

## Validation Study on Five Cytotoxicity Assays by JSAAE

### I.

#### Overview of the Study and Analyses of Variations of ED50 Values

Tadao Ohno\*<sup>1</sup>, Masumi Asakura<sup>2</sup>, Takumi Awogi<sup>3</sup>, Yoshihiro Futamura<sup>4</sup>, Akihiro Harihara<sup>5</sup>, Masato Hatao<sup>6</sup>, Chiharu Hattori<sup>7</sup>, Akira Hayasaka<sup>8</sup>, Makoto Hayashi<sup>9</sup>, Takashi Hayashi<sup>10</sup>, Zenzou Hirata<sup>4</sup>, Hiroshi Hori<sup>11</sup>, Hiroyoshi Hoshi<sup>12</sup>, Koichi Imai<sup>13</sup>, Ikuo Imazeki<sup>8</sup>, Takuya Ishibashi<sup>11</sup>, Hiroshi Itagaki<sup>6</sup>, Tatsuo Iwata<sup>14</sup>, Miyako Kakuma<sup>15</sup>, Shinya Kaneda<sup>16</sup>, Ikuo Kato<sup>5</sup>, Mayako Kato<sup>1</sup>, Tooru Kawahatsu<sup>17</sup>, Akio Kawakami<sup>18</sup>, Akemi Kazama<sup>19</sup>, Akihiko Kido<sup>20</sup>, Shigemi Kimura<sup>21</sup>, Michiyo Kitazawa<sup>22</sup>, Shigefumi Kogiso<sup>23</sup>, Hajime Kojima<sup>24</sup>, Mayumi Kotani<sup>25</sup>, Masahiro Kuramochi<sup>21</sup>, Daijyo Maki<sup>26</sup>, Masako Matsuda<sup>27</sup>, Chihomi Mitsuoka<sup>26</sup>, Satoru Miyazaki<sup>28</sup>, Fumio Mizuno<sup>20</sup>, Minako Mori<sup>29</sup>, Katsunori Morimoto<sup>30</sup>, Matsuko Moriyasu<sup>26</sup>, Keiya Nakajima<sup>29</sup>, Madoka Nakajima<sup>22</sup>, Masaaki Nakamura<sup>13</sup>, Masaki Nakamura<sup>4</sup>, Nahoko Nakano<sup>26</sup>, Shinobu Nakamura<sup>26</sup>, Akira I. Negami<sup>31</sup>, Masato Nishino<sup>19</sup>, Tamotsu Nishitomi<sup>20</sup>, Kenji Ohkoshi<sup>32</sup>, Yuuko Okamoto<sup>32</sup>, Takashi Omori<sup>33</sup>, Hiroshi Ono<sup>34</sup>, Makoto Ono<sup>15</sup>, Yuko Osanai<sup>15</sup>, Kaoru Saijo<sup>1</sup>, Yuka Sano<sup>35</sup>, Kyoko Saotome<sup>36</sup>, Kazunori Sasaki<sup>5</sup>, Tetsuji Sasaki<sup>1</sup>, Hidetaka Sato<sup>37</sup>, Shinichiro Sato<sup>38</sup>, Hiroyasu Shimada<sup>7</sup>, Shoichiro Shimogo<sup>6</sup>, Kazuyuki Shimonon<sup>16</sup>, Hiroshi Shionoya<sup>14</sup>, Hideaki Sugawara<sup>28</sup>, Yasuko Sugiki<sup>39</sup>, Shigeo Sugimoto<sup>40</sup>, Shigeru Sugimoto<sup>37</sup>, Junko Suzuki<sup>37</sup>, Kinya Takagaki<sup>4</sup>, Ken Takahashi<sup>41</sup>, Minoru Takizawa<sup>42</sup>, Chieko Tamaki<sup>23</sup>, Noriho Tanaka<sup>34</sup>, Jirou Taniya<sup>43</sup>, Noriko Teramoto<sup>1</sup>, Hisashi Torishima<sup>44</sup>, Toshiyuki Tsuchiya<sup>45</sup>, Motoo Uejima<sup>16</sup>, Hayao Ueno<sup>46</sup>, Yuji Ugai<sup>22</sup>, Shogo Wada<sup>47</sup>, Shinobu Wakuri<sup>34</sup>, Xinhai Wang<sup>1</sup>, Ikuo Watanabe<sup>29</sup>, Masami Watanabe<sup>48</sup>, Satoshi Yajima<sup>49</sup>, Yoshiki Yamagata<sup>19</sup>, Yuuzou Yamaguchi<sup>49</sup>, Osamu Yamakita<sup>40</sup>, Ryohei Yamamoto<sup>44</sup>, Masanori Yoshida<sup>46</sup>, Isao Yoshimura<sup>33</sup>, Kouichi Yuhki<sup>43</sup>, Shiro Yukiama<sup>40</sup>

<sup>1</sup>RIKEN Cell Bank, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba Science City, Ibaraki 305, <sup>2</sup>Japan Bioassay Laboratory,

---

\* 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, LDH assay, neutral red uptake assay, MTT assay, inter-laboratory validation, Draize eye irritation test

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.

2445 Hirasawa, Hadano-shi, Kanagawa 257, <sup>3</sup>Drug Safety Research Center, Tokushima Research Institute, Otsuka Pharmaceutical Co., Ltd., 463-10 Kagasuno, Kawauchi-cho, Tokushima-shi, 771-01, <sup>4</sup>Toxicology Laboratory, Mochida Pharmaceutical Co., Ltd., 342 Gensuke, Fujieda-shi, Shizuoka 426, <sup>5</sup>Shionogi Research Laboratories, Shionogi & Co., Ltd., 3-1-1 Futaba-cho, Toyonaka-shi, Osaka 561, <sup>6</sup>Shiseido Safety & Analytical Research Center, 1050 Nippa-cho, Kohoku-Ku, Yokohama-shi 223, <sup>7</sup>Developmental Research Laboratory, Daiich Pharmaceutical Co., Ltd., 1-16-13 Kita-Kasai, Edogawa-ku, Tokyo 134, <sup>8</sup>Chugai Pharmaceutical Co., Ltd., 3-41-8 Takada, Toshima-ku, Tokyo 171, <sup>9</sup>Division of Genetics and Mutagenesis, National Institute of Health Sciences (NIHS), 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158, <sup>10</sup>Ina Research Inc., 8047 Nishiminowa, Ina-shi, Nagano 399-45, <sup>11</sup>Tsuruga Institute of Biotechnology, Toyobo Co., Ltd., 10-24 Toyo-cho, Tsuruga-shi, Fukui 914, <sup>12</sup>Research Institute for the Functional Peptides, 11-26 Minamisanban-cho, Yamagata-shi 990, <sup>13</sup>Department of Biomaterials, Osaka Dental University, 1-5-31 Otemae, Chuo-ku, Osaka-shi 540, <sup>14</sup>Department of Drug Safety Research, R & D Division, Eisai Co., Ltd., Takehaya, Kawashima-cho, Hashima-gun, Gifu 501-61, <sup>15</sup>Beauty Care Products Laboratory, Nagase, Co., Ltd., 5-1 Kobuna-cho, Nihonbashi, Chuo-ku, Tokyo 103, <sup>16</sup>Naruto Research Institute, Otsuka Pharmaceutical Factory, Inc., 115 kuguhara, Tateiwa, Muya-cho, Naruto-shi, Tokushima 772, <sup>17</sup>Pharmaceutical Research Center, Meiji Seika Kaisha, Ltd., 760 Morooka-cho, Kohoku-ku, Yokohama-shi 222, <sup>18</sup>Safety Research Laboratories, Yamanouchi Pharmaceutical Co., Ltd., 1-1-8 Azusawa, Itabashi-ku, Tokyo 174, <sup>19</sup>Biochemical Development & Production, Oriental Yeast Co., Ltd., 4-4-1 Minamisuita, Suita-shi, Osaka 564, <sup>20</sup>Kashima Laboratory, Mitsubishi-kagaku Institute of Environmental and Toxicological Sciences, 14 Sunayama, Hasaki-machi, Kashima-gun, Ibaraki 314-02, <sup>21</sup>BML Inc., 1361-1 Matoba, Kawagoe-shi, Saitama 350, <sup>22</sup>Biosafety Research Center, Foods, Drugs and Pesticides, 582-2 Arahama, Shioshinden, Fukude-cho, Iwata-gun, Shizuoka 437-12, <sup>23</sup>Environmental Health Science Laboratory, Sumitomo Chemical Co., Ltd., 3-1-98 Kasugadanaka, Konohana-ku, Osaka-shi 554, <sup>24</sup>Biochemical Research Institute, Nippon Menard Cosmetic Co., Ltd., 2-7 Torimi-cho, Nishi-ku, Nagoya-shi 451, <sup>25</sup>R&D Planning HQ, Sunstar Inc., 3-1 Asahi-machi, Takatsuki-shi, Osaka 569, <sup>26</sup>Panapharm Laboratories Co., Ltd., 1285 Kurisaki-cho, Uto-shi, Kumamoto 869-04, <sup>27</sup>Research and Development Center, Duskin Co., Ltd., 4-19-8 Minamisuita, Suita-shi, Osaka 564, <sup>28</sup>Life Science Research Information Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-01, <sup>29</sup>The Center of Japan Biological Chemistry, 52 Fukue, Kaizu-cho, Kaizu-gun, Gifu 503-06, <sup>30</sup>Takara Belmont Co., Ltd., 5-1 Takamatsu, Kouzai-cho, Koga-gun, Shiga 520-32, <sup>31</sup>Department of Obstetrics and Gynaecology, Fukui Medical School, 23-3 Shimoaizuki, Matsuoka-cho, Yoshida-gun, Fukui 910-11, <sup>32</sup>Division of Fundamental Research, KOSÉ Co., 1-18-4 Azusawa, Itabashi-ku, Tokyo 174, <sup>33</sup>Department of Management Science, Faculty of Engineering, Science University of Tokyo, 1-3 Kagurazaka,

Shinjuku-ku, Tokyo 162, <sup>34</sup>Hatano Research Institute, Food and Drug Safety Center, 729-5 Ochiai, Hadano-shi, Kanagawa 257, <sup>35</sup>Biotechnology Research Laboratories, TAKARA SHUZO Co., Ltd., 3-4-1 Seta, Ohtsu-shi, Shiga 520-21, <sup>36</sup>Yokohama City Institute of Health, 1-2-17 Takigashira, Isogo-ku, Yokohama-shi 235, <sup>37</sup>Japan Food Research Laboratories, 6-11-10 Nagayama, Tama-shi, Tokyo 206, <sup>38</sup>Division of Radiation Medicine, National Institute of Radiological Sciences, 4-9-1 Anagawa, Inage-ku, Chiba-shi 263, <sup>39</sup>Menicon Co., Ltd., 5-1-10 Takamori-dai, Kasugai-shi, Aichi 487, <sup>40</sup>Drug Safety Laboratory, Taiho Pharmaceutical Co., Ltd., 224-2 Ebisuno, Hiraishi, Kawauchi-Cho, Tokushima-shi 771-01, <sup>41</sup>The Institute of Environmental Toxicology, 4321 Uchimoriya-cho, Mitsukaido-shi, Ibaraki 303, <sup>42</sup>Hakko Electric Machine Works Co. Ltd., 3055 Togura, Toguramachi, Hanishina-gun, Nagano 389-08, <sup>43</sup>Koken Bioscience Institute, Koken Co. Ltd., 2-11-12 Nakane, Meguro-ku, Tokyo 152, <sup>44</sup>Kurabo Industries Ltd., 14-41 Shimokida-cho, Neyagawa-shi, Osaka 572, <sup>45</sup>Safety Evaluation Center, Central Research Laboratory, Showa Denko K. K., 1-1-1 Ohnodai, Midori-ku, Chiba-shi 267, <sup>46</sup>Takeda Analytical Research Laboratories, Ltd., 2-17-85 Juso-honmachi, Yodogawa-ku, Osaka 532, <sup>47</sup>Wako Pure Chemical Industries, Ltd., 6-1 Takadacho, Amagasaki-shi, Hyogo 661, <sup>48</sup>Division of Radiation Biology, Department of Health Science, Faculty of Pharmaceutical Sciences, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki-shi 852, <sup>49</sup>Central Research Laboratory, Takasago International Co., 1-4-11 Nishiyawata, Hiratsuka-shi, Kanagawa 254, Japan

## Abstract

In October 1992, the Japanese Society of Alternatives to Animal Experiments (JSAAE) organized an inter-laboratory validation study on cytotoxicity assays for development of an alternative(s) to the *in vivo* Draize eye irritation test. The main purpose is to evaluate practicability of five proposed cytotoxicity assays with two cell lines each through a large scale inter-laboratory assessment. The five assays were colony formation (CF) assay, crystal-violet staining (CV) assay, lactate dehydrogenase release (LDH) assay, neutral red uptake (NR) assay, and MTT assay. They were selected because of their popularity in Japan. We chose six detergents as model chemicals for this first step validation. One of the six chemicals, which was revealed to be Tween 20 after breaking the code, was additionally included in the doubly-mask-coded chemical set for simultaneous evaluation of intra-laboratory variation. Technology transfer for all the assays was made cumulatively 118 times to less-experienced laboratories. Concomitantly, we performed the Draize test to confirm toxicity of the coded chemicals *in vivo*.

A total of 3,810 final data files including preliminary test results were submitted from 42 laboratories. Of 1,535 raw data files with final definitive assay results, 292 files were rejected because of not only apparent misunderstanding of the protocols provided by the Working Group but also for violating pre-set common rules. Acceptability of data files was also examined by a computer-assisted

logistic analysis program (LAP-JSAAE) with a six step-wise check code to detect abnormality in the data file. After generating ED50 values through the program, 5 data file sets of 7 tested chemicals were judged unreliable since the large differences in ED50 values were found for the same but differently coded test chemical, Tween 20. This clearly indicates considerable intra-laboratory variation. After excluding these data files, analyses of inter-laboratory variation were made on 969 data files with the box-whiskand-er plot analysis.

The important results of our study are as follows. (1) CF, CV, MTT, and NR assays are recommendable from the view point of performance of these assays. Performance rate of each assay was calculated on the number of finally accepted assay data files divided by the expected number of data files. The highest rate was for the CF assay with BALB/3T3 A31-1-1 cells followed by the CV assay with two cell lines. Lower performance rates were observed in the sub-divided LDH assays. The performance rate was considered to reflect simplicity of the method and labor needed for the assay; (2) From the view point of the intra-laboratory variation of the same but differently coded chemical Tween 20, medians of the  $\log(\text{ED}_{50})$  values of each assay were satisfactorily close; (3) After eliminating the sub-divided LDH assays which gave a small number of acceptable data files per assay and therefore resulted in unstable hinge-spread of  $\log(\text{ED}_{50})$  values, the CV assay with CHL cells and the MTT assay with SQ-5 cells were found to have given the smallest mean hinge-spread of  $\log(\text{ED}_{50})$  followed closely by the CF assay with HeLa S3 (SC) cells and the CV assay with HeLa S3 (SC) cells. These assays were therefore considered to give small inter-laboratory variation; (4) the CF assay with HeLa S3 (SC) cells resulted in the largest "power for distinction (PFD)" of toxicities between the least and the most toxic chemicals defined as the ratio of difference in medians of  $\log(\text{ED}_{50})$  values of two chemicals to the mean hinge-spread, followed by the CV assay with HeLa S3 (SC) cells. However, the CF assay is not necessarily advantageous as far as distinguishing moderately irritating chemicals from non-irritating chemicals; (5) Medians of  $\log(\text{ED}_{50})$  values enabled us to classify tested chemicals into at least three categories, namely, non-, moderately-, and highly-cytotoxic chemicals. Non- and highly-cytotoxic chemicals corresponded to non- and severe-irritants in the *in vivo* Draize test.

