Psychiatry Research xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

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

Personality traits and schizophrenia: evidence from a case–control study and meta-analysis

Kazutaka Ohi [a,d], Ryota Hashimoto [a,b,d,*], Yuka Yasuda [a,d], Motoyuki Fukumoto [a,d], Hidenaga Yamamori [a,c], Masao Iwase [a], Hiroaki Kazui [a], Masatoshi Takeda [a,b]

[a] Department of Psychiatry, Osaka University Graduate School of Medicine, Osaka, Japan
[b] Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Kanazawa University and Hamamatsu University School of Medicine, Osaka, Japan
[c] Department of Molecular Neuropsychiatry, Osaka University Graduate School of Medicine, Osaka, Japan
[d] CREST (Core Research for Evolutionary Science and Technology), JST (Japan Science and Technology Agency), Saitama, Japan

ARTICLE INFO

Article history:
Received 12 November 2010
Received in revised form 3 December 2011
Accepted 11 December 2011
Available online xxxx

Keywords:
Schizophrenia
Temperament and Character Inventory (TCI)
Personality trait
Association study
Meta-analysis

ABSTRACT

Personality is considered to be an important aspect of schizophrenia, primarily because it may influence patients' symptoms and social functioning. Specific personality traits are related to schizophrenia. The Temperament and Character Inventory (TCI) measures four traits of temperament – novelty seeking (NS), harm avoidance (HA), reward dependence (RD) and persistence (PS) – and three traits of character – self-directedness (SD), cooperativeness (CO) and self-transcendence (ST). We investigated associations between schizophrenia and personality traits using the TCI in a Japanese case–control sample (99 patients and 179 controls). Patients with schizophrenia scored higher on HA and ST and lower on NS, RD, SD and CO compared with controls in our case–control sample. We then performed a meta-analysis of samples from the published literature and our sample (384 patients and 656 controls). We found no evidence of heterogeneity among studies, except for NS in the overall population. Possible associations between personality traits (HA, RD, PS, SD, CO and ST) and schizophrenia were revealed. The effect sizes (Hedges' g) of the temperament traits were 0.98 for HA, −0.43 for RD and −0.23 for PS, and those of the character traits were −0.96 for SD, −0.47 for CO and 0.61 for ST. These findings suggest that patients with schizophrenia have a unique temperament and character profile compared with the general population.

1. Introduction

Schizophrenia is a complex psychiatric disease involving positive and negative symptoms and cognitive impairment that result in the disturbance of social functioning. However, positive and negative symptoms and cognitive variables appear to account for only 20% of the variance in social performance among patients with schizophrenia (Jaeger and Douglas, 1992; Lysaker and Bell, 1995). Thus, some additional variable, such as personality, may influence social functioning in schizophrenia. Personality is considered to be an important aspect of schizophrenia primarily because it may influence patients' symptoms and social functioning (Lysaker et al., 1998, 1999).

The Temperament and Character Inventory (TCI) is a well-established self-report questionnaire that measures four traits of temperament – novelty seeking (NS), harm avoidance (HA), reward dependence (RD) and persistence (PS) – as well as three traits of character – self-directedness (SD), cooperativeness (CO) and self-transcendence (ST) (Cloninger et al., 1993). The Tridimensional Personality Questionnaire (TPQ) is an older test of personality traits than the TCI. The TPQ measures three temperament traits (NS, HA and RD) (Cloninger, 1987). PS was included in the TPQ as part of the RD trait. TCI and TPQ scores are not directly comparable. Thus, in order to compare a seven-factor model of personality between patients with schizophrenia and controls, we focused on TCI in this study. Temperamental traits of personality may be mediated by pathophysiological mechanisms at a neurotransmitter level, i.e., involving dopamine, serotonin or norepinephrine (Cloninger, 1987). These traits are particularly relevant in schizophrenia because dopamine, serotonin and norepinephrine are involved in symptom expression and are the main targets of antipsychotic medications. However, the mechanisms underlying these genetic and molecular components of personality remain unexplained. Personality trait analyses using the TCI have reported consistent differences between patients with schizophrenia and controls (Guillem et al., 2002; Boeker et al., 2006; Calvo de Padilla et al., 2006; Hori et al., 2008; Smith et al., 2008; Cortes et al., 2009; Gonzalez-Torres et al., 2009). Although not all of

* Corresponding author at: Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Kanazawa University and Hamamatsu University School of Medicine, D3, 2-2, Yamadaoka, Suita, Osaka, 5650871, Japan. Tel.: +81 668793074; fax: +81 668793059.
E-mail address: hashimor@psy.med.osaka-u.ac.jp (R. Hashimoto).

0165-1781/$ – see front matter © 2012 Published by Elsevier Ireland Ltd.
doi:10.1016/j.psychres.2011.12.018

the results of these studies reached statistical significance, patients with schizophrenia had higher scores on ST and HA and lower scores on NS, PS, RD, SD and CO than controls. Studying related personality traits may help to improve our understanding of schizophrenia. However, the limited sample sizes of previous studies have made it difficult to reach a definitive conclusion. Two issues prevent firm conclusions from being drawn: low statistical power resulting from small sample sizes and heterogeneity of results because of possible confounding factors such as ethnicity, gender distribution and mean age in each study.

The current study was undertaken to determine whether there was an association between an abnormal personality trait structure and schizophrenia when patients were compared with healthy controls. Both a case–control study and a meta-analysis were performed. In addition, we examined the influences of potential confounding factors on our results.

2. Methods

2.1. Subjects in a Japanese case–control study

The subjects of this analysis consisted of 99 unrelated patients with schizophrenia [54.5% males (54/45); mean age ± S.D., 39.0 ± 12.1 years] and 179 unrelated healthy controls [48.6% males (87/92); mean age ± S.D., 36.7 ± 11.5 years]. The sex ratio and the mean age did not differ significantly between cases and controls (sex ratio: $\chi^2 = 0.90$, $P = 0.34$, mean age: $z = -1.48$, $P = 0.14$), but the years of education were significantly lower in patients with schizophrenia (14.1 ± 2.3) than in controls (15.4 ± 2.4) [$z = -4.12$, $P < 0.001$]. All subjects were biologically unrelated and of Japanese ethnicity and were recruited at Osaka University. Subjects were excluded from this analysis if they had neurological or medical conditions that could potentially affect the central nervous system, such as atypical headaches, head trauma with loss of consciousness, chronic lung disease, kidney disease, chronic liver disease, thyroid disease, active cancer, cerebrovascular disease, epilepsy or seizures. Cases were recruited from among both outpatients and inpatients at the university hospital. Each patient with schizophrenia was diagnosed by at least two trained psychiatrists according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria based upon structured clinical interviews. The trained psychiatrists obtained a consensus of diagnosis through discussions on unblinded assessments of the mental state of patients. Patients with schizophrenia complicated by co-morbid substance-related disorders or mental retardation were excluded. Controls were recruited through local advertisements. Psychiatrically, medically and neurologically healthy controls were evaluated using the structured DSM-IV Non-Patient Clinical Interview to exclude individuals who had current or past contact with psychiatric services. Written informed consent was obtained for all subjects after the procedures had been fully explained. This study was carried out in accordance with the World Medical Association's Declaration of Helsinki and approved by the Osaka University Research Ethics Committee.

2.2. Personality trait analysis and clinical assessment

Personality traits were assessed using the Japanese version of the TCI. The TCI is administered through a self-report questionnaire based on 240 items requiring true or false responses (Cloninger et al., 1993). We examined the main scores on the four temperament traits (HA, NS, RD and PS) and three character traits (SD, CO and ST) of the scale. The concepts of each trait are as follows: NS is the activation of behavior in response to novelty and signals of reward or relief of punishment. HA is the inhibition of behavior in response to signals of punishment or non-reward. RD is the maintenance of a behavior that was previously rewarded. PS is perseveration with behavior despite frustration and fatigue. SD is the concept of the self as an autonomous individual. CO is the concept of the self as an integral part of humanity or society. ST is the concept of the self as an integral part of the universe and its source (Cloninger et al., 1993). Current symptoms of schizophrenia were evaluated using the five syndrome models of the Positive and Negative Syndrome Scale (PANSS) (Lindenmayer et al., 1994).

2.3. Meta-analysis of studies using TCI

We first searched the studies for meta-analysis using PubMed with the search terms "TCI" or "temperament and character inventory" and "schizophrenia". Our searched data encompassed all publications up to June 2010. Additionally, references cited in the publications obtained were searched to identify additional potentially relevant studies that might not be listed in PubMed.

Next, studies were included in the meta-analysis if they met the following criteria: (1) published in a peer-reviewed journal in English; (2) compared patients with schizophrenia to healthy controls using TCI; (3) contained information on means and SDs for each trait of personality for patients and controls. Available information on country, age, sex, education, age at onset and duration of illness were also collected. Two raters independently verified the validity of the data in all of the included articles.

