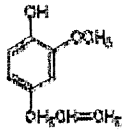
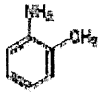
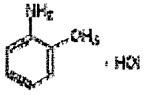
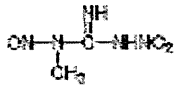
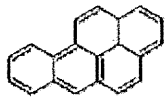
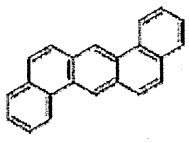
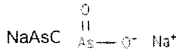
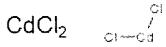
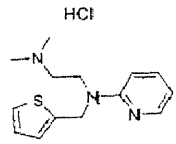
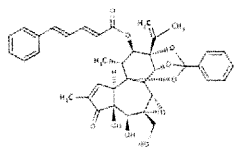
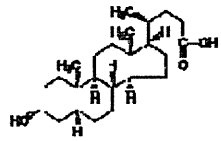
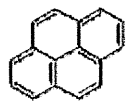
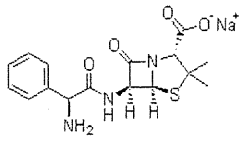
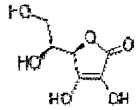
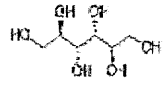
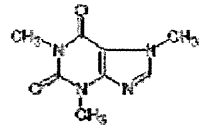
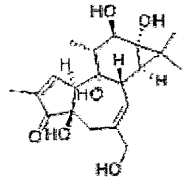
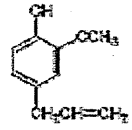
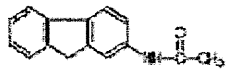
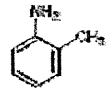
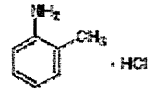
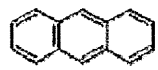
| # | Chemical Name | M.W. | Chemical structure | CAS | IARC [as May 2010] | Appear-ance | Water solublity | Solubility (Other solvents) | Vapor pressure (Volatile) | Chemical class |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | Eugenol | 164 |  | 97-53-0 | 3 (NC) | Liquid | 2.463X10+3 mg/L at 25C <poor> | Miscible with alcohol, chloroform, ether | 2.26X10-2 mm Hg at 25C <Maybe yes> | Phenylpropanoid cmpd |
| Additional information: other chemicals which were used in phase I validation study | | | | | | | | | | |
| | 2-Acetylaminofluorene | 223 |  | 53-96-3 | Not listed (C) | Solid | 144 mg/L at 25C (est) <very poor> | Sol in ether, acetic acid | (Maybe No) | Aromatic amine; Metaboloc activation |
| | o-Toluidine | 107 |  | 95-53-4 | 1 (C) | Liquid | Slightly sol; 1.66X10+4 mg/L at 25C <moderate (close to poor)> | Sol in alcohol, ether; | 2.60X10-1 mm Hg at 25C; 0.32 torr; Volatile with steam; Volatile organic cmpd (VOC) (Yes) | Aromatic amine |
| | o-Toluidine hydrochloride <as reference> | 144 |  | 636-21-5 | 1 (as o-Toluidine) (C) | Solid | Very sol <very good> | Sol in alcohol; insol in ether, benzene | as VOC (maybe Yes) | Aromatic amine |
| | Anthracene | 178 |  | 120-12-7 | 3 (NC) | Solid | 0.04-1.3 mg/L at 25C <very poor> | 1 gm dissolves in 67 mL absolute alcohol, 70 mL methanol, 62 mL benzene, 85 mL chloroform | 2.67X10-6 mm Hg at 25C (eat) (No) | Polycyclic aromatic hydrocarbon |

TP: Tumore promoter (Tanaka et al.,AATEX, 14, 831-848, 2009); C: Carcinogen; NC: Non-carcinogen
Chemical property data from HSDB except for methapyrilene HCL and mezerein.

Water solubility:
very good >1000 g/L; good 100-1000 g/L; moderate 10-100 g/L; poor 1-10 g/L; very poor <1 g/L

Table 2. Genotoxicity data on 16 coded test chemicals for Bhas 42 CTA with 96-well method in phase II validation study

| # | Chemical Name [Carcinogenicity]* | M.W. | Chemical structure | CAS | Bhas 42 [a] (6-well method) | | Genotoxicity Data [a] | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Initiation | Promotion | Ames | MLA | in vitro CA | in vivo MN/CA |
| 1 | N-Methyl-N´-nitro-N-nitrosoguanidine (MNNG) [C] | 147 |  | 70-25-7 | + | - | + | + | + | + |
| 2 | Benzo[a]pyrene [C] | 252 |  | 50-32-8 | + | - | + | + | + | + |
| 3 | Dibenz[a,h]anthracene [C] | 278 |  | 53-70-3 | + | - | + | + | + | + b) |
| 4 | Sodium arsenite [C] | 130 |  | 7784-46-5 | ± | + | - | + | + | + |
| 5 | Cadmium chloride [C] | 183 |  | 10108-64-2 | - | + | - | + | + | + c) |
| 6 | Methapyrilene hydrochloride [TP] | 298 |  | 135-23-9 | - | + | - | - | + | No data |
| 7 | Mezerein [TP] | 655 |  | 34807-41-5 | - | + | - | No data | No data | No data |
| 8 | Lithocholic acid [TP] | 377 |  | 434-13-9 | - | + | - | + | - | No data |
| 9 | Pyrene [NC] | 202 |  | 129-00-0 | + d) | + d) | + e) | + e) | +/- f) | - f) |
| 10 | epsilon-Caprolactam [NC] | 113 |  | 105-60-2 | - | - | - | - | - | - |
| 11 | Ampicillin sodium salt [NC] | 371 |  | 69-52-3 | - | - | - | - | - | - c) |
| 12 | L-Ascorbic acid [NC] | 176 |  | 50-81-7 | - | - | - | No data | - | + |

| # | Chemical Name [Carcinogenicity]* | M.W. | Chemical structure | CAS | Bhas 42 [a] (6-well method) | | Genotoxicity Data [a] | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Initiation | Promotion | Ames | MLA | in vitro CA | in vivo MN/CA |
| 13 | D-Mannitol [NC] | 182 |  | 69-65-8 | - | - | - | - | - | - |
| 14 | Caffeine [NC] | 194 |  | 58-08-2 | - | - | - | - g) | + | +/- |
| 15 | Phorbol [NC] | 364 |  | 17673-25-5 | - d) | - d) | No data | No data | - h) | No data |
| 16 | Eugenol [NC] | 164 |  | 97-53-0 | - | - | - | + | + | - |
| Additional information: other chemicals which were used in phase I validation study | | | | | | | | | | |
| | 2-Acetylaminofluorene [C} | 223 |  | 53-96-3 | + | + | + | + | + | + |
| | o-Toluidine [C] | 107 |  | 95-53-4 | - | + | + | +/- | + | +/- |
| | <o-Toluidine hydrochloride [C]> | <144> |  | <636-21-5> | | | | | | |
| | Anthracene [NC} | 178 |  | 120-12-7 | - | - | +/- | + | - | No data |

*: C, Carcinogesn; NC, Non-carcinogen; TP, Tumore promoter


a) Sakai et al., Mutation Res., 702, 100-122, 2010. (Some exceptions in genotoxicity data)
b) Morita et al., Mutation Res., 389, 3-122, 1997.
c) Kirkland et al., Mutation Res., 653, 99-108, 2008.
d) Tanaka et al., AATEX, 14, 831-848, 2009.
e) Gold and Zeiger, Handbook of carcinogenic potency and genotoxicity database, CRC press, Boca Raton, 1997.
f) 環境省、化学物質の環境リスク初期評価、第7巻、平成21年3月
g) Brambilla and Martelli, Mutation Res., 681, 209-229, 2009.
h) Ishidate et al., Mutation Res., 195, 151-213, 1988.

別紙４

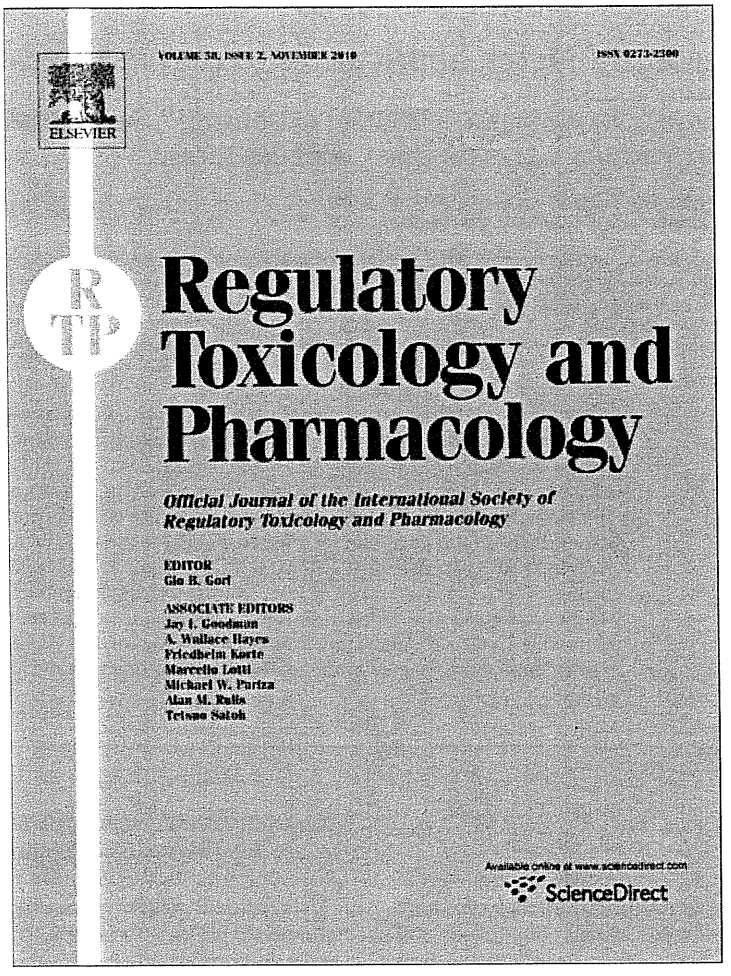
研究成果の刊行に関する一覧表

雑誌

| 発表者氏名 | 論文タイトル名 | 発表誌名 | 巻号 | ページ | 出版年 |
|---|---|---|---|---|---|
| Hasegawa R, Hirata-Koizumi M, Dourson ML, Parker A, Sweeney LM, Nishikawa A, Yoshida M, Ono A, Hirose A. | Proposal of new uncertainty factor application to derive tolerable daily intake. | Regul Toxicol Pharmacol | 58(2) | 237-42 | 2010 |
| Hajime Kojima | Commentary to the Discussion on Topic 3, "In Vitro Test Approaches with Better Predictivity" at the 5th International Workshop on Genotoxicity Testing (IWGT) | Genes and Environment | Vol. 32 No. 2 | 40-42 | 2010 |
| Hajime Kojima, Masahiro Takeyoshi, Takashi Sozu, Takumi Awogi, Kazunori Arima, Kenji Idehara, Yoshiaki Ikarashi, Yukiko Kanazawa, Eiji Maki, Takashi Omori, Atsuko Yuasa and Isao Yoshimura | Inter-laboratory validation of the modified murine local lymph node assay based on 5-bromo-2'-deoxyuridine incorporation | Journal of Applied Toxicology | Vol. 31 | 63-74 | 2010 |
| 小島 肇夫 | 動物実験の３Ｒにおける国内外の動向 | ドージンニュース | No. 138 | 1-9 | 2011 |
| Aoi Kimura, Naohiko Torigoe, Atsuro Miyata, and Masamitsu Honma | Validation of a simple In Vitro comet assay method using CHL cells | Genes and Environment | Vol.32 No.3 | 61-65 | 2010 |
| Yasuyuki Saitoh, Yoshikazu Harata, Fukutaro Mizuhashi, Madoka Nakajima and Nobuhiko Miwa | Biological safety of neutral-pH hydrogen- enriched electrolyzed water upon mutagenicity, genotoxicity and subchronic oral toxicity | Toxicol and Industrial Health | Vol. 26(4) | 203-216 | 2010 |