(6) Considering performance rate, inter-laboratory variation of data reflected on the mean hinge-spread, power for distinction of chemical cytotoxicity, time needed for an assay (i.e. the CF assay requires longer incubation time than others), and the data on the common cell line HeLa S3 (SC), we concluded that the CV assay is the most practical and recommendable as a part of alternatives to the *in vivo* Draize test.

Many problems that were revealed during the present validation study including human factors were discussed. The assay results will be further described in the ensuing articles in this issue, i. e., the calculation of ED50 values by LAP-JSAAE, problems on the CF, CV, LDH, MTT, and NR assays. A fact data base was constructed on the data files of this validation study which will be available on request.

## Introduction

Validation is necessary to confirm the practicability of any newly developed alternatives to *in vivo* experimentation. Many *in vitro* cytotoxicity assays have been proposed as alternatives to the Draize eye irritation test (Ekwall and Ekwall, 1988; Ekwall *et al.*, 1989; Balls *et al.*, 1990; Ekwall *et al.*, 1990; Bruner *et al.*, 1991; Spielmann *et al.* 1991) and to the acute systemic toxicity test (Clemenson *et al.*, 1996a, 1996b). However, well-organized, comparative, inter-laboratory validation studies on reported multiple cytotoxicity assays have not been carried out in Japan except for one which commenced in 1990 (Ohno, Y. *et al.*, 1994, 1995). Independently to this validation study, The Japanese Society of Alternatives to Animal Experiments (JSAAE) organized an early phase inter-laboratory validation study on five cytotoxicity assays in October 1992, since we assume that a proper battery of cytotoxicity assays should be the core alternative to the Draize test, and, as discussed in CAAT/ERGATT Workshop (Balls *et al.*, 1990), inter-laboratory assessment should consist of two major steps, an early phase and a definitive phase.

The purpose of the present study is to evaluate practicability of the five proposed cytotoxicity assays through a large scale inter-laboratory assessment in Japan without strict restriction on competence for participation, such as manufacturers of cosmetics, to avoid potential bias in the collected data. Performance of the assay is therefore one of the key markers of the present validation.

The five assays validated were the crystal-violet staining (CV) assay (Saotome *et al.*, 1989), the neutral red uptake (NR) assay (Borenfreund and Puerner, 1985), the MTT assay (Mosmann, 1983), the colony formation (CF) assay (Sasaki *et al.*, 1991), and the lactate dehydrogenase release (LDH) assay (Wang *et al.*, 1993). These were selected for their popularity and were expected to be standardized in Japan by the Working Group (The Office is located at RIKEN, Ibaraki) for this validation study organized under the Validation Committee of JSAAE (chaired

by H. Ono, Hatano Research Institute, Food and Drug Safety Center, Kanagawa). We thought that their protocols should be standardized in the near future since any one of them is highly likely to be selected as a component of the future official alternatives to the Draize test in Japan.

We selected the strategy where (1) any laboratory interested in the present validation study would be accepted as a participant, since it was impossible to expect as participants well-qualified laboratories which were able to carry out the expected assays with highly skillful techniques following good laboratory practice (GLP), (2) technology transfer on each assay would be given free of charge through 2-5 day hands-on workshop courses to less-experienced laboratories, (3) the collected data would be thoroughly reviewed and recalculated by expert toxicologists to determine whether or not there are any abnormalities or problems, (4) data files which passed this rigorous examination ("data cleaning") are then analyzed for performance rates of the assays from several view points and for the intra- and inter-laboratory variations of the 50% effective doses (ED50) values. (5) We also tried to evaluate the cytotoxicity assays with the newly defined "power for distinction" (PFD) among log(ED50) values of test chemicals.

(6) We chose six test chemicals with known irritancy based on the reported Draize test and set up a chemical bank for quality control and stable supply of the test chemicals. Each chemical was once given a mask-code number in the chemical bank. At this step, one of the coded chemicals was duplicated and the duplicate was given a new mask-code number. Thus, seven test chemicals were sent to The Working Group for the first validation study on cytotoxicity. They were again given new mask-code numbers (therefore, doubly masked), and sent to the participating laboratories. By this process, we expected to obtain intra-laboratory variation of data on a chemical together with inter-laboratory variation of data on 6 chemicals.

(7) We also selected 4 cell lines that are widely used and easily available in Japan. They were quality-controlled by an established cell bank. Additionally, rabbit cornea cells were selected

since the cells were primary-cultured in serum-free medium and supplied with the medium as a kit. The serum-free cultured cells have been generally considered as being more sensitive than those cultured in serum-containing medium (Torishima *et al.*, 1990, 1995). (8) A subline of HeLa was tested in every assay for mutual comparison of the five assay methods. (9) Simultaneously with these *in vitro* assays, the *in vivo* Draize test was carried out on the same coded seven chemicals for confirmation of the toxicity.

Common rules on experimentation for the five assays that were previously discussed and accepted in the orientation meeting before the start of this validation study are as follows;

- [1] Chemicals should be stored in the dark at 4 °C and should be diluted and recorded on a %(w/v) basis;
- [2] A 10-ml volume-certified flask should be used for initial preparation of a chemical solution except in the case where the initial concentration was less than 1%(w/v);
- [3] For an observation point\*, the data should be collected from 3 independent 96-multiwell plates. This is to make it clear that in an assay\*\* one plate corresponds to a single run of an expected test possible to draw a dose-response curve and the test was repeated 3 times for an assay. In a single plate, any number of replications for an observation point is allowed, but the plate must contain negative control (untreated) wells and, when indicated, positive control (definitively treated) wells.
- [4] Assays should be repeated until data for at least 3 observed points (a point as an average of replicate observations for a concentration of a chemical) ranging between 20% and 80% of the maximum effect were obtained.
- [5] After the assays, a floppy disk of raw data (observed values of optical density, not calculated percentages) written in a preset format should be submitted to the Working Group together with a print-out of a data file which corresponds to one assay.

The Working Group that included specialists on statistical analyses of collected data consisted of T. Ohno, M. Hayashi, H. Itagaki, M. Kato, S.

Miyazaki, T. Omori, H. Ono, K. Saijo, H. Sugawara, N. Tanaka, N. Teramoto, S. Wakuri and I. Yoshimura. A total of 50 laboratories has co-worked in this study and three staffs were required for the data cleaning and the other two for computer-software programming and statistical analyses of the results. After the preliminary data cleaning, a follow-up meeting was held in Tokyo on November 30, 1994, to discuss problems on this validation study. These problems are occasionally described in the following paragraphs.

This paper presents the overview of the first validation study, describes performance of the assays, and compares their variations on each tested chemical. General problems and points to be improved for further validation studies are discussed. In the following article (hereafter, abbreviated as Validation Article\*\*\*), we will describe the algorithm of the logistic analysis program LAP-JSAAE (see Validation Article II), and human errors found in the analysis (see Validation Article III). Moreover, five articles will follow to describe representative hand-plotted dose-response curves, precise descriptions on ED50 values, comments and discussions on each cytotoxicity assay (see Validation Article IV - VIII). Protocols for each assay will be given in the short articles in this issue. A fact data base constructed on the present results will be available on request.

## Materials and Methods

### *Chemicals tested*

The following 6 chemicals used in this study are given below. Since one of the six chemicals was added after mask-coding to determine intra-laboratory variation of the final data, a total of 7 chemicals was tested. These chemicals were selected to reflect a range from non-irritating to severely irritating as reported in *in vivo* Draize

---

\* : An observation point was based on the mean of raw data at a concentration of the tested chemical.

\*\* : An assay consists of 3 tests carried out in triplicate plates.

\*\*\* : A series of articles on validation study on five cytotoxicity assays by JSAAE consisting of I through VIII.

**Table 1.** Cell lines used in the first validation study

Assay	Cell lines	Characteristics	Number of participating laboratories	
			coworked*	data submitted**
Colony formation (CF)	HeLa S3 (SC)	Human cervix carcinoma	24	19
	BALB/3T3 A31-1-1	Mouse embryonic fibroblasts	24	22
Crystal violet staining (CV)#	HeLa S3 (SC)	Human cervix carcinoma	16	14
	CHL	Chinese hamster lung cells	16	14
LDH release				
LDH-1	HeLa S3 (SC)	Human cervix carcinoma	18	11
	SQ-5	Human lung squamous carcinoma	18	11
LDH-2A	HeLa S3 (SC)	Human cervix carcinoma	18	10
	SQ-5	Human lung squamous carcinoma	18	10
LDH-2B	HeLa S3 (SC)	Human cervix carcinoma	18	9
	SQ-5	Human lung squamous carcinoma	18	9
LDH-2C	HeLa S3 (SC)	Human cervix carcinoma	18	10
	SQ-5	Human lung squamous carcinoma	18	9
MTT assay (MTT)	HeLa S3 (SC)	Human cervix carcinoma	20	18
	SQ-5	Human lung squamous carcinoma	20	17
Neutral red uptake (NR)	HeLa S3 (SC)	Human cervix carcinoma	24	22
	NRCE	Rabbit corneal cells, serum-free	24	20

\* Number of participating laboratories involved in each assay.

\*\* Number of laboratories which submitted final data files to The Working Group.

# Characters in the parentheses of the column, Assay, indicate abbreviations used in the text.

tests (Draize, 1944). They are:

Tween 20 (#1); the maximum average score (MAS) at 10% is 5.7 to 6.0 (Catroux *et al.*, 1993; Bagley *et al.*, 1992a). In ocular irritancy classification, this chemical has been reported as 'non' (Watanabe *et al.*, 1989), 'minimal' (Gordon, 1992), 'less than mild' (Gordon and Bergman, 1988), 'mild' (Silverman and Pennisi, 1987), 'mild' at 100% (Rachui *et al.*, 1994).

Tween 80 (#2); MAS at 10% is 3.8 to 4.0 (Catroux *et al.*, 1993; Bagley *et al.*, 1992b), modified MAS is 4.00 (Bagley *et al.*, 1992a). In ocular irritancy classification, this chemical has been reported as 'non' (Watanabe *et al.*, 1989), 'no' (Marinovich *et al.*, 1990), 'minimal' (Gordon, 1992), 'less than mild' (Gordon and Bergman, 1988), 'mild' at 100% (Rachui *et al.*, 1994).

Sucrose fatty acid ester (#3); in ocular irritancy classification, this chemical has been reported as 'mild' (Watanabe *et al.*, 1989).

Propylene glycol (#4); modified MAS is 1.33 (Bagley *et al.*, 1992a). In ocular irritancy classification, this chemical has been reported as 'non' (Watanabe *et al.*, 1989), 'less than mild' at 15% (Gordon and Bergman, 1988), 'minimal' (Silverman and Pennisi, 1987), 'slightly irritant' (Guillot *et al.*, 1982).

Cetylpyridinium chloride monohydrate (#5); in ocular irritancy classification, this chemical has been reported as 'severe' (Gordon, 1992), 'moderate' at 0.5% (Gordon and Bergman, 1988). Otherwise as cetylpyridinium bromide, modified MAS at 10% is 89.67 (Bagley *et al.*, 1992a), MAS at 10% is 57.20 (Catroux *et al.*, 1993), thus as 'severe' (Marinovich *et al.*, 1990).

Sodium lauryl sulfate (#6); MAS at 10% is 37.0 (Bagley *et al.*, 1992a) and 37.30 and 22.30 (Catroux *et al.*, 1993), modified MAS at 15% is 59.17 (Bagley *et al.*, 1992b), maximum at 10% is 40 (Griffith *et al.*, 1980). This chemical has been reported as 'moderate' (Watanabe

*et al.*, 1989), 'moderate' (Marinovich *et al.*, 1990; Gordon, 1992), 'Moderate' at 10% (w/v) (Silverman and Pennisi, 1987), 'Moderate' at 10% (Rachui *et al.*, 1994).

Tween 20 (#7, equal to #1, an intrinsic masked reference chemical).

The # numbers in parentheses are doubly-mask-codes given before the transfer of the chemicals to each laboratory. The Chemical Bank was previously set up in The Research Laboratory of Wako Pure Chemical Industries Ltd. (Hyogo) which supplied all the chemicals except sucrose fatty acid ester purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo). A single lot of each chemical was used and original stocks of the lot were stored at 4°C in the Chemical Bank. These chemicals were sent to the 45 participating laboratories (two laboratories dropped out before receiving the chemicals) and also to Hatano Research Institute, Food and Drug Safety Center (Kanagawa) for the *in vivo* Draize test for confirmation of the Draize scores. Purity and chemical properties assured in the Chemical Bank revealed that these were stable through the whole period of the present validation study.

#### Cell lines

Two cell lines were used in each assay as shown in Table 1. HeLa S3 (SC) cells (a subline of HeLa) were used as a common cell line in all assays.

Each cell line except serum-free cultured NRCE was supplied from RIKEN Cell Bank as live cultures derived from a single lot. Eagle's minimum essential medium (MEM) supplemented with 10% bovine calf serum (CS) was used for HeLa S3 (SC) and CHL cells. MEM containing 10% fetal bovine serum (FBS) was used for SQ-5 and BALB/3T3 A31-1-1 cells. Since serum-lots are known to greatly influence the cloning efficiency of BALB/3T3 A31-1-1 cells, a specified lot of FBS (lot. U15600, NescoBio, Co., Tokyo) was purchased by the participating laboratories. Lots of CS and FBS, however, were freely chosen by each laboratory for assays with HeLa S3 (SC) cells and other cell lines. NRCE cells, the serum-free primary-

cultured Japanese white rabbit cornea epithelial cells, were provided in a commercially available frozen kit (Corne-Pack) together with the serum-free culture medium RCGM from Kurabo Industries Ltd. (Osaka).