2.4. Statistical analyses

Statistical analyses were performed using PASW Statistics 18.0 software (SPSS Japan Inc., Tokyo, Japan). Differences in clinical characteristics between patients and controls were analyzed using χ^2 tests for sex and the Mann–Whitney *U*-test for age. In the previous analyses of personality traits assessed with TCI, it was suggested that confounding factors could affect personality traits (Miettunen et al., 2007; Trouillet and Gana, 2008). Age and sex differences in some traits were identified in these analyses. In Japan, patients with schizophrenia have generally completed fewer years of education than healthy controls (Hori et al., 2008). Therefore, we analyzed the difference between patients with schizophrenia and healthy controls for each personality trait using a one-way analysis of covariance (ANCOVA) with age, sex and years of education as covariates. We further investigated correlations between personality traits and PANSS scores or chlorpromazine equivalents of total antipsychotic dosage (CPZ-eq) to examine the associations of some personality traits with this medical condition.

The meta-analyses were performed using the Comprehensive Meta-analysis Version 2.0 (CMA) software package (Borenstein et al., 2005). Effect sizes (Hedges' *g*) indexing the standardized difference in each score between patients with schizophrenia and healthy controls were calculated on the basis of reported statistics (the mean of the schizophrenia samples minus the mean of the healthy control groups, divided by the pooled SD) (Cooper and Hedges, 1994). Effect sizes are typically categorized as small ($g = 0.2$), medium ($g = 0.5$) or large ($g = 0.8$). In order to control for differences in sample size between studies when mean effect sizes were computed, studies were weighted according to estimates of inverse variance. To determine whether the mean effect sizes were statistically significant, the confidence interval (CI) and *z* transformation of the effect size were used. The effect sizes and 95% CIs are represented by forest plots. Cochran's *Q* statistical test was performed to assess possible heterogeneity between the individual studies. The effect sizes and 95% CIs were estimated under the random-effects model described by DerSimonian and Laird if there was evidence of heterogeneity ($P \leq 0.05$). Otherwise, the fixed-effects model described by Mantel–Haenszel was used ($P > 0.05$). Publication bias was assessed using Egger's regression asymmetry test with a funnel plot of the effect size against standard error in each study (Egger et al., 1997). To detect the potential effects of diverse study populations (Asian versus European) or inclusion criteria, subgroup analysis was performed by study location or by dropping one confounding study at a time, recalculating the overall effect size of the remaining studies and testing its significance. Finally, to estimate the impacts of the gender distribution (% male) and the mean age of the total sample on effect sizes, these moderator variables were evaluated using a meta-regression analysis. The mean age of the total sample was integrated from sample sizes, means and SDs by groups using an R software package. The significance level was set as two-tailed $P < 0.05$ for all statistical tests, with the exception of $P < 0.01$ for correlation analysis and $P < 0.10$ for Egger's test of publication bias. The significance level was adjusted using the Bonferroni correction for seven independent tests.

3. Results

3.1. Japanese case–control analysis

Supplementary Table S1 illustrates the comparisons between patients with schizophrenia and healthy controls for each TCI personality trait in a Japanese population. One-way ANCOVA revealed significant differences between cases and controls for each personality trait (NS: $F_{1,273} = 30.9$, $P = 6.50 \times 10^{-8}$, $\eta^2 = 0.102$, HA: $F_{1,273} = 56.9$, $P = 6.75 \times 10^{-13}$, $\eta^2 = 0.173$, RD: $F_{1,273} = 15.6$, $P = 1.00 \times 10^{-4}$, $\eta^2 = 0.054$, SD: $F_{1,273} = 34.8$, $P = 1.09 \times 10^{-8}$, $\eta^2 = 0.113$, CO: $F_{1,273} = 9.1$, $P = 2.80 \times 10^{-3}$, $\eta^2 = 0.032$ and ST: $F_{1,273} = 23.9$, $P = 1.71 \times 10^{-6}$, $\eta^2 = 0.081$). Patients with schizophrenia scored higher than controls on HA and ST and lower on NS, RD, SD and CO. No significant difference in PS was found between patients and controls. We also investigated correlations between TCI and PANSS scores because some studies have reported that personality traits are correlated with psychopathology in patients with schizophrenia (Guillem et al., 2002; Hori et al., 2008; Smith et al., 2008). Patients showed significant positive correlations between HA and both the negative symptom domain ($r = 0.30$, $P = 0.0027$) and the depression/anxiety domain ($r = 0.32$, $P = 0.0017$) and negative correlations between SD and both the negative symptom domain ($r = -0.33$, $P = 8.81 \times 10^{-4}$) and the depression/anxiety domain ($r = -0.41$, $P = 3.67 \times 10^{-5}$) (Table 1). No correlations were

Please cite this article as: Ohi, K., et al., Personality traits and schizophrenia: evidence from a case–control study and meta-analysis, Psychiatry Res. (2012), doi:10.1016/j.psychres.2011.12.018

Table 1
Correlations between TCI scores and PANSS scores or CPZ-eq in patients with schizophrenia.

	NS	HA	RD	PS	SD	CO	ST
Positive[a]	−0.05	0.01	−0.10	<0.01	−0.05	−0.01	0.12
Negative[a]	−0.16	**0.30****	−0.21	−0.18	**−0.33*****	−0.16	−0.07
Cognitive[a]	−0.02	<0.01	−0.14	0.03	−0.16	0.01	0.21
Excitement[a]	0.07	−0.02	−0.05	−0.03	−0.12	−0.01	0.11
Depression/anxiety[a]	−0.17	**0.32****	−0.13	−0.01	**−0.41*****	−0.09	0.07
CPZ-eq (mg/day)[b]	0.11	−0.03	0.02	−0.05	−0.04	−0.02	0.17

TCI, Temperament and Character Inventory; PANSS, Positive and Negative Syndrome Scale; NS, novelty seeking; HA, harm avoidance; RD, reward dependence; PS, persistence; SD, self directedness; CO, cooperativeness; ST, self transcendence; CPZ-eq: Chlorpromazine equivalents of total antipsychotics. Significant P values <0.01 are in boldface.
[a] Five syndrome model of PANSS proposed by Lindenmayer et al. (1994). Each score represents Pearson's r (d.f. = 94).
[b] Pearson's r (d.f. = 97).
** $P<0.01$.
*** $P<0.001$.

identified between other TCI scores and PANSS scores. In addition, no correlation was observed between TCI score and CPZ-eq (Table 1).

3.2. Meta-analysis to estimate the differences between patients with schizophrenia and healthy controls

We found 34 relevant articles in PubMed using the search terms "TCI" or "temperament and character inventory" and "schizophrenia". Of these, a total of seven studies, including ours, met the inclusion criteria for our meta-analysis (with a total of 384 patients and 656 controls) (Guillem et al., 2002; Boeker et al., 2006; Hori et al., 2008; Smith et al., 2008; Cortes et al., 2009; Gonzalez-Torres et al., 2009). Although Gonzalez-Torres et al. (2009) included some patients with schizophrenia-spectrum disorders, such as schizoaffective disorder or schizotypal personality disorder, and Cortes et al. (2009) applied gender-corrected scores for each trait in their study, we included these two studies in the meta-analysis. The characteristics of the combined samples are shown in Table 2. There was heterogeneity in traits among studies only for NS ($I^2=86.2$, $P=9.05\times10^{-8}$). Thus, the fixed-effects model was applied for all traits except for NS (for which a random-effects model was used). The corresponding forest plots for each trait are presented in Fig. 1. We found significant differences in personality traits (HA, RD, PS, SD, CO and ST) between patients with schizophrenia and healthy controls in the combined sample (HA: Hedges' $g=0.98$, $P<1.00\times10^{-16}$, RD: $g=-0.43$, $P=2.20\times10^{-10}$, PS: $g=-0.23$, $P=7.67\times10^{-4}$, SD: $g=-0.96$, $P<1.00\times10^{-16}$, CO: $g=-0.47$, $P=5.61\times10^{-12}$ and ST: $g=0.61$, $P<1.00\times10^{-16}$). These results remained significant even after Bonferroni correction (all corrected threshold $P<0.0071$). There was no difference in NS scores between patients and controls ($g=-0.28$, $P=0.14$). Patients with schizophrenia scored higher on HA and ST and lower on RD, PS, SD and CO than controls. Marginally significant publication bias was observed for RD ($P=0.015$) and CO ($P=0.03$), but no evidence for publication bias was found for the other traits (Supplementary Fig. S1). We further analyzed subgroups by study location (Supplementary Table S2) and by excluding a possible confounding study (Supplementary Table S3). In the subgroup analysis, the effect size for NS was affected by study location. For Asian populations, we found no evidence for heterogeneity of NS between studies, and we detected a significant difference between patients and controls ($g=-0.72$, $P=1.09\times10^{-13}$). Conversely, for European populations, we found evidence of heterogeneity in NS among studies ($I^2=76.6$, $P=0.002$) and no difference between patients and controls ($g=-0.08$, $P=0.70$). The effect sizes for traits other than NS were not affected by study location. On the other hand, none of the effect sizes were affected by excluding a confounding study. We examined the impacts of potential moderator variables, including the gender distribution and mean age of the total sample. Gender distribution was a significant moderator of effect size for HA ($z=2.70$, $P=0.0069$, corrected $P=0.048$) and PS ($z=-2.32$, $P=0.02$, corrected $P=0.14$) (Supplementary Fig. S2). HA results remained significant even after excluding the study by Cortes et al. (2009), which applied gender-corrected scores ($z=2.69$, $P=0.0071$, corrected $P=0.0497$). The effect size for HA was positively correlated with the proportion of males in the study sample. Age could not be considered as a moderator (Supplementary Fig. S3).

