Altern. Animal Test. EXperiment. 5, 39-58 (1998)

Validation Study on Five Cytotoxicity Assays by JSAAE
II.
Statistical Analysis

Takashi Omori[1], Kaoru Saijo[2], Mayako Kato[2], Makoto Hayashi[3], Hiroshi Itagaki[4], Satoru Miyazaki[5], Tadao Ohno[2*], Hideaki Sugawara[5], Noriko Teramoto[2], Noriho Tanaka[6], Shinobu Wakuri[6], and Isao Yoshimura[1]

(The Working Group for data analyses in this validation study)

[1]Department of Management Science, Faculty of Engineering, Science Univeristy of Tokyo, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162, [2]RIKEN Cell Bank, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyodai, Tsukuba Science City, Ibaraki 305, [3]Shiseido Safety & Analytical Research Center, 1050 Nippa-cho, Kohoku-ku, Yokohama-shi 223, [4]Division of Genetics and Mutagenesis, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158, [5]Life Science Research Information Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-01, [6]Hatano Research Institute, Food and Drug Safety Center, 729-5 Ochiai, Hadano-shi, Kanagawa 257, Japan

Abstract

The validation study organized by JSAAE (JSAAE study) brought to light three main statistical problems, namely, (1) how to obtain a reliable estimate of ED50 from each experiment, (2) how to check the availability of the ED50 estimate obtained from each experiment, and (3) how to evaluate the feasibility of assays included in the JSAAE study.

To resolve the first problem, the authors devised a computer program (LAP-JSAAE) on SAS which incorporated a non-linear least squares method to obtain estimates of ED50 based on a logistic model for raw measurements. To resolve the second problem, the authors set several criteria for checking data which had previously been treated with manual adjustments for trivial errors such as de-

* Correspondence : The Working Group of JSAAE for the 1st Validation Study on Cytotoxicity Assays, c/o T. Ohno, RIKEN Cell Bank, 3-1-1 Koyadai, Tsukuba Science City, Ibaraki 305, Japan

Key words: alternatives, chemical irritancy, colony formation assay, crystal-violet staining assay, cytotoxicity assay, Draize eye irritation test, ED50, inter-laboratory validation, LDH release assay, logistic model, MTT assay, neutral red uptake assay, non-linear least squares method

Abbreviations: CF, colony formation; CV, crystal-violet staining; ED50, 50% effective dose; FRLA, factor for remaining LDH activity; JSAAE, Japanese Society of Alternatives to Animal Experiments; JSAAE study, the validation study organized by JSAAE; LDH, lactate dehydrogenase release; MTT, 3-(4,5-dimethylthiazol-2-yl)2,5-diphenyl tetrazolium bromide; NR, neutral red uptake; PFD, power for distinction.

Received: July 4, 1997. Accepted: September 10, 1997.

scriptions out of format. They are related to detecting extraordinary large observations, inspecting excessively wide confidence intervals obtained by LAP-JSAAE, checking whether observed responses, for at least one dose, are within the range of 20% and 80%, assessing the lack of fit to the logistic model, and so on. To resolve the last problem, the authors devised the "power-for-distinction" (PFD) which was defined as the ratio of the range of medians to the mean value of the hinge-spreads for log(ED50), where medians and hinge-spreads are from inter-laboratory variation and the range and the mean are from chemical response variation.

After establishing the methods, the authors performed data analysis for the JSAAE study and concluded that the crystal-violet staining assay (CV) with HeLa S3 (SC) cells and the colony formation assay (CF) with HeLa S3 (SC) cells are reliable in the sense that they give high values of the PFD.

1. Introduction

In 1992, the Japanese Society for Alternatives to Animal Experiments (JSAAE) organized a validation study, the JSAAE study, to evaluate the feasibility of five cytotoxicity assays (CF, CV, LDH, MTT, and NR) as a part of alternatives for the Draize eye irritation test (Ohno *et al*., 1995). The principal results obtained in this study were reported in the preceding article (Validation Article I) in this issue, where the 50% effective dose (ED50) was the index used to assess the toxicity of chemicals.

The authors encountered some statistical issues during the course of data analysis in this JSAAE study, and accordingly, were forced to address them before the completion of data analysis. Through intensive discussions within the working group composed of the authors of this paper, the issues were reduced to the following three problems: (1) how to obtain a reliable estimate of ED50 from each experiment, (2) how to check the availability of the ED50 estimate obtained from each experiment, and (3) how to evaluate the feasibility of assays included in the JSAAE study.

We devised statistical methods to answer these problems, and subsequently applied them to analyse the data gathered in the JSAAE study. The results obtained and conclusions drawn were described in Validation Article I. In what follows, we explain these devised methods along with computer programs named LAP-JSAAE and make some remarks on their application to such data as in this study.

2. Data in the JSAAE study

In the JSAAE study, the toxicity of six chemicals, one of which was dupulicated under a double-mask code to assess the reproducibility of results, was measured by four cytotoxicity assays (CF, CV, MTT, and NR) and the LDH-release assay sub-divided into four assays (LDH-1, LDH-2A, LDH-2B, LDH-2C) using two cell lines as has been described (see Validation Article I in this issue).

The committee for the JSAAE study specified that one data file for each experiment with a chemical in each assay using a cell line be constructed by all laboratories that participated in the study. Therefore, one value of ED50 was assumed to have been obtained from each data file corresponding to one experiment. Each data file contained various data such as the date of the experiment, the name of laboratory, and so on. The essential details of raw data from each assay necessary to estimate ED50 are shown in Tables 1 through 4. The principal data in the four assays excluding CF were measurements of optical density (OD) while those in CF were numbers of colonies, reflecting the proportion of survived or damaged cells incubated in the corresponding well.