#### Cytotoxicity assays

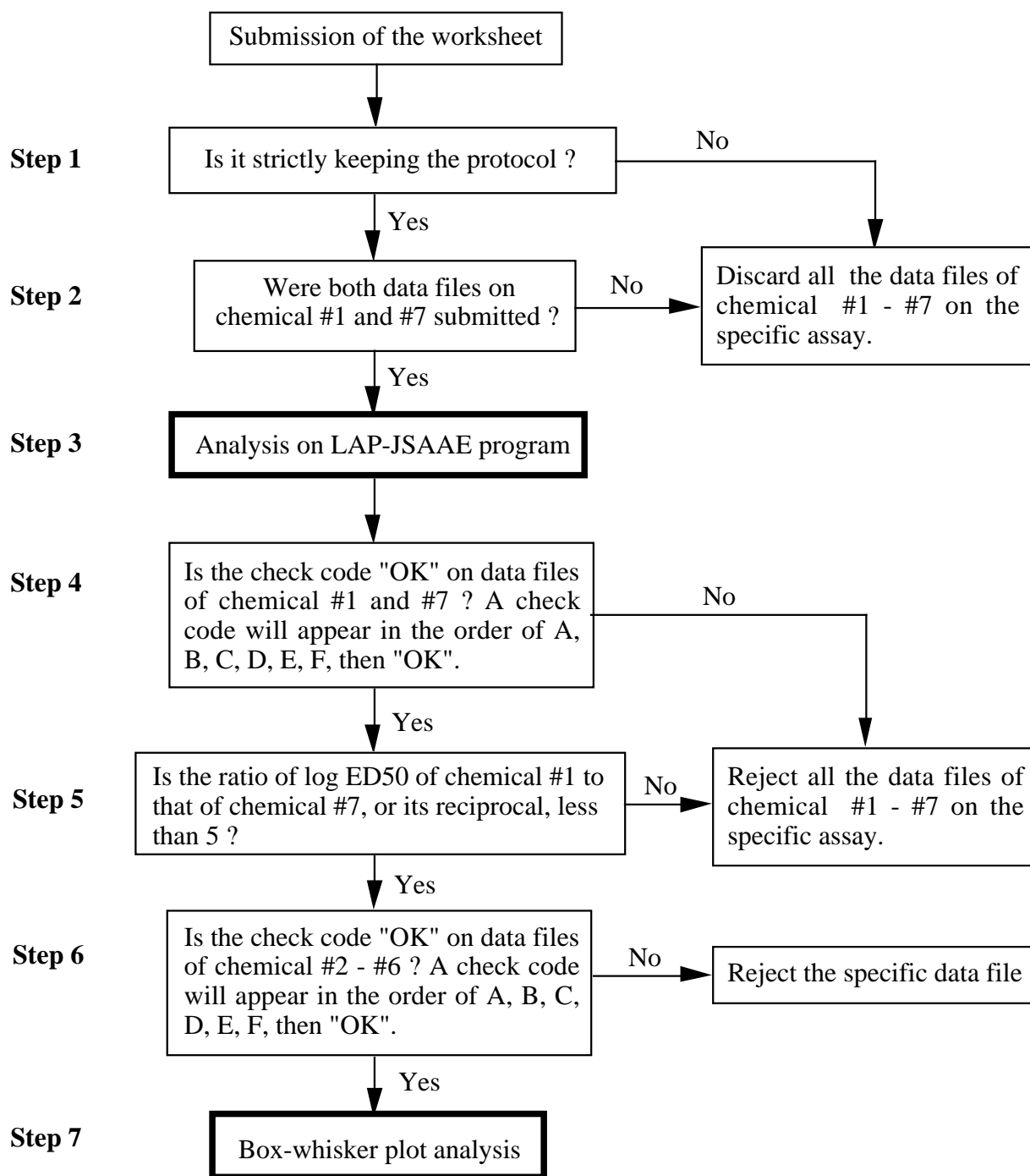
Precise protocols on each assay will be described in subsequent articles in this issue written by the Working Group. The following paragraphs are briefly digested descriptions of the assays. Manufacturers of plastic plates and other lab-wares were not specified but were freely selected by the participants.

Colony formation (CF) assay has been done to determine directly the proliferation ability of individual cells (Sasaki and Tanaka, 1991). Two kinds of cells, BALB/3T3 A31-1-1 and HeLa S3 (SC) were used. Cells were thoroughly dissociated and seeded into 60-mm dishes. After 18-24 hr culture, 20 µl of appropriately diluted test chemical solution was added. Incubation was continued without medium change at 37°C under 5% CO<sub>2</sub> in air for 7 days for BALB/3T3 A31-1-1 cells and 13 days for HeLa S3 (SC) cells. Then the cells were fixed and stained with Giemsa solution. Colonies with 50 or more cells were counted.

For crystal-violet staining (CV) assay (Saotome *et al.*, 1989; Itagaki *et al.*, 1991), 100 µl samples of suspensions of HeLa S3 (SC) (10,000 cells/well) or CHL (4,000 cells/well) cells in logarithmic growth phase were seeded in each well of 96-well plates containing appropriately diluted chemicals in 100 µl of culture medium. The cells were cultured at 37°C under 5% CO<sub>2</sub> in air for 3 days, and then fixed with 25% glutaraldehyde for 10 minutes, and stained with 0.4% crystal violet for 30 minutes. The plates were then dried and OD590 values were determined.

Lactate dehydrogenase release (LDH) assay has been described in detail in previous papers (Sasaki *et al.*, 1992; Wang *et al.*, 1993). Briefly, SQ-5 cells or HeLa S3 (SC) cells were seeded into 96-well culture plates at 2,000 cells/well and 4,000 cells/well, respectively, suspended in 100 µl of culture medium which contain 5% FBS to





**Fig. 1.** 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 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 interval 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.

reduce the background LDH activity in the medium, and pre-cultured at 37°C under 5% CO<sub>2</sub> in air for 24 hr. Then the cells were treated with appropriately diluted solutions of test chemicals in PBS(-) for 20 min at room temperature (assay series LDH-1), for 2 hrs (assay series LDH-2A) or for 48 hrs (assay series LDH-2B and LDH-2C) in culture medium at 37°C. LDH activity was determined with a commercially available enzymatic test kit (Kyokuto Pharmaceutical Inc., Tokyo). Cell lysis and growth-inhibition by the treatments were determined from OD560 values of the culture supernatants (LDH-1, LDH-2A, LDH-2C) or cell layer (LDH-2B). The possible direct inhibition of LDH activity by the chemicals was first tested, and if inhibition was detected, the factor for remaining LDH activity (FRLA) at each concentration of the test chemicals was used to normalize the cytotoxicity data.

For the MTT assay [MTT, 3-(4,5-dimethylthiazol-2-yl)2,5-diphenyl tetrazolium bromide], HeLa S3 (SC) cells (3,000 cells/well) and SQ-5 cells (3,000 cells/well) were pre-cultured in 100 µl/well of culture medium in 96-well plates for overnight. Then the same volume of 2-fold concentrated test solution was added and the culture was continued for 48 hr. MTT metabolism was determined according to Mosmann (1983).

For the neutral red uptake assay (NR) (Borenfreund and Puerner, 1985; Hockley and Baxter, 1986), NRCE cells (2,500 cells/well) were pre-cultured in 100 µl/well of serum-free RCGM-CaCl<sub>2</sub> medium at 37°C under 5% CO<sub>2</sub> in air for 72 hr. HeLa S3 (SC) cells (4,000 cells/well) were pre-cultured in 100 µl/well of MEM-10%CS medium at 37°C under 5% CO<sub>2</sub> in air for 24 hr. Both cell lines were incubated for 48 hr with test chemicals. After uptake of NR and fixation, OD540 values derived from neutral red were measured and the cell survival was calculated.

### *Technology transfer*

After the initial orientation held in Tokyo on October 30, 1992, technology transfer of the five cytotoxicity assays was carried out according to the protocols (standard operating procedures) at

RIKEN on the 5 assays; at Hatano Research Institute, Food and Drug Safety Center, on the CF and NR assays; at Nagasaki University on the 5 assays; at Yokohama City University on the CV assay, and at Kurabo Industries Ltd. on the NR assay with NRCE cells in serum-free culture. Participants from each laboratory in this validation study received training in one of these technology-transfer courses (2-5 days) except those who had already received training in the 5 assays at RIKEN in July 1992. After the training courses, each laboratory received cell lines for the assays in the present study.

### *Data handling at The Working Group*

Fig. 1 illustrates the flow of collected data files handled by The Working Group. One data file written in the style of a pre-set worksheet corresponded to one assay on a chemical carried out by a laboratory. Data files stored on floppy disks were also submitted to The Working Group. Each data file was examined to determine whether the assay followed the common rules and/or if it contained any problem pertaining to the protocol used. Any data file found to contain a serious problem against even only one of the common rules or the protocol was rejected and no further analyses were done on it. If the data file of chemical #1 or #7 was not included in the set of data files for the 7 chemicals on a particular assay submitted from a laboratory or if the data file was rejected for any reason, the set of data files for the 7 chemicals on that particular assay from that laboratory was discarded.

Acceptable files were analyzed by the logistic analysis program, LAP-JSAAE (see details in the Validation Articles II in this issue : the source code of the program is available on request). If any check code from Code-A - Code-F (see below) was attached by the program to the data file of chemical #1 or #7, the set of data files for the 7 chemicals on a particular assay submitted from the laboratory was rejected. If the ratio of ED50 values of chemicals #1 to #7 (or #7 to #1: the smaller ED50 was taken as the denominator) resulted in a figure of 5 or more, again, the set of data files of the 7 chemicals on

the assay submitted from the laboratory was rejected. If any check code was attached by the program to a specific data file of chemicals #2 - #6, the specific data file of the chemical in the assay submitted from the laboratory was rejected. Data files that passed through the above scrutiny were further submitted to box-and-whisker plot analysis.

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 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 responses of 200% or more where the negative controls were set at 100%.

Code-E : Of the 95% confidence interval 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.

#### *Analysis by box-and-whisker plot*

To evaluate the inter-laboratory variations, we used the method of box-and-whisker plot (Tukey, 1977), since the distribution of ED50 values among participating laboratories was rather skew in the upper tail and frequently contained outliers. As a result, the median and the hinge-spread of log(ED50) were calculated for each chemical and each assay with a cell line to represent the central tendency and the amount of variation of ED50s among laboratories, where the hinge-spread was defined as the difference between the upper and lower hinges, i.e., the inter-quartile range (see Validation Article-II).

With this context, the capability of each assay to distinguish irritancy of chemicals was evaluated by the range of medians among chemicals standardized by the mean of hinge-spreads, which was referred to as the "power for distinction"(PFD). Thus, PFD was defined as

$$\text{PFD} = [\text{max}(\text{median of log(ED50)}) - \text{min}(\text{median of log(ED50)})] / [\text{mean of hinge-spreads}]$$

where the mean in the denominator was calculated among the chemicals.

#### *In vivo Draize test*

The Draize eye irritation test (Draize, 1944) was carried out to evaluate the eye irritancy of chemicals at Hatano Research Institute, Food and Drug Safety Center. Using 0.1 ml of a 10% aqueous solution, the test chemical was applied into the conjunctival sac of one eye of each of 3 rabbits and no irrigation was performed following the application. Effects on the eyes were graded periodically according to the classic scoring system (Draize, 1944). The average scores for cornea, iris and conjunctiva and the average of the total scores were recorded at the time of maximal responses. If the eye irritancy was found to be too high or too low to quantify, the test was repeated with ten-times lower or higher concentration of the test chemical, respectively.

## **RESULTS**

#### *Organization and processing of the present validation study*

After calling on laboratories to participate in the present validation study by September 1992, 47 laboratories had registered. Every laboratory was allowed to participate in any assays they selected without any prerequisites, but they were required to state their selected assays before the start of the validation study. Two laboratories did not attend the orientation meeting held in October 1992, in which common rules for assays were discussed and accepted.

Among the remaining 45 laboratories, three laboratories began to co-work in this study but two of them were not able to submit any data files by the deadline that was initially set at March 19, 1993, but deferred until August 31, 1993. The other one had carried out LDH assay they participated in, but was beset by a mycoplasma contamination in their test cell lines during maintenance culture; other laboratories did not have this problem. Thus 3 laboratories were prompted to withdraw from the validation study. The number of qualified laboratories was there-

**Table 2.** Number of participating laboratories that received the technology transfer

Assay	Coworked laboratories	Technology transfer			
		(1) Received and practiced	(2) Received but not practiced, used technician	(3) Not received but practiced*; used technician	(4) Did not submit any data (but received)
CF	24	15	1	6	2 (1)
CV	16	6	1	7	2 (2)
LDH-1	18	9	1	1	7 (5)
LDH-2A	18	8	1	1	8 (6)
LDH-2B	18	7	1	1	9 (7)
LDH-2C	18	8	1	1	8 (6)
MTT	20	10	2	6	2 (1)
NR	24	16	1	5	2 (2)
Total**	156	79	9	28	40 (30)

\* Laboratories which have been well experienced in each assay.

\*\* Cumulative number of laboratories. Each laboratory was allowed to participate in multiple assays. The actual number of co-worked laboratories which carried out the assays was 45.

fore trimmed down to 42. Their data files were analyzed based on the following description.

The Chemical Bank and the Cell Bank were set for this validation study in Wako Pure Chemical Industries, Ltd. (Hyogo) and in RIKEN Cell Bank (Ibaraki), respectively. These banks provided quality-controlled chemicals and cell lines except the serum-free cultured cell line, NRCE, which was directly provided in a commercially available kit together with the serum-free culture medium from Kurabo Industries Ltd. (Osaka).

Technology transfer on the assays was carried out for the participating laboratories through 2-5 day courses as described in Materials and Methods. We assumed that this transfer contributed to the adjustment of technical expertise of, at least, the attendants to the course. As shown in Table 2, the cumulative number of 79 laboratories in a total of 156 co-worked laboratories carried out the assays by those who received the technology transfer. However, in 9 laboratories the assays were carried out not by researchers who received the technology transfer but by technicians who did not directly receive the technol-

ogy transfer (the qualities of these technicians are unknown). Twenty-eight laboratories that are well experienced in all assays did not attend the training. There were 40 laboratories which did not submit any data, 30 of which had received the technology transfer.

Through a questionnaire, we asked a total of 90 participants about the length of their experience in cell culture. Among the 61 who had over one year's experience, 34 received the technology transfer, while among 5 with less than one year's experience, 4 received the technology transfer.

#### *Cell lines*

Two cell lines per assay (Table 1) were sent to participating laboratories as live cultures (except NRCE cells that have been frozen) between November 24 and December 1, 1992. HeLa S3 (SC), a subline of HeLa, was set as a common cell line in the 5 assays since it has been adopted in the standardized protocol for quality control of Eagle's minimum essential medium labeled with the Japanese Industrial Standard (JIS) K-

3602. Other cell lines were selected because of their wide-spread use in corresponding assays.

Since a single lot of cells was sent to each laboratory, no essential difference of characteristics was observed in the cells during the present validation study except in Lab-15 where mycoplasma contamination was found as mentioned above.

#### *Common rules for assays and data collection*

The five cytotoxicity assays were carried out from November 1992 to August 1993. From the expected 2,184 data files of final definitive assays where one data file corresponded to one assay (i.e., triplicate tests carried out in three separate plates) on a chemical, we collected 1,535 data files, although a total of 3,810 data files were submitted including 2,275 data files of preliminary assays (Table 3).

Participating laboratories were expected to submit data files written in the preset format on computer software. However, many of the submitted data formats were different from the preset ones. Due to the wide variety of data formats, The Working Group should have examined precisely and thoroughly all the input positions in every data file. The data files were finally shaped up to the format for a personal computer software program, EXCEL version 4.0 (Microsoft Co. Ltd.) with considerable data-cleaning work.

Although no violation of the common rules [1] and [2] was observed, violation of the common rule [3] was found. There were 98 data files for the CV, LDH, MTT, and NR assays from Lab-24 which were derived from one-plate tests (not 3 independent-plate tests), and 68 data files on the LDH and MTT assays from Lab-23 obtained from tests in which negative- and/or positive-control measurements had been done in separate (not in the same) plates. Assays in which cell lines were cultured in the medium with different type of serum and/or basal medium comprised 77 data files. These data files were not accepted for the calculation of ED50 values.

A laboratory (Lab-2) carried out the CV assay with two cell lines on the 7 test chemicals

using two plates each. Lab-17 performed the CF assay with BALB/3T3 A31-1-1 cells using 3 dishes for an observation point although 4 dishes were required. For these two rare cases, however, The Working Group included their data files as possibly acceptable candidate data files since measurements for an observation point in duplicate plates or triplicate dishes allowed us to detect their intra-assay variations which were revealed to be generally far smaller than the inter-laboratory variation.