4. Discussion

This is the first report that pools studies on comparisons of Cloninger's temperament and character traits between patients with schizophrenia and healthy controls. Our results suggest that patients with

Table 2
Demographic information included in the meta-analysis.

Authors	Country	Schizophrenia ($n=384$)						Control ($n=656$)				Diagnostic criteria
		N	Age	Sex (%M)	EY	AAO	DOI	N	Age	Sex (%M)	EY	
Guillem et al. (2002)	Canada (Montreal)	52	37.3 (9.6)	71.2	12.7 (2.6)	25.9 (8.1)	11.0 (8.2)	25	35.2 (7.8)	52.0	17.2 (4.1)	DSM-IV
Boeker et al. (2006)	Germany (Magdeburg)	22	29.7 (4.5)	45.5	11.2 (1.3)	N/A	N/A	22	27.9 (2.8)	45.5	11.8 (1.1)	DSM-IV
Hori et al. (2008)	Japan (Tokyo)	86	41.7 (11.3)	61.6	13.4 (2.3)	24.2 (7.5)	17.5 (10.9)	115	41.4 (11.3)	61.7	16.2 (3.0)	DSM-IV
Smith et al. (2008)	USA (Missouri)	35	22.9 (3.3)	82.9	11.6 (1.9)	N/A	N/A	63	21.0 (3.6)	46.0	13.2 (2.8)	DSM-IV
Cortes et al. (2009)	Spain (Reus)	29	30.5 (7.4)	86.2	11.5 (3.7)	23.4 (6.8)	6.7 (5.7)	188	40.9 (10.9)	50.0	N/A	DSM-IV
Gonzalez-Torres et al. (2009)	Spain (Bilbao)	61	31 (25–37)[a]	63.9	12 (10–14)[a]	N/A	N/A	64	33.5 (27–41)[a]	26.6	14 (13–15)[a]	DSM-IV
Ohi et al. (present study)	Japan (Osaka)	99	39.0 (12.1)	54.5	14.1 (2.3)	24.6 (9.0)	14.4 (11.1)	179	36.7 (11.5)	48.6	15.4 (2.4)	DSM-IV

%M: %male, EY: education years, AAO: age at onset (years), DOI: duration of the illness (years), N/A: not applicable. Means (S.D.).
[a] Means (ranges) are shown.

Please cite this article as: Ohi, K., et al., Personality traits and schizophrenia: evidence from a case–control study and meta-analysis, Psychiatry Res. (2012), doi:10.1016/j.psychres.2011.12.018

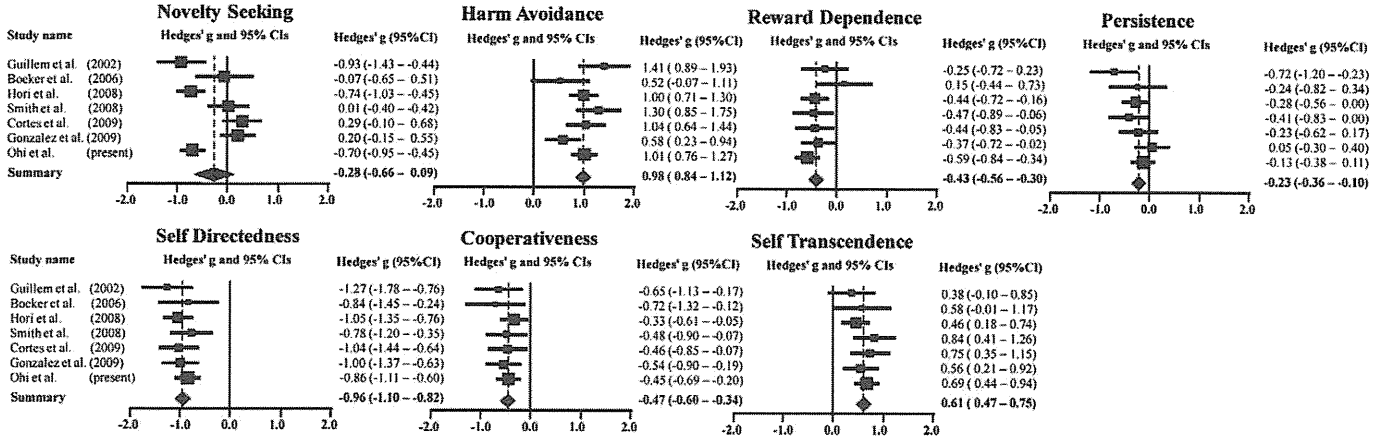

Fig. 1. Forest plots demonstrating the effect size estimates of the differences in each trait between patients with schizophrenia and healthy controls. The results are presented using effect sizes with 95% CIs in forest plots for each study. The diamond in the bottom portion represents the pooled effect size with a 95% CI. A positive effect size means that patients with schizophrenia score higher than healthy controls, while a negative effect size means that patients score lower than controls.

schizophrenia have distinguishable personality traits relative to healthy controls. Subgroup analysis showed that, with the exception of NS, the differences between patients and controls were not affected by including a potentially confounding study or by study location. Meta-regression analysis also showed that the differences were not affected by moderators such as age or sex, except for the effect of gender distribution on HA.

The main outcome of this study is the identification of unique differences in temperament and character between patients with schizophrenia and healthy controls. Patients with schizophrenia scored higher on HA and ST and lower on RD, PS, SD and CO compared with healthy controls. The effect sizes for the temperament traits were 0.98 for HA, −0.43 for RD and −0.23 for PS, and those of the character traits were −0.96 for SD, −0.47 for CO and 0.61 for ST. The average effect sizes for personality traits ranged from small (PS) to medium (RD, CO and ST) to large (HA and SD). Our meta-analysis indicates that there was little variation in effect sizes for personality traits between studies, despite differences in protocol, location, gender distribution and mean age. As these differences may have influenced the results, we will discuss the influences of these potentially confounding factors on the identified associations between schizophrenia and specific traits of personality.

The subjects and methods used in the studies included in this meta-analysis were similar across studies. All patients included in each study were diagnosed according to DSM-IV criteria. Only patients with schizophrenia and healthy controls were included in each study. These studies used raw scores of each personality trait assessed by TCI. However, there were two exceptions: one study included some patients with schizophrenia spectrum disorders, such as schizoaffective disorder or schizotypal personality disorder (Gonzalez-Torres et al., 2009). Another study used gender-corrected scores for each trait (Cortes et al., 2009). It could be argued that including such studies could introduce heterogeneity into the results. However, a subgroup analysis excluding these studies from the meta-analysis showed that their results did not alter the overall findings.

As personality traits assessed by the TCI among the general population vary across cultures (Pelissolo and Lepine, 2000; Brandstrom et al., 2001), it is possible that cultural differences in personality affected the results of the studies. To examine this possibility, we performed a subgroup analysis by study location (Asia versus Europe). There was no cross-cultural difference in the traits of personality, with the exception of NS. We detected a significant difference in NS between patients and controls in Asian populations, while we did not detect this difference in European populations. Patients with schizophrenia showed lower NS scores than controls in Asian populations, suggesting a cross-cultural difference. However, it should be noted that we found significant heterogeneity among the studies undertaken in European populations when studies in Asian populations were excluded. This result means that possible confounding factors still exist in European studies of NS.

Trouillet and Gana (2008) showed a substantial decrease in NS with increasing age. However, our results were not affected by the mean age of the study population, suggesting that the differences in personality traits between patients and controls are stable and unrelated to the mean age of the subjects in each study. As the proportion of males in each study's overall sample increased from 42.8% to 67.0%, the effect size for HA increased from 0.6 to 1.3. A previous meta-analysis reported that females scored higher on HA than males in healthy populations (Miettunen et al., 2007). By contrast, male patients with schizophrenia scored significantly higher on HA than female patients did (Hori et al., 2008). As a result, if the proportion of males in the total sample were to increase, it would create a substantial difference in HA between patients with schizophrenia (male > female) and healthy controls (male < female) as patients scored higher on HA than controls.