To formulate the devised methods, it was convenient to express the data by variables. We de-

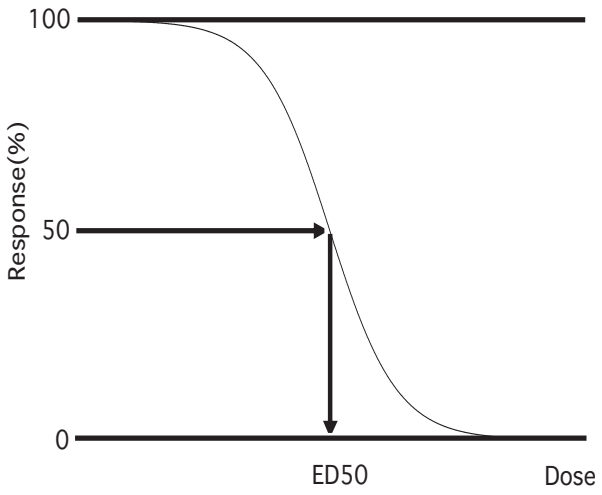

Fig. 1. Illustration of the relationship between dose-response curve and ED50

noted the OD measurement or the number of colonies by *z* with subscripts such as *i*, *j*, *B*, *N*, and *P*, corresponding to the *i-th* dose (or concentration) level d_i of the test chemical, the *j-th* repetition, the blank, i.e., vacant-cell-well, the negative control, i.e., zero-dose level, and the positive control, i.e., a toxic chemical specified in the protocol, respectively. Further, we denoted the number of dose levels, excluding negative control, positive control, and blank, by *a* and the number of repetitions of measurement by *r*. Numeric values of these variables given in Table 1 are such that z_{B1}=0.057, z_{N1}=0.528, d_1=0.005, z_{11}=0.680, z_{12}=0.255, a = 7, r = 6 and, in Table 2, z_{N1}=0.096, z_{P1}=0.349, d_1=0.25, z_{11}=0.101, and so on.

In the case of CF, the number of cells inoculated in each well was also necessary in order to estimate the proportions of survived cells, and hence was denoted by a variable such as s_B or s_i. Note that they were regarded as error-free values in the data analysis of the JSAAE study because it was believed that their variabilities did not affect the results. Likewise in case of the LDH assay, the factor for remaining LDH activity (FRLA) in the *i-th* dose, i = 1, 2, ... , a is necessary and hence denoted by c_i, the meaning of which is explained in Validation Article VI.

In the case of LDH-2A, 2B, and 2C, three interrelated data sheets, which were distinguished by superscripts A, C, and S, were used in part or simultaneously as is explained in Appendices A and B. The superscript "A" simply implies the first sheet, "C", the cell layer, and "S", the supernatant of the solution.

3. ED50 estimation

3.1 *Difficulty inherent in the conventional ED50 estimation*

In this section, we deal only with the case of the CV assay to explain the problems to be resolved, leaving the case of other assays to Appendices, since the issues in question were essentially the same for all assays.

In the field of *in vitro* cytotoxicity testing, the relationship of the survived proportion of incubated cells with the applied dose of chemical is referred to as the dose-response relationship or the dose-response curve, and the toxicity of the chemical tested is assessed by ED50 which is conceptually defined as the dose inflicting damage on 50% of incubated cells or, in other words, maintiaining 50% of survived cells as is seen in Fig. 1. Practically, however, the true dose-response curve as well as ED50 cannot be known exactly and, therefore, have to be estimated from experimental data.

In the case of the CV asssay, the response, i.e., the survived proportion p_i (%) of incubated cells treated with dose d_i, is conventionally calculated by Equation (1) where the $\bar{z}$ refers to the mean of the observed z values at each dose (i), blank (B), or negative control (N). Formulae for other assays are given in Appendix A.

To estimate ED50 from this type of dose-response data $\{(d_i, p_i); i = 1, 2, ... , a\}$, it is quite reasonable to assume a suitable function as the dose-response curve and fit it to the data. The dose corresponding to 50% response on the fitted function was taken to be the ED50 value. As for the suitable function and the method of curve fitting, a probit function, *i.e.*, the cumulative normal distribution function and a probit analy-

Table 1. An example of raw data in a CV assay

z_{ij}'s are OD measurements and d_i's are doses.

In this case, the number of dose levels is a = 7 and the number of repetition is r = 6. The same form of data as CV is used in the case of the NR assay.

File: 20CVC7.TXT	OD measurement		
Dose (% in W/V)	Plate 1	Plate 2	Plate 3
Blank	$z_{B1} = 0.057$	$z_{B2} = 0.052$	$z_{B3} = 0.064$
Negative control	$z_{N1} = 0.528$	$z_{N2} = 0.513$	$z_{N3} = 0.582$
$d_1 = 0.005$	$z_{11} = 0.680$	$z_{13} = 0.255$	$z_{15} = 0.592$
	$z_{12} = 0.581$	$z_{14} = 0.435$	$z_{16} = 0.444$
$d_2 = 0.01$	$z_{21} = 0.525$	$z_{23} = 0.313$	$z_{25} = 0.494$
	$z_{22} = 0.526$	$z_{24} = 0.456$	$z_{26} = 0.455$
...	...	...	...
$d_7 = 0.06$	$z_{71} = 0.060$	$z_{73} = 0.060$	$z_{75} = 0.060$
	$z_{72} = 0.059$	$z_{74} = 0.059$	$z_{76} = 0.061$

Table 2. An example of raw data in a MTT assay

z_{ij}'s are OD measurements and d_i's are doses.

In this case, the number of dose levels is a = 12 and the number of repetition is r = 3.

File: 3MH3.TXT	OD measurement		
Dose (% in W/V)	Plate 1	Plate 2	Plate 3
Negative control	$z_{N1} = 0.096$	$z_{N2} = 0.076$	$z_{N3} = 0.050$
Positive control	$z_{P1} = 0.349$	$z_{P2} = 0.419$	$z_{P3} = 0.701$
$d_1 = 0.25$	$z_{11} = 0.101$	$z_{12} = 0.057$	$z_{13} = 0.063$
$d_2 = 0.10$	$z_{21} = 0.555$	$z_{22} = 0.547$	$z_{23} = 0.684$
...	...	...	...
...	...	...	...
$d_{10} = 0.01$	$z_{101} = 0.465$	$z_{102} = 0.480$	$z_{103} = 1.128$
$d_{11} = 0.001$	$z_{111} = 0.525$	$z_{112} = 0.480$	$z_{113} = 0.737$
$d_{12} = 0.0001$	$z_{121} = 0.529$	$z_{122} = 0.472$	$z_{123} = 0.720$

sis method have been conventionally used as specified in the initial protocol for each assay. Applying the protocol to collected data, however, we found that reasonable estimates of ED50 could not be obtained in some cases such as shown in Fig. 2; in one case, the applied method did not yield any estimate and in another case, the acquired estimate was far from a reasonable value in the sense that it deviated from the visually plausible one as is shown in Fig. 3. After some examinations and discussions, we understood that the failure resulted from the inappro-

Table 3. An example of raw data in a CF assay

z_{ij}'s are number of colonies, d_i's are doses, and s_i's are number of seeds.
In this case, the number of dose levels is a = 6 and the number of repetition is r = 4.