| | | | | | |
|---|---|---|---|---|---|
| Kenichi Masumura, Yasuteru Sakamoto, Megumi Ikeda, Yasuo Asami, Tetsuya Tsukamoto, Hironobu Ikehata, Yuichi Kuroiwa, Takashi Umemura, Akiyoshi Nishikawa, Masae Tatematsu, Tetsuya Ono, and Takehiko Nohmi | Antigenotoxic effects of p53 on spontaneous and UVB-induced deletions in the epidermis of gpt delta transgenic mice | Environ. Mol. Mutagen. | 52 | 244-252 | 2011 |
| Alexander Sheh, Chung wei Lee, Kenichi Masumura, Barry H. Rickman, Takehiko Nohmi, Gerald N. Wogan, James G. Fox, and David B. Schauer | Mutagenic potency of Helicobacter pylori in the gastric mucosa of mice is determined by sex and duration of infection | Proc. Natl. Acad. Sci. | Vol. 107 no. 34 | 15217-15222 | 2010 |
| Naomi Toyoda-Hokaiwado, Tomoki Inoue, Kenichi Masumura, Hiroyuki Hayashi, Yuji Kawamura, Yasushi Kurata, Makiko Takamune, Masami Yamada, Hisakazu Sanada, Takashi Umemura, Akiyoshi Nishikawa, and Takehiko Nohmi | Integration of *in vivo* genotoxicity and short-term carcinogenicity assays using F344 *gpt* delta transgenic rats: *in vivo* mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene structural isomers | Toxicol. Sci. | 114(1) | 71-78 | 2010 |
| Takeshi Morita, James T. MacGregor and Makoto Hayashi | Micronucleus assays in rodent tissues other than bone marrow | Mutagenesis | Vol. 26 | 223-230 | 2011 |

| | | | | | |
|---|---|---|---|---|---|
| Sheila Galloway, Elisabeth Lorge[2], Marilyn J. Aardema, David Eastmond, Mick Fellows, Bob Heflich, David Kirkland, Dan D. Levy, Anthony Lynch, Daniel Marzin, Takeshi Morita, Maik Schuler, Günter Speit | Workshop summary: Top concentration for in vitro mammalian cell genotoxicity assays; and Report from working group on toxicity measures and top concentration for in vitro cytogenetics assays (chromosome aberrations and micronucleus) | Mutation Res. | In press | | 2011 |



Regulatory Toxicology and Pharmacology 58 (2010) 237–242

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

# Proposal of new uncertainty factor application to derive tolerable daily intake

Ryuichi Hasegawa [a], Mutsuko Hirata-Koizumi [a], Michael L. Dourson [b], Ann Parker [b], Lisa M. Sweeney [b], Akiyoshi Nishikawa [a], Midori Yoshida [a], Atsushi Ono [a], Akihiko Hirose [a,*]

[a] National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
[b] Toxicology Excellence for Risk Assessment, 2300 Montana Avenue, Suite 409, OH 45211, USA

ARTICLE INFO

*Article history:*
Received 2 February 2010
Available online 16 June 2010

*Keywords:*
Chemical risk assessment
Uncertainty factor
Probabilistic approach
Subdivision of UF

ABSTRACT

We propose new uncertainty factors (UFs) and a new subdivision of default factors in chemical risk assessment using a probabilistic approach based on the latest applicable information. Rounded values of 150 for mice, 100 for hamsters and rats, and 40 for rabbits, monkeys and dogs for inter- and intra-species differences ($UF_{AH}$) were derived from the probabilistic combination of two log-normal distributions. Further calculation of additional UFs when chronic data ($UF_S$) or NOAEL ($UF_L$) are lacking was conducted using available log-normal distribution information. The alternative $UF_S$ and $UF_L$ values of 4 are considered to be appropriate for both cases where data are lacking. The default contributions of inter-species difference ($UF_A$) and intra-species difference ($UF_H$) to the $UF_{AH}$ of 100 for hamsters and rats as an example are considered to be 25 and 4, respectively. The $UF_A$ of 25 was subdivided into $25^{0.6}$ (i.e., 7.0) for pharmacokinetics (PK) ($UF_{A,PK}$) and $25^{0.4}$ (i.e., 3.6) for pharmacodynamics (PD) ($UF_{A,PD}$), and the $UF_H$ of 4 was evenly subdivided into $4^{0.5}$ (i.e., 2) ($UF_{H,PK}$ and $UF_{H,PD}$), to account for chemical-specific difference data between humans and laboratory animals for PK and/or PD. These default UFs, which come from actual experimental data, may be more appropriate than previous default UFs to derive tolerable daily intake values.



## 1. Introduction

Principle uncertainty factors (UFs) consisting of inter-species differences (or extrapolation from laboratory animals to humans, referred to as "$UF_A$") and intra-species differences (human variability, referred to as "$UF_H$") have commonly been used when extrapolating from animal experimental data to human risk values in chemical risk assessment. The current combined default UF of 100 ($10_A \times 10_H$) for extrapolation from animal data was introduced in the US in 1954 (Lehman and Fitzhugh, 1954) for food contaminants with a rationale for its suitability for environmental contaminants provided by Dourson and Stara (1983) years later. The physical size of laboratory animals is variable, with animals as small as mice to larger animals like dogs. In some cases the size difference results in more than a 500-fold difference in body weight indicating that some type of variable adjustment might be needed, rather than just a 10-fold factor.

Body surface area correction, (human body weight/animal body weight)$^{1/3}$ was the first data supported size adjustment (Freireich et al., 1966). It has been applied to cancer endpoints in US Environmental Protection Agency (US EPA) assessments and was also used in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) residual solvent guidelines (Connelly et al., 1997). Recently, allometric scaling according to caloric demand or metabolic size, (human body weight/animal body weight)$^{1/4}$ was introduced as a more appropriate adjustment (Schneider et al., 2004) and is currently used by US EPA in cancer risk assessment (US EPA, 1992, 2005a). Size adjustment might be more appropriately based on allometric scaling as discussed by Falk-Filipsson et al. (2007). However, the use of allometric scaling in non-cancer endpoints remains untested by US EPA and other organizational assessments. The caloric demand adjustment factor for a mouse (0.030 kg) or a dog (16 kg) compared to a human (70 kg) based on body weight is 7 or 1.4, respectively, which is significantly lower than the default of 10. However, Schneider et al. (2004) demonstrated that caloric demand scaling was effective for predicting median differences between humans and animals on the basis of body weight in maximal tolerated dose (MTD) ratios of anti-cancer drugs, and also calculated the combined geometric standard deviation (GSD) of the empirical distribution.

Useful experimental data are quite limited for human intra-species differences, specifically variability between different ages (Dourson and Stara, 1983; Dourson et al., 1996, 2002). However, some insights can be gained from experimental animal work. For

* Corresponding author. Address: Division of Risk Assessment, Biological Safety Research Center, National Institute of Health Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Fax: +81 3 3700 1408.
*E-mail address:* hirose@nihs.go.jp (A. Hirose).

0273-2300/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2010.06.006

example, a recent comparative investigation of no observed adverse effect levels (NOAELs) in repeat-dose studies of newborn and young rats for 18 chemicals was conducted (Hasegawa et al., 2007). The data provided the median and distribution of NOAEL ratios.

The default UF of 100 for inter- and intra-species differences is typically applied by multiplication of $10_A$ and $10_H$. However, the default UF of 100 is not always appropriate to use. For example, multiplication of two log-normal distributions for inter- and intra-species differences also produces a log-normal distribution, and simple multiplication such as $10 \times 10$ causes overestimation if both individual values are in the 95th percentile. Kodell and Gaylor (1999) recommended standard statistical techniques that could be used to estimate the upper tolerance limits on the distribution of sums which can also be used for other UFs (e.g., the ratio of subchronic to chronic NOAELs). Swartout et al. (1998) also addressed this problem and gave hypothetical examples of UF combinations.

Another method for division of the default UF of 100 ($10_A \times 10_H$) for inter- and intra-species differences was proposed by Renwick (1993). He proposed a subdivision of these UFs into two parts, pharmacokinetics (PK) and pharmacodynamics (PD). Based on his analysis of experimental data and physiological parameters between animals and humans, the contribution ratios of PK and PD are 60:40 for inter-species differences and 50:50 for intra-species differences, leading to $10^{0.6}$ (4.0) $\times$ $10^{0.4}$ (2.5) and $10^{0.5}$ ($\sim$3.2) $\times$ $10^{0.5}$ ($\sim$3.2), respectively (IPCS, 1994). When chemical-specific data for the differences between animals and humans for PK and/or PD are available, the data should be used to develop chemical-specific adjustment factors instead of the default PK/PD factors (WHO, 2005). However, the default subdivision factors should be re-estimated if animal size-specific UFs are adopted as inter-species differences.

In this article, we propose new default UFs by a probabilistic approach using appropriate log-normal distribution data, taking animal size into consideration. We also propose development of new default values according to animal size for the subdivision of inter- and intra-species differences.

