Altern. Animal Test. EXperiment. 5, 59-73 (1998)

Validation Study on Five Cytotoxicity Assays by JSAAE
III.
Quality of Collected Data Files

Takashi Omori[1], Kaoru Saijo[2], Mayako Kato[2], Hiroshi Itagaki[3], Makoto Hayashi[4], Satoru Miyazaki[5], Tadao Ohno[2*], Hidiaki 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 University 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 (NIHS), 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

In the inter-laboratory validation study organized by JSAAE to evaluate the practicability of five cytotoxicity assays as feasible alternatives to the Draize eye irritation test, we have carried out thorough data cleaning before data analysis in order to be assured of the reliability and correctness of the results of the study, since so many errors on protocol understanding and data recording were found which were assumed to have primarily originated from the participating laboratories. We call these errors “human errors.” In this article, various human errors were categorized and analyzed for their possible causes such as lack of description of serum that might have led to the use of mismatched serum types. Based on our experience on data cleaning, we also described proposals with key points that should be taken into consideration at the planning stage of further validation studies to assure the quality of data and to attain reliable results.

* 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, cytotoxicity assay, colony formation assay, crystal-violet staining assay, Draize eye irritation test, inter-laboratory validation, LDH release assay, MTT assay, neutral red uptake assay, quality of data

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; LDH, lactate dehydrogenase release; MAS, maximum average score; 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.

Introduction

In 1992, the Japanese Society for Alternatives to Animal Experiments (JSAAE) organized a validation study to evaluate the feasibility of five cytotoxicity assays as possible alternatives to the Draize eye irritation test (Ohno *et al.*, 1995). The principal results obtained in this study were reported in the first article (Validation Article I) in this issue. Since one of the objectives of this study was to compare the practicability of each assay, the Validation Committee of JSAAE (chaired by H. Ono, Hatano Research Institute, Food and Drug Safety Center) did not select the participating laboratories by their capability, and examined in detail the quality of the collected data files from two aspects, namely, (1) the necessity of data cleaning in this kind of study from a data management viewpoint and (2) an evaluation of the reliability of data files from a statistical viewpoint. The first aspect is important in further improvement of the protocols for each assay and development of the method of technology transfer to obtain reliable data. The second aspect is useful in the evaluation of the effectiveness of each assay when the data are well managed.

Examination of the collected data files from the above two aspects was extremely difficult due to the wide variation in data type, the large amount of data, and the large number of participating laboratories. After more than two years of laborious work, The Working Group (composed of the authors of this article) discovered many noticeable points. In this article, we describe the flaws of the collected data, their categorization, and trials to determine causes of these flaws considered to be "human errors" during the raw data handling in each laboratory. In addition, we suggest directions to prevent recurrence of these flaws in future studies.

Errors Found in the Data Files

When this validation study began, we instructed each laboratory to input the data according to a specified format (hereafter referred to as worksheet format) and to save it on a floppy disk. We requested that both data files on the floppy disk and their print-outs be sent to us. Fig. 1 shows the explanation for the worksheet format handed to each laboratory. We assumed that, by using data files constructed according to the worksheet format, we would be able to manage the collected data and to carry out a smooth data analysis. However, when the data files were collected, we found that almost all of them did not comply with the requested format and, therefore, did not produce reliable results. We have carried out thorough data cleaning by checking one by one all the data files.

When any input-errors or blanks in the worksheet were found, we confirmed them with the laboratory concerned before making any correction to the inputs. Judging from the collected data files, many laboratories seemed to have thought that it was unnecessary to record data which would not be used in the data analysis, such as mean, SD, and name of the cell line and the chemical tested. Since the first two items are not essential for the data analysis, we also ignored these records. However we could not disregard the last two items because they were essential to identify the source of the data. Information about the type of serum and medium used was also necessary because we found data files in which the recorded serum was different from that specified in the protocol.

Various errors in the data files were found during the data cleaning process. In order to grasp the entire range of errors, we categorized them into the following seven types.

1) *Out-of-format inputs*

After the data collection, we immediately found that many data were not recorded according to the worksheet format we requested. The data files contained the following items:

(1) More than one value of blanks, negative- and/or positive-controls were inputted instead of their single mean value (an example, Fig. 2).
(2) Data were inputted in wrong places (Fig. 3).
(3) The worksheet was different from the requested format. Some laboratories submitted the data files as the 96-well plate form (Fig. 4).

Fig. 1 Explanation of a sample worksheet format

(4) Extra mean values of OD were inputted together with the observed raw values of OD (Fig. 5), or an extra calculation column was added (Fig. 6).

