265	55	1.5	100

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
DU-10 5.0	5	177	67	3	289	122	1.6	NC[5]
DU-11B 5	5	984	210	5	1362	561	1.4	100
DU-11B 25	5	984	210	5	639	449	0.6	100
DU-11B 50	5	984	210	5	651	531	0.7	100
DU-11B 100	5	984	210	5	1016	1032	1.0	100
DU-11C 5	5	769	310	5	1168	472	1.5	100
DU-11C 25	5	769	310	5	871	217	1.1	100
DU-11C 50	5	769	310	5	719	133	0.9	100
DU-11C 100	5	769	310	5	1113	300	1.4	100
DU-12 1	5	617	265	5	479	132	0.8	100
DU-12 5	5	617	265	5	749	378	1.2	100
DU-12 25	5	617	265	5	477	253	0.8	100
DU-12 50	5	617	265	5	872	497	1.4	100
DU-13A 5	5	621	455	5	284	67	0.5	100
DU-13A 25	5	621	455	5	276	93	0.4	100
DU-13A 50	5	621	455	5	322	167	0.5	100
DU-13A 100	5	621	455	5	370	56	0.6	100
DU-13B 1	5	578	161	5	703	450	1.2	100
DU-13B 10	5	578	161	5	551	179	1.0	100
DU-13B 50	5	578	161	5	413	117	0.7	100
DU-13B 100	5	578	161	5	376	201	0.7	100

Abbreviations: DU = DuPont; N = number of animals per dose group; NC = not calculated; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] Although N = 4 for the experimental group, the responses in this particular group clearly would have shown 100% concordance between the outcomes for N = 5 and N = 4.

[4] These SIs are significantly ($p < 0.05$) different from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

[5] Agreement could not be assessed, since N < 4.

Table B-13 Experiments Conducted at EFfCI Laboratories

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Fumaric Acid 5	5	327	85	5	419	126	1.3	100
Fumaric Acid 10	5	327	85	5	742	284	2.3[3]	100
Fumaric Acid 25	5	327	85	5	479	201	1.5	100
Linoleic Acid 10	5	223	133	5	326	176	1.5	100
Linoleic Acid 25	5	223	133	5	1567	303	7.0[3]	100
Linoleic Acid 50	5	223	133	5	2025	601	9.1[3]	100
Linoleic Acid 10	5	223	133	5	699	301	3.1[3]	56 (14/25)
Linoleic Acid 25	5	223	133	5	2075	344	9.3[3]	100
Linoleic Acid 50	5	223	133	5	2290	1174	10.3[3]	100
Maleic Acid 10	5	327	85	5	2186	934	6.7[3]	100
Maleic Acid 25	5	327	85	5	5262	686	16.1[3]	100
Maleic Acid 50	5	327	85	5	5244	2304	16.0[3]	100
Octinol 10	5	1120	512	5	6327	1446	5.6[3]	100
Octinol 25	5	1120	512	5	9833	2523	8.8[3]	100
Octinol 50	5	1120	512	4	12594	1250	11.2[3]	100[4]
Oleic Acid 10	5	223	133	5	581	408	2.6[3]	84 (21/25)
Oleic Acid 25	5	223	133	5	3336	1688	14.9[3]	100
Oleic Acid 50	5	223	133	5	1550	897	6.9[3]	100
Squalene 10	5	223	133	5	839	245	3.8[3]	100
Squalene 25	5	223	133	5	1536	209	6.9[3]	100
Squalene 50	5	223	133	5	1821	327	8.2[3]	100
Succinic Acid 5	5	327	85	5	376	146	1.1	100
Succinic Acid 10	5	327	85	5	407	113	1.2	100
Succinic Acid 25	5	327	85	5	420	243	1.3	100
Undecylenic Acid 10	5	223	133	5	556	140	2.5[3]	80 (20/25)
Undecylenic Acid 25	5	223	133	5	736	250	3.3[3]	84 (21/25)
Undecylenic Acid 50	5	223	133	5	991	149	4.4[3]	100

Abbreviations: EFfCI = European Federation for Cosmetics Ingredients; N = number of animals per dose group; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] These SIs are significantly ($p < 0.05$) different from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

[4] Although N = 4 for the experimental group, the responses in this particular group clearly would have shown 100% concordance between the outcomes for N = 5 and N = 4.