Several studies have assessed for clinical symptoms in their participants, particularly patients with schizophrenia. Two studies including our study examined correlations between the TCI and the clinical symptoms using the PANSS (Hori et al., 2008), while others examined the correlations using the Scale for Assessment of Positive Symptoms (SAPS)/Scale for Assessment of Negative Symptoms (SANS) (Guillem et al., 2002; Smith et al., 2008). We found that patients with schizophrenia showed significant positive correlation between HA and negative symptom domain, and negative correlation between SD and the negative symptom domain. Guillem et al. (2002) found that patients showed significant positive correlations between psychotic dimension and both NS and ST, and negative correlation between the psychotic dimension and SD. Hori et al. (2008) found that patients showed significant positive correlation between positive symptom domain and ST, and negative correlations between negative symptom domain and RD, SD and CO. Smith et al. (2008) found that patients with schizophrenia showed significant positive correlations between NS and negative symptom domain, HA and both positive and negative symptom domains, and ST and positive symptom domain. They also found that healthy subjects showed significant positive correlation between HA and positive symptom domain. These results were inconsistent across studies. Each of these studies showed that patients with schizophrenia had different types of psychopathology correlated with several personality traits. The differences in findings could be due to the variation in sample across studies and/or the use of different measures across studies. On the other hand, Guillem et al. (2002), Smith et al. (2008), and Hori et al. (2008), except for the present study, found correlations between ST and positive/psychotic symptom domain. This appears to be a consistent finding across studies and across methods (i.e., PANSS and SAPS). Future research could benefit by examining this relationship further.

Please cite this article as: Ohi, K., et al., Personality traits and schizophrenia: evidence from a case–control study and meta-analysis, Psychiatry Res. (2012), doi:10.1016/j.psychres.2011.12.018

This study has several limitations. Publication bias was found for RD and CO. As there are relatively few published studies comparing patients with schizophrenia to healthy controls, a much larger sample size would be needed to definitively test the associations between these traits and schizophrenia. In all studies, personality traits were assessed in patients after the onset of symptoms. Careful interpretation of our results is called for because we did not consider whether the findings reflect pre-clinical personality traits versus pre- or post-therapeutic personality traits. Further study is needed to clarify the effects of premorbid state or treatment on personality traits.

As personality assessment is based on self-report, it is not an objective measurement. However, our results showed consistent differences in personality traits between patients with schizophrenia and controls, suggesting that the results of subjective personality assessment are easy to obtain and yield reliable data. In conclusion, our data suggest that patients with schizophrenia have a unique temperament and character profile compared with the general population.

Acknowledgments

We thank all subjects who participated in this study. This work was funded in part by grants-in-aid from the Japanese Ministry of Health, Labor and Welfare (H19-kokoro-002 and Comprehensive Research on Disability Health and Welfare), the Japanese Ministry of Education, Culture, Sports, Science and Technology, CREST of JST, grant-in-aid for scientific research on priority areas – Research on Pathomechanisms of Brain Disorders – from the MEXT and Japan Foundation for Neuroscience and Mental Health. The study sponsors had no further role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Appendix A. Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.psychres.2011.12.018.

References

Boeker, H., Kleiser, M., Lehman, D., Jaenke, L., Bogerts, B., Northoff, G., 2006. Executive dysfunction, self, and ego pathology in schizophrenia: an exploratory study of neuropsychology and personality. Comprehensive Psychiatry 47, 7–19.

Borenstein, M., Hedges, L., Higgins, J., Rothstein, H., 2005. Comprehensive Meta-analysis Version 2. Biostat, Englewood NJ.

Brandstrom, S., Richter, J., Przybeck, T., 2001. Distributions by age and sex of the dimensions of temperament and character inventory in a cross-cultural perspective among Sweden, Germany, and the USA. Psychological Reports 89, 747–758.

Calvo de Padilla, M., Padilla, E., Gonzalez Aleman, G., Bourdieu, M., Guerrero, G., Strejilevich, S., Escobar, J.I., Svrakic, N., Cloninger, C.R., de Erausquin, G.A., 2006. Temperament traits associated with risk of schizophrenia in an indigenous population of Argentina. Schizophrenia Research 83, 299–302.

Cloninger, C.R., 1987. A systematic method for clinical description and classification of personality variants. A proposal. Archives of General Psychiatry 44, 573–588.

Cloninger, C.R., Svrakic, D.M., Przybeck, T.R., 1993. A psychobiological model of temperament and character. Archives of General Psychiatry 50, 975–990.

Cooper, H., Hedges, L., 1994. Potentials and limitations of research synthesis. In: Cooper, H., Hedges, L. (Eds.), The Handbook of Research Synthesis. Russell Sage Foundation Publications, New York.

Cortes, M.J., Valero, J., Gutierrez-Zotes, J.A., Hernandez, A., Moreno, L., Jariod, M., Martorell, L., Vilella, E., Labad, A., 2009. Psychopathology and personality traits in psychotic patients and their first-degree relatives. European Psychiatry 24, 476–482.

Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315, 629–634.

Gonzalez-Torres, M.A., Inchausti, L., Ibanez, B., Aristegui, M., Fernandez-Rivas, A., Ruiz, E., Fernandez, E., Bayon, C., 2009. Temperament and character dimensions in patients with schizophrenia, relatives, and controls. The Journal of Nervous and Mental Disease 197, 514–519.

Guillem, F., Bicu, M., Semkovska, M., Debruille, J.B., 2002. The dimensional symptom structure of schizophrenia and its association with temperament and character. Schizophrenia Research 56, 137–147.

Hori, H., Noguchi, H., Hashimoto, R., Nakabayashi, T., Saitoh, O., Murray, R.M., Okabe, S., Kunugi, H., 2008. Personality in schizophrenia assessed with the Temperament and Character Inventory (TCI). Psychiatry Research 160, 175–183.

Jaeger, J., Douglas, E., 1992. Neuropsychiatric rehabilitation for persistent mental illness. Psychiatric Quarterly 63, 71–94.

Lindenmayer, J.P., Bernstein-Hyman, R., Grochowski, S., 1994. A new five factor model of schizophrenia. Psychiatric Quarterly 65, 299–322.

Lysaker, P., Bell, M., 1995. Negative symptoms and vocational impairment in schizophrenia: repeated measurements of work performance over six months. Acta Psychiatrica Scandinavica 91, 205–208.

Lysaker, P.H., Bell, M.D., Kaplan, E., Bryson, G., 1998. Personality and psychosocial dysfunction in schizophrenia: the association of extraversion and neuroticism to deficits in work performance. Psychiatry Research 80, 61–68.

Lysaker, P.H., Bell, M.D., Kaplan, E., Greig, T.C., Bryson, G.J., 1999. Personality and psychopathology in schizophrenia: the association between personality traits and symptoms. Psychiatry 62, 36–48.

Miettunen, J., Veijola, J., Lauronen, E., Kantojarvi, L., Joukamaa, M., 2007. Sex differences in Cloninger's temperament dimensions — a meta-analysis. Comprehensive Psychiatry 48, 161–169.

Pelissolo, A., Lepine, J.P., 2000. Normative data and factor structure of the Temperament and Character Inventory (TCI) in the French version. Psychiatry Research 94, 67–76.

Smith, M.J., Cloninger, C.R., Harms, M.P., Csernansky, J.G., 2008. Temperament and character as schizophrenia-related endophenotypes in non-psychotic siblings. Schizophrenia Research 104, 198–205.

Trouillet, R., Gana, K., 2008. Age differences in temperament, character and depressive mood: a cross-sectional study. Clinical Psychology & Psychotherapy 15, 266–275.

Please cite this article as: Ohi, K., et al., Personality traits and schizophrenia: evidence from a case–control study and meta-analysis, Psychiatry Res. (2012), doi:10.1016/j.psychres.2011.12.018

Toxicology in Vitro 25 (2011) 796–809

Contents lists available at ScienceDirect

Toxicology in Vitro

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

Validation study of the Short Time Exposure (STE) test to assess the eye irritation potential of chemicals

Hitoshi Sakaguchi [a,*], Naoko Ota [b], Takashi Omori [c], Hirofumi Kuwahara [d], Takashi Sozu [e], Yumi Takagi [f], Yutaka Takahashi [a], Kouko Tanigawa [g], Miki Nakanishi [g], Tsuneaki Nakamura [h], Takashi Morimoto [i], Shinobu Wakuri [j], Yuko Okamoto [g], Mayumi Sakaguchi [b], Takumi Hayashi [d], Takayuki Hanji [f], Shinichi Watanabe [h]