2) *Inconsistency between the electronic files and the print-outs*

We found that 83 data files showed inconsistency between their electronic files in simultaneously submitted floppy disks and the print-outs.

3) *Lack of essential data*

Many data files lacked essential data for analysis. The missing data found were:

(1) Type of serum

Fig. 2. An example of a data file with out-of-format inputs
More than one value of blanks, negative- and/or positive-controls were inputted instead of their single mean value.

Fig. 3. An example of a data file with out-of-format inputs
Data were inputted in wrong places.

Fig. 4. An example of a data file with out-of-format inputs

The worksheet was different from the requested format. Some laboratories submitted the data files as the 96-well plate form.

Fig. 5. An example of a data file with out-of-format inputs

Extra mean values of OD were inputted together with the observed raw values of OD.

Fig. 6. An example of a data file with out-of-format inputs

Extra calculation column was added.

Fig. 7. An example of a data file with lack of essential data

Even though data were inputted in the former worksheet of (cell layer + supernatant) (a, upper), no data were inputted in the latter worksheet of (supernatant) (b, lower) which should have been carried out as the simultaneous test in LDH-2B and -2C assays.

Fig. 8. Examples of a data files with violation of the protocols or the common rules.

Unpaired chemical concentrations in the worksheets of (cell layer + supernatant) (a, left) and (supernatant) (b, right) in LDH-2B and -2C assays.

Fig. 9. An example of a data file with violation of the protocols or the common rules.

Values for the response rates (viability) were different from those defined in the assay protocol. In this assay, the viability is defined as (the test - blank) / (negative control - blank). However, the viability was calculated using the positive-control instead of the blank in this example.

Fig. 10. An example of a data file with violation of the protocols or the common rules.

The same FRLA value was recorded at every concentration of the chemical tested in the LDH assay.

Fig. 11. An example of a data file with abnormal data
A ring-in chemical concentration was inputted in the ordered concentration range.

Fig. 12. An example of a data file with abnormal data
The negative- and positive-control values were abnormal in a CV assay. In the protocols for the CV assay, a description about positive-control is not written although, in the worksheet of the CV assay, the name, Posi. Cont., was retained because the worksheet was made by each laboratory basically by copying the worksheet of the LDH-1 assay (see Validation Article V in this issue). In this example, the positive-control values were inputted besides those of blank and negative-control. This positive-control may be inferred to be the negative-control.

Fig. 13. An example of a data file with abnormal data
Abnormal data without a numerical value was inputted.

Fig. 14. An example of a data file with abnormal data
The FRLA values were too large. Based on our experience, FRLA usually ranged from 0.1 to 2.0.

(2) Final concentration of the test chemical
(3) Blank values (found in CV, LDH, NR assays)
(4) Negative-control values (found in CF, LDH, and MTT assays but not in CV and NR assays)
(5) Positive-control values (found in LDH and MTT assays)
(6) Number of seeded cells per dish (found in CF assay)
(7) FRLA values (factor for remaining LDH activity) (found in LDH assay)
(8) Even though data were inputted in the worksheet of (cell layer + supernatant), no data were inputted in the worksheet of (supernatant) which should have been carried out as the simultaneous test in LDH-2B and -2C assays (Fig. 7).

4) *Violation of the protocols or the common rules*

Data files that violated the protocols or the common rules included the following:
(1) Unpaired chemical concentrations in the worksheets of (cell layer + supernatant) and (supernatant) in LDH-2B and -2C assays (Fig. 8).
(2) Values for the response rates (viability) were different from those defined in the assay protocol. Fig. 9 shows an example in the NR assay. In this assay, the viability is defined as (the test - blank) / (negative control - blank). However, in this example, the viability was calculated using the positive-control instead of the blank.
(3) The same FRLA value was applied to the two cell lines in the LDH assay.
(4) Medium used was different from that specified in the protocol.
(5) Serum used was different from that specified in the protocol.
(6) The same FRLA value was recorded at every concentration of the chemical tested in LDH assay. Fig. 10 illustrates an example where all FRLA values are 1.0. Since FRLA should be obtained by determining the direct inhibitory activity of the chemical at the tested concentration, we assumed that this laboratory did not carry out the test for FRLA determination.
(7) Only one plate was used while the protocol specified the use of three independent plates in an assay.