Other apparent misunderstandings of the protocols were observed. Such misunderstandings were evident in 49 data files that consisted of the following : 1) the correction factor for calculation of original LDH activity resulting from direct inhibition on LDH activity of each chemical tested only with HeLa S3 (SC) cells was applied to the tests with SQ-5 cells; 35 data files; 2) the chemical concentrations tested in each LDH-2B and LDH-2C assay should have been the same since these assays were to be carried out in tandem but the concentrations were different in 12 data files. Two files were short of LDH-2C assay. Taken altogether, 292 files were judged unacceptable before calculation of the dose of 50% effectiveness (ED50) (Table 3, column c).

On the common rule [4], only 797 data files recorded 3 or more observation points between 20% and 80% of the maximum effect, that comprising 64% of the 1,243 possibly acceptable candidate files (Table 4). The lowest was 46% of possibly acceptable candidate files observed in the LDH-2A assay with HeLa S3 (SC) cells, while the highest was 82% in the LDH-2B assay with SQ-5 cells. Although hand-plotted dose-response curves of data drawn for each laboratory (See subsequent Validation Articles IV - VIII for each assay in this issue) did not conform to the common rule [4], we found that many of them which did not contain 3 observed points between 20-80% of the maximum effect, but 2 (238 files) or 1 (122 files), were practically informative to calculate ED50 values.

Therefore The Working Group reconsidered

---

\* Source code of LAP-JSAE is available on request.

**Table 3.** Data files of final definitive tests submitted to The Working Group

Assay	Cells	Number of data files					Performance rate of		
		a*	b	c	d	e	runs <sup>\$</sup> b/a	candidate files <sup>\$\$</sup> d/a	finally accepted e/a
CF	HeLa S3 (SC)	168	130	14	116	100	77%	69%	60%
	BALB/3T3 A31-1-1	168	149	0	149	126	89	89	75
CV	HeLa S3 (SC)	112	98	14	84	75	88	75	67
	CHL	112	97	14	83	78	87	74	70
LDH-1 <sup>#</sup>	HeLa S3 (SC)	126	70	13	57	39	56	45	31
	SQ-5	126	70	20	50	25	56	40	20
LDH-2A	HeLa S3 (SC)	126	68	14	54	26	54	43	21
	SQ-5	126	67	28	39	28	53	31	22
LDH-2B	HeLa S3 (SC)	126	62	21	41	18	49	33	14
	SQ-5	126	62	28	34	20	49	27	16
LDH-2C	HeLa S3 (SC)	126	68	14	54	26	54	43	21
	SQ-5	126	61	21	40	25	48	32	20
MTT	HeLa S3 (SC)	140	123	28	95	89	88	68	64
	SQ-5	140	117	28	89	83	84	64	59
NR	HeLa S3 (SC)	168	154	28	126	108	92	75	64
	NRCE	168	139	7	132	103	83	79	61
Total number of files		2184	1535	292	1243	969			

\* a, expected; b, submitted; c, unacceptable; d, acceptable before ED50 calculation; e, finally accepted by the logistic analysis program, LAP-JSAAE, and the intra-laboratory variation analysis (see Table 5 and 7).

\$ runs = b/a x 100

\$\$ candidate files = d/a x 100, accepted by The Working Group before ED50 calculation.

# Different series in the LDH release assay, see Materials and Methods.

The figures in the table were updated and amended after our presentation in the meeting INVITOX'94 (Zurich, Switzerland, 1994) to which data were submitted and published in the proceedings (Ohno, T., *et al.*, 1995). Definition of performance rates were changed in order to reflect reality from that described in the proceedings. Other than these final test data files, 157 laboratories submitted 2,275 data files of preliminary tests.

that, by using a computer-assisted program for statistical analysis to calculate ED50, these files should be accepted if dose-response curves were successfully drawn with regular sigmoidal shape. The curves were calculated by the logistic analysis program, named LAP-JSAAE\*, which was newly developed and based on the general statistical data analysis software, SAS (SAS Institute Japan Co., Tokyo) (see Validation Article II in this issue for the algorithm). However, 86 data files containing no observation point between 20-80% of the maximum effect were discarded since 47 of them were found unable to give any reasonable sigmoidal dose-response curves by the

LAP-JSAAE program. This reconsideration on the common rule [4] increased the number of acceptable files to 1,157 (93% of the candidate 1,243 files).

#### Human errors

Fig. 2 illustrates numbers and percentages of data files in which errors, originating from data handling before the data submission (defined as "human errors"), were found in the data cleaning process. Although precise description will be made later (see the Validation Article III in this issue), cumulatively 1,742 data files were

**Table 4.** Number of files including different number of observed points\* between 20 - 80 % effectiveness

Assay	Cells	Candidate files	Between 20 - 80 % effectiveness				
			3 points or more	%**	2 points	1 point	0 points
CF	HeLa S3 (SC)	116	68	59%	28	11	9
	BALB/3T3 A31-1-1	149	95	64%	32	12	10
CV	HeLa S3 (SC)	84	56	67%	17	7	4
	CHL	83	59	71%	14	7	3
LDH-1	HeLa S3 (SC)	57	37	65%	5	8	7
	SQ-5	50	34	68%	5	3	8
LDH-2A	HeLa S3 (SC)	54	25	46%	11	9	9
	SQ-5	39	20	51%	5	9	5
LDH-2B	HeLa S3 (SC)	41	22	54%	7	6	6
	SQ-5	34	28	82%	3	3	0
LDH-2C	HeLa S3 (SC)	54	35	65%	7	6	6
	SQ-5	40	25	63%	14	1	0
MTT	HeLa S3 (SC)	95	75	79%	12	5	3
	SQ-5	89	59	66%	20	5	5
NR	HeLa S3 (SC)	126	92	73%	26	5	3
	NRCE	132	67	51%	32	25	8
Total		1243	797	64%	238	122	86

\* An observed point corresponds to mean of raw data derived from replicate wells for an observation of effectiveness of a chemical.

\*\* % of candidate files

confirmed one by one by The Working Group as possessing errors. The highest error (44%) was the lack of essential data into the data files.

#### *Rejected data files*

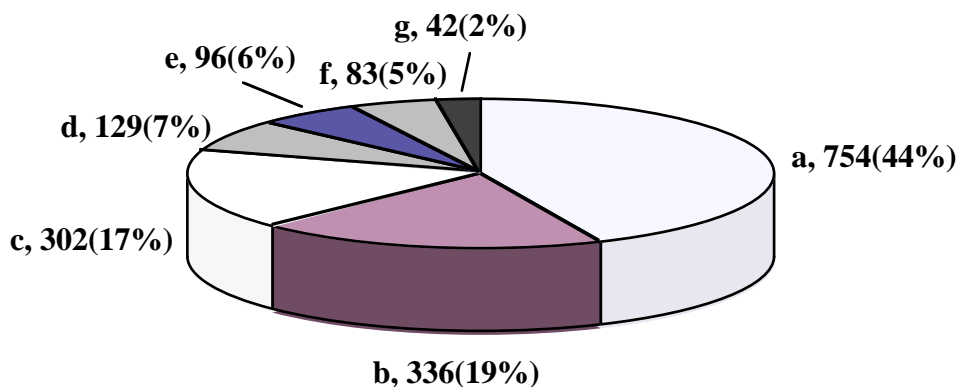
As described in Materials and Methods, after checking the abnormality of data files of chemical #1 and #7 (both are Tween 20) and intra-laboratory variation of ED50 values on these two coded chemicals, the check codes worked to detect “rejectable” data files. Table 5 shows the results. Priorities of these check codes are in the order of A, B (for LDH assay only), C, D, E, and F.

If any error was found in one of the data files of chemical #1 and #7 before the ED50 calculation, detection of the intra-laboratory variation became impossible for that specific assay. The data file set of chemicals #1 - #7 became accordingly meaningless. As a whole, 133 data files were thus rejected before the calculation of

ED50 values. The most frequent rejection happened in the LDH-2A and -2B assays with HeLa S3 (SC) cells where 21 data files each were rejected in 54 and 41 possibly acceptable candidate files, respectively. With analyses on the intra-laboratory variation of ED50 values determined as the ratio of ED50 values of chemical #1 and #7 (see below for the explanation of Table 6), 5 more data file sets (29 files) were rejected because the ratio exceeded 5.

In the data files for chemicals #2 - #6, a total of 26 data files resulted in failure to calculate an ED50 value (Code-A). Drawing of a logistic curve was impossible in these data files due to large variation of the raw data. An example of this type of variation is shown in Fig. 3.

In the LDH assay, the observed LDH activity is occasionally reduced or enhanced by the test chemicals. The observed raw data for LDH activity were therefore corrected using these inhibitory- or stimulatory-factors, i. e., the factor for remaining LDH activity (FRLA), as de-



**Fig. 2.** Data files in which errors were found in the data cleaning process.

a, Lack of essential data. b, Out-of-format inputs. c, Violation of the protocol or common rules. d, Simple mis-recording 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 print-outs, the data stored in the submitted floppy-disks, and the worksheet. g, others. Total files are 1742.

scribed in Materials and Methods and more precisely in the Validation Article VI on the LDH assay in this issue. However, if the correction factors FRLA fell below 0.1 or above 2.0 that were apparently out of linear range of the possible correction of the observed LDH activity to the original LDH activity in a sample, the correction should be considered essentially impossible. The LAP-JSAAE program marked these data files with the Code-B. Four files in the subdivided LDH assays were therefore rejected (Table 5). An example of this type is shown in Fig. 4 as the raw data plot.

Although calculation met the fundamental requirements of the logistic analysis, as described above, the data files containing no observed point between 20-80% of the maximum effect were rejected (indicated with Code-C in Table 5). There were, at this step, 47 files\*. Also, those files with raw data that included a response of 200% or more, where that of negative controls was set at 100%, were rejected (Code-D in Table 5). Five data files were rejected for this reason.

In 7 data files, the upper limit of the 95% confidence interval of ED50 values was over 100 times the lower limit (Code-E in Table 5). An example of a raw data plot is shown in Fig. 5. At most, 3 files in the NR assay with the serum-free cultured NRCE cells were marked with the Code-E.

We set the upper limit of the RMS value at 10, which, roughly speaking, corresponds to the robustness of the drawn curve (see the subsequent Validation Article II on statistical analyses for its definition) at 10. If the value is 10 or more, the estimated logistic curve should be considered unstable because of a shortage of rigid data for estimation of the logistic curve. Thus, 23 files were rejected for this reason (Code-F in Table 5). As a result, 969 files appeared to be finally acceptable for further analyses of the present inter-laboratory validation study (Tables 3 and 5). Therefore, as a whole, 37% of the submitted data files of the final definitive assays were discarded before the box-and-whisker plot analysis.

#### *Intra-laboratory variation*

The chemical #1 and #7 are the same chemical, Tween 20, derived from the same lot. Since samples of this chemical were sent to each laboratory after doubly-mask-coding, the difference in ED50 values of chemical #1 and #7 was considered to reflect intra-laboratory variation of each assay. To examine this variation, we calculated the ratio of ED50 values of chemical #1

\* : This number does not contain the files that have been rejected under the check code A and B. It is, therefore, different from the number of data files shown in Table 4.



**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	Code-A	Code-B	Code-C	Code-D	Code-E		Code-F <sup>§</sup>
CF	HeLa S3 (SC)	116	0	6 <sup>&amp;</sup>	2		7	0	1	0	100
	BALB/3T3 A31-1-1	149	7	4 <sup>&amp;</sup>	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 <sup>&amp;</sup>	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 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 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 that 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 interval of ED50, the upper limit is over 100 times the lower limit.

§ Code-F indicates that 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.

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

and #7, wherein the smaller ED50 value was taken as the denominator (Table 6).

In 164 data file sets of the chemical #1 and #7 paired in each assay, 75 data file sets (46%) had a ratio in the range of 1.0 to less than 1.1. Only 16 data file sets showed a ratio of 2.0 or more and 159 data file sets (97%) had a ratio less than 5.0. In one extreme case, Lab-24 submitted data files for the CF assay with HeLa S3 (SC) cells that resulted in a ratio of over 100.

We defined a ratio of 5.0 or more as an indication of extraordinary large intra-laboratory variation. Such a data file set in the assay from

the indicated laboratory should, therefore, be eliminated, i.e., Lab-24 in the CF assay with HeLa S3 (SC) cells, Lab-4 in the CF assay with BALB/3T3 A31-1-1 cells, Lab-7 in the LDH-1 assay with SQ-5 cells, Lab-31 in the NR assay with HeLa S3 (SC) cells, and Lab-1 in the NR assay with NRCE cells (Table 6). The above consideration resulted in 159 remaining data file sets which were thought to be acceptable.

As seen in Table 6, four data file sets resulted in a ratio of over 2.0 out of 21 CF assays with BALB/3T3 A31-1-1 cells, and of 18 NR assays with serum-free NRCE cells 3 data sets had

**Fig. 3.** An example of data files that resulted in calculation-failure (Code A) by the logistic analysis program, LAP-JSAAE

Because of extremely large variation of data, drawing of a dose-response curve was impossible in the file of MTT assay for chemical #4 with HeLa S3 (SC) cells.

**Fig. 4.** An example of data files in LDH-2A assay which used the correction factor, FRLA, below 0.1 (Code B)

The data file of LDH-2A assay with HeLa S3 (SC) cells used the correction factor 0.09 for the highest concentration of the chemical #5.

**Fig. 5.** An example of data files in NR assay with NRCE cells resulted in that the upper limit of 95% confidence interval of ED50 was over 100 times the lower limit (Code E)

ED50 ratios exceeding 2.0. The CV and MTT assays, each with 2 cell lines, that were conducted in 11 - 14 trials resulted in ratios of less than 1.7. In the NR assay with HeLa S3 (SC) cells, the ratios observed were within 1.0-1.4 except for Lab-31.

#### *Performance rate of the assays*

From the data files that survived the above-mentioned scrutiny, it was possible to calculate the performance rate of each assay as shown in

Table 3. Laboratories performed the CF, CV, NR, and MTT assays at rates of 77 - 92% on a running basis (column b/a in Table 3). On the other hand, subdivided LDH assays revealed lower performance rates of 48 - 56%. On the possibly acceptable candidate file basis (column d/a in Table 3), the CF, CV, NR, and MTT assays maintained relatively higher performance rates of 64 - 89%. Especially in the CF assay with BALB/3T3 A31-1-1 cells, none of the submitted data files was seen to be unacceptable before calculation of ED50 values.