File: 40CH6-3.TXT		Number of colonies			
	Seed				
Dose (% in W/V)		Dish 1	Dish 2	Dish 3	Dish 4
Negative control	$s_N = 100$	$z_{N1} = 92$	$z_{N2} = 77$	$z_{N3} = 100$	$z_{N4} = 106$
$d_1 = 0.012$	$s_1 = 100$	$z_{11} = 0$	$z_{12} = 0$	$z_{13} = 0$	$z_{14} = 0$
$d_2 = 0.011$	$s_2 = 100$	$z_{21} = 0$	$z_{22} = 2$	$z_{23} = 1$	$z_{24} = 1$
$d_3 = 0.010$	$s_3 = 100$	$z_{31} = 3$	$z_{32} = 2$	$z_{33} = 3$	$z_{34} = 2$
$d_4 = 0.009$	$s_4 = 100$	$z_{41} = 49$	$z_{42} = 38$	$z_{43} = 56$	$z_{44} = 53$
$d_5 = 0.008$	$s_5 = 100$	$z_{51} = 61$	$z_{52} = 69$	$z_{53} = 67$	$z_{54} = 58$
$d_6 = 0.007$	$s_6 = 100$	$z_{61} = 83$	$z_{62} = 84$	$z_{63} = 81$	$z_{64} = 80$

Table 4. An example of raw data in a LDH-1 assay

z_{ij}'s are OD measurements, d_i's are doses, and c_i's are factors of remaining LDH activity. In this case, the number of dose levels is a = 6 and the number of repetition is r = 6. The form of data in LDH-A, B and C is the same as this one.

File: 46AH1-2.TXT		OD measurement		
Dose (% in W/V)	Factor	Plate 1	Plate 2	Plate 3
Blank		$z_{B1} = 0.189$	$z_{B2} = 0.309$	$z_{B3} = 0.245$
Negative control	1.0	$z_{N1} = 0.190$	$z_{N2} = 0.300$	$z_{N3} = 0.252$
Positive control	1.0	$z_{P1} = 0.665$	$z_{P2} = 0.781$	$z_{P3} = 0.676$
$d_1 = 0.200$	$c_1 = 1.43$	$z_{11} = 1.195$	$z_{13} = 1.335$	$z_{15} = 1.271$
		$z_{12} = 1.129$	$z_{14} = 1.258$	$z_{16} = 1.192$
$d_2 = 0.125$	$c_2 = 1.42$	$z_{21} = 1.161$	$z_{23} = 1.297$	$z_{25} = 1.235$
		$z_{22} = 1.078$	$z_{24} = 1.201$	$z_{26} = 1.115$
... ...	...	... ...		
$d_6 = 0.050$	$c_5 = 1.41$	$z_{61} = 0.427$	$z_{63} = 0.555$	$z_{65} = 0.474$
		$z_{62} = 0.659$	$z_{64} = 0.739$	$z_{66} = 0.695$

priateness of the adopted method for the collected data.

As is well known (Finney, 1971, 1985a), the probit analysis method comprises a variety of algorithms for practical use, a typical one being the maximum likelihood method based on the assumption of binomial distribution and another one being the graphical method such as the Litchfield-Wilcoxon method (Litchfield & Wilcoxon, 1949). All of them, however, have a common feature in that they are essentially based on the line fitting with proper weight on the re-

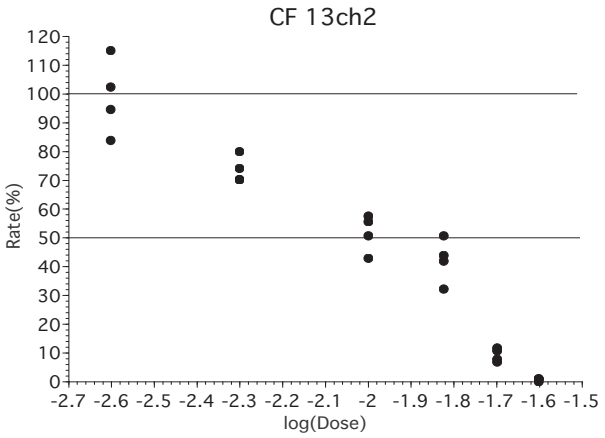

Fig. 2. An example of data file on which a probit analysis method failed to yield ED50 estimate.

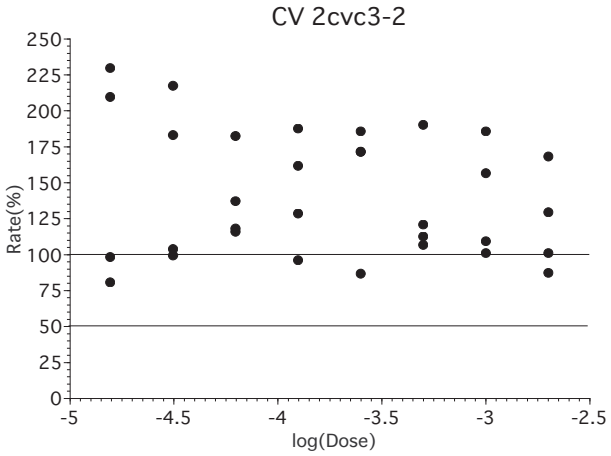

Fig. 3. An example of data file with which estimated ED50 by a probit analysis method is far different from a visually-obtained estimate.

The log(ED50) estimated by a probit analysis method is -12.6.

sponse values obtained through a transformation such as the probit transformation. This common feature, without doubt, brought us the above-mentioned difficulty because some of response values were outside the range of 0% and 100%, although they were assumed to be within this range. If we would apply the probit method to the collected data ignoring this feature, we should have removed such measurements ignoring the risk of biases in estimation. In reality, however, it was not acceptable and, accordingly, we were forced to devise another method to obtain reasonable ED50 estimates.

3.2 *Proposed method of ED50 estimation*

3.2.1 *Logistic model on the raw data*

Since the basic concept of the proposed method is essentially the same in all assays, we explained our idea using the case of the CV assay with a logistic function exemplified by the dose-response curve, leaving the details in other cases to Appendix B.

The difficulty mentioned above was obviously due to the improper recognition of errors such that the measurements, especially for the blank and control, are not affected at all, or at most in negligibly small orders, by measurement errors and, therefore, responses calculated by the formula (1) can neither be less than 0% nor greater than 100%.

In reality, however, they are affected by many sources of errors such as the number of cells incubated, time duration of processing, diluting of the concentration of the test chemical, measurement of OD, etc. Their influence is not at all negligible. Actually, the variability of measurements within the same dose is rather large, and the assumption of small errors is not correct as is seen in Fig. 2 and 3. Consequently, it became necessary to construct a suitable model on measurements incorporating the errors such as the one characterized above.