5) *Abnormal data*

We suspected that wrong data may have been inputted in the following cases.
(1) A ring-in chemical concentration was inputted in the ordered concentration range (Fig. 11), although concentrations are not necessarily recorded in sequential order.
(2) The negative- and positive-control values were abnormal. Fig. 12 illustrates an example of a CV assay where the negative-control value is much smaller than the positive-control value (although, in the worksheet of the CV assay, the name, Posi. Cont., was retained because the worksheet was made by each laboratory basically by copying the worksheet of the LDH-1 assay; see Fig. 1). However, in the protocols for the CV assay, a description about positive-control is not written (see Validation Article V in this issue). In this example (Fig. 12), the positive-control values were inputted besides the value of the blank and negative-control. This positive-control may be inferred to be the negative-control.
(3) Abnormal data without a numerical value was inputted. Fig. 13 illustrates an example of an abnormal data likely to be an input error.
(4) The FRLA value was too large (Fig. 14). Based on our experience, FRLA usually ranged from 0.1 to 2.0.

6) *Errors reported from the laboratories*

The following errors were reported voluntarily from the laboratories:
(1) Wrong colony numbers or OD values were inputted.
(2) Wrong blank values were inputted.
(3) Wrong cell name was inputted.
(4) Wrong FRLA values were inputted.
(5) Concentrations recorded in the worksheet were not the final values.

7) *Others*

In addition to the various errors described

Table 1. Sum of submitted, accepted and corrected data files

* Sum of submitted data files
$ Sum of finally accepted data files
# Sum of corrected data files in finally accepted data files

above, we found data files that contained the following errors:

(1) Nothing was recorded on the submitted floppy disk.
(2) The floppy disk was protected and, therefore, we could not read the data.
(3) Some data were put in brackets.
(4) The data from plate 1 were exactly the same as those from plate 2.
(5) Recorded mean values were different from those calculated from the raw data.
(6) The data of a chemical were exactly the same as those of a different chemical.
(7) Number of colonies recorded in the data file was not an integer in a CF assay.

8) *Frequency of errors*

As a result of the above data cleaning, errors were found in 955 out of 1535 submitted data files. Table 1 shows the sum of the submitted data files, the finally accepted data files, and the corrected data files for each laboratory and for each assay (Submitted data files and finally accepted data files are defined in Validation Article I in this issue). To our regret, the rigorous data cleaning resulted in 509 data files (53%) out of 969 finally accepted data files requiring correction.

We corrected the above errors one by one, although we rejected the data files that violated the common rule [3] as described in Validation Article I. A total number of 1742 sets were man-

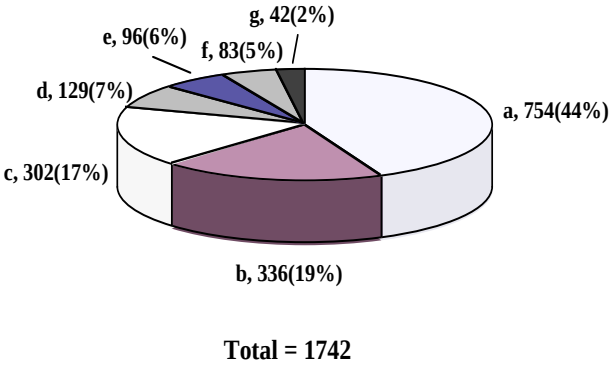

Fig. 15. Frequency of errors found in the data cleaning process (defined as "human errors")

a, Lack of essential data. b, Out-of-format inputs c, Violation of the protocol or common rules. d, Simple misrecording reported from the laboratories. e, Abnormal data amended by the laboratory before any notation from The Working Group. f, Different data were written in the printouts, the data stored in the submitted floppy-disks, and the worksheet. g, others.

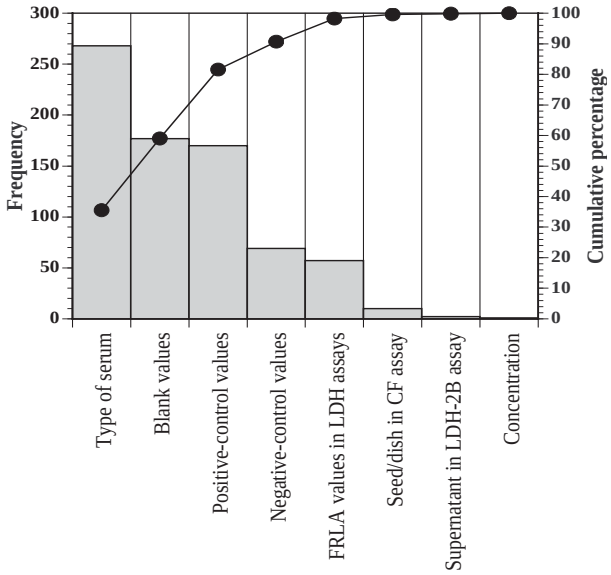

Fig. 16. Pareto chart for lack of essential data

aged by The Working Group. Fig. 15 shows the frequency of each type of error found in the data files. Three major errors found in the data files were lack of essential data (44%), out-of-format inputs (19%), and violation of protocols or common rules (17%). These three composed almost 80% of the total.