**Table 6.** Number of data file sets which resulted in the ratio of ED50 values of chemicals #1 and #7 deviating from 1

Assays	CF		CV		LDH-1		LDH-2A		LDH-2B		LDH-2C		MTT		NR		Total file sets
	HeLa	BALB	HeLa	CHL	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	NRCE	
Ratio*																	
1.0 - <1.1	8	7	4	7	1	1	4	3	2	1	1	2	8	5	12	9	75
1.1 - <1.2	3	1	5	2	2	1		1				2	2	2	2	3	26
1.2 - <1.3	3	3	1		1								2	2	1	2	15
1.3 - <1.4		2	1				2							3	1	1	10
1.4 - <1.5	1	1		1						1	2			1			7
1.5 - <1.6		1		1		1	1		1								5
1.6 - <1.7		1		1									2				4
1.7 - <1.8		1			1						1						3
1.8 - <1.9					1												1
1.9 - <2.0	1				1												2
2.0 - <5.0		3			1			1		1	2	1				2	11
5.0 - <10						1									1	1	3
10 - <100		1															1
>100	1																1
Total file sets	17	21	11	12	8	6	5	5	3	3	6	5	14	13	17	18	164

The file sets with the ratio of 2.0 or more

Assays	CF		CV		LDH-1		LDH-2A		LDH-2B		LDH-2C		MTT		NR	
	HeLa	BALB	HeLa	CHL	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	SQ-5	HeLa	NRCE
Labo No.	24	4			9	7		7		45	29	45			31	1
Ratio**	<b>173</b>	<b>13.0</b>			3.29	<b>7.11</b>		2.7		2.83	4.28	3.65			<b>9.15</b>	<b>7.76</b>
Labo No.		12								42						18
Ratio		4.45								2.00						3.06
Labo No.		30														30
Ratio		2.78														2.25
Labo No.		45														
Ratio		2.01														

# Cell names were partly abbreviated as follows. HeLa, HeLa S3 (SC); BALB, BALB/3T3 A31-1-1.

\* Ratio of ED50 values of chemicals #1 and #7. The smaller ED50 value derived from #1 or #7 in each assay was taken as the denominator.

\*\* Extraordinary figures refer to the ratio of ED50s of chemical #1 and #7 (or #7 and #1) equivalent to 5.0 or more as indicated with bold figures.

However, since a relatively large number of unacceptable files (274 files) was detected by the logistic analysis program LAP-JSAAE as shown in Table 5, each assay lowered the performance rate on finally accepted file basis by 59 - 75% for the CF, CV, NR, and MTT assays and by 14 - 31% for the subdivided LDH assays. In that, LDH-2B with HeLa S3 (SC) cells performed on the finally accepted basis as low as 14% of 126 data files initially expected from

18 laboratories. The total number of files finally accepted was 969, that is 44% of 2,184 initially expected files, 63% of 1,535 submitted files as final definitive assays, and 78% of 1,243 possibly acceptable candidate files.

In Fig. 6, we visualized distribution of acceptability of data files from each laboratory. Boxes with a thick background pattern represent files not submitted, unaccepted before the ED50 calculation, or finally rejected. Boxes with a thin



dot pattern (for example, Lab-44 in CV assay shown in Fig. 6b) are the files finally accepted but their descriptions of original “final” concentrations of the tested chemicals were miswritten at the submission of raw data to The Working Group.

Boxes containing a wide thick-patterned area in the subdivided LDH assay sections indicate that many laboratories did not perform the assay. In contrast, sections with many white boxes such as in the CV assay with two cell lines and the CF assay with BALB/3T3 A31-1-1 cells indicate that the assays were performed relatively efficiently. In the CF assay with HeLa S3 (SC) cells carried out by Lab-36, all the log(ED50) values calculated from 6 accepted data files resulted in outliers of the box-and-whisker plot analysis. Also, 5 outliers from Lab-28 were observed in the NR assay with the serum-free cultured NRCE cells.

In Fig. 6, failure of assays on chemical #4 was apparent especially in the CF assay. With HeLa S3 (SC) cells, 16 data files from 24 expected data files were not submitted or were finally rejected. Also with BALB/3T3 A31-1-1

cells, 16 data files were not submitted or were finally rejected and 2 data files resulted in outliers of log(ED50) values in the box-and-whisker plot analysis. This tendency was also observed for the LDH-1 assay (Fig. 6c) but not in the CV, MTT, and NR assays.

#### *Typical dose-response curves*

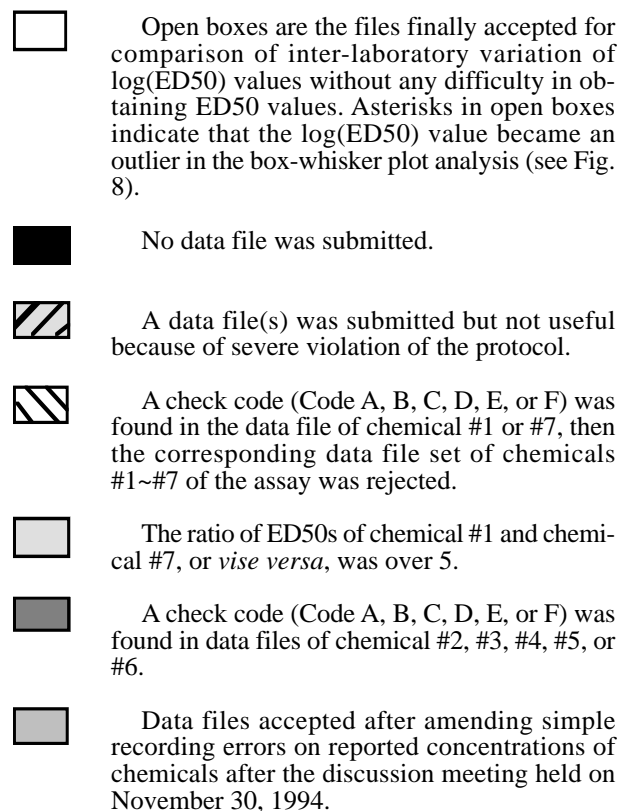
Although the performance rate of the LDH-2B assay was low (Table 3), Fig. 7 shows a representative example of dose-response curves of the chemicals in LDH-2B with the HeLa S3 (SC) cell line as determined by Lab-45. The curves are essentially parallel between the cytotoxicity levels of 20% and 80%, therefore relative cytotoxicity of the 7 test chemicals can be represented by ED50.

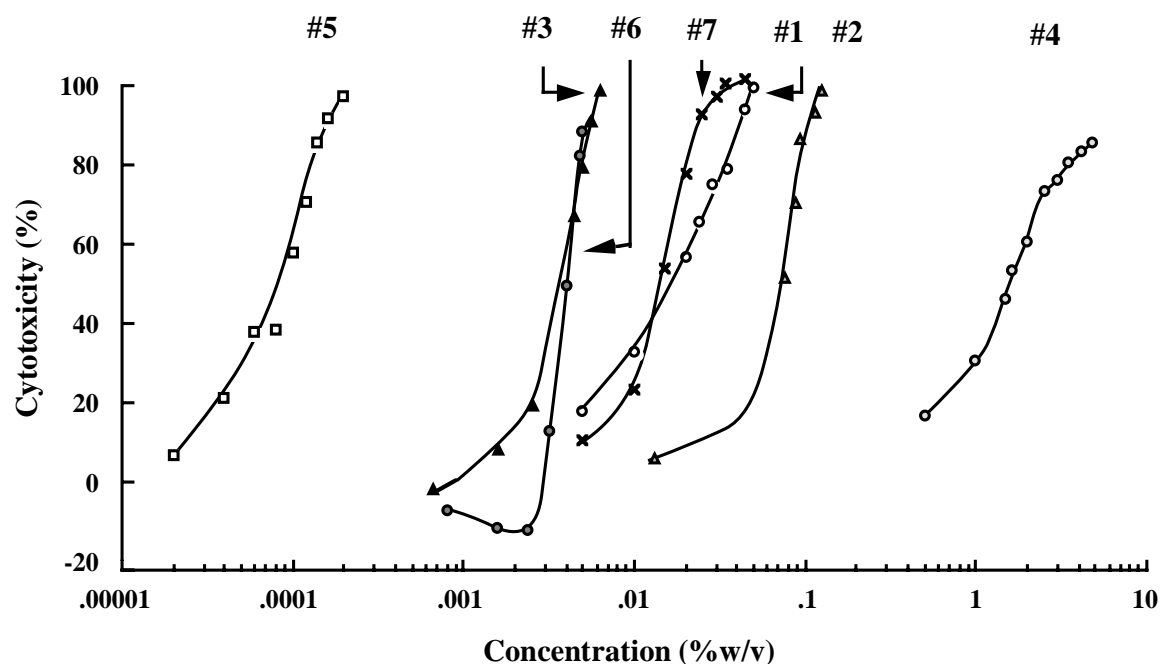
#### *Box-and-whisker plots of log(ED50) values and inter-laboratory variation*

While arithmetic means and standard deviation are usually used to summarize a set of measurements, they are inappropriate to represent our data, since the ED50 values were irregularly distributed and the distribution among laboratories was highly asymmetric, thus outliers were likely to exist. In this situation, statistical insight suggests the use of percentiles such as median and quartiles, invoking the use of box-and-whisker plot as has been described by Tukey (1977). In the box-and-whisker plot (Fig. 8a - g), the vertical line drawn in each box implies the median and, accordingly, represents an average tendency among laboratories, while the length of each box implies the inter-quartile range, i.e., the hinge-spread, and, accordingly, represents the inter-laboratory variation.

All log(ED50) values, their mean and standard deviations, and their coefficient of variations will be described in subsequent Validation Articles IV - VIII written on each assay. In Fig. 8a on chemical #1, relatively small box sizes were observed in the CV and in LDH-2A assays with 2 cell lines. We did not plot the box for the LDH-2B assay since finally accepted data files were less than 5 per chemical tested. For

**Fig. 6.** Acceptability of data files from each laboratory





**Fig. 7.** Typical dose-response curves of the chemicals in LDH-2B with the common HeLa S3 (SC) cell line from Lab-45

the LDH-2C assay with HeLa S3 (SC) cells, however, a large box and an extraordinary long upper whisker were drawn from 6 data files. The same assay on chemical #7 (the same chemical as #1, Tween 20), shown in Fig. 8b, revealed that the median was almost attached to the left edge of the box (lower hinge) and the box lost the upper whisker but was accompanied by an upper outlier.

The CF assay with BALB/3T3 A31-1-1 cells resulted in the lowest median values of  $\log(\text{ED}_{50})$  on chemical #1, #7, #2, and #5 (Fig. 8a, b, c, and f, respectively), while for chemical #4 the lowest median value of  $\log(\text{ED}_{50})$  was obtained in the CF assay with HeLa S3 (SC) cells (Fig. 8e). However, for chemical #3 and #6, the lowest median value of  $\log(\text{ED}_{50})$  was seen in the NR assays with the serum-free cultured NRCE cells (Fig. 8d and g). Two upper outliers were observed in the NR assay with NRCE cells for chemical #6 (Fig. 8g).

#### *Evaluation of the cytotoxicity assays*

To compare inter-laboratory variation of the CF assays with two cell lines, the means of the

hinge-spreads were calculated on the seven chemicals (Table 7). The mean hinge-spread on the CF assay with HeLa S3 (SC) cells and with BALB/3T3 A31-1-1 cells was 0.24 and 0.48, respectively. The latter was twice the former, suggesting that the assay with BALB/3T3 A31-1-1 cells has a tendency for greater variation of  $\text{ED}_{50}$  values than the assay with HeLa S3 (SC) cells.

This concept was extended to other assays with different cell lines. However, it is well known that, when sample size of the box-and-whisker plot analysis is too small, box size will become unstable and inappropriate for statistical analysis. Calculation was therefore not made on mean hinge-spreads of the sub-divided LDH assays, since their sample sizes, i.e., numbers of finally accepted data files, were apparently too small on each chemical, except the LDH-1 assay with HeLa S3 (SC) cells for chemical #1 and #7 (Table 7). Thus, the smallest in calculated mean hinge-spreads was 0.22 for the CV assay with CHL cells and for the MTT assay with SQ-5 cells, followed by 0.24 for the CF assay with HeLa S3 (SC) cells and for the CV assay with HeLa S3 (SC) cells. The largest, 0.48, was seen

**Fig. 8.** Box-whisker plots of cytotoxicity data on each chemical. Code number of the chemical tested was given under each figure.

**Fig. 8(Continued).** Box-whisker plots of cytotoxicity data on each chemical. Code number of the chemical tested was given under each figure.



**Fig. 8(Continued).** Box-and-whisker plots of cytotoxicity data on each chemical. Code number of the chemical tested was given under each figure.

for the CF assay with BALB/3T3 A31-1-1 cells.

Considering the practicability of a cytotoxicity assay, it is desirable that the most cytotoxic chemicals should give the lower ED50 value and the least cytotoxic ones should give the higher ED50 value. Differences in these ED50 values of the assay will serve as an indicator to distinguish individual chemicals from the view point of cytotoxicity.

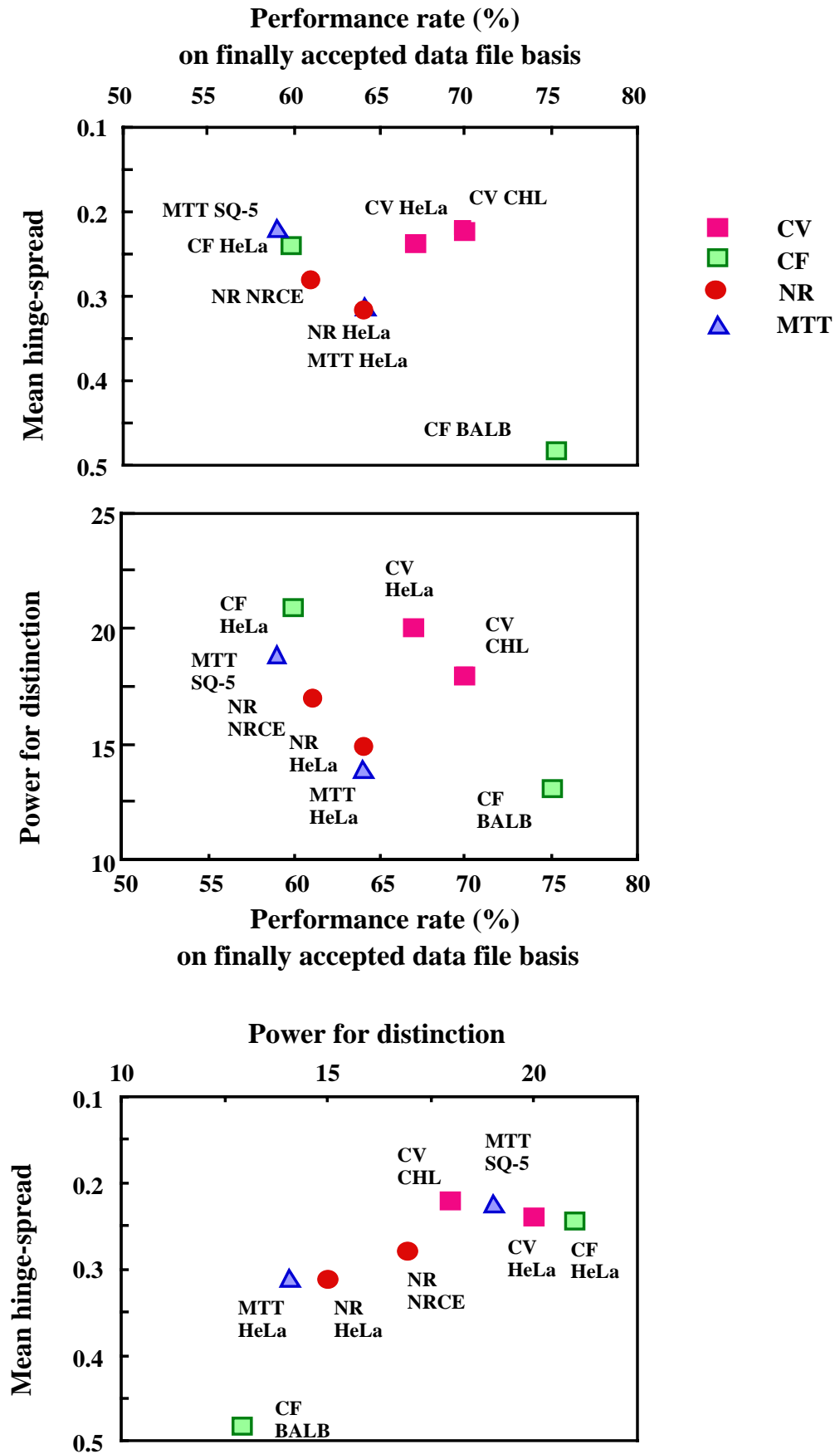
Comparing the difference of median  $\log(\text{ED}_{50})$  values of the least toxic chemical #4 and the most severely toxic chemical #5 among the assays, the CF assay with BALB/3T3 A31-1-1 cells and the CV assay with CHL cells resulted in the largest and the smallest difference, respectively (Table 7). Then the difference was divided with the corresponding mean hinge-spread to normalize the effect of variation of  $\log(\text{ED}_{50})$  values. This was defined as “power for distinction” (PFD) (Table 7, last column) of cytotoxicity of chemicals. The CF assay with HeLa S3 (SC) cells gave the largest PFD of 21.0. CV assay with HeLa S3 (SC) cells revealed also a high PFD with a value of 20.4. The lowest PFD, 13.4, belonged to the CF assay with BALB/3T3 A31-1-1 cells.