## 2. Data for each uncertainty

### 2.1. Inter-species difference data

Eight publications featuring chemical toxicity comparisons between humans and laboratory animals for anti-cancer drug toxicity were located. The first study by Freireich et al. (1966) showed MTD differences between humans and five animal species (mice, hamsters, rats, monkeys and dogs) in the analysis of 18 drugs. Recently, Schneider et al. (2004) extracted correlated human and animal data sets for 63 anti-cancer drugs from six additional publications (Goldsmith et al., 1975; Schein et al., 1979; Travis and White, 1988; Rozencweig et al., 1981; Grieshaber and Marsoni, 1986; Paxton et al., 1990) to demonstrate that caloric demand scaling was a suitable adjustment factor for the differences of inter-species median MTDs. Schneider et al. (2004) also derived a GSD of 3.23 from the combined distribution of all MTD ratios for humans versus the five animal species stated above.

Alternatively, inter-species differences in susceptibility could be derived based on the differences in NOAELs rather than MTDs. Schneider et al. (2004) also analyzed inter-species differences for pesticide NOAELs between mice/rats, rats/dogs and mice/dogs, providing further support to the caloric demand adjustment. Therefore, the median and GSD derived from MTD ratios of anti-cancer drugs might be equivalent to those based on NOAEL ratios between humans and animals. No other publications featuring an estimated direct comparison of chemical toxicity between humans and animals were identified.

### 2.2. Intra-species difference data

The NOAEL ratio of a sensitive subpopulation compared to that of the general population is a source of uncertainty for intra-species differences in risk assessment (Dourson et al., 2002). Occasionally the sensitive subpopulations are directly addressed in the risk assessment. For example, the Reference Dose (RfD) for nitrate on the Integrated Risk Information System (IRIS) used methemoglobinemia in children as the critical effect, therefore an intra-species UF may not be needed (US EPA, 2009). However, in most cases, the sensitive subpopulations are only considered protected with the use of an intra-species UF. Generally, infants, pregnant women, the elderly and other specified groups are considered high-susceptibility groups, although exceptions are not uncommon. For example, Tylenol overdose is more of a problem in adults than in children because the toxic metabolite is more readily formed in adults. Effects during pregnancy and gestation are considered to be adequately evaluated in the reproductive/developmental toxicity studies, while lifetime toxicity studies cover the potential for effects to the elderly. For a well-tested chemical, the only remaining sensitive subpopulation to be protected by an intra-species UF are infants. Currently, there are no experimental animal test guidelines intended for direct exposure of neonatal animals to chemicals. Other specified groups may include patients exhibiting hepatic or renal dysfunction and persons with a specific genetic background. These subpopulations need specific risk management and should not be the target population for a chemical risk assessment for public health because it is possible that their susceptibility to specific chemicals may be unexpectedly high owing to significantly reduced metabolism or excretion of toxic substances.

The comparative data between human adults and children/infants was assessed by many scientists. Glaubiger et al. (1981) compared MTDs in patients for 17 anti-cancer drugs demonstrating that children's MTDs were 50% higher than those of adults, indicating that children were less sensitive. Calabrese (1985) investigated the variation in physiological response to exogenous stress in humans, and judged that 80–95% of the variation in a human group for a given agent was less than 10-fold. Hattis et al. (1987) analyzed 101 PK parameter data sets for 49 substances (mostly medications) and showed that 96% of the human variation was also less than 10-fold. Ginsberg et al. (2002) compared PK in adults and children using a database of approximately 45 medications, and showed that the half-lives of medications for 1-week to 2-month old infants were twice as long as the half-lives in adults. Hattis et al. (2003) also showed significantly longer half-lives of medications in infants and children compared to adults.

Animal data has also been reviewed. Dourson and Stara (1983) analyzed acute rat toxicity data for 490 substances reported by Weil (1972). They concluded that the $LD_{50}$/non-lethal dose ratio for 92% of the chemical substances would be less than 10. In a meeting abstract, Sheehan and Gaylor (1990) stated that the $LD_{50}$ of 238 substances in adult rats was about 2.6 times higher than the $LD_{50}$ in newborn rat pups, and the $LD_{50}/LD_{50}$ ratio for 86% of substances was less than 10. Calabrese (2001) showed that the $LD_{50}$ in younger animals was within a 10-fold range of older animals for 86.3% of 313 substances. Charnley and Putzrath (2001) examined the influence of age on carcinogenesis caused by chemicals, but were unable to reach a clear conclusion. Similarly, the US EPA considered the effect of age in their most recent guidelines for carcinogen risk assessment. They estimated the geometric mean ratio of early-life to adult cancer potencies was 10.4 based on repeated and lifetime exposure data in the available scientific literature for six chemicals acting through a mutagenic mode of action

(US EPA, 2005b; Barton et al., 2005). As for chemicals causing cancer through other modes of action, the ratio was 3.4 for lifetime exposure (5 chemicals) and 2.2 for repeated exposure (6 chemicals).

The quantitative human and experimental animal data for severe endpoints and kinetic parameters are useful. However, a study design similar to the repeat-dose exposure studies used in risk analysis would be ideal to derive an intra-species UF. The UF is applied to the NOAEL derived from the results of repeated dose toxicity studies, therefore a comparative analysis of NOAELs from repeat-dose toxicity studies of newborn and young rats for 18 chemicals was considered more appropriate (Hasegawa et al., 2007). In this study, Hasegawa et al. (2007) strictly compared the NOAEL ratios for newborn and young rats in a repeat-dose study. The NOAEL ratios were log-normally distributed. The ratio median was 3, and 5 was equivalent to 94.4% of the whole data set, from which the GSD can be calculated (see below).

### 2.3. Data for supplemental uncertainty factors[1]

The appropriate adjustment from short-term NOAEL to lifetime NOAEL for risk assessment was evaluated using 33 data sets of subchronic (3 months)/chronic (2 years) NOAELs in rats and mice reported by Weil and McCollister (1963) and 68 additional data sets from analyses of published reports or papers that we previously summarized (Hasegawa, 1991). Comparison of NOAELs from published 3-month and 2-year repeated dose toxicity studies, unpublished data). The combined data sets yielded a median of 1.7 with a GSD of 3.30. If only a LOAEL was identified, the median LOAEL/NOAEL ratio of 3.5 with a GSD of 1.82 from Abdel-Rahman and Kadry (1995) from other chemicals can be adapted as an UF for this area, with the usual upper bound value of 10. However, it is recognized that the application of the benchmark dose approach is usually more appropriate in cases where only a LOAEL is available, and as such this UF is not used as frequently.

## 3. Calculation of new uncertainty factors based on experimental data by probabilistic approach, an example of rats

The distribution of both inter- and intra-species differences is log-normal because each component consists of the NOAEL ratios for two groups. If the default values of 10 are used, simple multiplication of 10 by 10, resulting in 100, leads to overestimation for the 95th percentile of the combined distribution, more appropriately it should be 51, as shown by Monte Carlo simulation (Swartout et al., 1998). Generically, the *N*th percentile of a log-normal distribution can be expressed as *N*th percentile = Exp [LN (median) + $\alpha_n \times$ LN (GSD)]. For the 95th percentile, $\alpha_n = 1.645$. The equation for the combination of two log-normal A and B distributions can be shown as follows: 95th percentile of (A × B) = Exp [LN ($\text{median}_A$) + LN ($\text{median}_B$) + $1.645 \times ((\text{LN}(\text{GSD}_A))^2 + (\text{LN}(\text{GSD}_B))^2)^{0.5}$] (Kodell and Gaylor, 1999).

Inter-species differences were calculated using an analytical method presented by Schneider et al. (2004). A median of 4 was reported for the caloric demand adjustment, rounded from $3.76 = (70/0.35)^{1/4}$ (70 kg human body weight and 0.35 kg that of rats). A GSD of 3.23 was adopted from a combined distribution of MTD ratio for humans versus the 5 animal species previously described. For the 95th percentile, $\alpha_n = 1.645$.

LN (95th percentile) = LN (4) + 1.645 × LN (3.23)

95th percentile = UF (95%) = Exp [1.39 + 1.645 × 1.17] = 27.5.

Intra-species differences were calculated using rat young/newborn NOAEL ratios in repeat-dose toxicity studies (Hasegawa et al., 2007). The median was 3 for 18 data sets and 5 was equivalent to 94.4% of all the data sets. For the 94.4th percentile, $\alpha_n = 1.590$.

LN (5 as 94.4th percentile) = LN (3) + 1.590 × LN (GSD)
Rearranging,
LN (GSD) = (1.61 − 1.10)/1.590 = 0.321
Therefore,
GSD = Exp [0.321] = 1.38
95th percentile = UF (95%) = Exp [1.10 + 1.645 × 0.321] = 5.09.

From the above data for inter- and intra-species differences, the combined $UF_{AH}$ was calculated as follows:

$\text{LN}(4) + \text{LN}(3) + 1.645 \times ((\text{LN}(3.23))^2 + (\text{LN}(1.38))^2)^{0.5} = 1.39 + 1.10 + 1.645 \times (1.17^2 + 0.321^2)^{0.5} = 4.48$
Exp[4.48] = 88.7.

For adjustment of short-term NOAEL to lifetime NOAEL, all 101 data sets of subchronic NOAEL/chronic NOAEL were used. The median was 1.7 with 10 equivalent to 93.1% of all the data sets. For the 93.1th percentile, $\alpha_n = 1.483$.

LN (10 as 93.1th percentile) = LN (1.7) + 1.483 × LN (GSD)
Rearranging,
LN (GSD) = (2.30 − 0.531)/1.483 = 1.20
Therefore,
GSD = Exp [1.20] = 3.30
95th percentile = UF (95%) = Exp [0.531 + 1.645 × 1.20] = 12.1.

From the above UF calculations, the combined $UF_{AHS}$ was calculated as follows:

$1.39 + 1.10 + 0.531 + 1.645 \times (1.17^2 + 0.321^2 + 1.20^2)^{0.5} = 5.82$
Exp[5.82] = 337.

If a benchmark dose approach cannot be applied, an additional UF should be applied when using LOAEL data. The LOAEL/NOAEL ratio for 24 chemicals was reported by Abdel-Rahman and Kadry (1995). The median was 3.5 and 10 was equivalent to 96% of the whole data. For the 96.0th percentile, $\alpha_n = 1.751$.

LN (10 as 96.0th percentile) = LN (3.5) + 1.751 × LN (GSD)
Rearranging,
LN (GSD) = (2.30 − 1.25)/1.751 = 0.600
Therefore,
GSD = Exp [0.600] = 1.82
95th percentile = UF (95%) = Exp [1.25 + 1.645 × 0.600] = 9.39.