[a] *Kao Corporation, Safety Science Research Laboratories, 2606 Akabane, Ichikai-Machi, Haga-Gun, Tochigi 321-3497, Japan*
[b] *Pola Chemical Industries, Inc., Quality Research Department, 560 Kashio-cho, Totsuka-ku, Yokohama-shi, Kanagawa 244-0812, Japan*
[c] *Doshisha University, Faculty of Culture and Information Science, 1-3 Tatara Miyakodani, Kyotanabe City, Kyoto 610-0394, Japan*
[d] *Kanebo Cosmetics Inc., Quality Management Department, 3-28, 5-chome, Kotobuki-cho, Odawara-shi, Kanagawa 250-0002, Japan*
[e] *Kyoto University, Department of Biostatistics Kyoto University School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan*
[f] *Pias Corporation, Safety Research Laboratories, 1-3-1 Murotani, Nishi-ku, Kobe-shi, Hyogo 651-2241, Japan*
[g] *KOSÉ Corporation, Fundamental Research Laboratories, 1-18-4, Azusawa, Itabashi-ku, Tokyo 174-0051, Japan*
[h] *LION Corporation, Human & Environmental Safety Evaluation Center, 100 Tajima, Odawara-shi, Kanagawa 256-0811, Japan*
[i] *Sumitomo Chemical Co., Ltd., Environmental Health Science Laboratory, 1-98, Kasugadenaka 3-chome, Konohana-ku, Osaka-shi, Osaka 554-8558, Japan*
[j] *Food and Drug Safety Center, Laboratory of Cell Toxicology, 729-5, Ochiai, Hadan-shi, Kanagawa 257-8523, Japan*

ARTICLE INFO

Article history:
Received 4 November 2010
Accepted 26 January 2011
Available online 1 February 2011

Keywords:
Eye irritation
Alternative method
Validation
STE
SIRC cells
GHS

ABSTRACT

Short Time Exposure (STE) test is a cytotoxicity test in SIRC cells (rabbit corneal cell line) that assesses eye irritation potential following a 5-min chemical exposure. This validation study assessed transferability, intra- and inter-laboratory reproducibility, and predictive capacity of STE test in five laboratories (supported by Japanese Society for Alternatives to Animal Experiments). Sodium lauryl sulfate, calcium thioglycolate, and Tween 80 were evaluated, in triplicate, using 5%, 0.5%, and 0.05% concentrations in physiological saline, to confirm transferability. Good transferability was noted when similar mean relative viabilities and rank classifications were obtained in all five laboratories and were comparable to data from test method developing laboratory. Good intra- and inter-laboratory reproducibility was obtained with four assay controls (three solvents and one positive control), and four assay controls and 25 chemicals, respectively. STE irritation category based on relative viability of a 5% solution of 25 blinded test chemicals showed good correlation with Globally Harmonized System (GHS) categories (NI; I: Cat. 1 and 2). The STE prediction model, using relative viability of the 5% and 0.05% solutions, provided an irritation rank (1, 2, or 3) that had a good correlation (above 80%), or predictive capacity, with GHS irritation ranks in all laboratories. Based on these findings, the STE test is a promising alternative eye irritation test that could be easily standardized.

1. Introduction

Eye irritation potential of chemicals has been evaluated mainly by the rabbit Draize test for many years. Recently, animal welfare considerations and enforcement of a new regulations like the banning of cosmetics in animal eye irritation tests in the EU (Directive 2003/15/EC, 2003) are drawing much attention in reducing or replacing animal experiments with alternative methods. In September 2009, the bovine corneal opacity and permeability (BCOP) assay and isolated chicken eye (ICE) assay were adopted by the Organisation for Economic Cooperation and Development (OECD) for identifying ocular corrosives and severe irritants, i.e., OECD Test Guideline 437 or 438, respectively (OECD, 2009a,b). However, no in vitro test has been approved by OECD so far to evaluate minimal to moderate eye irritation potential. The STE test is a Short Time Exposure cytotoxicity test that uses SIRC cells (rabbit corneal cell line) to evaluate minimal, moderate or severe eye irritation potential (Takahashi et al., 2008). This test

Abbreviations: BCOP, bovine corneal opacity and permeability; CT, calcium thioglycolate; DMSO, dimethyl sulfoxide; ECETOC, European Centre for Ecotoxicology and Toxicology of Chemicals; ECVAM, European Centre for the Validation of Alternative Methods; GHS, Globally Harmonized System; HCE, human corneal epithelial; I, irritant; ICE, isolated chicken eye; JSAAE, Japanese Society for Alternatives to Animal Experiments; MTT, methylthiazolydiphenyl-tetrazolium bromide; NI, non-irritant; OECD, Organisation for Economic Cooperation and Development; REML, Restricted Maximum Likelihood; SLS, sodium lauryl sulfate; SD, standard deviation; STE, Short Time Exposure; TW80, Tween 80.

* Corresponding author. Tel.: +81 285 68 7447; fax: +81 285 68 7452.
E-mail address: sakaguchi.hitoshi@kao.co.jp (H. Sakaguchi).

0887-2333/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tiv.2011.01.015

uses cell viability as an endpoint after 5 min of exposure and solves the problems associated with conventional cytotoxicity tests. For example, the STE test has the advantage of being able to evaluate the eye irritation potential of water insoluble chemicals using mineral oil as solvent. It is also very simple to use and provides rapid results. In collaborative research in three laboratories evaluating the STE test, similar test results were obtained for positive and negative controls, indicating that the STE test has excellent "transferability" (Takahashi et al., 2009). In addition, in an evaluation of 51 chemicals, high reproducibility and a good predictive capacity were found in each of the three laboratories (inter-laboratory accuracy was 98.0% or higher; Takahashi et al., 2009). However, each of the three laboratories was not blinded to the identity of the test chemicals. In the present validation study conducted by the Japanese Society for Alternatives to Animal Experiments (JSAAE), the number of laboratories performing evaluations was increased to five laboratories to confirm the transferability of the STE test using three standard chemicals. The usefulness of the alternative STE test for evaluating eye irritation was further determined by establishing the intra- and inter-laboratory reproducibility of four assay controls, and inter-laboratory reproducibility and predictive capacity of STE test (i.e., agreement with the irritation category and rank of the GHS) of the 25 blinded test chemicals.

2. Materials and methods

2.1. Organization for STE test validation

The Executive Committee formed a number of committees to ensure appropriate execution of the study. The STE Test Validation Executive Committee comprised of members of the JSAAE (Validation Committee members), biostatisticians, and representatives of the participating validation laboratories. The STE Test Validation Executive Committee undertook deliberations and made decisions regarding the study plan, test protocol, confirmed the data analysis results, and published the study results. The Validation Management Team was organized under the STE Test Validation Executive Committee and comprised of the Chief Manager, Test Chemical Manager, Biostatistician, Experiment Record Manager, Accountant, and Test Development Manager. This committee managed the operations of the validation study. The five participating laboratories are as follows:

Laboratory 1 (Lab. 1), KOSÉ Corporation, Fundamental Research Laboratories.
Laboratory 2 (Lab. 2), Kanebo Cosmetics Inc., Quality Management Department.
Laboratory 3 (Lab. 3), Pias Corporation, Safety Research Laboratories.
Laboratory 4 (Lab. 4), LION Corporation, Human & Environmental Safety Evaluation Center.
Laboratory 5 (Lab. 5), Pola Chemical Industries, Inc., Quality Research Department.

The participating laboratories conducted the experiments, printed out the absorbance measured during the experiment, and inputted the value into the data sheet. The printout and the electronic file were then submitted to the Biostatistician, who then confirmed the consistency of the data printed out with the data inputted into the electronic file. If the values differed, inquiries were made to the laboratories involved, and the value was changed to reflect the correct value.

2.2. Globally Harmonized System (GHS)

The GHS is a system by which, according to globally standardized rules, a chemical is classified as to its type and degree of hazard and labeling so that the information can be understood easily when conveyed in a material safety data sheet for that chemical (United Nations, 2003). For eye irritation, there are several classifications which are mainly based on Draize test results using rabbits: irreversible eye effects (Category 1), reversible eye effects (Categories 2A, 2B, or 2A–B), and not-classified (not an eye irritant; hereafter non-irritant (NI)). In the present report, GHS Categories 1 and 2 are combined and termed as irritants (I). In addition, similar to the approach of Goethem et al. (2006), analysis was performed and chemicals were classified into three rank levels (Category 1, Category 2, and NI) of eye irritation.

2.3. Test chemicals

In the pretest, three standard chemicals (SLS, sodium lauryl sulfate; CT, calcium thioglycolate; and TW80, Tween 80) were evaluated to confirm transferability. SLS and TW80 were purchased from Sigma–Aldrich (St. Louis, MO, USA) and CT was purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan).

In the present validation study, 25 chemicals were selected. Most chemicals had already being used in for the prevalidating of other alternative eye irritation test methods, for which results have been reported (Goethem et al., 2006). In addition, the other chemicals were selected for which data have been published by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC, 1998), Gautheron et al. (1992), or Ohno et al. (1999). The details on these 25 test chemicals, such as the code numbers, chemical names, CAS numbers, suppliers, purities, and GHS rankings, are shown in Table 1. In terms of the 25 chemicals selected, the corresponding categories were as follows: five chemicals were Category 1; seven chemicals were Category 2; and 13 chemicals were non-irritants (NI). The test chemicals covered the whole range of eye irritation potencies and represented different chemical classes. The test chemicals used were encoded by Chemical Manager and participating laboratories were blinded to the identification of the test chemicals.

2.4. Solvent selection and sample preparation

The work flow from solvent selection to sample preparation is described in Fig. 1.