When the performance rates, mean-hinge spreads, and PFDs were displayed in 2-dimensional figures (Fig. 9), CV assays with the two cell lines shared the right upper corner in Fig. 9a and b, although in Fig. 9c the CF assay with HeLa S3 (SC) cells and the MTT assay with SQ-5 cells were also grouped together with CV assays with the two cell lines.

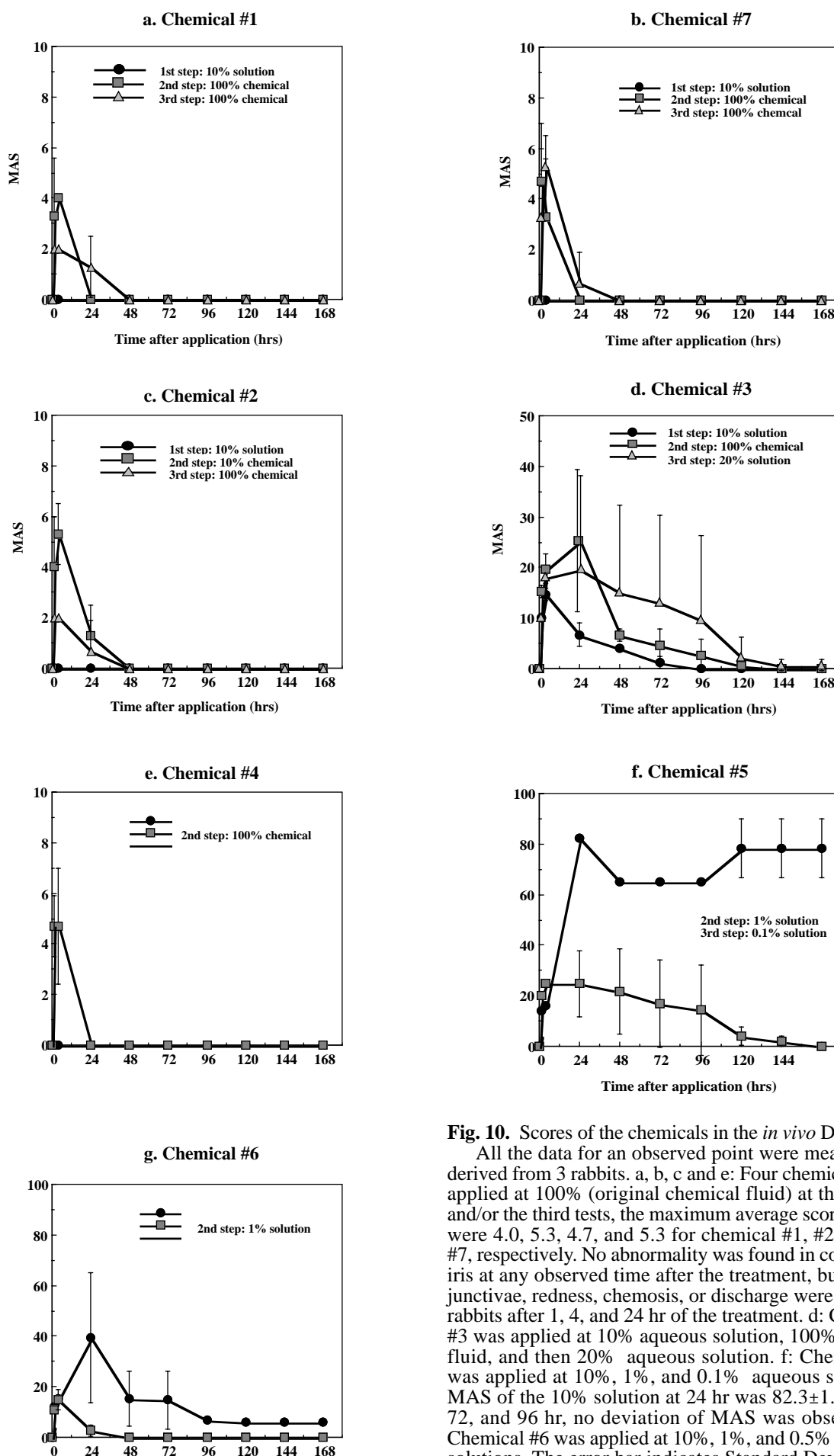
#### *In vivo Draize test*

We have conducted an *in vivo* Draize test to confirm the toxicity of the 7 doubly-mask-coded chemicals in parallel with the cytotoxicity assays. At the first test in which the chemicals were applied to rabbit eyes as 10%(w/v) aqueous solution, chemicals #1, #2, #4, and #7 showed no irritating effect on cornea, iris, and conjunctiva. The maximum average score (MAS) was therefore 0. When chemicals were applied at 100% (original undiluted chemical fluid) at the second and the third tests, the MAS were 4.0, 5.3, 4.7, and 5.3 for chemical #1, #7, #2, and #4, respectively (Fig. 10a, b, c, and e, respectively). No abnormality was found in cornea or iris at any observed time after the treatment, but in con-

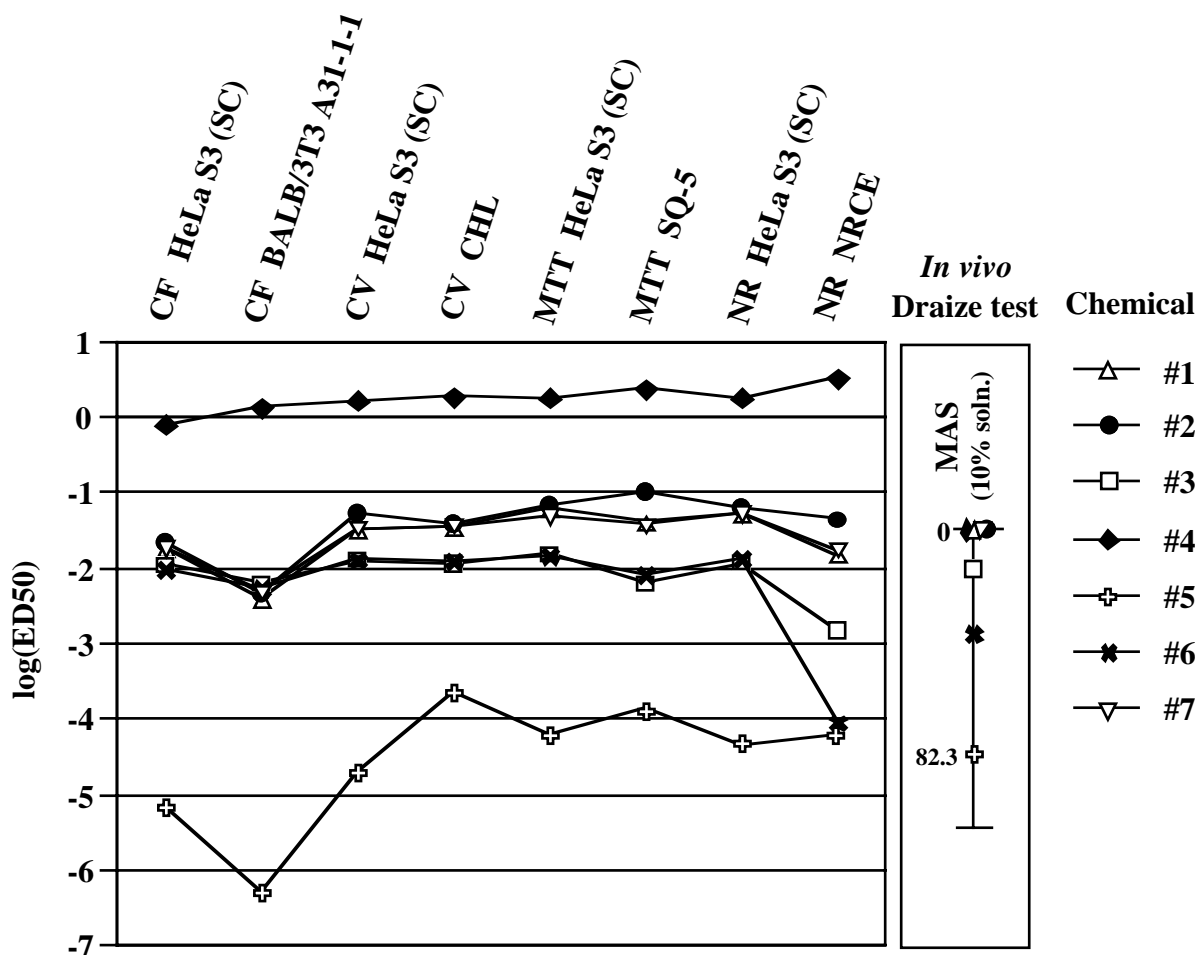




**Fig. 9.** Two dimensional displays of performance rate, mean hinge-spread, and the power for distinction (PFD)  
 a, performance rate vs. mean hinge-spread. b, performance rate vs. PFD. c, mean hinge-spread vs. PFD. Note that mean hinge-spread was plotted up-side down.



**Fig. 10.** Scores of the chemicals in the *in vivo* Draize test. All the data for an observed point were mean values derived from 3 rabbits. a, b, c and e: Four chemicals were applied at 100% (original chemical fluid) at the second and/or the third tests, the maximum average score (MAS) were 4.0, 5.3, 4.7, and 5.3 for chemical #1, #2, #4, and #7, respectively. No abnormality was found in cornea and iris at any observed time after the treatment, but in conjunctivae, redness, chemosis, or discharge were found in rabbits after 1, 4, and 24 hr of the treatment. d: Chemical #3 was applied at 10% aqueous solution, 100% original fluid, and then 20% aqueous solution. f: Chemical #5 was applied at 10%, 1%, and 0.1% aqueous solutions. MAS of the 10% solution at 24 hr was  $82.3 \pm 1.2$ . At 48, 72, and 96 hr, no deviation of MAS was observed. g: Chemical #6 was applied at 10%, 1%, and 0.5% aqueous solutions. The error bar indicates Standard Deviation of the observed MAS.



**Fig. 11.** Correspondence of the median log(ED50) values measured by the *in vitro* cytotoxicity assays to the MAS measured in the *in vivo* Draize test

Medians of log(ED50) were plotted on each *in vitro* assays. The MAS value, 0, of non-irritants, i.e., chemical #1, #2, and #7, was adjusted to the mean of median log(ED50) values of these chemicals, -1.54. Then the MAS of chemical #5 at 10% solution, 82.3, was adjusted to the position where the mean of median log(ED50) values of chemical #5 was -4.54.

junctiva, redness, chemosis, or discharge were found in rabbits 1, 4, and 24 hr after the treatment.

For chemical #3 at the first test, 10% aqueous solution induced no change in cornea or iris, but caused redness and chemosis in all the rabbits 1-48 hr after the treatment. The MAS was 14.7. With 100% chemical application at the second test, the MAS increased to 25.3. When 20% aqueous solution was applied to rabbit eyes at the third test, one of the 3 rabbits exhibited opacity of cornea for 24 - 96 hr after treatment and injection of iris for 24 - 120 hr after treatment. The MAS was 19.7 (Fig. 10d).

Chemical #5 irritated all the 3 rabbits treated with 10% aqueous solution after 24 hr. Rabbits developed opacity of cornea and injection of iris.

Redness of conjunctiva, chemosis, and discharge were observed through the whole test period. The MAS was 82.3. With 1% and 0.1% aqueous solutions, these values decreased to 25.0 and 4.0, respectively (Fig. 10f).

Chemical #6 at 10% aqueous solution induced opacity of cornea in 2 of 3 rabbits after 24 hr and in one rabbit after 48 hr. All the rabbits developed redness of conjunctiva and chemosis through the whole test period. The MAS was 39.3. At 1% and 0.5% aqueous solution, the values decreased to 14.7 and 4.0, respectively (Fig. 10g).

From these results, chemical #1 (Tween 20), #2 (Tween 80), #4 (propylene glycol), and #7 (Tween 20) were judged “non-irritant”; chemical #3 (sucrose fatty acid ester), “irritant”;

chemical #5 (cetylpyridinium chloride monohydrate), “severe irritant”; chemical #6 (sodium lauryl sulfate), “moderate irritant”. These results roughly corresponded to the log(ED50) values measured by the *in vitro* cytotoxicity assays.

As shown in Fig. 11, medians of log(ED50) can be possibly classified into 4 groups, namely, (1) medians of chemical #4 which were distributed near log(ED50) values of 0.0 in every assay; (2) those of chemical #2 and #1 (and therefore #7) of log(ED50) between -2.0 and -1.0 except that of the CF assay with BALB/3T3 A31-1-1 cells which resulted in the median slightly lower than -2.0; (3) those of chemical #3 and #6 consistently near -2.0, although chemical #6 (sodium lauryl sulphate) exhibited an extremely low median log(ED50) in the NR assay with serum-free cultured NRCE cells; (4) those of chemical #5, the severe irritant in the Draize test, near -4.0 or lower (Fig. 11).

When the MAS value of non-irritants, i.e., MAS = 0, was arbitrarily adjusted to the mean of median log(ED50) values of group 2 chemicals (-1.54) and the MAS of chemical #5 at 10% solution (82.3) to the mean median values of group 4 (-4.54), the mean median values of group 3 (-2.17) corresponded roughly to the MAS values of ‘mild’ irritant chemicals #3 and #6. However, the median log(ED50) values of chemical #3 and #6 from the NR assay with serum-free cultured NRCE cells did not match the position of MAS values of these two chemicals.

#### *Fact data base construction*

A database was developed that allows the collaborators to cross-refer the data produced in this project. Since most of the members use stand-alone DOS machines or Macintosh in their laboratories, Excel version 4.0 was selected for the database management system that absorbs the difference of the platform. A user is able to search the database by using a combination of four data items such as the identification number of a laboratory, a range of ED50 values, a test chemical, and a type of proposed cytotoxicity assay. In the event of a query, the file name which contains the raw data submitted by a par-

ticipating laboratory is also displayed together with the four data items and reliability information. Thus the detailed experimental results could be referred to at once and with ease with point and click of the PC mouse. This fact data base is available at the homepage below.