From the above UF calculations, the combined $UF_{AHSL}$ was calculated as follows:

$1.39 + 1.10 + 0.531 + 1.25 + 1.645 \times (1.17^2 + 0.321^2 + 1.20^2 + 0.6^2)^{0.5} = 7.24$
Exp [7.24] = 1400.

## 4. Summary of combined uncertainty factors for six animal species by probabilistic approach

All fundamental values for the median, GSD and UF (95%) are shown in Table 1. The median for inter-species differences was derived using caloric demand adjustment from the standard human and animal body weights and rounded to a simple value. The

[1] The uncertainty factor used by several organizations for missing certain studies in the database (e.g., Dourson et al., 1992, 2002) was not considered here at this time, as it is being studied for applicability in Japan.

**Table 1**
Median, GSD and UF (95%) of inter-species differences for 6 animal species and other uncertainties.

| | Median | GSD | UF (95%) |
|---|---|---|---|
| Inter-species differences (caloric demand)[a] | | | |
| Mice to humans | 6.95 → 7 | | 48.2 |
| Hamsters to humans | 4.86 → 5 | | 34.4 |
| Rats to humans | 3.76 → 4 | 3.23 | 27.5 |
| Rabbits to humans | 2.04 → 2 | | 13.8 |
| Monkeys to humans | 1.77 → 1.8 | | 12.4 |
| Dogs to humans | 1.44 → 1.4 | | 9.63 |
| Intra-species differences[b] | 3.0 | 1.38 | 5.09 |
| Subchronic to chronic[b] | 1.7 | 3.30 | 12.1 |
| LOAEL to NOAEL[b] | 3.5 | 1.82 | 9.39 |

[a] Use of caloric demand and distribution from MTD ratios of 63 anti-cancer drugs between humans and 5 animals given by Schneider et al. (2004). Medians were calculated as caloric demand adjustment ((human body weight/animal body weight)$^{1/4}$) on the bases of body weight: humans = 70 kg, mice = 0.03 kg, hamsters = 0.125 kg, rats = 0.35 kg, rabbits = 4 kg, monkeys = 7 kg and dogs = 16 kg.
[b] Calculation details are shown in the previous section.

GSD for inter-species differences was obtained by combining the distribution of all the MTD data sets. This distribution may contain some additional, but unquantifiable, conservatism since humans are more heterogeneous than laboratory animals; thereby inflating the upper limits. The 95th percentile of UFs for six laboratory animal species ranged from approximately 10–50, a 5-fold difference.

All possible cases of UFs for six laboratory animal species were calculated by a probabilistic approach (Table 2) using the values from Table 1. The $UF_{AH}$ for each animal is calculated by combining inter- and intra-species differences. We propose a rounded $UF_{AH}$ of 150 for mice, 100 for hamsters and rats, and 40 for rabbits, monkeys and dogs. Additional single UFs for either subchronic to chronic ($UF_S$) or LOAEL to NOAEL ($UF_L$) extrapolation, resulted in a 3.8-fold increase for the $UF_{AHS}$ from the $UF_{AH}$ and a 4.4-fold increase for the $UF_{AHL}$ from the $UF_{AH}$, giving UFs approximately 4-fold higher than the $UF_{AH}$ in either case. Finally, the four combined UFs, $UF_{AHSL}$, when chronic data and NOAEL are lacking, resulted in a 16-fold increase from the $UF_{AH}$. All the UFs obtained by Monte Carlo simulation, based on the default UF of 10, are slightly lower than our proposed UFs for rats. Simple multiplication of the default value of 10, resulted in much larger values than all three or four combined UFs ($UF_{AHS}$, $UF_{AHL}$, $UF_{AHSL}$) for all animals.

## 5. Application of subdivision and replacement of uncertainty factors for inter- and intra-species differences (chemical-specific adjustment factors)

In the present article, we propose animal size-specific inter-species UFs and new combined UFs ($UF_{AH}$) by using probabilistic approaches. For the cases of hamsters and rats, $UF_{AH}$ is set at 100 but the contributions of inter- and intra-species differences are not equal. The application of the same default subdivision factor shown by Renwick (1993) is not appropriate, if the $UF_A$ values of Table 2 are used as the basis of the assessment. However, the concept established by Renwick (1993) is appropriate because we also recommend that actual and reliable experimental data for PK or PD differences should be incorporated into the risk assessment processes wherever possible. Therefore, we subdivided the new $UF_A$ to determine the contribution ratio of inter- and intra-species differences. In the case of hamsters and rats, the average $UF_A$ is approximately 30 (hamsters = 34.4 and rats = 27.5) and the intra-species difference is 5.09, (calculated above from the Hasegawa et al. (2007) data), resulting in a ratio contribution of ~6:1. The $UF_{AH}$ for hamsters and rats is set at 100, which can be divided into factors of 25 and 4, according to the above ratio of 6:1. Considering the contribution ratios of PK and PD as 60:40 for inter-species differences and 50:50 for intra-species differences, 25 will be subdivided into $25^{0.6}$ (7.0) for PK and $25^{0.4}$ (3.6) for PD, and 4 will be evenly subdivided into $4^{0.5}$ (2) (Table 3).

Similar approaches can be used elsewhere. For example, the mice $UF_{AH}$ of 150 can be divided into 38 and 4, then 38 will be subdivided into $38^{0.6}$ (9.0) for PK and $38^{0.4}$ (4.3) for PD. For rabbits, monkeys and dogs, the $UF_{AH}$ of 40 can be divided into 10 and 4, then 10 will be subdivided into $10^{0.6}$ (4.0) for PK and $10^{0.4}$ (2.5) for PD.

If actual data for the difference between humans and animals for PK and/or PD are available, those data can be used as chemical-specific adjustment factors instead of respective default subdivision factors.

**Table 2**
Combined UFs for six animal species by probabilistic approach (95th percentile), Monte Carlo simulation and simple multiplication of UF 10.

| Species | $UF_A$ | $UF_{AH}$ | $UF_{AHS}$ | $UF_{AHL}$ | $UF_{AHSL}$ |
|---|---|---|---|---|---|
| Mice | 48.2 | 155 | 589 | 684 | 2440 |
| Hamsters | 34.4 | 111 | 421 | 488 | 1740 |
| Rats | 27.5 | 88.7 | 337 | 391 | 1400 |
| Rabbits | 13.8 | 44.3 | 168 | 195 | 698 |
| Monkeys | 12.4 | 39.9 | 152 | 176 | 628 |
| Dogs | 9.63 | 31.0 | 118 | 137 | 488 |
| All animals | | | | | |
| Monte Carlo[a] | 10 | 51 | 234 | 234 | 1040 |
| Default[b] | 10 | 100 | 1000 | 1000 | 3000 |

A, inter-species differences; H, intra-species differences; S, subchronic to chronic; L, LOAEL to NOAEL.
[a] Data from Swartout et al. (1998).
[b] Note that US Environmental Protection Agency (USEPA) combines the default values of 4 UFs into 3000, because of the generally conservative nature of combining 10-fold factors that are each somewhat conservative (Dourson, 1994).

## 6. Discussion

The proposed written document to address chemical safety assessment methodology is needed because officially agreed upon guidelines do not exist in Japan. For this purpose, the latest scientific information has been collected to reduce the uncertainty in the risk assessment process. It would be more reliable for UFs to be estimated on the basis of actual experimental data rather than use conventional default UFs. Furthermore, the values are more representative of the data if they are developed using statistical components such as the median with distribution of differences rather than point estimates. A tolerable daily intake can be derived by probabilistic approaches, using the median or geometric mean (GM) and GSD to combine two or more distributions.

Recently, Falk-Filipsson et al. (2007) reviewed a wide variety of assessment factors in various historical and scientific ranges, including guidelines from national and international bodies. They reported that "over-conservatism" should be avoided by using a probability distribution for the various assessment factors. However, such an approach was only applied to the UF for inter-species

**Table 3**
Subdivision of uncertainty factors for inter- and intra-species differences.

| Species | $UF_{AH}$ | $UF_A$<br>$UF_H$ | Subdivision<br>PK × PD |
|---|---|---|---|
| Mice | 150 | 38 | 9.0 × 4.3 |
| | | 4 | 2 × 2 |
| Hamsters | | 25 | 7.0 × 3.6 |
| Rats | 100 | 4 | 2 × 2 |
| Rabbits | | 10 | 4.0 × 2.5 |
| Monkeys | 40 | 4 | 2 × 2 |
| Dogs | | | |

**Table 4**
Median or GM with GSD for each uncertainty in four different methodologies.

| | Inter-species differences | | Intra-species differences | | Subchronic to chronic | | LOAEL to NOAEL | |
|---|---|---|---|---|---|---|---|---|
| | Median/GM | GSD | Median/GM | GSD | Median/GM | GSD | Median/GM | GSD |
| Baird et al. (1996) (GM) | AF[a] | 4.9 | 2.7 | 2.3 | 2.0 | 2.1 | 3.4 | 1.70 |
| Swartout et al. (1998) | 10[b] | | 10[b] | | 10[b] | | 10[b] | |
| Kodell and Gaylor (1999) (median) | 1 | 5.27 | 1 | 5.15 | 2 | 3.67 | 3.5 | 1.82 |
| Present experiment (median) | AF[c] | 3.23 | 3.0 | 1.38 | 1.7 | 3.30 | 3.5 | 1.82 |

[a] Adjustment factor for each animal on the basis of body surface correction.
[b] Use of 10 for every traditional default factor.
[c] Adjustment factor for each animal on the basis of caloric demand.

differences because appropriate distribution data for intra-species differences could not be located.

This study is the fourth trial following those of Baird et al. (1996), Swartout et al. (1998) and Kodell and Gaylor (1999) to use a probabilistic approach to estimate UFs for chemical risk assessment. Table 4 shows the median/GM and GSD for the four methodologies and Table 5 shows combined UFs for inter- and intra-species differences, and two other uncertainties. Swartout et al. (1998) estimated four UFs by Monte Carlo simulation using a traditional default UF of 10 for each uncertainty. Baird et al. (1996) also performed Monte Carlo simulation with specific software, but used actual data instead of default values. On the other hand, Kodell and Gaylor (1999) used standard statistical techniques, as we do here. Key differences in the three methodologies result from the original data used for inter- and intra-species extrapolation. For inter-species differences, the data used in this assessment are considered appropriate because the data are a direct comparison between humans and animals (Schneider et al., 2004). However, Baird et al. (1996) used comparative data within laboratory animals from pesticide safety studies (Dourson et al., 1992) and Kodell and Gaylor (1999) used toxicity comparisons of marine-life $LD_{50}$ (Calabrese and Baldwin, 1995).