As a suitable model, we decided to use a logistic model, not on responses p, but on raw measurements z as Equation (2) through (5), where β_1, β_2, correspond to a location and a scale parameter; β_3 and β_4 are parameters corresponding to a blank and a negative control, respec-

tively, and ε's represent random errors. In this model, the log ED50 and ED50 are defined as by Equation (6) or Equation (7).

Note that, although the models for the other assays should have been modified from the above modeling according to the definition of responses explained in Appendix A, the function $f(d_i; \beta_1, \beta_2)$ was assumed to be common for all assays and, therefore, the estimation procedure in other assays was essentially the same as the one for CV except for the modification given in Appendix B.

3.2.2 *Non-linear least squares method for ED50 estimation*

Since there were various sources of measurement error which were thought to be additive and continuous in nature, it was reasonable to assume a normal distribution of errors. We decided to use the method of non-linear least squares to estimate parameters b_1, b_2, b_3 and b_4, *i.e.*, to use as estimates of b's the values of them which minimize Q defined by Equation (8).

Without doubt, it is natural to estimate ED50 by inserting acquired estimates of β_1 and β_2 into the definition of ED50 given by the formula (7), which is the proposed method of ED50 estimation.

3.2.3 *Delta method for constructing the confidence interval*

The precision of the ED50 estimate obtained from each experiment should be evaluated using a confidence interval for ED50. Three methods, namely, 1) the delta method (D-method), 2) the Fieller's method (F-method), and 3) the likelihood ratio method (L-method), can be considered for constructing the confidence interval, among which we adopted the D-method due to the reason discussed later in Chapter 8.

Let the estimates of β_1 and β_2 obtained by the proposed method be b_1 and b_2, respectively. Then the logarithm of the estimate of ED50, say logED, can approximately be expanded as Equation (9), which yields an approximate formula for the variance of logED as Equation (10).

By obtaining estimates of the variance and covariance of b_1 and b_2 through linear approximation in the non-linear least squares method (cf. SAS manual), an estimate of var(logED), say S^2, can be calculated by inserting them into the above formula. From it, we can obtain an approximate 95% confidence interval of logED50 as Equation (11), subsequently yielding the confidence interval of ED50 as Equation (12).

4. The method of data check

Before evaluating assays based on ED50s obtained from collected data files, we were compelled to check whether each data file should be deleted or not. This is because, even after careful cleaning and correction of data as described in Validation Article III in this issue, there remained some files with abnormal behavior, which led us to a suspicion that some data files might have been obtained through inappropriate experimentation. Since the existence of data files from inappropriate experimentation does not reflect the capability of the assay as alternatives to animal experiments but the impracticability of the assay, we excluded such data files from the evaluation of the capability of assays. Consequently, we decided to provide criteria for exclusion of data files as a “data check” in the JSAAE study after the correction of trivial errors such as mis-recording or out-of- format description of measurements.

Based on intensive discussions, we reduced the possible inappropriateness of experimenta-

tion using to the following checkpoints:

Check point expressed as Code-A

If LAP-JSAAE failed to yield any ED50 value, we regarded it as evidence of experiment failure, denoting with "FAILED" in Code-A, otherwise with "MET". A typical example, in which LAP-JSAAE outputs "FAILED" is shown in Fig. 3. We excluded the data files with "FAILED" output from the evaluation of assays. Note that, even when ED50 is obtained, the data file was excluded when the estimated value of b_2 was negative, because it was theoretically unacceptable in case of CV as well as other assays except LDH; in case of LDH, the data file was deleted when the estimated value of β_2 was positive.

Check point expressed as Code-B

If the correction factor is greater than 2.0 or less than 0.1 in case of LDH, we regarded it as evidence of inappropriate experimentation and excluded the corresponding data file from the evaluation of assays, because then the correction of the observed LDH activity to the original LDH activity is essentially meaningless as explained in Validation Article VI. Note that this checkpoint is applied only to the LDH assay.

Check point expressed as Code-C

In the above-mentioned model, q's in Equation(13), where the β_3 and β_4 replaced with their estimates, were observed responses corresponding to p_i in Equation (1), and, therefore, in this paper we refer to their average within the same dose d_i as the observed mean response.

If there was no dose, at which the observed mean response was within the range of 20% and 80%, we regarded it as evidence of inappropriate choice of doses in the experiment and excluded the corresponding data file from the evaluation of assays, outputting in Code-C, the number of doses with the mean response within this range. Note that, although the initial protocol had required the selection of doses in each experiment so as to hold, at least, three doses with observed mean response within this range, it was difficult maintaining this strict requirement. We therefore relaxed the restriction in order to utilize more data files, and prepared a new condition of keeping at least one dose in this range, which was required because an inappropriate ED50 was obtained when this condition was violated as is shown in Fig. 4.

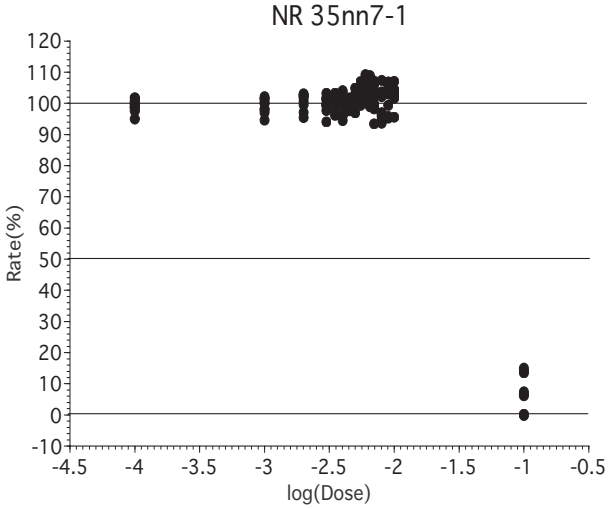

Fig. 4 An example of data file with no observed mean response within the range of 20% and 80%
The log(ED50) estimated by a probit analysis method is -12.6.

Check point expressed as Code-D

If the maximum observed response was greater than 200% in a given dose in an experiment, we regarded it as evidence of excessive errors in measurement and excluded the corresponding data file from the evaluation of assays, expressing it with "X" in Code-D, otherwise with "OK".

Check point expressed as Code-E

If the width of the induced 95% confidence interval of ED50 was excessive, *i.e.*, the upper confidence limit was greater than the lower limit by 100 times, we regarded it as evidence of inappropriate experimentation and excluded the corresponding data file from the evaluation of assays.