First, a 5% (w/w) solution of a test chemical was prepared using physiological saline (Otsuka Pharmaceutical Co., Ltd., Tokyo, Japan) as solvent and the dissolution pattern of the chemical was observed. If the chemical dissolved or homogeneously dispersed (Notes 1 and 2), physiological saline was chosen as the solvent for this chemical.

When the chemical did not dissolve or homogeneously disperse in physiological saline, physiological saline containing 5% (w/w) dimethyl sulfoxide (DMSO, Sigma–Aldrich) was tried as the solvent. If the chemical dissolved or homogeneously dispersed in physiological saline containing 5% (w/w) DMSO, this became the solvent for this chemical.

When the chemical did not dissolve or homogeneously disperse in physiological saline containing 5% (w/w) DMSO, mineral oil (Sigma–Aldrich) was tried as the solvent. If the chemical dissolved or homogeneously dispersed in the mineral oil, a solution of 5% (w/w) chemical was made for the evaluation. If the chemical did not dissolve or homogeneously disperse in mineral oil, the chemical was not evaluated in this test.

Table 1
Information on 25 test chemicals. Code number, test chemical name, CAS number, supplier, purity, and GHS rank of 25 test chemicals are shown. TCI: Tokyo Chemical Industry Co., Ltd., Wako: Wako Pure Chemical Industries, Ltd. ∗1: GHS class: Globally Harmonized System of Classification and Labeling of Chemicals (United Nations, 2003) NI: non-irritant; cat 2 (category 2): irritating to the eye; cat 1 (category 1): irreversible effects on the eye.

Code	Test chemical	CAS No.	Supplier	Purity (%)	GHS
A	3-Methoxy-1,2-propanediol	623-39-2	TCI	>98	NI
B	Polyethyleneglycol 400	25322-68-3	TCI	–	NI
C	Glycerol	56-81-5	Wako	>99	NI
D	Tween 20	9005-64-5	Sigma	–	NI
E	Ethanol	64-17-5	Wako	99.8	cat 2
F	Sodium hydroxide	1310-73-2	Wako	>96	cat 1
G	Triton X-100	9002-93-1	Sigma	SigmaUltra	cat 1
H	Cetylpyridinium bromide	140-72-7	TCI	>96	cat 1
I	Benzalkonium chloride	8001-54-5	Sigma	–	cat 1
J	Methyl amyl ketone	110-43-0	Wako	⩾98	NI
K	2-Methyl-1-pentanol	105-30-6	TCI	⩾98	cat 2
L	*n*-Hexanol	111-27-3	Aldrich	98	cat 2
M	3,3-Dimethylpentane	562-49-2	Aldrich	99	NI
N	Methyl cyclopentane	96-37-7	TCI	⩾98	NI
O	Methyl isobutyl ketone	108-10-1	TCI	⩾99	NI
P	Toluene	108-88-3	Wako	⩾99	NI
Q	1-Octanol	111-87-5	Wako	⩾98	cat 2
R	2-Ethyl-1-hexanol	104-76-7	Wako	⩾98	cat 2
S	Acetone	67-64-1	Wako	⩾99	cat 2
T	Cyclohexanol	108-93-0	Sigma	99	cat 1
U	*n*,*n*-Dimethylguanidine sulfate	598-65-2	TCI	⩾97	NI
V	2-Ethylhexyl *p*-dimethyl-amino benzoate	21245-02-3	Aldrich	98	NI
W	Gluconolactone	90-80-2	TCI	⩾98	NI
X	Methyl ethyl ketone	78-93-3	TCI	⩾99	cat 2
Y	Propylene glycol	57-55-6	Wako	⩾99	NI

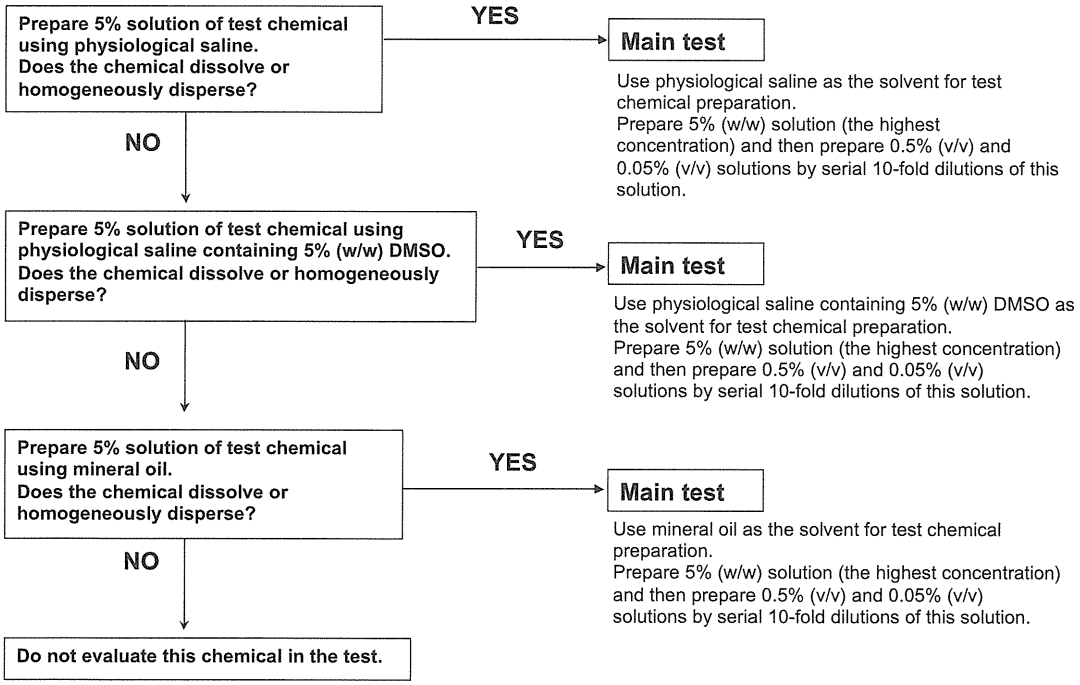

Fig. 1. Solvent selection and sample preparation. Three solvents were used in the STE test. Selection on solvent was dependent on solubility, physiological saline, physiological saline containing 5% (w/w) DMSO or mineral oil and was selected based on the process described in this diagram.

Note (1) A test chemical was considered homogeneously dispersed in a fluid and when this condition was maintained for 5 min or longer.
Note (2) Dissolution was aided by vortexing, sonication, or warming, as appropriate.

2.5. Assay controls

Three solvents (physiological saline, physiological saline with 5% (w/w) DMSO, and mineral oil) and positive control (0.01% SLS) were evaluated in each assay to monitor intra- and inter-laboratory reproducibility and to provide the necessary data for defining possible acceptance criteria.

2.6. Cell culture

SIRC cells were purchased from American Type Culture Collection (Cat. No. CCL60; Lot No. 3981569) by each laboratory. Cells were used between 3 weeks and 3 months after initiation of culture or up to 25th passage.

SIRC cells were cultured in a culture flask (37 °C, 5% CO_2) in Eagle MEM media (Sigma–Aldrich) containing 10% (v/v) fetal bovine serum (Invitrogen Corp., Carlsbad, CA, USA), 2 mM L-glutamine (Invitrogen Corp.), 50–100 U/mL of penicillin (Invitrogen Corp.), and 50–100 μg/mL streptomycin (Invitrogen Corp.). Confluent cells were dispersed in the culture flask to single cells by using trypsin–EDTA solution (Sigma–Aldrich); they were then transferred into culture flask or seeded onto 96-well plates (flat bottom, Tissue Culture Treated Polystyrene Sterile, Corning Inc., Lowell, MA, USA) for 200 μL at 3.0×10^4 cells/mL. After incubating (37 °C, 5% CO_2) for 4 days (or 1.5×10^4 cells/mL for 5 days), the cells reached confluence. Since two pre-culture conditions were used, we calculated the number of cells after pre-culture based on doubling time and confirmed with a microscope that both cell conditions reached confluence. Additionally we measured absorbance in wells for each pre-culture condition and confirmed that the values of absorbance were within a constant range.

$$\text{Cell Viability} = \frac{\text{Absorbance of Test Sample (Absorbance of Test Sample} - \text{Blank)}}{\text{Absorbance of Solvent Control (Absorbance of Solvent Control} - \text{Blank)}} \times 100$$

2.7. The Short Time Exposure (STE) test

The STE test was developed by the Kao Corporation (Takahashi et al., 2008). An overview of the STE test method is shown in Fig. 2. The procedures are described in greater detail below.