A similar analysis can be done for intra-species differences. This assessment used comparative NOAEL data from newborn and young rat repeat-dose studies as a sensitive subpopulation compared to the general population (Hasegawa et al., 2007). However, the other groups (Baird et al., 1996; Kodell and Gaylor, 1999) used lethality distribution data from acute toxicity studies (Dourson and Stara, 1983).

The different methodologies resulted in similar $UF_{AH}$ values for Kodell and Gaylor (1999) and Baird et al. (1996), but were different from Swartout et al. (1998), as shown in Table 5. However, the Baird et al. (1996) $UF_{AH}$ does not include a scaling adjustment factor, thus the median of inter-species differences of Baird's data was calculated as 1. As presented in this assessment, the body surface area correction factor, such as 13.3 for mice, 5.8 for rats, and 1.6 for dogs, should be used to reduce the uncertainty. This assessment calculated the expected UFs for rats using Baird et al. (1996) data (found in Table 4) and using the standard statistical techniques described in the previous sections of this paper. The results of these calculations are shown as "Baird et al., 1996 Our Calc" in Table 5.

**Table 5**
Combined UFs at 95% confidence limit by four methodologies.

| UFs | Baird et al. (1996)[a] | Baird et al. (1996)[b]Our Calc | Swartout et al. (1998)[c] | Kodell and Gaylor (1999)[c] | Present study[b] | Default[c] |
|---|---|---|---|---|---|---|
| $U_{AH}$ | 50 | 300 | 51 | 46 | 89 | 100 |
| $U_{AHS}$ | 126 | 764 | 234 | 161 | 337 | 1000 |
| $U_{AHL}$ | 192 | 1156 | 234 | 184 | 400 | 1000 |
| $U_{AHSL}$ | 484 | 2920 | 1040 | 629 | 1400 | 3000 |

[a] Not including inter-species scaling.
[b] Specific to rats.
[c] For all laboratory animals.

The calculated values were almost six times larger for each UF than those without the scaling adjustment factor (Baird et al., 1996 in Table 5). The calculated UFs in this assessment are relatively similar to Swartout et al. (1998) and much smaller than the default UF values.

The actual data used for our probabilistic estimation of the four UFs are considered suitable at this moment, and the combined $UF_{AH}$ values for several commonly used laboratory animal species were given by standard statistical techniques (Table 2). However, as a rounded value is preferred for risk assessment, we propose size-specific UFs of 150 for mice, 100 for hamsters and rats, and 40 for rabbits, monkeys and dogs. As for other UFs such as $UF_{AHS}$, $UF_{AHL}$ and $UF_{AHLS}$, the average uncertainty values for each ($UF_{AHS}/UF_{AH}$, $UF_{AHL}/UF_{AH}$ and $UF_{AHLS}/UF_{AH}$) were 3.8, 4.4 and 15.7, respectively. Therefore, we propose to uniformly use a factor of 4 when a NOAEL ($UF_L$) and/or chronic data ($UF_S$) is lacking.

The application of an alternative subdivision of UFs should be considered in order to address the new concept of including animal size-specific UFs in the contribution of inter- and intra-species differences. The values of the new subdivision described in this study may be too precise, but this is inevitable, because the contribution of inter- and intra-species differences is definitively different. When further data on human and animal PK/PD differences are available, a more practical risk assessment can be implemented.

## 7. Conclusions

We propose an animal size-specific UF for $UF_{AH}$ of 150 for mice, 100 for hamsters and rats, and 40 for rabbits, monkeys and dogs, for inter- and intra-species differences using a probabilistic approach. An additional default factor of 4 could be applied for either lack of chronic data or lack of a NOAEL. In addition to the proposed animal size-specific UFs, new subdivided PK/PD default factors for each animal are also proposed according to the different contribution of inter- and intra-species differences.

## Conflict of interest statement

The authors declare that there are no conflicts of interest.

## Acknowledgments

The present study was supported by Grants-in-Aid from Food Safety Commission, Japan (No. 0704) and developmental research funds from Toxicology Excellence for Risk Assessment, a non-profit toxicology R&D organization. The authors acknowledge Drs. Takeo Nomi and Masami Yamada (National Institute of Health Sciences, Japan) for their valuable comments.

## References

Abdel-Rahman, M.S., Kadry, A.M., 1995. Studies on the use of uncertainty factors in deriving RfDs. Hum. Ecol. Risk Assess. 1, 614–624.

Baird, S.J.S., Cohen, J.T., Graham, J.D., Shlyakhter, A.I., Evans, J.S., 1996. Noncancer risk assessment: a probabilistic alternative to current practice. Hum. Ecol. Risk Assess. 2, 79–102.

Barton, H.A., Cogliano, V.J., Flowers, L., Valcovic, L., Setzer, R.W., Woodruff, T.J., 2005. Assessing susceptibility from early-life exposure to carcinogens. Environ. Health Perspect. 113, 1125–1133.

Calabrese, E.J., 1985. Uncertainty factors and interindividual variation. Regul. Toxicol. Pharmacol. 5, 190–196.

Calabrese, E.J., 2001. Assessing the default assumption that children are always at risk. Hum. Ecol. Risk Assess. 7, 37–59.

Calabrese, E.J., Baldwin, L.A., 1995. A toxicological basis to derive generic interspecies uncertainty factors for application in human and ecological risk assessment. Hum. Ecol. Risk Assess. 1, 555–564.

Charnley, G., Putzrath, R.M., 2001. Children's health, susceptibility, and regulatory approaches to reducing risks from chemical carcinogens. Environ. Health Perspect. 109, 187–192.

Connelly, J.C., Hasegawa, R., McArdle, J.V., Tucker, M.L., 1997. ICH guideline residual solvents. Pharmaeuropa 9 (Suppl. 1), S1–S68.

Dourson, M.L., 1994. Methods for establishing oral reference doses (RfDs). In: Mertz, W., Abernathy, C.O., Olin, S.S. (Eds.), Risk Assessment of Essential Elements. ILSI Press, Washington DC, pp. 51–61.

Dourson, M., Charnley, G., Scheuplein, R., 2002. Differential sensitivity of children and adults to chemical toxicity. II. Risk and regulation. Regul. Toxicol. Pharmacol. 35, 448–467.

Dourson, M.L., Felter, S.P., Robinson, D., 1996. Evolution of science-based uncertainty factors in noncancer risk assessment. Regul. Toxicol. Pharmacol. 24, 108–120.

Dourson, M.L., Knauf, L.A., Swartout, J.C., 1992. On reference dose (RfD) and its underlying toxicity data base. Toxicol. Ind. Health 8, 171–189.

Dourson, M.L., Stara, J.F., 1983. Regulatory history and experimental support of uncertainty (safety) factors. Regul. Toxicol. Pharmacol. 3, 224–238.

Falk-Filipsson, A., Hanberg, A., Victorin, K., Warholm, M., Wallén, M., 2007. Assessment factors—applications in health risk assessment of chemicals. Environ. Res. 104, 108–127.

Freireich, E.J., Gehan, E.A., Rall, D.P., Schmidt, L.H., Skipper, H.E., 1966. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. Cancer Chemother. Rep. 50, 219–244.

Ginsberg, G., Hattis, D., Sonawane, B., Russ, A., Banati, P., Kozlak, M., Smolenski, S., Goble, R., 2002. Evaluation of child/adult pharmacokinetic differences from a database derived from the therapeutic drug literature. Toxicol. Sci. 66, 185–200.

Glaubiger, D.L., von-Hoff, D.D., Holcenberg, J.S., Kamen, B., Pratt, C., Ungerleider, R.S., 1981. The relative tolerance of children and adults to anticancer drugs. Front. Radiat. Ther. Oncol. 16, 42–49.

Goldsmith, M.A., Slavik, M., Carter, S.K., 1975. Quantitative prediction of drug toxicity in humans from toxicology in small and large animals. Cancer Res. 35, 1354–1364.

Grieshaber, C.K., Marsoni, S., 1986. Relation of preclinical toxicology to findings in early clinical trials. Cancer Treat. Rep. 70, 65–72.

Hasegawa, R., 1991. Comparison of NOAELs from published 3-month and 2-year repeated dose toxicity studies, unpublished data.

Hasegawa, R., Hirata-Koizumi, M., Dourson, M., Parker, A., Hirose, A., Nakai, S., Kamata, E., Ema, M., 2007. Pediatric susceptibility to 18 industrial chemicals: a comparative analysis of newborn with young animals. Regul. Toxicol. Pharmacol. 47, 296–307.

Hattis, D., Erdreich, L., Ballew, M., 1987. Human variability in susceptibility to toxic chemicals—a preliminary analysis of pharmacokinetic data from normal volunteers. Risk Anal. 7, 415–426.

Hattis, D., Ginsberg, G., Sonawane, B., Smolenski, S., Russ, A., Kozlak, M., Goble, R., 2003. Differences in pharmacokinetics between children and adults—II. Children's variability in drug elimination half-lives and in some parameters needed for physiologically-based pharmacokinetic modeling. Risk Anal. 23, 117–142.

IPCS (International Programme on Chemical Safety), 1994. Derivation of guidance values for health-based exposure limits. Environmental Health Criteria No. 170: Assessing human health risks of chemicals. World Health Organization, Geneva.

Kodell, R.L., Gaylor, D.W., 1999. Combining uncertainty factors in deriving human exposure levels of noncarcinogenic toxicants. Ann. NY Acad. Sci. 895, 188–195.

Lehman, A.J., Fitzhugh, O.G., 1954. 100-fold margin of safety. Assoc. Food Drug Off. US Quant. Bull. 18, 33–35.

Paxton, J.W., Kim, S.N., Whitfield, L.R., 1990. Pharmacokinetic and toxicity scaling of the antitumor agents amsacrine and CI-921, a new analogue, in mice, rats, rabbits, dogs, and humans. Cancer Res. 50, 2692–2697.

Renwick, A.G., 1993. Data-derived safety factors for evaluation of food additives and environmental contaminants. Food Addit. Contam. 10, 275–305.

Rozencweig, M., Von Hoff, D.D., Staquet, M.J., Schein, P.S., Penta, J.S., Goldin, A., Muggia, F.M., Freireich, E.J., DeVita Jr., V.T., 1981. Animal toxicology for early clinical trials with anticancer agents. Cancer Clin. Trials 4, 21–28.

Schein, P., Davis, R.D., Carter, S., Newman, J., Schein, D.R., Rall, D.P., 1979. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. Clin. Pharmacol. Ther. 11, 3–40.