Check point expressed as Code-F

If the root-mean-square error (RMS) after fitting the logistic function was too large, *i.e.*, RMS is greater than 10%, we regarded it as evidence of inappropriateness of the assumed dose-response curve or the logistic model and excluded

the corresponding data file from the evaluation of assays, although it might occur in exceptional cases discussed later in Chapter 8. Here, RMS is formally expressed by Equation(14), where the parameters are replaced by their estimates. It is clear that RMS is an index which indicates the degree of deviation of the observed dose-response curve from the logistic model. In other words, a large value of RMS implies lack of fit of the logistic model to the data.

5. Criteria to evaluate the validity of assays

Three points were considered to be important in the evaluation of validity of each assay. Firstly, the result obtained by the assay must be effective in predicting the result of the *in vivo* animal experiment. In other words, ED50s measured through the assay must be highly correlated with Draize test scores on a variety of chemicals to be assessed. Secondly, ED50 values for the same assay must be reproducible, that is, almost the same values should be obtained regardless of which laboratory performed the assay. This is expected to allow only inter-laboratory variations among laboratories. Thirdly, the assay must be robust in the sense that there should rarely be room for misoperation in the conduct of the assay, which is assured by the simplicity of operation and realized by the number of files available to estimate ED50 values in the JSAAE study.

It was quite disadvantageous to evaluate the predictability of *in vivo* results from each assay where only six chemicals were included in the JSAAE study and half of them were non-irritant. For such small numbers of chemicals, the correlation coefficient was not a good measure of predictability of each assay even when the rank correlation instead of Pearson correlation was used. Therefore we had to devise another index for the evaluation.

As is clearly seen from dose-response curves shown in Validation Articles VI - VIII in this issue, the distribution of ED50 estimates on each assay, cell line, and chemical was not a normal one; outliers in both larger and smaller sides frequently appeared even after data cleaning and data check. This suggested the use of such statistical tools as the box-and-whisker plot proposed by Tukey (Tukey, 1977) to analyse data in this JSAAE study. Along this line of thought, we decided to use medians instead of means and hinge-spreads instead of standard deviations to evaluate the central tendency and the reproducibility of ED50 values in each assay, where the hinge-spread is defined by the difference of the upper quartile and the lower quartile.

Using these statistics, we devised an index to evaluate the capability of assays as alternatives to animal experiments, being the ratio of the range of medians among the chemicals over the mean hinge-spread taken on all chemicals. It is formally defined in Equation(15), and referred to as the “power-for-distinction (PFD)” index in the present study.

6. LAP-JSAAE

We developed a computer program named the “logistic analysis program by the Japanese Society of Alternatives to Animal Experiments (LAP-JSAAE)", which works on SAS, a widely

```
*************************************************************;
*                          CV_V6.SAS                        *;
*                                              LAP-JSAAE  Ver.6.2   *;
*                                                          9/5/1996  *;
*************************************************************;

OPTIONS LINESIZE=90 PAGESIZE=40 NODATE;

FILENAME FI_IN 'a:\6cvh1.out';

DATA YOMI1;
 YOMIGYO=2;
 CALL SYMPUT('YOMI',YOMIGYO);

%INCLUDE 'a:\cv_61.sas';

PROC PLOT DATA=IN12 HPCT=90 VPCT=90 NOLEGEND;
 PLOT PLATE*LOG_DOSE;

%INCLUDE 'a:\cv_62.sas';

PROC PLOT DATA=PARA3 HPCT=90 VPCT=90 NOLEGEND;
 PLOT P_PRED*LOG_DOSE='P' P_RATIO*LOG_DOSE='Y'
    / OVERLAY VREF=0 50 100;

PROC PRINT DATA=ED9 NOOBS LABEL;
 VAR F_NAME N P CODE_A CODE_C CODE_D CODE_E CODE_F
    ED50 LOW_ED50 UP_ED50 ;
RUN;
```

Fig. 5 Initial control program for LAP-JSAAE on SAS
The log(ED50) estimated by a probit analysis method is -12.6.

used statistical software package. The source program of LAP-JSAAE for which copyright is jointly held by Takashi Omori and JSAAE, together with an operation manual in Japanese, can be acquired upon request.

The LAP-JSAAE essentially consists of three parts, namely, 1) the ED50 estimation, 2) the data check, and 3) the graphic representation, and is optionally controlled with an initial setting as shown in Fig. 5. (see SAS manuals for specific options.)

An example of output with an input is shown in Fig. 6. In this output, “File” is the data file name to be analyzed which is given by the user, “N” is the total number of measurements in the data, and “P” is the number of dose levels included in the experiment excluding the negative and positive controls and blank. “A”, “C”, etc. are the output of check codes. The values in “LOW_ED50” and “UP_ED50” are the lower and upper limit of the 95% confidence interval of ED50, respectively. The value in “Code_F” is the value of RMS as already explained.

7. Results of statistical analysis

The collected data files were processed following the flow chart given in Fig. 7 (also shown