Cause of Errors

Causal relationships, if found, between the above mentioned errors and the assays would be very important to the construction of a concrete set of protocols for further validation studies. The most frequent error, lack of essential data, accounts for 44% of the total (Fig. 15).

Fig. 16 is a Pareto chart for types of the missing essential data. Clearly, the lack of description on the serum used was the most frequent source of error. Since the present protocols indicated the use of a specified type of serum, many laboratories have judged that it was not necessary to input the type of serum used. However, if the laboratory used serum different from the specified type, growth rates of the cells will be largely affected. Then, as a result, the chemical toxicity could be different from those observed in the other laboratories. Therefore we believe that inputting serum type is essential in the confirmation of whether the assay has been carried out correctly or not.

The next common error was lack of blank and positive-control values (Table 2, see Validation Article I on the definition of the submitted data files).

The percentages of lack of blank values in LDH assays and lack of positive-control values in MTT assays are apparently higher than others. The protocols for LDH sub-assays did not clearly indicate that blank values must be used when the response rates (cytotoxicity, growth inhibition and killing index) are calculated (see Validation Article VI in this issue). In the MTT assay, definition of the positive control in the protocol was different from that of other assays (see Validation Article VII in this issue).

Out-of-format inputs account for 19% of the total errors. During the efforts of confirming this, we knew that many laboratories misunderstood the worksheet format, ignored the worksheet format, and made mistakes in the inputting of data. If every laboratory had checked by themselves whether their data files followed the indicated format, these errors would have drastically de-

Table 2. Total number of data files which lack essential data for each assay

Assay	a	b	c	d	e	f	g	h	Total	Submitted
CF	39(14%)	0(0%)	-	27(10%)	-	10(4%)	-	-	76	279
CV	14(7%)	0(0%)	14(7%)	0(0%)	-	-	-	-	28	195
LDH-1	45(32%)	0(0%)	40(29%)	14(10%)	14(10%)	-	14(10%)	-	127	140
LDH-2A	42(31%)	0(0%)	40(30%)	14(10%)	14(10%)	-	14(10%)	-	124	135
LDH-2B	35(28%)	1(1%)	26(21%)	0(0%)	14(11%)	-	15(12%)	2(2%)	93	124
LDH-2C	28(22%)	0(0%)	43(33%)	0(0%)	14(11%)	-	14(11%)	-	99	129
MTT	37(15%)	0(0%)	-	14(6%)	114(48%)	-	-	-	165	240
NR	28(10%)	0(0%)	14(5%)	0(0%)	-	-	-	-	42	293
Total	268	1	177	69	170	10	57	2	754	1535

a, Type of serum; b, Concentration; c, Blank values; d, Negative-control values; e, Positive-control values; f, Seed/dish in CF assay; g, FRLA values in LDH assays; h, Supernatant in LDH-2B assay. Number in the parenthesis is the percentage of the submitted data files which lack necessary data.

creased. When the present validation study started, we handed to each laboratory only the worksheet with inset explanation for the LDH-1 assay (Fig. 1). Details on how to input data on the worksheet was announced without written manuals. If we had handed the manual with detailed explanation and examples of the worksheet for each assay, many input errors would have been prevented. It seems that excessive optimism for the worksheet handling caused the errors of this type.

Violations of the protocols or the common rules accounted for 17% of the total errors, even though technology transfer had been carried out. By asking the laboratories after the data analyses, we realized that misunderstanding of the protocols happened frequently, especially in the laboratories that carried out multiple assays. The protocols with slightly different styles of description and formats were not necessarily understandable.

From these considerations, we concluded that insufficient understanding of protocols and common rules, incorrect interpretation of the protocols and the worksheets, and simple input errors were the major causes of the “human-errors”.

Proposals for Further Studies

As shown in the Table 1, almost all laboratories had submitted data files that required some type of correction. Therefore, to be assured of reliable results, we considered the step of data cleaning as essential in further studies. However, in the present validation study, the data cleaning took more than 2 years to complete. Since it is undesirable to spend a long time for the data cleaning in any validation studies, we pointed out the possible causes of delay in the present validation study as follows:

(1) Since each laboratory was allowed to name freely the data files and to input data by using its own computer software, we needed to rename all the data files and convert them to an MS-DOS text before any data analyses were carried out.

(2) During the technology transfer, we taught participating laboratories how to conduct each cytotoxicity assay but not how to record data on the worksheet.