Schneider, K., Oltmanns, J., Hassauer, M., 2004. Allometric principles for interspecies extrapolation in toxicological risk assessment–empirical investigations. Regul. Toxicol. Pharmacol. 39, 334–347.

Sheehan, D.M., Gaylor, D.W., 1990. Analysis of the adequacy of safety factors. Teratology 41, 590–591.

Swartout, J.C., Price, P.S., Dourson, M.L., Carlson-Lynch, H.L., Keenan, R.E., 1998. A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. 18, 271–282.

Travis, C.C., White, R.K., 1988. Interspecific scaling of toxicity data. Risk Anal. 8, 119–125.

US EPA, Environmental Protection Agency, 1992. Draft report: a cross-species scaling factor for carcinogen risk assessment based on equivalence of mg/kg$^{3/4}$/day. Fed. Regist. 57, 24152–24172.

US EPA, Environmental Protection Agency, 2005a. Guidelines for carcinogen risk assessment. Office of Research and Development, Washington, DC. <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=116283> (accessed 1 February 2010).

US EPA, Environmental Protection Agency, 2005b. Supplemental guidance for assessing susceptibility from early-life exposure to carcinogens. Risk Assessment Forum, Washington, DC. <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=160003> (accessed 8 April 2010).

US EPA, Environmental Protection Agency, 2009. Integrated Risk Information System (IRIS). National Center for Environmental Assessment. <http://www.epa.gov/iris> (accessed 19 January 2010).

Weil, C.S., 1972. Statistics vs safety factors and scientific judgment in the evaluation of safety for man. Toxicol. Appl. Pharmacol. 21, 454–463.

Weil, C.S., McCollister, D.D., 1963. Relationship between short- and long-term feeding studies in designing an effective toxicity test. Agric. Food Chem. 11, 486–491.

WHO, 2005. Harmonization project document No.2: chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration–response assessment. World Health Organization Geneva.

Genes and Environment, Vol. 32, No. 2 pp. 40–42 (2010)

*Letter to the editor*

# Commentary to the Discussion on Topic 3, "*In Vitro* Test Approaches with Better Predictivity" at the 5th International Workshop on Genotoxicity Testing (IWGT)

**Hajime Kojima**[1]

Japanese Center for the Validation of Alternative Methods (JaCVAM), National Institute of Health Sciences, Tokyo, Japan

(Received March 17, 2010; Revised April 20, 2010; Accepted April 21, 2010)

The 5th International Workshop on Genotoxicity Testing (IWGT) was held on August 17–19, 2009, at Biozentrum of the University of Basel, Switzerland, prior to the 2009 International Conference on Environmental Mutagens (ICEM), Florence, Italy. In this workshop, approximately 200 participants from government, industry and academia, mainly from the USA, EU, Canada and Japan, discussed the following six topics: 1) suitable maximum concentrations for tests with mammalian cells (this group was divided into two subgroups: 1a) *in vitro* chromosome aberrations and micronuclei, and 1b) mammalian cell gene mutations); 2) photogenotoxicity testing requirements; 3) *in vitro* test approaches with better predictivity; 4) improvement of *in vivo* genotoxicity assessment, i.e., the link to standard toxicity testing; 5) use of historical control data in the interpretation of positive results; and 6) suitable follow-up risk assessment testing for *in vivo* positive results.

Dr. Toshio Kasamatsu (Kao Co.) and I were invited to the group on topic 3, "*In Vitro* Test Approaches with Better Predictivity". Over the course of seven presentations within the group, we discussed new test methods based on the following background. In sharp contrast with other groups, a new subgroup (led by Stefan Pfuhler, Procter & Gamble: P&G) was convened to develop consensus recommendations for choosing a better test systems to improve the predictivity of *in vitro* tests. This group reviewed current studies investigating whether certain cell types were more susceptible to give irrelevant positive results *in vitro*.

The background for the discussion is as follows:

1) The high rate of false positive results in the current battery of *in vitro* tests–as high as 80% when mammalian cell assays are combined (i.e., chromosome aberration assay and mouse lymphoma assay) (1);
2) Seventh Amendment to EU Cosmetics Directive: Marketing and testing ban on ingredients tested *in vivo*, which came into force March 2009 (2);
3) REACH (Registration, Evaluation, Authorization and Restriction of Chemical) issue. The REACH Regulation gives greater responsibility to industry to manage the risks from chemicals and to provide safety information on the substances (3).

Particularly, the European testing ban resulting from the Seventh Amendment to the EU Cosmetics Directive may result in valuable compounds being unnecessarily discarded from European cosmetic markets. In the field of mutagenicity/genotoxicity, validated alternative methods are available, and *in vivo* studies were prohibited in EU territory after March 11, 2009 (2). In spite of the high rate of false positive results in the current battery of *in vitro* tests, the Scientific Committee on Consumer Safety (SCCS) (4), a European regulatory agency, recommends a battery of three *in vitro* assays: a bacterial reverse mutation test; an *in vitro* mammalian cell gene mutation test; and either an *in vitro* micronucleus test or *in vitro* mammalian chromosome aberration test. False positive results may be due to experimental conditions that have no relevance to *in vivo* situations. It is hard to confirm or deny the genotoxicity potential of candidate cosmetic ingredients based on results of the current *in vitro* tests. Therefore, we hope to develop promising new approaches for *in vitro* testing that can be used in place of *in vivo* studies to reduce the high incidence of false positive results of existing *in vitro* testing.

First, the group addressed the following questions:

Choice of cell line:

→ Are there restrictions on the choice of cells used for testing?
→ Should there be recommendations about the choice of cell line? Can this recommendation be made now, or are more data needed?

[1]Correspondence to: Hajime Kojima, Japanese Center for the Validation of Alternative Methods (JaCVAM), National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Tel: +81-3-3700-9874, Fax: +81-3-3700-1145, E-mail: h-kojima@nihs.go.jp

New approaches:

→ How do we rate the status of the genotoxicity assays performed on 3-dimensional human skin equivalents (3-D skin)?

→ What is the applicability domain of that assay?

→ What kind of data would the team like to see before considering this as a valid assay?

Presentations

1) Which cell lines should we be using?

✓ Paul Fowler (Covance) presented "Reduction of false positives in *in vitro* genetic toxicology testing: Importance of cell selection"; the "false positives" project, which compares rates of false positives with V79, CHO, CHL, TK6, HepG2, L5178Y, and HuLy obtained from COLIPA; European trade association for cosmetic, toiletry and perfumery industry;

✓ Mick Fellows (AstraZeneca) presented "Data generated in conjunction with the *in vitro* micronucleus test (MNvit) OECD guideline finalisation";

✓ Azeddine Elhajouji (Novartis) presented "Comparison of V79, L5178Y, TK6 and primary human lymphocytes for micronucleus induction"; and

✓ Ludowig Le Hegarat (Agence Francaise De Securite Sanitaire Des Aliments) presented "The HepaRG cell line, a predictive model in genetic toxicology?"

2) Promising new approaches

✓ Rodger Curren (Institute for *In vitro* Science: IIVS) presented "A novel micronucleus assay conducted in reconstructed human skin"; micronucleus data generated in 3-D skins (COLIPA 3-D skin project);

✓ Gladys Ouédraogo (L'Oreal) presented "The comet assay on 3-D skins (COLIPA 3-D skin project)";

✓ Hajime Kojima presented "Proposal of protocol for comet assay using a three-dimensional epidermal model".

From these presentations, we reached a consensus, which were reflected in the following statements in the final IWGT comments:

1. Data were presented indicating that p53-compromised rodent cell lines over-estimate genotoxic potential in the micronucleus test. Therefore, IWGT suggests *in vitro* micronucleus or chromosome aberration assay using p53-competent cells.

The data presented at the IWGT from the OECD MNvit test showed that all cell types correctly identify clastogens and aneugens. However, the data of the compounds that gave false results in the MNvit test suggested that there was a great diversity of the responses of the various cell types. The group agreed, based on the similarity of MNvit and chromosome aberration assays, that the diversity would also be observed in the chromosome aberration assay. HepaRG is a promising model, in that the cells appear to have better phase I and II metabolizing potential than the other cell lines. However, further evaluation is required in order to confirm the value of this model for genotoxicity testing.

2. It has been demonstrated that cell line stability and source can affect the outcome of genotoxicity assays. Therefore, IWGT recommends adhering to good cell culture practices, characterizing all new cells, checking regularly for drift, and working from low-passage stocks. It would be useful to compile a common genotoxicity cell bank with fully characterized stocks of all cells.
3. Genotoxicity testing in 3-D skins (micronucleus test and comet assay) is a promising new *in vitro* test for chemicals applied to the skin.

The advantage of the model is that it resembles the properties of human skin (barrier function, metabolism), and that the route of exposure is relevant for dermally applied chemicals (e.g., cosmetics). The data presented show that the micronucleus assay by 3-D skin is further advanced; furthermore, inter-and intra-lab reproducibility has been demonstrated. IWGT agreed that the comet assay should be further evaluated. The comet assay is seen as a valuable addition, as it is not dependent on cell proliferation and covers a wider spectrum of DNA damage. The metabolic capacity needs to be further evaluated (this work is ongoing). It would be valuable to capture the kinetics of penetration and toxicity in order to establish the ideal sampling time(s) for the comet assay. Recommendations on the use of appropriate vehicles should be established. It was agreed that 3-D skin, once validated, will be useful for following up on positive results from standard *in vitro* assays for dermally applied chemicals. The applicability domain will be established once the validation is completed.

Considering the EU situation and the international expansion of animal welfare laws, we must avoid progressively adopting *in vivo* genotoxicity testing. Therefore, I think that the consensus in this group was made "according to script", at least with regard to European colleagues. Against European drastic transformation, my concern on the safety evaluation is that the trend may push out *in vivo* testing in the cosmetic field. We must send out a warning about the current trend to avoid all *in vivo* tests and emphasize importance of re-evaluating cell lines and validating studies on new approaches to testing *in vitro*. I think further discussion is needed to established well-balanced *in vitro* and *in vivo* tests for evaluation of genotoxic risk of chemicals to human.

## References

1 Kirkland D, Kasper P, Muller L, Corvi R, Speit G. Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A follow-up to an ECVAM workshop. Mutat Res. 2008; 653: 99–108.
2 Commission Staff Working Documents, Time tables for the phasing-out of animal testing in the framework of the 7th Amendment to the Cosmetics Directive (Council Directive 76/768/EEC), EN, SEC 2004; 1210: 1–8.
3 JRC. 2008. http://www.vet.uu.nl/nca/userfiles/other/REACH
4 Scientific Committee on Consumer Safety (SCCS)/1294/10, Memorandum on "Alternative test methods in human health safety assessment of cosmetic ingredients in the European Union". 2009.