Fig. 6. An example of input and output of LAP-JSAAE

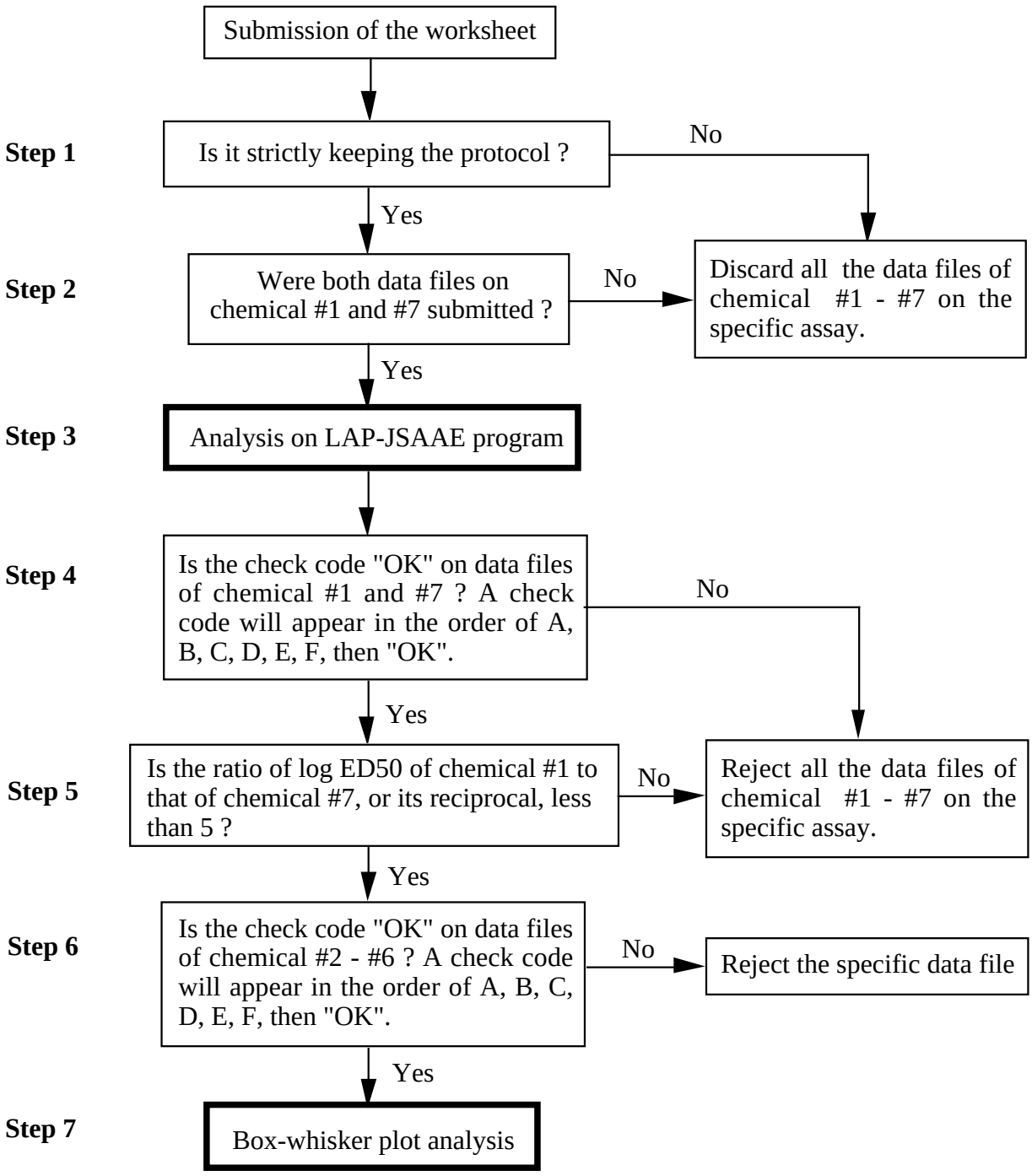

Fig. 7. Data file processing in the present study

Definitions of check codes are as follows:

Code-A : Calculation of ED50 values failed because of wide variation of data.

Code-B : In the LDH-release assay, LDH activity is directly inhibited or stimulated by the test chemical with the correction factor of below 0.1 or above 2.0, respectively. With these factors, correction of the observed LDH activity to the original LDH activity in a sample is essentially meaningless.

Code-C : No observed point between 20-80% of the maximum effect was found in the data file.

Code-D : Data included response of 200% or more (negative controls set at 100%).

Code-E : Of the 95% confidence limits of ED50, the upper limit was over 100 times that of the lower limit.

Code-F : RMS is 10 or more. RMS indicates the degree of deviation of observed dose-response relationship from the logistic model. See details in the following Validation Article II in this issue.

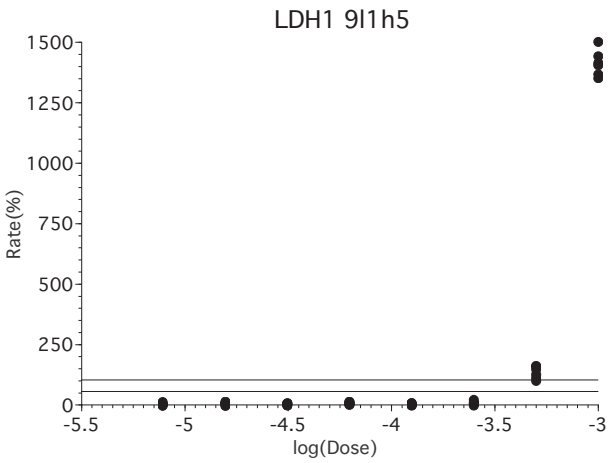

Fig. 8. An example data file deleted at the check point B

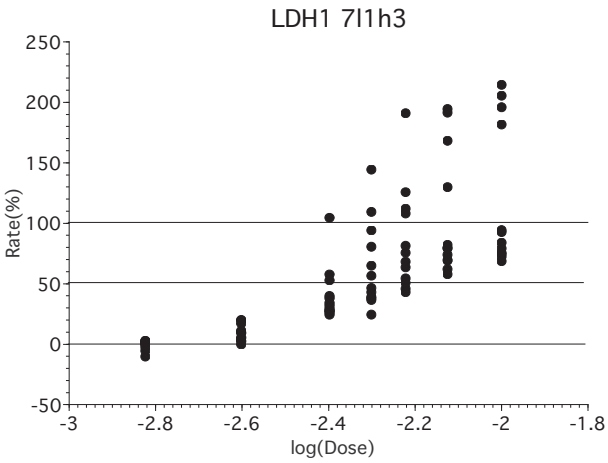

Fig. 9. An example data file deleted at the checkpoint D

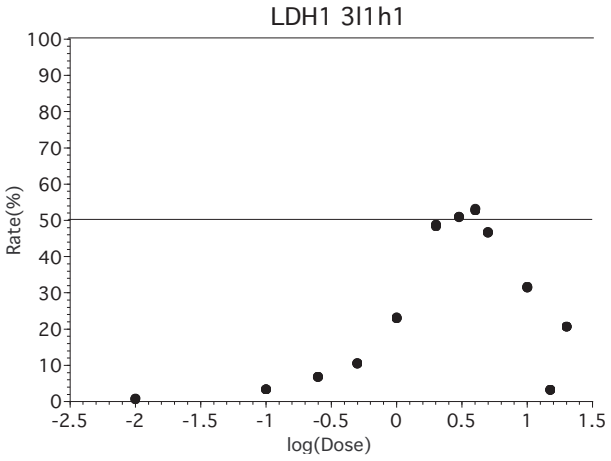

Fig. 10. An example data file deleted at the checkpoint E

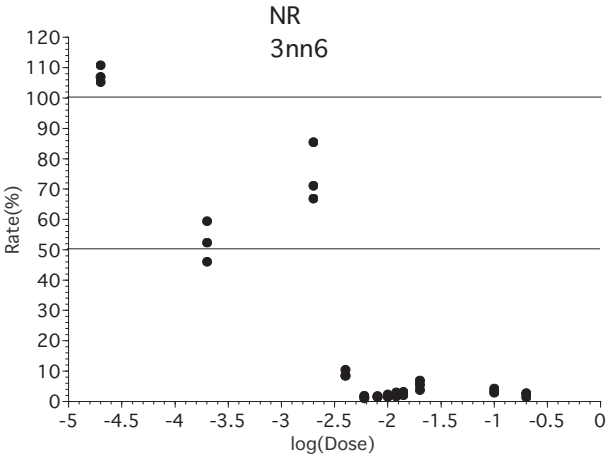

Fig. 11. An example data file deleted at the checkpoint F

as Fig. 1 in Validation Article I. in this issue). As explained in Validation Article III, much laborious work had been done during this study to inspect and correct trivial mistakes. We applied LAP-JSAAE to data which had been subjected to such cleaning and summarized the results in Table 5 (The same table shown in Validation Article I. in this issue, Table 5). Typical examples of data files excluded from the evaluation of assays are shown in Figs. 8 - 11.