Table B-14 Experiments Conducted at BAuA Laboratories

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Yellow E-JD 3442 1	5	70	21	5	70	19	1.0	100
Yellow E-JD 3442 3	5	70	21	5	52	9	0.8	100
Yellow E-JD 3442 9	5	70	21	5	60	32	0.9	100
Yellow E-JD 3442 15	5	70	21	5	61	16	0.9	100
CI Reactive Red 231 1	5	70	21	5	334	147	4.8[3]	100
CI Reactive Red 231 3	5	70	21	5	234	78	3.4[3]	88 (22/25)
CI Reactive Red 231 9	5	70	21	5	305	121	4.4[3]	100
CI Reactive Red 231 15	5	70	21	5	317	105	4.6[3]	100
P-46 1	5	70	21	5	167	86	2.4[3]	100
P-46 3	5	70	21	5	175	73	2.5[3]	96 (24/25)
P-46 9	5	70	21	5	135	39	1.9[3]	100
P-46 15	5	70	21	5	175	45	2.5[3]	100
CI Reactive Yellow 174 1	5	70	21	5	288	62	4.1[3]	100
CI Reactive Yellow 174 3	5	70	21	5	231	70	3.3[3]	80 (20/25)
CI Reactive Yellow 174 9	5	70	21	5	385	242	5.5[3]	100
CI Reactive Yellow 174 15	5	70	21	5	539	114	7.8[3]	100
Navy 14 08 723 1	5	70	21	5	353	54	5.1[3]	100
Navy 14 08 723 3	5	70	21	5	335	116	4.8[3]	100
Navy 14 08 723 9	5	70	21	5	398	102	5.7[3]	100
Navy 14 08 723 15	5	70	21	5	361	90	5.2[3]	100
Dispersionsrot 2754 1	5	70	21	5	68	27	1.0	100
Dispersionsrot 2754 3	5	70	21	5	65	19	0.9	100
Dispersionsrot 2754 9	5	70	21	5	67	40	1.0	100

Abbreviations: BAuA = Federal Institute for Occupational Safety and Health (Germany); N = number of animals per dose group; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] These SIs are significantly ($p < 0.05$) different from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

Table B-15 Experiments Conducted at Dow AgroSciences Laboratories

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Formulation 29 5	6	567	305	6	1036	663	1.8	100
Formulation 29 25	6	567	305	6	913	200	1.6	100
Formulation 29 100	6	567	305	6	823	373	1.5	100
Formulation 30 5	6	536	258	6	947	253	1.8[3]	100
Formulation 30 25	6	536	258	6	3839	736	7.2[3]	100
Formulation 30 100	6	536	258	6	7269	1014	13.6[3]	100
Formulation 31 5	6	385	121	5	393	223	1.0	100
Formulation 31 25	6	385	121	5	724	215	1.9[3]	100
Formulation 31 100	6	385	121	6	696	262	1.8[3]	100
Formulation 32 5	6	332	346	6	2136	737	6.5[3]	100
Formulation 32 25	6	332	346	6	14833	6139	44.7[3]	100
Formulation 32 100	6	332	346	6	22965	5480	69.3[3]	100
Formulation 33 5	6	672	249	6	479	194	0.7	100
Formulation 33 25	6	672	249	6	913	496	1.4	100
Formulation 33 100	6	672	249	6	843	303	1.3	100
Formulation 34 5	6	385	121	6	713	331	1.9	100
Formulation 34 25	6	385	121	6	528	227	1.4	100
Formulation 34 100	6	385	121	6	581	216	1.5	100
Formulation 35 5	6	332	346	6	360	294	1.1	100
Formulation 35 25	6	332	346	6	383	158	1.2	100
Formulation 35 100	6	332	346	6	412	317	1.3	100
Formulation 37 1	6	744	359	6	1008	525	1.4	100
Formulation 37 5	6	744	359	6	1999	1687	2.7	56[4]
Formulation 37 15	6	744	359	6	5586	4162	7.5[3]	100
Formulation 38 5	6	889	520	6	960	515	1.1	100
Formulation 38 25	6	889	520	6	4098	1541	4.6[3]	100
Formulation 38 100	6	889	520	6	11232	2102	12.7[3]	100
Formulation 39 1	6	627	256	6	1076	268	1.7[3]	100
Formulation 39 5	6	627	256	6	1551	650	2.5[3]	84[5]
Formulation 39 25	6	627	256	6	2083	259	3.3[3]	73[6]
Formulation 40 1	5	821[7]	263	6	1481	621	1.8	100
Formulation 40 5	5	821[7]	263	6	2316	401	2.8[3]	73 (55/75)
Formulation 40 25	5	821[7]	263	6	4646	1833	5.7[3]	100
Formulation 41 5	6	1017	325	6	1936	1024	1.9[3]	100
Formulation 41 25	6	1017	325	6	1891	1133	1.9	100
Formulation 41 100	6	1017	325	5	5653[7]	2750	5.6[3]	100
Formulation 49 5	5	626[7]	298	6	442	250	0.7	100