Research Article

Journal of AppliedToxicology

Received: 10 April 2010, Revised: 27 May 2010, Accepted: 28 May 2010 Published online in Wiley Online Library: 30 July 2010

(wileyonlinelibrary.com) DOI 10.1002/jat.1567

# Inter-laboratory validation of the modified murine local lymph node assay based on 5-bromo-2′-deoxyuridine incorporation

Hajime Kojima,[a]* Masahiro Takeyoshi,[b] Takashi Sozu,[c] Takumi Awogi,[d] Kazunori Arima,[e] Kenji Idehara,[f] Yoshiaki Ikarashi,[a] Yukiko Kanazawa,[g] Eiji Maki,[h] Takashi Omori,[i] Atsuko Yuasa[j] and Isao Yoshimura[k]

**ABSTRACT: The murine local lymph node assay (LLNA) is a well-established alternative to the guinea pig maximization test (GPMT) or Buehler test (BT) for the assessment of the skin sensitizing ability of a drug, cosmetic material, pesticide or industrial chemical. Instead of radioisotope using in this method, Takeyoshi M. *et al.* (2001) has developed a modified LLNA based on the 5-bromo-2′-deoxyuridine (BrdU) incorporation (LLNA:BrdU-ELISA). The LLNA:BrdU-ELISA is practically identical to the LLNA methodology excluding the use of BrdU, for which a single intraperitoneal injection of BrdU is made on day 4, and colorimetric detection of cell turnover. We conducted the validation study to evaluate the reliability and relevance of LLNA:BrdU-ELISA.**

**The experiment involved 7 laboratories, wherein 10 chemicals were examined under blinded conditions. In this study, 3 chemicals were examined in all laboratories and the remaining 7 were examined in 3 laboratories. The data were expressed as the BrdU incorporation using an ELISA method for each group, and the stimulation index (SI) for each chemical-treated group was determined as the increase in the BrdU incorporation relative to the concurrent vehicle control group. An SI of 2 was set as the cut-off value for exhibiting skin sensitization activity.**

**The results obtained in the experiments conducted for all 10 chemicals were sufficiently consistent with small variations in their SI values. The sensitivity, specificity, and accuracy of LLNA:BrdU-ELISA against those of GPMT/BT were 7/7 (100%), 3/3 (100%), and 10/10 (100%), respectively. Copyright © 2010 John Wiley & Sons, Ltd.**

**Keywords:** 5-bromo-2′-deoxyuridine (BrdU); inter-laboratory validation study; local lymph node assay (LLNA); skin sensitization; stimulation Index (SI)

## INTRODUCTION

Skin sensitization (i.e. allergic contact dermatitis, ACD) is an immunologically mediated cutaneous reaction to a drug, cosmetic material, pesticide or industrial chemical. It is known that the detection and evaluation of the immune-based adverse effects that are collectively referred to as hypersensitivity reactions are very difficult tasks, particularly during the drug approval process, because of the lack of adequate non-clinical models and the low incidence rate of reactions (Hastings, 2001). However, there are several adequate and predictive methods for modeling ACD. For several decades, tests involving guinea pigs, such as the guinea pig maximization test (GPMT) or the Buehler test (BT), have been used for assessing the skin sensitization potential of chemicals (OECD, 1992). In addition, a mouse model for assessing the relative sensitization potential is a well-established alternative method for determining whether a chemical causes ACD. Although GPMT and BT can be viewed as phenomenological methods in which the clinical signs are modeled, local lymph node assay (LLNA) and the mouse ear swelling test are based on a mechanistic understanding of immune-based contact dermatitis (Gad *et al.*, 1986; Hastings, 2001). In addition, these methods also offer important animal welfare benefits. In these assays, the use of LLNA has been successfully validated by several studies (Basketter and Scholes, 1992; Basketter *et al.*, 1996, 2002; Gerberick *et al.*, 2000; Haneke *et al.*, 2001). Recently, it has been formally adopted by the Organization for Economic Co-operation and Development (OECD), according to the guidelines for testing

*Correspondence to: *H. Kojima, Japanese Center for the Validation of Alternative Methods, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan.*
*E-mail: h-kojima@nihs.go.jp*

[a]*National Institute of Health Sciences, Tokyo, Japan*

[b]*Chemicals Evaluation and Research Institute, Kitakatsushika, Saitama, Japan*

[c]*Osaka University, Osaka, Japan*

[d]*Otsuka Pharmaceutical Co. Ltd, Naruto, Tokushima, Japan*

[e]*Taisho Pharmaceutical Co. Ltd, Saitama, Japan*

[f]*Daicel Chemical Industries Ltd, Himeji, Hyogo, Japan*

[g]*Food and Drug Safety Center, Hadano, Kanagawa, Japan*

[h]*Biosafety Research Center, Foods, Drugs and Pesticides, Iwata, Shizuoka, Japan*

[i]*Kyoto University School of Public Health, Kyoto, Japan*

[j]*Fuji Film Co. Ltd, Ashigara, Kanagawa, Japan*

[k]*Tokyo University of Science, Tokyo, Japan*

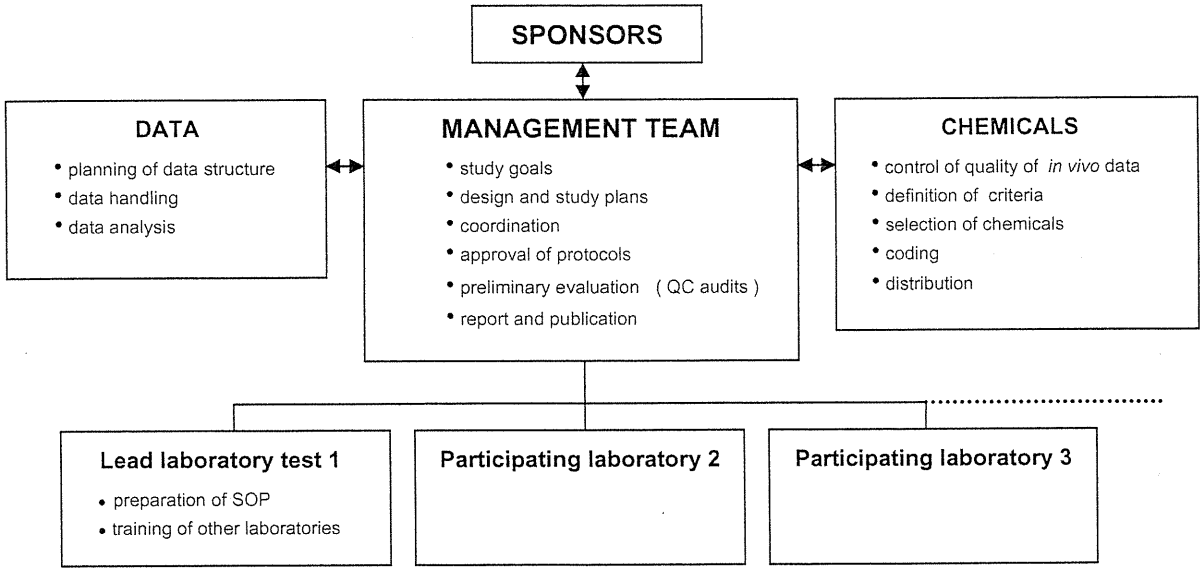



**Figure 1.** The organization of validation management structure.

chemicals 406 and 429 (OECD, 1992, 2002), and it is accepted by the EU and USA as a suitable method for classifying the skin sensitizing ability of chemicals (Basketter *et al.*, 2005; Dean *et al.*, 2001; Sailstad *et al.*, 2001). The LLNA is specifically designed to identify contact allergens. The assay is not intended to facilitate the detection of low molecular weight chemicals associated with systemic sensitization or drug allergies (Kimber, 2001). However, an investigation, which was designed to explore the ability of LLNA to identify pharmaceutical process intermediates known to cause contact allergy in humans, provided evidence that the assay is a useful method for hazard identification (Durand *et al.*, 2003). Furthermore, the use of the method, along with the use of GPMT and BT, is recommended for the determination of the skin sensitization potential of new drugs (FDA, 2002).

The original LLNA uses [3H]-methyl thymidine to measure lymphocyte proliferation. The use of this method is hindered, particularly in Japan, because such a radioisotope (RI)-based method requires special facilities and handling procedures. Several authors have been conducting investigations for the development of an alternative non-RI method for performing LLNA (Dearman *et al.*, 1999; Ehling *et al.*, 2005a, b; Hatao *et al.*, 1995; Idehara *et al.*, 2008, Lee *et al.*, 2002; Omori *et al.*, 2008; Takeyoshi *et al.*, 2001, 2005, 2006; Yamashita *et al.*, 2005).

One method, the LLNA:BrdU-ELISA proposed by Takeyoshi *et al.*, is a modification of the original LLNA that involves assessing lymphocyte proliferation using ELISA (enzyme-linked immunosorbent assay) to measure 5-bromo-2'-deoxyuridine (BrdU) incorporation instead of measuring the radioactivity produced by [3H]-methyl thymidine incorporation. This modification is also one of the most promising non-radioisotopic LLNA modifications that has been recently peer-reviewed by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Peer Review Panel Meeting for New Versions and Applications of the Murine Local Lymph Node Assay (ICCVAM 2009).

Although the methodology of the LLNA:BrdU-ELISA is essentially the same as the original LLNA, the data available were insufficient for validating the assay method for inter-laboratory reproducibility. Therefore, we conducted the inter-laboratory validation studies for LLNA:BrdU-ELISA. Prior to the studies, a preliminary study was conducted using only the positive control chemical, namely, 50% hexyl cinnamic aldehyde (HCA). After that, the first study testing 12 chemicals in nine laboratories yielded a large variation among tests. It was considered that extreme variation in the stimulation index (SI) was produced by the very small BrdU labeling index of the negative control. Then we determined that protocol revisions were necessary. We describe the phase II study in which the revised protocol was used. The objectives of the study were (1) to evaluate the extent of inter-laboratory variation with regard to LLNA:BrdU-ELISA and (2) to ascertain whether the results of LLNA:BrdU-ELISA are comparable with those of LLNA and GPMT/BT.

## METHODS

### Organization

This study was organized by researchers belonging to the validation management team (VMT) for the validation of the assay. The structure of the VMT is shown in Fig. 1. The VMT comprised representatives from each experimental laboratory, toxicologists as the chemical selectors and distributors of the chemicals and materials, biostatisticians and the study manager. All the experiments were performed by the toxicologists of the experimental laboratories. In this study, nine experimental laboratories with sufficient experience in the use of the LLNA and/or its modifications participated. Research teams of all the experimental laboratories obtained ethical approval for each standard operational procedure conducted in their laboratories. After the preliminary and pre-validation studies by the VMT, seven laboratories, excluding Chemicals Evaluation and Research Institute, Japan (CERI) and Meiji Seika Kaisha Ltd, participated in the final validation study.