Table 5 shows that the CV assay with HeLa S3 (SC) cells achieved the best score in PFD and in the number of data files available for the evaluation of assays; the second best score was for the MTT assay with HeLa S3 (SC) cells.

8. Discussion

8.1 PROC PROBIT for in vivo bioassay and LAP-JSAAE

Although the proposed method realized in the LAP-JSAAE uses a logistic function as the dose-response curve in place of a probit curve in the conventional probit analysis, the difference is no more essential than the difference between LAP-JSAAE and the probit analysis (Finney, 1985b). In fact, we can assume a logistic function as the dose-response curve for any *in vivo* bioassay and apply PROC PROBIT in SAS to the data entirely in the same manner as in the case of a probit curve, obtaining, in most cases, quite similar results on the same data.

The essential difference between the two methods lies in the estimation procedure; a non-linear least squares method is used in LAP-JSAAE, whereas a maximum likelihood method

Table 5. Number of files rejected or finally accepted by the logistic analysis program, LAP-JSAAE.

Assay	Cells	Candidate files	Rejected files								Finally accepted files
			Chemical #1 or #7 with Code A* ~ F	Ratio of ED50 of #1/#7 (#7/#1) is 5 or more	Chemical #2 - #6 with**						
					Code A	Code B	Code C	Code D	Code E	Code F $	
CF	HeLa S3 (SC)	116	0	6 &	2		7	0	1	0	100
	BALB/3T3 A31-1-1	149	7	4	3		7	0	0	2	126
CV	HeLa S3 (SC)	84	7	0	1		1	0	0	0	75
	CHL	83	0	0	2		1	1	0	1	78
LDH-1	HeLa S3 (SC)	57	7	0	2	1	6	1	0	1	39
	SQ-5	50	14	5 &	1	0	4	0	0	1	25
LDH-2A	HeLa S3 (SC)	54	21	0	2	2	1	0	0	2	26
	SQ-5	39	7	0	1	0	3	0	0	0	28
LDH-2B	HeLa S3 (SC)	41	21	0	0	0	0	0	0	2	18
SQ-5		34	14	0	0	0	0	0	0	0	20
LDH-2C	HeLa S3 (SC)	54	14	0	6	1	3	0	2	2	26
	SQ-5	40	7	0	2	0	0	0	1	5	25
MTT	HeLa S3 (SC)	95	0	0	2		2	0	0	2	89
	SQ-5	89	0	0	2		4	0	0	0	83
NR	HeLa S3 (SC)	126	7	7	0		3	0	0	1	108
	NRCE	132	7	7	0		5	3	3	4	103
Total files		1243	133	29	26	4	47	5	7	23	969

* Code A. Calculation of an ED50 value failed because of wide variation of data.

** Calculation met the requirement by the logistic analysis program LAP-JSAAE but with the code B, D, E, or F.

Code B indicates that, in the LDH-release assay, LDH activity is directly stimulated or inhibited by the test chemical with the correction factor of over 2.0 or under 0.1, respectively. With these factors, correction of the observed LDH activity to the original LDH activity in a sample is essentially meaningless.

Code-C indicates that no observed point was found between 20-80% of the maximum effect in the data file. Note that the numbers in this column do not contain the files which have been rejected under the check code A and B, therefore they are different from the number of data files shown in Table 4.

Code D indicates that data include response of 200% or more where that of negative control was 100%.

Code E indicates that, of the 95% confidence limits of ED50, the upper limit is over 100 times the lower limit.

$ Code F indicates that RMS was 10 or more. On the difinition of RMS, see the subsequent article on statistical analyses and LAP-JSAAE.

& The data file set on this assay did not include files on one or two of chemicals #2 - #6.

based on the assumption of binomial distribution or a modified version of it is used in the probit analysis. The reason why we adopted a non-linear least squares method for LAP-JSAAE lies in the characteristic of random variations or measurement errors in observed data. They are reasonably regarded to be normally distributed, being affected by many sources of variation in continuous type such as time setting, dilution of solution, volume measurement by pipette, OD measurement and so on. For data with such types of variation, the appropriate method is not a maximum likelihood method based on the binomial distribution as is usually assumed for *in vivo* assays, but a non-linear least squares method based on the normal distribution. The adoption of the latter automatically removed the difficulty of the probit analysis method which came from possible responses outside of the range of 0% and 100%.

8.2 Appropriateness of the logistic model

Although an ogive curve such as a probit curve or logic function is widely used as the dose-response curve, it should not be admitted without validation. In fact, some assays eventually