<http://wdcm.nig.ac.jp/validation.html>

## **Discussion**

### *General consideration*

Since the present validation study was planned to determine the practicability of the proposed cytotoxicity assay methods, no restriction was set on the competence of participating laboratories. However the submitted data files were rigorously examined and the performance of the assays was seriously considered. Results shown in Table 3 suggest that the CF, CV, NR, and MTT assays, but not the LDH assay, are recommendable from the view point of performance rates based on the finally accepted data files. These four assays have essentially the same characteristic, i. e., determining the extent of cell survival and/or growth. On the other hand, the LDH assay measures simultaneously, as shown by the subdivided LDH-2 assays, not only the survival and/or growth of cells but also cell death through LDH activity released from the cells (LDH-2C assay). Besides the determination of FRLA, this characteristic of the LDH assay made the procedures apparently more complicated than the other 4 assays and, therefore, was assumed to have resulted in low performance rates (Table 3). This suggests that simplicity of the assay method is an important component to give optimum performance rate.

The CV assay with CHL cells and the MTT assay with SQ-5 cells resulted in the smallest mean hinge-spread (Table 7). These assays may be proposed as the most useful from the view point of data variation. However, from the view point of the newly defined PFD of cytotoxicities for chemicals (see the next section for its meaning), the CF assay with HeLa S3 (SC) cells may also be the most useful assay if one compares the least and the most toxic chemicals

(Table 7). This assay, however, was not powerful enough to distinguish the toxicities of slightly- to moderately-toxic chemicals, #1, #7, #2, #3, and #6, as shown in Fig. 11. Next to the CF assay, the CV assay with HeLa S3 (SC) cells gave the second smallest mean hinge-spread and the second highest PFD (Table 7) with the ability to distinguish the toxicities of slightly- to moderately-toxic chemicals (Fig. 11).

The CF assay required 8 or 14 days for a run, i. e., one day for preculture and 7 or 13 days for treatment with chemicals and colony formation, although actual labor for the assay requires only 3 days, i.e., the preculture, the treatment, and colony counting. The size distribution of colonies is additionally informative. Contrary to this long period assay, the CV assay with HeLa S3 (SC) cells, the second best suggested from Table 7 and Fig. 11, requires 3-day assay period and 2-day actual labor. Although cells were treated for only 2 days in the MTT and NR assays, these assays require a 24-hr preculture (except the NR assay with NRCE cells that require 72-hr preculture) for a total of 3 days with actual labor, which is equal to the entire period required for the CV assay. Therefore we think that the CV assay is simpler than the CF, MTT, and NR assays.

These points and the localization displayed in Fig. 9 prompted us to consider that the CV assay will give the most robust inter-laboratory results and therefore is the most practical cytotoxicity assay at least for surfactants as tested in the present validation study.

#### *Power for distinction*

The following is the background explaining the reason for the adoption by The Working Group of the newly defined “power for distinction” (PFD) in the evaluation of cytotoxicity in an assay. In detecting cytotoxicity, *in vitro* cytotoxicity assays have generally higher sensitivity, on the basis of chemical concentration in test solutions, than the *in vivo* Draize eye irritation test. A typical example is shown by chemical #5 in Fig. 7 and Fig. 10 (less than 0.0002% and 0.1%, respectively). Highly sensitive detection

of cytotoxicity is feasible with any one of the 5 cytotoxicity assays for this purpose when compared to the *in vivo* Draize test. Therefore, the absolute sensitivity of cytotoxicity among the 5 assays is not necessarily important.

Considering two chemicals, the larger the difference in the median toxicity values of the two is important in the distinction of toxicity of these two chemicals. However, if variation of data for each of the two chemicals is also large, the distinction of their toxicities will be difficult. On the contrary, if variation of data is small enough, the assay will be useful for distinction of toxicity of the two chemicals.

#### *Participating laboratories*

Many laboratories claimed that the initially set duration for the cytotoxicity assays was too short for completion of sufficiently repeated assays, even though the deadline for data file submission was postponed for five months. Considering the first step validation study and manpower cost in each assay and since the participants were called on a volunteer basis, we could not insist on them to carry out all the assays under the principle of good laboratory practice (GLP) as recommended by the regulatory organizations. Other reasons for the failure to follow GLP included a variety of laboratory equipment and the insufficiency of experience in GLP in most of the participating laboratories. These may have affected the efficiency of data file acceptance and the variation of ED50 values. If we had selected only the participants who are highly skillful and sufficiently experienced, the above mentioned conclusion could have been slightly different.

We have partly covered this weak point by replicating one of the 6 test chemicals in the 7 doubly-mask-coded samples and by requiring the participants to carry out an assay on a chemical with 3 separate plates including negative- and/or positive-control wells in each plate. The former contributed to detection of the intra-laboratory variation of data (Table 6), and the latter contributed to detect technical stability in each laboratory. Researchers or technicians who at-

tended the technology transfer course were made aware that, when they found an extraordinary difference in the control data in one of the 3 plates, they should repeat the assay until the data fall in a range of relatively small difference than before which can be considered acceptable. Since many laboratories (corresponding to 64% of possibly acceptable candidate data files) repeated assays until they observed a dose-response curve which included 3 or more observed points falling between 20-80% of the maximum cytotoxic effect (Table 4), although still many other laboratories (corresponding to 36% of possibly acceptable candidate data files) could not complete repeating the assays satisfactorily, we thought that most of the participating laboratories trained themselves until they obtained stable results among the 3 plates in an assay. So far, in the data files examined by The Working Group, the differences of observed negative- or positive-control values among the 3 plates in an assay were sufficiently small compared to the differences of those control values among the laboratories (data not shown. See Validation Articles IV - VIII on each assay). After the discussion meeting held on November 30, 1994, The Working Group accepted simple recording errors on reported concentrations of chemicals. At the final box-and-whisker plot analyses, the jump-off log(ED50) values among data files in each assay were excluded and further analyses were concentrated on the log(ED50) data that fell in the hinge-spread. By these strategies, as many errors as possible were thought to be excluded.

#### *Chemicals, cell lines and protocol of the assays*

In the present validation study, the following six test chemicals with diverse Draize scores reported in the literature were selected (see Materials and Methods): 3 non-irritants, 1 irritant, 1 moderate irritant, and 1 severe irritant. This unbalanced selection was reflected in the scores of the Draize test done on the same lot of the chemicals. For ease of further statistical analyses, there should have been a more balanced selection of chemicals including more severe irri-

tants.

Although we did not indicate in the protocol that the test chemicals should be prepared just before use, the 6 test chemicals are all stable at room temperature. In case of unknown chemicals in the assays, timing will be an important factor in the preparation of test chemical solutions.

What is more important is that the cells in preculture should be carefully kept in log phase. Culture of HeLa S3 (SC) cells was stably performed in every laboratory since this cell line was familiar and can be easily handled by almost all participants. We were not, however, able to confirm this point from submitted data files except the LDH-2B and -2C assays. From the preliminary experiments, population doubling times for HeLa S3 (SC) cells and SQ-5 cells were determined to be 20.3 hr and 20.8 hr, respectively. Therefore, in the 48 hr treatment period in the LDH-2B and -2C assays, the number of cells in the negative (untreated) control wells should increase by about 5-fold.

Unexpectedly, growth of the cell lines in control wells was very different among laboratories, resulting in large variability in negative control data in the inter-laboratory assays (data not shown, see Validation Articles IV - VIII on each assay). These variations may have influenced the results of assays.

The Working Group considered that observation of initial cell density and a reference chemical should have been set at a specified concentration showing near 50% effectiveness in each assay to determine whether the assay was valuable from the view points of cell growth control and the effectiveness of a standard toxicant. Furthermore, in the LDH assay, we should have been more careful in determining LDH activity in FBS-containing culture medium. Some lots of FBS increased the background level of LDH in preliminary assays and therefore reduced cytotoxic sensitivity of the assay.

To our surprise, many laboratories could not perform LDH assays (Table 3, Fig. 6). The LDH-1 assay which includes 20 min incubation at room temperature is the shortest one in the assays validated. This assay can be repeated 2 or 3



times in a day. However, presumably because this assay requires the determination of FRLA (the factor of direct effect on LDH activity of the test chemical at each concentration used), participants may have considered it tedious, thus resulting in non submission of so many data files.

The LDH-2A and LDH-2B assays (and therefore the LDH-2C assay coupled with LDH-2A and 2B) should be carried out on the same lot of tests, since calculation of ED50 for the LDH-2B assay in which cells were treated for 48 hrs with chemicals utilizes the total LDH activity of the cells in negative control wells of the LDH-2A assay as the initial cellular LDH activity. This point was not clearly made in the protocol. In the present study, only 6 data file sets were found to have come from assays which have been carried out in the same lot of the test. Many sets of LDH-2B assays used the initial cellular LDH activity derived from a different test for the LDH-2A assay (for details, see Validation Article VI on LDH assay in this issue). The Working Group accepted this utilization since the initial LDH activity per number of cells from different lots of tests gave very small variation of data.

The CV assay includes steps of cellular attachment to the culture surface, cell proliferation, and protein synthesis in the medium containing chemicals over a 3-day incubation (Saotome *et al.*, 1989). The attachment step is not included in other assays. However, this characteristic of the CV assay seemed to not strongly influence the log(ED50) values observed in the present study since the medians of log(ED50) values were not necessarily lower than those observed in the MTT and NR assays with HeLa S3 (SC) cells (Table 7).

The Working Group was aware later that, if cell-bound crystal violet dye was extracted with 70% methanol before the OD590 measurement, as the formazane formed in the MTT assay was also extracted, then the observed values became free from occasional uneven localization of cells in a well. This point will improve variation of the observed data.

NRCE cells, the primary cultured rabbit cornea cells, sometimes ceased to increase in number over the culture period indicated in the in-

struction sheet attached in the commercial kit, probably because these cells enter the so-called "crisis". Therefore in some laboratories, shortage of cells was experienced for extensive repeating of the NR assays with NRCE cells in one kit. The cost was found to be a potential problem in the NR assay using this kit.

Concerning the length of treatment of cells with chemicals, the CF assay is the longest (7 or 13 days) and the LDH-1 assay is the shortest (20 min). Effect of duration of the chemical treatments was markedly reflected in the mean values of log(ED50) for chemicals #1, #7, #2, and #5 (Fig. 8a, b, c, and f, respectively). However, for chemical #3 and #6, this effect was not observed (Fig. 8d and g), presumably because of different mechanism of toxic action.

Cytotoxicity assay with serum-free cultured cells has been described as generally more sensitive than those with cells cultured in serum-containing medium (Torishima *et al.*, 1990; Torishima *et al.*, 1995). In the present assays, this tendency was observed only on 4 out of 6 chemicals tested (not on chemical #4 and #5) when the mean log(ED50) of the NR assay with NRCE cells were compared to those of the NR assay with HeLa S3 (SC) cells (Table 7, Fig. 8e and f). Especially with chemical #6, approximately 100-fold higher sensitivity was observed in the NR assay with NRCE cells than in assays with cells cultured in regular serum-containing medium (Fig. 8g). This reason should be at least partly ascribed to the strong affinity of chemical #6, sodium lauryl sulfate, to serum proteins.

### *Technology transfer*

In the initial stage of this study, The Working Group asked the participants their extent of cell culture experience. However, so far as we observed during the technology transfer, qualitatively speaking, extensive experience in cell culture could not necessarily serve to guarantee that one will obtain smaller variations in assay results. We found a problem soon after the beginning of collection of data that in some laboratories the trainees did not carry out the assays themselves but the technicians under their supervi-

sion. In such cases, it was doubtful whether these persons were as skillful as the representative trainees. There was a considerable number of laboratories whose staffs were not skillful enough to carry out the popular and easy cytotoxicity assays (Table 3 and Fig. 6). The Working Group assumes this was partly because the participants were called on a volunteer basis. It is desirable by some arrangements to adopt a minimum threshold level of skill on cytotoxicity assays to avoid the human errors (Fig. 2) as much as possible in future validation studies.

### *Common rules*

On the common rule [4], all the participating laboratories initially thought that calculation of ED50 values and data analyses would be easy from final data files if they contained 3 observed points between 20% and 80% of the maximum effectiveness on each chemical. Direct reading of ED50 values should have been possible from hand-plotted dose-response curves. However, only 64% of data files could provide such data and so many laboratories could not (Table 4); perhaps time was over-spent and work remained incomplete in repeating assays many times to determine the suitable narrow dose range of each chemical to fit the desired effectiveness. Then, considering the practicability of assays, we have loosened strict application of the common rule [4] and rescued a number of data files by passing through the LAP-JSAAE program as described in Results. However, the data files which contain no observed point between 20% and 80% of maximum effectiveness were discarded since precision of the calculated ED50 value will be impaired.

### *Data collection*

Because of the difference in application softwares and difference of versions of the software used, The Working Group had to invest a major effort to clean up the raw data. On collection of data, this point must be improved, if possible, through on-line systems such as the Internet in future validation studies. On the final

definitive test of assays, The Working Group judged a file as the final definitive one recorded with the latest date. This point should have been clearly indicated in each protocol of assays.

Simple mistakes on data recording were realized by the corresponding laboratories after the discussion meeting held November 30, 1994. In the present validation study, we accepted correction of this type of simple miswriting of concentrations after the discussion meeting. Such a stance rescued 42 files. However, human errors should have been excluded in the validation study as much as possible. For this purpose, several check points had to be set during assays, before the data submission, and during raw data cleaning.

### *Abnormal data*

Wang and Ohno (1995) described that some types of chemicals such as ethylene glycol increases the LDH activity of cells to more than that of control cells in low-dose ranges. Slightly above this low dose range, the chemical showed inhibition of cell growth with a sigmoidal dose-response. We could not judge whether the increase in the LDH activity that indicated cell growth stimulation has actually occurred. Although we could not find typical responses of this type in the present study, such a phenomenon should be considered on the statistical curve fittings in the dose-response of new chemicals in *in vitro* assays.

### *Performance rates*

From the viewpoint of practice, simplicity of the protocol seemed very influential on the performance rate of an assay and ultimately on the variation of log(ED50) values (Table 3, Table 7). The apparently lower performance rates observed in every sub-divided LDH assay compared to other 4 assays (Table 3, Fig. 6) indicated that the LDH assay seemed immature practically to be carried out in a laboratory with average technical expertise.

If performed skillfully (for example, Fig. 7), the LDH-2A, B, and C assays done in one series

of practice were more informative on toxicological characteristics of chemicals than the other 5 assays (Wang *et al.*, 1993; Wang and Ohno, 1995). The rapid cell lysis was detectable by LDH-2A, and after 48-hr incubation, both attached cells and lysed cells were quantifiable in the same culture by LDH-2B and LDH-2C, respectively. However, LDH assays required assessment of direct effect of chemicals on the LDH activity. In these assays, determination of the factor FRLA, for correction of observed LDH activity to original LDH activity at each concentration of the test chemical, forced laboratories to carry out complicated assay procedures. On the other hand, the other 4 assays included no step to correct initially observed data. In this context, on the basis of runs, the NR assay with HeLa S3 (SC) cells seemed the easiest among the 16 assays since it was run at the highest performance rate, 92% (Table 3), while, in the LDH assay, especially LDH-2C with SQ-5 cells, showed the lowest performance rate, 48% (Table 3). To compensate for this weak point, use of automated laboratory processors (robots) is strongly recommended (Wang *et al.*, 1993). Once the assay conditions were set precisely, the robot can be expected to reproduce data in the assay with smaller variations than in the assay performed manually (data not shown).