Journal of AppliedToxicology

### Technology Transfer

A one-day technology-transfer seminar was held by the LLNA:BrdU-ELISA developer, which was attended by at least one researcher (or technician) from each experimental laboratory. Participants learned the method for conducting the assay according to the standard protocol.

### Test Phases and Schedule

Prior to the study, a preliminary study was conducted by researchers from all the experimental laboratories, whic used only the positive control chemical, 50% HCA. Two phases of the validation study were conducted. In phase I (pre-validation study), we examined the reliability and the reproducibility of the test protocol. Inter- and intra-laboratory variability and transferability were assessed using suitable statistical analysis from testing 12 chemicals in nine laboratories from August to December 2006. However the variability of the SI values for the positive control was too large, the VMT determined that protocol revisions were necessary without analyzing the data for the 12 chemicals. This report details the phase II study, in which 10 new chemicals were tested by seven laboratories from September to December 2007 using the revised protocol.

### Chemical Selection and Allocation

The chemical selectors chose 20 candidate chemicals that were previously used in LLNA and whose test results had been documented (Basketter and Scholes, 1992; Basketter *et al.*, 1998, 1999a, b, 2000; Gerberick *et al.*, 2004; Haneke *et al.*, 2001; Kimber *et al.*, 1998; Loveless *et al.*, 1996). On the basis of these literature data and solubility of the chemicals, the chemical selectors selected vehicles and prepared three fixed doses (low, medium and high) for each chemical. Subsequently, the chemicals were transported from the chemical and material distributors to the experimental laboratories.

In this study, 10 of the 20 candidate chemicals were selected and classified as strong, mild or weak sensitizers or non-sensitizers on the basis of original LLNA data. In order to reduce the number of animals used, pairs comprising groups treated with three chemicals and the corresponding vehicle control group were employed. In other words, in each laboratory, three chemicals were simultaneously tested with one negative control and one positive control for every experiment. Of the 10 chemicals, three were dispatched to all seven participating experimental laboratories, and the remaining seven were randomly allocated to the laboratories by a biostatistician and dispatched to three experimental laboratories.

So that the research teams could not predict the severity of the effects for each chemical, all the chemicals were coded into alphabetic characters for blinded distribution, and indicated as low, medium and high in terms of the dose. Each research team prepared solutions using the distributed solvents before experiments. However, prior to the study, the researchers and toxicologists of the respective laboratories were informed of the identity of the 20 candidate chemicals and the corresponding control vehicles. This was done in order to ensure the safety of the chemists performing the experiments (e.g. with regard to proper disposal of the chemicals) and to prevent any anxiety that they would experience while handling unknown chemicals.

### Standard Protocol for LLNA:BrdU-ELISA

The standard protocol for the assay was prepared prior to the preliminary test and refined before the commencement of the study.

#### *Animals*

Young adult female mice (nulliparous and non-pregnant) of the CBA/JN strain were used at age 8–12 weeks. Healthy animals in good general condition on arrival were quarantined for more than 5 days. During the quarantine and the acclimation period, clinical signs, body weights and excrement of the animals were monitored. Animals confirmed to be in good health with favorable body weight gains by a person in charge of animal management during the quarantine and acclimation period were allocated to groups by a stratified randomization or other appropriate methods before the start of the study. Animals were identified by marking the tail with colored marker, ear tags or other appropriate methods.

The animals were housed in an animal room maintained at a temperature of 22 ± 3°C and a relative humidity of 30–70%. The rooms were artificially lighted for 12 h daily, and the animals were given free access to conventional laboratory diet and drinking water.

#### *Animal experiment*

A minimum of four successfully treated animals was used per dose group, with a minimum of three consecutive doses of the chemical, and one group each for the negative (vehicle) control and positive control. A 25 µl dose of test solution was applied to the dorsum of both ears of the mice for three consecutive days using a microvolume pipette. A single intraperitoneal injection of 0.5 ml of BrdU solution (5 mg/mouse/injection) was given to the mice 48 h after the final application. BrdU solution was prepared before administration and stored in a freezer below –20°C until use.

Clinical signs were observed at least once per day. Body weights were measured on the day of the first application and on the day lymph nodes were collected. Approximately 24 h after BrdU injection, the auricular lymph nodes were removed. The lymph nodes were carefully dissected and trimmed of fascia and fat, weighed and stored individually in a 1.5 ml centrifuge tube at –20°C until BrdU-ELISA measurement.

#### *Preparation of Lymph Nodes (LN) Cell Suspension*

From the discussion of the results on the phase I, we determined that the mean absorbance of the negative (vehicle) control group should be within 0.1–0.2. Because the absorbance depends on the combination of assay apparatus and the target volume of cell suspension, every laboratory had to decide their own optimal target volume of LN cell suspension in advance so that the absorbance value of the negative control group would be within 0.1–0.2. The volume would be expected to be approximately 15 ml. The volume of LN cell suspension of the all test animals was adjusted to the optimized volume.

A small amount (*ca* 0.3 ml) of physiological saline was added to the centrifuge tube that contained the collected LN, and the LN was crushed with a disposal plastic pestle to make the cell



suspension. The cell suspension was passed through a #70 nylon mesh and adjusted to the target volume in a 50 ml tube.

*Assay Flow (BrdU-ELISA)*

The incorporation of BrdU into LN cells was determined using a commercial cell proliferation assay kit (Roche Diagnostics GmbH, Roche Applied Science, 68298 Mannheim, Germany; catalog no. 11 647 229 001) after the lymph node was crushed and suspended in physiological saline. The absorbance was defined as the BrdU labeling index.

The cell suspension (100 μl) was added to the wells of a flat-bottom microplate (three wells per sample) after being homogenized by mixing thoroughly with a Vortex mixer. Simultaneously, three blank wells were prepared by adding 100 μl of physiological saline. After adding all samples and blank solution to wells, the plate was centrifuged at 300 *g* for 10 min.

Three-quarters of the volume of the supernatants was removed. In this step, great care was taken not to aspirate the LN cells. The assay plate was dried completely in a hot-air oven.

Two hundred microliters of Fix-Denat solution was added into each well and allowed to stand for 30 min at room temperature. Then the Fix-Denat solution was completely removed.

One hundred microliters of anti-BrdU antibody working solution was added and allowed to stand for 1 h. Then the anti-BrdU antibody working solution was completely removed.

Two hundred microliters of wash solution was added to each well, and the well was washed by pippetting 10 times. Then the wash solution was completely discarded. The wash step was repeated twice (three times in total).

One hundred microliters of tetramethylbenzidine substrate solution was added and allowed to stand for 15 min at room temperature in a dark place. The absorbance at 370 nm was measured with a reference wavelength of 492 nm. When using stop solution (1 M sulfuric acid, 25 μl/well), the absorbance at 450 nm was measured with a reference wavelength of 690 nm.

The BrdU labeling index and SI are defined as follows: without stop solution, BrdU labeling index = (ABS370-ABSblank370) – (ABS490-ABSblank490); with stop solution, BrdU labeling index = (ABS450-ABSblank450) – (ABS650-ABSblank650). The BrdU labeling index was determined for each test animal. SI is the ratio of the mean BrdU labeling index for the treated group to the mean for the concurrent vehicle control group

The positive/negative decision was made based on the criteria of SI ≥ 2, rather than SI ≥ 3, because the dynamic range of the LLNA:BrdU-ELISA is narrower than scintillation counting of incorporated [3H]-methyl thymidine and it produces lower SI values compared with the original LLNA.

*Acceptance criteria for each experiment*

Under the optimized assay conditions described, the mean absorbance range of the negative control is 0.1–0.2. Where the absorbance was higher than 0.2, the stock solutions of the assay were diluted and the absorbance was re-measured. Furthermore, the SI for the positive control group (50% HCA) was equal to or greater than 2. If not, data derived from the experiment were not employed for evaluation.

### Adherence to Good Laboratory Practices

The studies were not conducted under full compliance with Good Laboratory Practices (GLP). However, all the laboratories were equipped to perform, and competent with, GLP. In addition, all the laboratories that participated in the inter-laboratory validation studies used the same experimental protocol and took part in a one-day seminar in which the protocol and execution of the test method were explained. Also, the same commercial kit, test materials and the same dose of each coded substance were used in all the laboratories.

### Quality Control Check

A formatted file for the entry of the experimental data and information was prepared using Microsoft Excel. The file was distributed to the experimental laboratories prior to the experiment. After all testing was completed, all records and documents were checked by the chief and biostatistician of the VMT. If there were missing or strange points, the laboratory was asked for all documents to be submitted to the VMT.

### Database

A biostatistician created a database containing the body weight on the start and final days, LN weight and BrdU incorporation data obtained from each mouse in all of the experimental laboratories. For comparison, data from studies on the original LLNA were collected and included in the database.

### Statistical Methods

In order to demonstrate the variability within the SI values, the confidence interval of the SI values was calculated (Omori and Sozu, 2007). A variance component, $\tau^2$, estimated by a random effect model for the log-transformed SI, was used as a measure of the inter-laboratory variations; this is similar to the meta analysis technique used in clinical studies (Normand, 1999). Using the above-mentioned random effect model, we estimated the weighted average as an overall estimate of the SI value for each chemical dose. The EC2 was defined as the estimated concentration that yielded an SI of 2 from the dose–response curve. If the response for an absorbance was not clearly dose-related, but the Si was greater than or equal to 2 for two doses, then it was considered to be positive. We defined this rule as the positive criterion. The EC2 of the weighted average was estimated and classified into the appropriate chemical category (Gerberick *et al.*, 2004). Finally, the sensitivity, specificity, accuracy, positive predictivity and negative predictivity were calculated as measures of relevance on the basis of the weighted averages in order to assess the concordance of the LLNA:BrdU-ELISA results with the LLNA or GPMT/BT results (OECD, 2005).

## RESULTS

### Chemical Selection

Table 1 shows the selected chemicals, the results of LLNA and GPMT/BT as references, and the LLNA:BrdU-ELISA results. The GPMT/BT results for chemical F (glutaraldehyde) are not listed in Table 1 because the data were not available at the time the list was prepared.

### Body Weights

Table 2 summarizes the body weight statistics observed on days 1 and 6 in each laboratory. No substantial inter-laboratory varia-