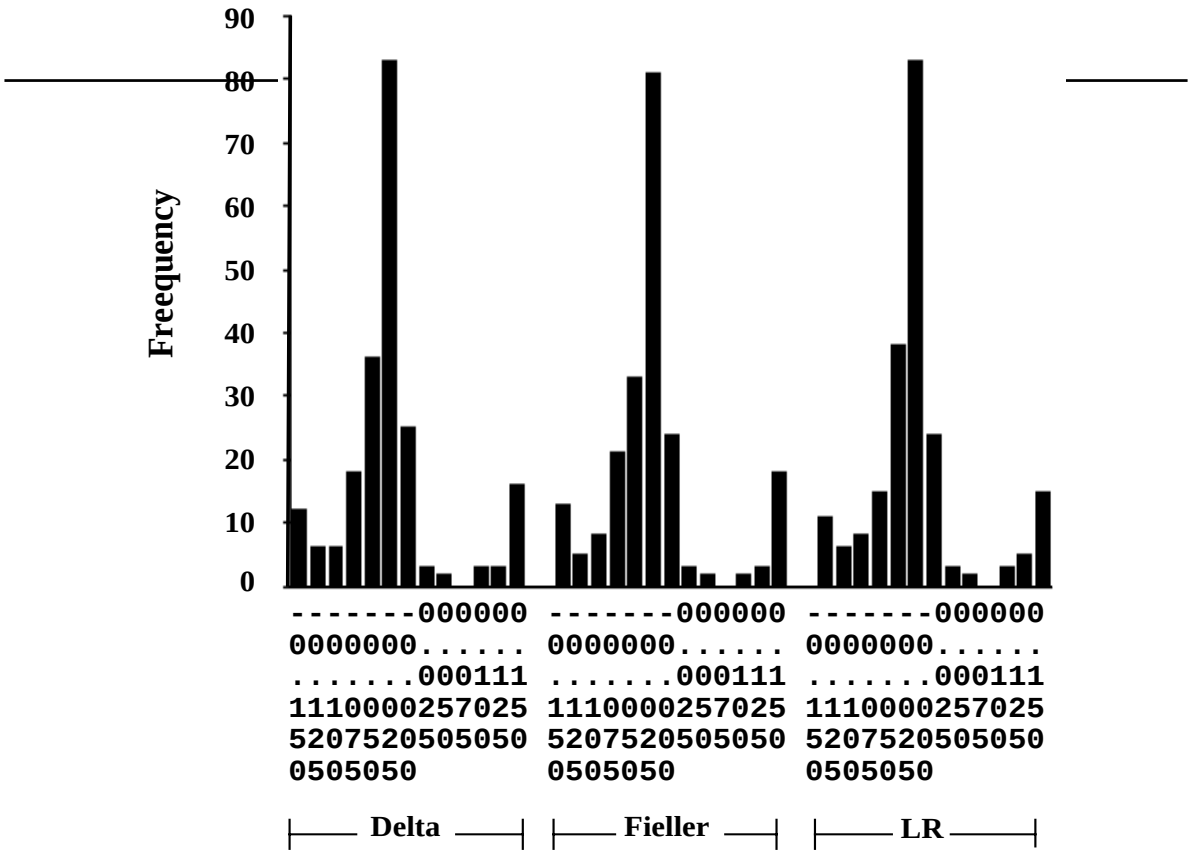

Fig. 12. Frequency distribution of the length of confidence intervals

Although there is no remarkable difference between the histograms of interval length, the number of data files which did not yield confidence intervals within a reasonable time was one for D-method, nine for F-method, and twenty five for L-method.

show a dose-response curve with a peak in a low dose region for some kinds of chemicals, examples of which are shown by Wang *et al.* (Wang *et al.*, 1995). If such a phenomenon occurs, the corresponding data file would be improperly excluded from the evaluation of assays through data check by the checkpoint F. We investigated the possibilities of this kind of erroneous action of the JSAAE study and confirmed that the RMS criterion did not encounter such errors.

8.3 Three methods to construct confidence intervals

There are three available methods in literature to construct a confidence interval for ED50 from a data file. They are the delta method (Kotz & Johnson, 1985), the method proposed by Fieller (1967), and the method based on the likelihood ratio test (Williams, 1986), which are referred to as the D-method, the F-method, and the L-method, respectively. Although all of these are asymptotically equivalent as is well known, their behavior is different in small sample sizes such as in our study. Williams (1986) and Alho and Valtonen (1995) compared their nature through a simulation study based on the binomial distribution, leading to the conclusion that the L-method is superior in the sense that, in many cases, it leads to shorter intervals keeping nominal confidence levels. We, however, examined their differences by applying the three methods to our data which were not binominally distributed; the results are summarized in Fig. 12. This demonstrated the superiority of the D-method, because the confidence interval was easily obtained in all cases, whereas the other two methods failed to reach rapid convergence of iteration in some cases and the length of acquired intervals did not show any remarkable tendency as was asserted by Williams (1990). Thus, we

implemented the D-method into LAP-JSAAE as the method to construct confidence intervals.

8.4 Homoscedasticity and sources of variability

In some cases, the variability among measurements on the same dose looks greater for low doses. This is probably due to the lack of required fineness of manipulation with low doses. In other words, if this is true, the homoscedasticity is violated. The criterion Q in the least squares method must be modified so that different weights can be multiplied on each term of squares, which subsequently creates a problem of how to determine the optimum weighting. This problem is difficult to answer and is left for future studies.

In any assay, each experiment required more than one plate because there are possible sources of bias commonly affecting wells in the same plate. In the data analysis in this study, however, these biases were ignored as a whole as long as the data file was accepted after the data check. The modification of the method to incorporate plate-depnedent biases will be considered in future studies.

Acknowledgments

This study was supported in part by funds from JSAAE, a Special Coordination Funds from Promoting Science and Technology from The Science and Technology Agency of Japan, and the Grant-in-Aid for Scientific Research C07680327 of the Ministry of Education, Science, Sports, and Culture of Japan.

References

Alho J. M. and Valtonen E. (1995): Interval estimation of inverse dose-response. Biometrics, 51, 491-501.

DeGroot M. H., Ferber R., Frankel M. R., Seneta E., and Watson G. S. (1985): Encyclopedia of Statistics. John Wiley & Sons, New York.

Fieller E. C. (1940): The biological standardization of insulin. JRSS Supplement, VII, 1-45.

Fieller E. C. (1954): Some problem in interval estimation. JRSS B, 16, 175-185.

Finney, D. J. (1971): Probit Analysis, 3rd edition, Cambridge University Press.

Finney, D. J. (1985a): The median lethal dose and its estimation, Archives of Toxicology, 56, 215-218.

Finney, D. J. (1985b): The estimation of the ED50 for a logistic response curve. Sankhya, 12, 121-136.

Litchfield J. T.Jr. and Wilcoxon F. (1949): A simplified method of evaluating dose-effect experiment, Journal of Pharmacology and Experimental Therapeutics, 96, 99-114.

Ohno T., Itagaki H., Tanaka N., and Ono H. (1995): Validation study on five different cytotoxicity assays in Japan - An intermediate report. Toxicology in vitro, 9, 571-576.

SAS Institute inc. SAS/STAT User's Guide, Version 6, Fourth Edition.

Tukey T. W. (1977): Exploratory data analysis. Addison-Wesley.

Wang X., Liu S.Q., Chen S., Ge X.R., Saijo K., and Ohno T. (1995): In vitro cytotoxicity of 20 MEIC chemicals with “quantum-type” release of rhodamine 123 detected by flow cytometry. Alternatives to Animal Testing and Experimentation, 3, 94-105.

Williams D. A. (1986): Interval estimation of the median lethal dose. Biometrics, 42, 641-645.

Appendix for Validation Article II.