Before cells were exposured to test chemical, medium in each well was fully removed by suction using a suction tube. Cells were not rinsed prior to treatment. The cells in the 96-well plate were exposed to 200 μL of 5%, 0.5%, and 0.05% test chemical solutions prepared using physiological saline, physiological saline with 5% (w/w) DMSO or mineral oil for 5 min. Three wells in a 96-well plate were used for each concentration. After exposure, the cells were washed with phosphate buffered saline (−) [PBS (−); Takara Bio Inc., Siga, Japan] twice, and 200 μL of methylthiazolydiphenyl-tetrazolium bromide (MTT, Sigma–Aldrich) solution (0.5 mg MTT/mL of medium) was added. After a 2-h reaction time, MTT formazan was extracted with 0.04 N HCl–isopropanol for 60 min, and the absorbance of MTT formazan in the extract was measured at 570 nm with a plate reader (Lab. 1: μQUANT, BioTek Instrument Inc., Vermont, USA; Lab. 2: Viento® XS, DS Phama Biomedical Co., Ltd., Osaka, Japan; Lab. 3: Sunrise Rainbow Thermo, Tecan Japan Co., Ltd., Kanagawa, Japan; Lab. 4: MultiskanJX, Thermo Fisher Scientific Co., Ltd., Kanagawa, Japan; Lab. 5: Microplate Reader Model 680, Bio-Rad laboratories, Inc., Tokyo, Japan). The ratio (%) of MTT formazan absorbance for each test chemical to the absorbance of MTT formazan for control represented cell viability (triplicate determinations) using the following formula.

The mean of three wells for each test concentration was calculated for all materials. This was the mean cell viability for one independent experiment. If the mean value of the test sample absorbance of three wells is lower than the mean value of the solvent control absorbance, the cell viability value is considered a negative value and cell viability is designated as 0. A total of three independent experiments were conducted for each concentration of a test chemical, and the calculated overall mean of three independent experiments (hereafter termed the "relative viability") was used in the final analysis. After obtaining the relative viability in the STE test, which used a 5 min exposure to test chemicals with physiological saline, physiological saline with 5% DMSO, mineral oil, and 0.01% SLS, a category and rank classification were determined as described in Sections 2.9 and 2.10.

2.8. Validity criteria

For the STE test, there are four criterions for validity. The 1st, 2nd, and 3rd criterion are applied to each experiment. The 4th cri-

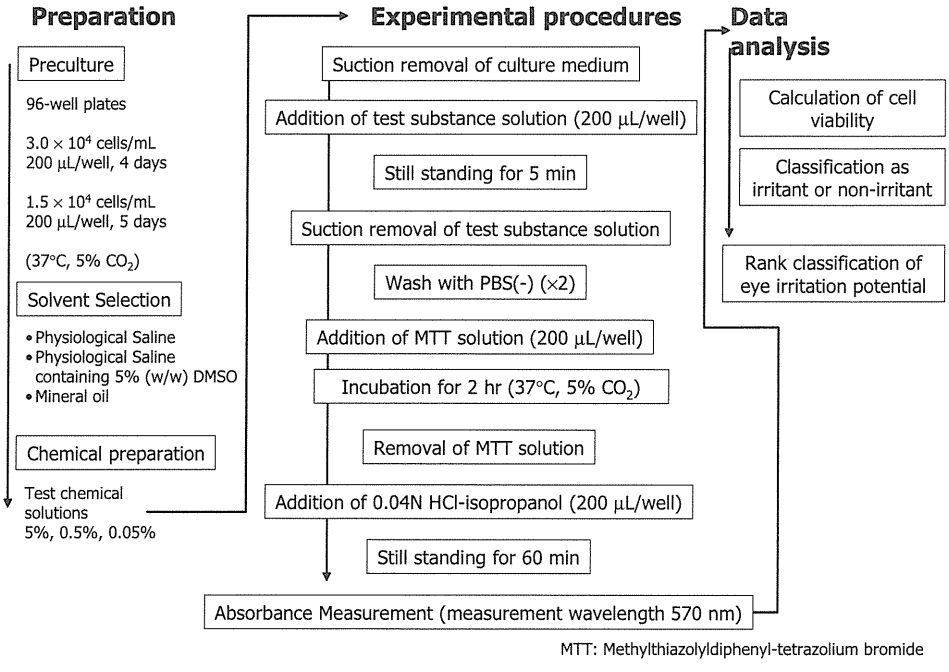

Fig. 2. STE test procedure. STE test consists mainly of three parts: preparation of cells and test chemicals, experimental procedure, and data analysis.

terion is applied to the outcome of three independent experiments. These four criteria are as follows:

Criterion 1: The absorbance after subtracting the blank of the medium operation control is 0.3 or higher.
Criterions 2a through 2c are solvent control criteria.
Criterion 2a: The cell viability in physiological saline is 80% or higher when the cell viability of the medium control is 100%.
Criterion 2b: The cell viability in physiological saline containing 5% (w/w) DMSO is 80% or higher when the cell viability of the medium control is 100%.
Criterion 2c: The cell viability in mineral oil is 80% or higher when the cell viability of the medium control is 100%.
Criterion 3: The cell viability of positive control (0.01% SLS) is in the range of 21.1–62.3% (i.e., within the range of the mean cell viability ± 2 standard deviation (SD), 41.7 ± 2 × 10.3%, according to background data (83 independent experiments) of test method developing laboratory).
Criterion 4: The standard deviation (SD) of the relative viability obtained with three experiments performed independently for each test chemical concentration is less than 15%. If SD is greater than 15%, a new determination will need to be made using the results from another three independent experiments until this criterion is met.

2.9. Category classification as irritant (I) or non-irritant (NI)

The relative viability after adding the test chemical and exposing for 5 min is the endpoint of the STE test. Under the 5% exposure condition, chemicals with relative viability values greater than 70% were considered non-irritant (NI), while those with relative viability values of 70% or less were considered irritant (I).

2.10. Rank classification

The eye irritation rank classification for STE tests is shown in Fig. 3. Under the 5% exposure condition of STE test, a score of 0 is given when relative viability is greater than 70%, and a score of 1 is given when relative viability is 70% or less. Under the 0.05% exposure condition of STE test, when relative viability is greater than 70%, the score is 1; when relative viability is 70% or less, the score is 2. The 5% and 0.05% scores are added together, and

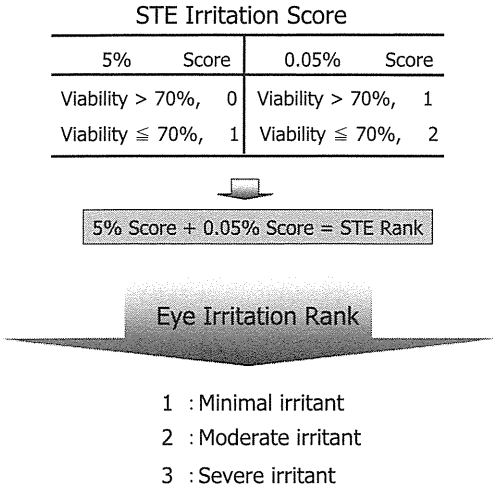

Fig. 3. Eye irritation rankings in STE test. Rank classification of eye irritation potential was carried out by the prediction model. The procedure for scoring and ranking a test chemical was described in Section 2.10. Rank 1, 2, and 3 correspond to "minimal irritant", "moderate irritant" and "severe irritant", respectively.

the strength of the chemical's eye irritation is classified based on the value obtained. A total score of 1 was classified as minimal irritant and given a Rank of 1; a total score of 2 was Rank 2 or classified as moderate irritant, and a total score of 3 resulted in a classification of severe irritant, a Rank of 3.

2.11. Pretest with standard chemicals

To confirm transferability for the STE test, three standard chemicals (SLS, CT, and TW80) were evaluated. Each laboratory prepared 5%, 0.5%, and 0.05% solutions using physiological saline as the solvent and three experiments were then performed for each concentration. For each chemical, relative viabilities for three concentrations and the rank based on results for the 5% and 0.05% solutions were compared to the background data obtained from the test method developing laboratory (Kao Corporation).

2.12. Intra-laboratory reproducibility

For three standard chemicals and 25 blinded chemicals, three independent experiments were evaluated only when the test was satisfied for the 1st through 4th criterion of the STE test. To analyze for intra-laboratory reproducibility, it was thought that the number of three was not appropriate. Therefore, four assay controls (physiological saline, 5% (w/w) DMSO in physiological saline, mineral oil, and 0.01% SLS) that were included in all experiments were used to analyze for intra-laboratory reproducibility. The primary and secondary analyses are outlined below.

2.12.1. Primary analysis

The SD of the mean cell viabilities for each assay control (physiological saline, physiological saline containing 5% (w/w) DMSO, mineral oil, and 0.01% SLS (positive control)) were obtained from each laboratory. The value of the SD in each laboratory was analyzed.

2.12.2. Secondary analysis

The SD of the square root of intra-laboratory variance is obtained by decomposing the total variance into the inter- and intra-variance components based on a random-effect model. This is calculated by Restricted Maximum Likelihood (REML) method. The value of the SD by REML method in each laboratory was analyzed.