#### *In vivo Draize test*

Reported toxicities of the test chemicals were confirmed by the *in vivo* Draize eye irritation test (Fig. 10). Unfortunately however, Tween 20 (the chemical to detect intra-laboratory variation that has been chosen under a doubly-masked system) showed no toxicity *in vivo*. It was therefore impossible for us to know the intra-laboratory variation in the *in vivo* Draize test.

The assay methods examined in the present validation study should be correlated with the *in vivo* Draize test. However, since the actual number of chemicals is only 6 in which half are non-irritants, and there being only one severe irritant, estimation of a statistically reliable correlation was difficult. A fairly correlative tendency was observed between the median of log(ED50)

values and the MAS values at 10% chemical solution in the Draize test (Table 7 and Fig. 11).

Chemical #6, sodium lauryl sulfate, has been widely used as a reference chemical in toxicological studies *in vivo* and *in vitro* (Kojima *et al.*, 1995). The MAS of chemical #6, 39.3, in the present *in vivo* Draize test was apparently distinguishable from the MAS of chemical #3, 14.7. At present, however, we were not able to explain why log(ED50) values of chemical #3 and #6, except in the NR assay with serum-free cultured NRCE cells, were not clearly distinguishable *in vitro* as seen in Fig. 11. This implies that the *in vitro* assay is not necessarily able to reflect exactly the *in vivo* Draize test.

Balls *et al.* (1995) described that, after the large scale international validation study on alternatives to the Draize test, none of the nine alternative tests including the NR assay on cell viability met any of the four performance targets (wherein the results of the alternative tests and the modified MAS obtained in the Draize test were examined statistically for a significant relationship between them). Their *in vivo* Draize test data were obtained from the data bank of the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) which stores data from many different Draize tests carried out on different occasions and in different places. Therefore, large variability of the data among the *in vivo* Draize tests carried out in different laboratories must be taken into account, but still Balls *et al.* (1995) described their disappointing results “with the possible exception of predicting the irritancy of surfactants”. The corresponding results for the possible exception were derived from the NR assay on cell viability tested with surfactants. Our present results are consistent with their conclusion on surfactants.

#### *Statistical analyses*

In Fig. 6, open boxes are the files finally accepted for comparison of inter-laboratory variation of log(ED50) values obtained without any difficulty. If required and if possible, we read ED50 values from hand-plotted dose-response curves (see the subsequent issue with Validation

Articles IV - VIII on each assay). However, in this study, we calculated ED50 values by the logistic analysis program LAP-JSAAE. When we tried direct reading of ED50 values from hundreds of hand-plotted dose responses, although most of the ED50 values were very close to those derived from the LAP-JSAAE program, some of the hand-plotted curves contained observed points showing only lower or higher than 50% effectiveness and therefore did not cross over the 50% effectiveness level. Two or more ED50 values were readable from a hand-plotted curve since the curve crossed over the 50% effectiveness level twice or more. The LAP-JSAAE program enabled us to avoid these difficulties. In the coming validation study, a statistical analysis program such as the LAP-JSAAE should be given to each participating laboratory beforehand in order to greatly reduce the number of rejected data files.

#### *Fact database construction*

On the final definitive data files, we have developed a fact data base on a personal computer. We acquired a powerful tool for data dissemination on the network, World Wide Web (WWW). It will be useful for researchers worldwide after we have converted the database to the one on WWW. In WWW, on-demand-analysis will be also feasible on each assay by use of the common browser software.

#### *Conclusion*

From the results shown in the present study, i.e., performance rate, inter-laboratory variation of data represented by mean hinge-spread, and the power for distinction of chemical cytotoxicity, and time needed for the assays, we consider that the CV assay is the most practical and recommendable assay as a part of alternatives to the *in vivo* Draize eye irritation test.

#### **Acknowledgments**

This work was supported in part by a fund from JSAAE and in part by Special Coordination

Funds for Promoting Science and Technology from The Science and Technology Agency of Japan.

#### **References**

- Bagley, D. M., Botham, P. A., Gardner, J. R., Holland, G., Kreiling, R., Lewis, R. W., Stringer, D. A., and Walker, A. P. (1992a) Eye irritation: Reference chemicals data bank. *Toxicol. in Vitro*, 6, 487-491.
- Bagley, D. M., Bruner, L. H., deSilva, O., Cottin, M., O'Brien, K. A. F., Uttley, M., and Walker, A. P. (1992b) An evaluation of five potential alternatives *in vitro* to the rabbit eye irritation test *in vivo*. *Toxicol. in Vitro*, 6, 275-284.
- Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rozenkranz, H., Schmid, B., Spielman, H., Stamatii, A., and Walum, E. (1990) Report and recommendations of the CAAT/ERGATT Workshop on the validation of toxicity test procedures. *ATLA*, 18, 313-337.
- Balls, M., Botham, P. A., Bruner, L. H. and Spielman, H. (1995) The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicol. in Vitro*, 6, 871-929.
- Borenfreund, E. and Puerner, J. A. (1985) Toxicity determined *in vitro* by morphological alterations and neutral red absorption. *Toxicol. Lett.*, 24, 119-124.
- Bruner, L. H., Kain, D. J., Roberts, D. A., and Parker, R. D. (1991) Evaluation of seven *in vitro* alternatives for ocular safety testing. *Fundamental Appl. Toxicol.*, 17, 136-149.
- Catroux, P., Rougier, A., Dossou, K. G., and Cottin, M. (1993) The silicon microphysiometer for testing ocular toxicity *in vitro*. *Toxicol. in Vitro*, 7, 465-469.
- Clemedson, C., McFarlane-Abdulla, E., Andersson, M., Barile, F. A., Calleja, M. C., Chesne, C., Clothier, R., Cottin, M., Curren, R., Daniel-Szolgay, E., Dierickx, P., Ferro, M., Fiskesjo, G., Garza-Ocanas, L., Gomez-Lechon, M. J., Gulden, M., Isomaa, B., Janus, J., Judge, P., Kahru, A., Kemp, R. B., Kerszman, G., Kristen, U., Kunimoto, M., Karenlampi, S., Lavrijsen, K., Lewan, L., Lilius, H., Ohno, T., Persoone, G., Roguet, R., Romert, L., Sawyer, T. W., Seibert, H., Shrivastava, R., Stamatii, A., Tanaka, N., Torres-Alanis, O., Voss, J.-W., Wakuri, S., Walum, E., Wang, X., Zucco, F., and Ekwall, B. (1996a) MEIC evaluation of acute systemic toxicity. Part I. Methodology of 68 *in vitro* toxicity assays used to test the first 30 reference chemicals. *ATLA*, 24, 251-272.
- Clemedson, C., McFarlane-Abdulla, E., Andersson, M., Barile, F. A., Calleja, M. C., Chesne, C., Clothier, R., Cottin, M., Curren, R., Dierickx, P., Ferro, M., Fiskesjo, G., Garza-Ocanas, L., Gomez-Lechon, M. J., Gulden, M., Isomaa, B., Janus, J., Judge, P., Kahru, A., Kemp, R. B., Kerszman, G., Kristen, U., Kunimoto, M., Karenlampi, S., Lavrijsen, K., Lewan, L., Lilius,

- H., Malmsten, A., Ohno, T., Persoone, G., Pettersson, R., Roguet, R., Romert, L., Sandberg, M., Sawyer, T. W., Seibert, H., Shrivastava, R., Sjoström, M., Stamatii, A., Tanaka, N., Torres-Alanis, O., Voss, J.-W., Wakuri, S., Walum, E., Wang, X., Zucco, F., and Ekwall, B. (1996b) MEIC evaluation of acute systemic toxicity. Part II. *In vitro* results from 68 toxicity assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. *ATLA*, 24, 273-311.
- Draize, J., Woodard, G., and Calvery, H. (1944) Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Pharmac. Exp. Ther.*, 82, 377-390.
- Ekwall, B. and Ekwall, K. (1988) Comments on the use of diverse cell systems in toxicity testing. *ATLA*, 15, 193-200.
- Ekwall, B., Bondesson, I., Castell, J. V., Gomez-Lechon, M. J., Hellberg, S., Hogberg, J., Jover, R., Ponsoda, X., Romert, L., Stenberg, K., and Walum, E. (1989) Cytotoxicity evaluation of the first ten MEIC chemicals: Acute lethal toxicity in man predicted by *in vitro* toxicity in five cellular assays and by oral LD50 tests in rodents. *ATLA*, 17, 83-100.
- Ekwall, B., Gomez-Lechon, M. J., Hellberg, S., Bondesson, I., Castell, J. V., Jover, R., Hogberg, J., Ponsoda, X., Romert, L., Stenberg, K., and Walum, E. (1990) Preliminary results from the Scandinavian multicenter evaluation of *in vitro* cytotoxicity (MEIC) *Toxic. in Vitro*, 4, 688-691.
- Gordon, V. C. and Bergman, H. C. (1988) External evaluation of the EYTEX and EYTEX-MPA system in twelve laboratories to determine the correlation of EYTEX results with *in vivo* results. *Progress in In Vitro Toxicology* (ed. by Goldberg, A. M.), *Alternative Methods in Toxicology*, 6, 325-331.
- Gordon, V. C. (1992) The scientific basis of the EYTEX system. *ATLA*, 20, 537-548.
- Griffith, J. F., Nixon, G. A., Bruce, R. D., Reer, P. J., and Bannan, E. A. (1980) Dose-response studies with chemical irritants in the albino rabbit eye as a basis for selecting optimum testing conditions for predicting hazard to the human eye. *Toxicol. Appl. Pharmacol.*, 55, 501-513.
- Guillot, J. P., Gonnet, J. F., and Clement, C. (1982) Evaluation of the ocular-irritation potential of 56 compounds. *Food Chem. Toxicol.*, 20, 573-582.
- Hockley K. and Baxter D. (1986) Use of the 3T3 cell-neutral red uptake assay for irritants as an alternative to the rabbit (Draize) test. *Food Chem. Toxicol.*, 24, 473-475.
- Itagaki, H., Hagino, S., Kobayashi, T., and Umeda, M. (1991) An *in vitro* alternative to Draize eye-irritation test: evaluation of the crystal violet staining method. *Toxic. in Vitro* 5, 139-143.
- Kojima, H., Hanamura, A., Sato, A. and Konishi, H. (1995) Comparison of cytotoxicity of surfactants on a range of mammalian cells cultured under various conditions. *Altern. Animal Test. Experiment.*, 3, 111-120.
- Marinovitch, M., Tragni, E., Corsini, A., and Galli, C. L. (1990) Quantification of *in vitro* cytotoxicity of surfactants: Correlation with their eye irritation potential. *J. Toxicol.-Cut. & Ocular Toxicol.*, 9, 169-178.
- Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, 65, 55-63.
- 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.
- Ohno, Y., Kaneko, T., Kobayashi, T., Inoue, T., Kuroiwa, Y., Yoshida, T., Momma, J., Hayashi, M., Akiyama, J., Atsumi, T., Chiba, K., Endo, T., Fujii, A., Kakishima, H., Kojima, H., Masamoto, K., Masuda, M., Matsukawa, S., Ohkoshi, K., Okada, J., Sakamoto, K., Takano, K., and Takanaka, A. (1994) First-phase validation of the *in vitro* eye irritation tests for cosmetic ingredients. *In Vitro Toxicol.*, 7, 89-94.
- Ohno, Y., Kaneko, T., Kobayashi, T., Inoue, T., Kuroiwa, Y., Yoshida, T., Momma, J., Hayashi, M., Akiyama, J., Atsumi, T., Chiba, K., Endo, T., Fujii, A., Kakishima, H., Kojima, H., Masamoto, K., Masuda, M., Matsukawa, S., Ohkoshi, K., Okada, J., Sakamoto, K., Takano, K., Suzuki, T., and Takanaka, A. (1995) First-phase inter-laboratory validation of the *in vitro* eye irritation tests for cosmetic ingredients: (1) Overview, organization and results of the validation study. *Altern. Animal Test. Experiment.*, 3, 123-136.
- Rachui, S. R., Robertson, W. D., Duke, M. A., and Heinze, J. (1994) Predicting the ocular irritation potential of surfactants using the *in vitro* SKIN2 Model ZK1200. *J. Toxicol.-Cut. & Ocular Toxicol.*, 13, 215-220.
- Saotome, K., Morita, H., and Umeda, M. (1989) Cytotoxicity test with simplified crystal-violet staining method using microtitre plates and its application to injection drugs. *Toxicol. in Vitro*, 3, 317-321.
- Sasaki, K., Tanaka, N., Watanabe, M., and Yamada, M. (1991) Comparison of cytotoxic effects of chemicals in different cell types. *Toxicol. in Vitro*, 5, 403-406.
- Sasaki, T., Kawai, K., Saijo, K., and Ohno, T. (1992) Detergent cytotoxicity: simplified assay of cytolysis by measuring LDH activity. *Toxicol. in Vitro* 6, 451-457.
- Silverman, J. and Pennisi, S. (1987) Evaluation of Tetrahymena thermophila as an *in vitro* alternative to ocular irritation studies in rabbits. *J. Toxicol.-Cut. & Ocular Toxicol.*, 6, 33-42.
- Spielmann, H., Gerner, I., Kalwei, S., Moog, R., Wirnsberger, T., Krauser, K., Kreiling, R., Kreuzer, H., Lupke, N. -P., Miltenburger, H. G., Muller, N., Murmann, P., Pape, W., Siegemund, B., Spengler, J., Steiling, W., and Wiebel, F. (1991) Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. *Toxicol. in Vitro*, 5, 539-542.
- Torishima, H., Arakawa, H., Matsui, S., and Watanabe, M. (1990) Application of normal human epidermal keratinocytes in serum-free medium as an alternative

- to the Draize ocular irritating test. *Altern. Animal Test. Experiment.*, 1, 20-26.
- Torishima, H., Yamamoto, R., and Watanabe, M. (1995) Neutral red assay using normal rabbit corneal epithelial cells grown in serum-free medium as an alternative to the Draize irritation test. *Altern. Animal Test. Experiment.*, 3, 29-36.
- Tukey, J. W. (1977) *Exploratory data analysis*. Addison-Wesley, Reading, Massachusetts, pp. 27-56.
- Wang, X., Sasaki, T., Matsudo, T., Saijo, K., and Ohno, T. (1993) Correlation of *in vitro* toxicities of MEIC chemicals determined by lactate dehydrogenase release assay with *in vivo* toxicities to animals and human. *Altern. Animal Test. Experiment.*, 2, 115-126.
- Wang, X. and Ohno, T. (1995) Typing of MEIC chemicals according to their toxicokinetic modes of action by lactate dehydrogenase-release assay. *In Vitro Toxicol.*, 8, 55-63.
- Watanabe, M., Watanabe, K., Suzuki, K., Nikaido, O., Ishii, I., Konishi, H., Tanaka, N., and Sugahara, T. (1989) Use of primary rabbit cornea cells to replace the Draize rabbit eye irritancy test. *Toxicol. in Vitro*, 3, 329-334.