(3) Although the response rate for the LDH-2B assay is calculated by using OD values of blank, negative- and positive-controls obtained in the LDH-2A assay, the sample worksheets of the LDH-2B&C assays shown initially to each laboratory did not contain a column where the data from the LDH-2A assay must be recorded. Therefore, new worksheets were sent to each laboratory with an appended column for the LDH-2A assay data.

(4) Some laboratories tested more than eight con-

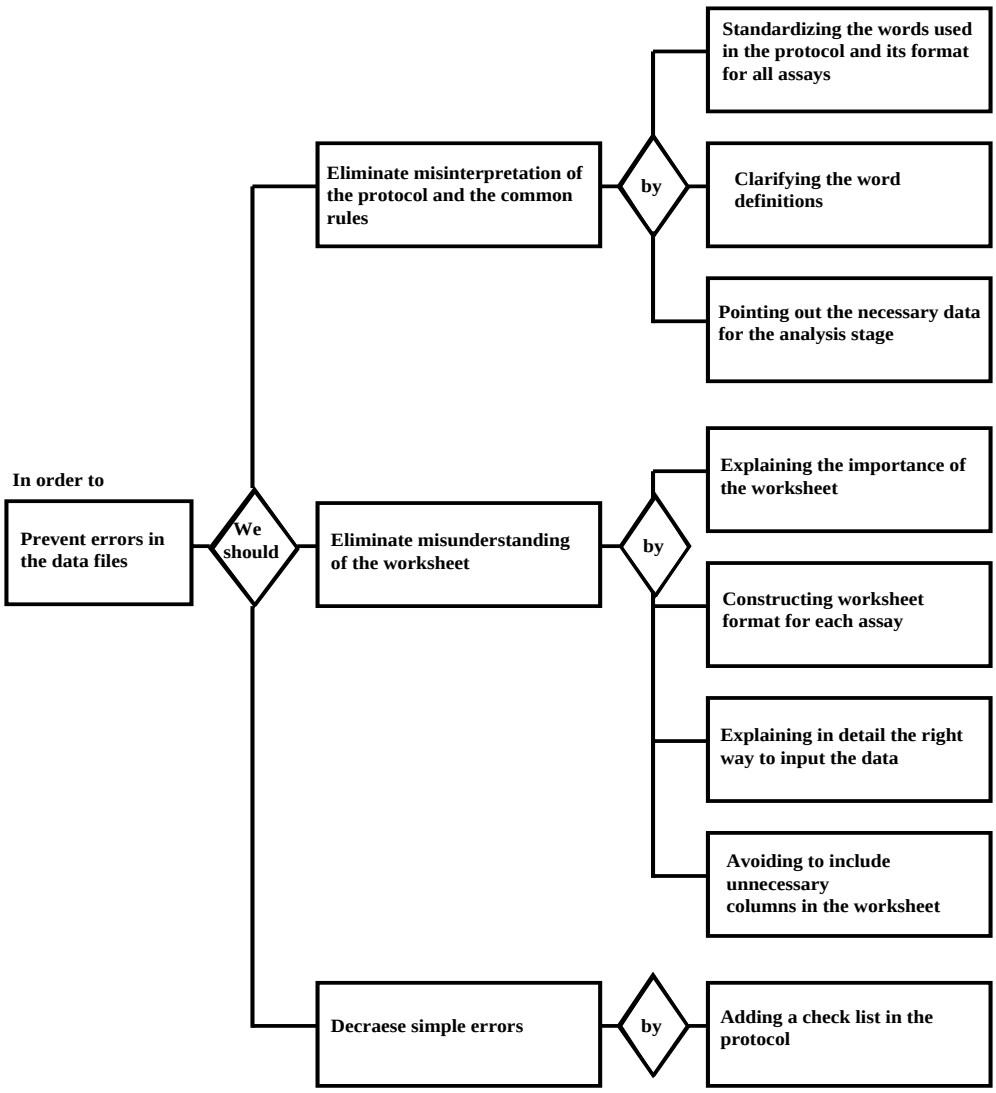

Fig. 17. Check points for data management

centrations of a chemical in an assay, and therefore submitted the worksheet divided into two sections for one assay. We then needed to rejoin the two worksheets.

As a result of the current data cleaning, we suggested check points for data management as shown in Fig. 17. These points should be considered at the planning stage of further studies.

Furthermore, to actually attain this, we consider that it is not enough that only those who collect and analyze the data conduct the data cleaning. We recommend that the planner of a validation study should set in the protocol some type of feedback assurance system of raw data within the laboratory before their data file submission.

Reference

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