Study[1]	Control N	Control Mean	Control SD	Experimental N	Experimental Mean	Experimental SD	SI	Agreement (%)[2]
Formulation 49 25	5	626[7]	298	6	880	444	1.4	100
Formulation 49 100	5	626[7]	298	5	2958	489	4.7[3]	100
Formulation 50 5	6	1208	882	6	796	183	0.7	100
Formulation 50 25	6	1208	882	6	786	436	0.7	100
Formulation 50 100	6	1208	882	6	9439	4239	7.8[3]	100
Formulation 51 5	6	863	526	6	1346	537	1.6	100
Formulation 51 25	6	863	526	6	3893	2120	4.5[3]	96 (215/225)
Formulation 51 100	6	863	526	6	2084	1725	2.4	66[8]
Formulation 53 2.5	5	392[7]	159	6	596	317	1.5	100
Formulation 53 7.5	5	392[7]	159	6	1240	987	3.2[3]	52[9]
Formulation 53 15	5	392[7]	159	4	2609	1494	6.7[3]	100[10]
Formulation 54 5	6	438	143	6	551	357	1.3	100
Formulation 54 25	6	438	143	6	502	262	1.2	100
Formulation 54 100	6	438	143	6	1016	583	2.3	93 (209/225)
Formulation 55 5	6	529	238	6	781	602	1.5	100
Formulation 55 25	6	529	238	6	1348	947	2.5[3]	68[11]
Formulation 55 100	6	529	238	6	1972	758	3.7[3]	90 (202/225)
Formulation 56 5	6	529	238	6	1726	831	3.3[3]	57[12]
Formulation 56 25	6	529	238	6	3217	1996	6.1[3]	100
Formulation 56 100	6	529	238	2	2064	21	3.9[3]	NC[13]

Abbreviations: N = number of animals per dose group; NC = not calculated; SD = standard deviation; SI = stimulation index

[1] Test substance and dose tested (%)

[2] Agreement (%) between N = 5 and N = 4 for the Ratio Rule. When agreement is less than 100%, numbers in parentheses or footnoted indicate the proportion of the total number of N = 4 and N = 5 dose group combinations that agree with respect to whether SI < 3 or SI > 3. This is calculated by multiplying the proportion of N = 5 dose groups yielding SI > 3 with the proportion of N = 4 dose groups yielding SI > 3 and then adding the product of the proportion of N = 5 dose groups yielding SI < 3 with the proportion of N = 4 dose groups yielding SI < 3.

[3] These SIs are significantly ($p < 0.05$) different from 1 based on a Student's *t* test applied to the logged disintegrations per minute data.

[4] 56% = (26/36 x 142/225) + (10/36 x 83/225)

[5] 84% = (35/36 x 194/225) + (1/36 x 31/225)

[6] 73% = (33/36 x 175/225) + (3/36 x 50/225)

[7] Data reflects elimination of one control outlier (4258) in Formulation 40, one dosed group outlier (428) in Formulation 41, one control outlier (3) and one dosed group outlier (6273) in Formulation 49, and one control outlier (3172) in Formulation 53.

[8] 66% = (29/36 x 172/225) + (7/36 x 53/225)

[9] 52% = (4/6 x 42/75) + (2/6 x 33/75)

[10] Although N = 4 for the experimental group, the responses in this particular group clearly would have shown 100% concordance between the outcomes for N = 5 and N = 4.

[11] 68% = (31/36 x 168/225) + (5/36 x 57/225)

[12] 57% = (26/36 x 150/225) + (10/36 x 75/225)

[13] Agreement could not be assessed, since N < 4.