In general, the range of permissible SD for analyses of intra-laboratory reproducibility was difficult to determine. In the analysis in the present validation study, since the unit of analysis is the mean of cell viability of three experiments, the SD is the value obtained by dividing 15% (Criterion 4) by the square root of 3, or equal to 8.7%. Therefore, we discuss the degrees of intra-laboratory reproducibility of relative viability with reference to 17.4% and 26.1%. These values are derived by multiplying two times the 8.7 value (17.4%) and three times the 8.7 value (26.1%) and using Criterion 4 as the basis.

2.13. Inter-laboratory reproducibility

In general, intra-laboratory reproducibility is better than inter-laboratory reproducibility. Therefore, we applied the same analysis method used for determining intra-laboratory reproducibility to the analysis for determining inter-laboratory reproducibility. Criterion 4 will be used for the inter-laboratory reproducibility analysis and the criterion states: "The standard deviation of the cell viability obtained with three experiments performed independently for each test sample concentration is less than 15%."

As described in Section 2.7, there is no repetition for relative viability in 25 test chemicals at each concentration. Therefore, in-

ter-laboratory reproducibility can be obtained by calculating the SD of each relative viability for five laboratories for each chemical at each concentration. On the other hand, since there are repetitions for four assay controls (physiological saline, physiological saline containing 5% (w/w) DMSO, mineral oil, and 0.01% SLS (positive control)), the SD was obtained by separating the total dispersion into inter-laboratory components. For this calculation, the REML method was used. The primary and secondary analyses are shown below:

2.13.1. Primary analysis

The SD of the relative viability was determined for 25 test chemicals at 5% and 0.05% as just described.

For the mean cell viabilities of the physiological saline, physiological saline with 5% (w/w) DMSO, mineral oil, and 0.01% SLS (positive control), the SD of the square root of the intra-laboratory variance was determined when the total dispersion was decomposed into inter- and intra-laboratory variance components.

2.13.2. Secondary analysis

The differences in STE category and rank classification between laboratories were confirmed. The inter-laboratory reproducibility of the category classification was analyzed by calculating the κ-value, which is a measure of agreement between the category classifications according to the different laboratories.

2.13.3. Others

For each test chemical, a scatter plot was constructed of mean cell viability (with a reference line drawn at 70%, the cut-off value for classifying STE rank), with 5% concentration on the vertical axis and 0.05% concentration on the horizontal axis.

2.14. Predictive ability

For each laboratory, the primary analysis provided the correspondence between the irritation categories of I and NI, as determined by the STE test at the 5% concentration and the two categories of I and NI by GHS. A 2 × 2 contingency table was provided by each laboratory for this purpose.

For the secondary analysis, the correspondence was obtained between the rank classification obtained with the STE test and the GHS rank (i.e., NI, Category 2 and Category 1). A 3 × 3 contingency table was provided by each laboratory for this purpose. The following are details of the primary and secondary analyses.

2.14.1. Primary analysis

The sensitivity, specificity, positive predictivity, negative predictivity, and accuracy for category classification were outlined through the 2 × 2 contingency table at each laboratory.

2.14.2. Secondary analysis

The accuracy for rank classification value was outlined through the 3 × 3 contingency table at each laboratory.

3. Results

3.1. Transferability data for three standard chemicals

Table 2 shows mean cell viability ± SD at three test concentrations and the eye irritation ranks calculated from the prediction model (i.e., results of the 5% and 0.05% test solutions) for the three standard chemicals obtained by the five laboratories. The background data for these chemicals obtained by the laboratory that developed the test method are also shown for comparison. The SLS had a cell viability of 7% or less; TW80 was 90% or higher at all three concentrations; and CT was 20% or less at a concentration of 5%, around 70% at a concentration of 0.5%, and 85% or higher at a concentration of 0.05%. Similar cell viability values were obtained from all five laboratories, and these cell viability values were similar to the background data of the test method developing laboratory (Table 2). Also, ranking of each chemical is identical among all labs including the test method developing laboratory, i.e., SLS, CT and TW80 were classified as Rank 3, 2, and 1, respectively (Table 2).

3.2. Intra-laboratory reproducibility data for four assay controls

The mean cell viability and SD from 9 to 11 experiments for four assay controls (physiological saline, physiological saline with 5% (w/w) DMSO, mineral oil, and 0.01% SLS (positive control)) from each laboratory are shown in Table 3. The mean value and SD of the cell viability for five laboratories are also shown in Table 3.

The SD of mean cell viability for the physiological saline in each laboratory ranged from 3.0% to 4.6%. The SD of physiological saline with 5% (w/w) DMSO was from 2.4% to 4.4%, and the SD of mineral oil was from 3.1% to 5.7%. The SD of mean cell viability for the physiological saline, physiological saline with 5% (w/w) DMSO, and mineral oil – where the mean cell viability was near 100% – was small. Each mean cell viability for 0.01% SLS in the five laboratory ranged from 34.2% to 47.6%; however, the SD for each laboratory was from 2.9% to 7.3% (Table 3). These SD were also thought as relatively small.

Table 2
Transferability data for three standard chemicals in STE test. Each value represents the mean cell viability ± SD% (n = 3 for Labs. 1–5, n = 18 for developing lab).

	Concentration	Lab. 1	Lab. 2	Lab. 3	Lab. 4	Lab. 5	Background data (developing lab)
SLS	5%	2.2 ± 0.7	2.1 ± 0.9	0.0 ± 0	0.0 ± 0	4.5 ± 4.0	0.2 ± 0.3
	0.50%	4.4 ± 0.7	2.8 ± 0.3	0.4 ± 0.6	1.6 ± 2.6	5.8 ± 5.1	0.4 ± 0.8
	0.05%	3.1 ± 2.2	1.6 ± 0.3	0.2 ± 0.4	0.4 ± 0.6	6.3 ± 2.7	0.4 ± 0.3
	Rank	3	3	3	3	3	3
Calcium thioglycollate	5%	18.5 ± 6.8	12.8 ± 3.6	13.7 ± 4.3	14.9 ± 1.3	17.5 ± 8.2	10.8 ± 2.2
	0.50%	64.3 ± 8.9	72.8 ± 5.9	72.0 ± 10.6	66.8 ± 12.9	76.3 ± 3.4	64.7 ± 10.1
	0.05%	107.1 ± 10.5	106.8 ± 7.7	99.0 ± 11.7	118.0 ± 3.0	87.9 ± 2.0	101.1 ± 5.2
	Rank	2	2	2	2	2	2
Tween 80	5%	110.3 ± 6.2	101.7 ± 1.9	102.5 ± 15.1	117.1 ± 4.7	103.4 ± 14.0	101.3 ± 8.0
	0.50%	101.3 ± 1.3	106.1 ± 6.0	98.7 ± 10.4	108.4 ± 2.3	109.5 ± 2.9	95.5 ± 5.5
	0.05%	103.8 ± 3.3	99.2 ± 3.6	99.0 ± 5.4	102.2 ± 1.9	95.7 ± 2.0	98.7 ± 5.6
	Rank	1	1	1	1	1	1

Table 3
Intra-/inter-laboratory reproducibility of mean cell viability for four assay controls (three solvents and positive control). The mean cell viability (%) and SD (%) are given for each laboratory for each of the solvents (physiological saline, physiological saline containing 5% (w/w) DMSO and mineral oil) and positive control (0.01% SLS) used in the study. Each laboratory repeated each experiment 9–11 times. In addition, the overall mean value ± SD% of the five laboratories for each assay control is also provided.

Laboratory		Lab. 1	Lab. 2	Lab. 3	Lab. 4	Lab. 5	Mean ± SD%
Number of experiment		11	9	12	9	10	
Physiological saline	Mean (%)	94.2	94.1	94.2	96.3	92.9	94.3 ± 1.2
	SD (%)	3.3	3.6	4.6	3.0	4.3	
Physiological saline containing 5% (w/w) DMSO	Mean (%)	91.7	92.8	94.7	91.2	92.3	92.5 ± 1.4
	SD (%)	4.4	2.9	2.9	2.4	3.0	
Mineral oil	Mean (%)	98.7	102.6	93.2	103.9	96.7	99.0 ± 4.4
	SD (%)	4.1	5.7	3.9	3.3	3.1	
0.01% SLS	Mean (%)	47.6	46.2	34.2	43.2	39.8	42.2 ± 5.4
	SD (%)	2.9	5.7	3.6	7.3	3.7	

Table 4
Intra-/inter-laboratory SD calculated by REAL method for three solvents and positive control.

		Physiological saline	5% (w/w) DMSO	Mineral oil	0.01% SLS
Intra	SD (%)	3.8	3.2	4.1	4.8
Inter	SD (%)	0.0	1.0	4.2	5.2

The intra-laboratory SD estimated by REML method with the random-effects model for physiological saline, physiological saline with 5% (w/w) DMSO, mineral oil, and 0.01% SLS are shown in Table 4 and these SD values were from 3.2% to 4.8%.

3.3. Inter-laboratory reproducibility data for four assay controls

The mean value and SD of the mean cell viability for four assay controls (physiological saline, physiological saline with 5% (w/w) DMSO, mineral oil, and SLS 0.01% (positive control)) from five laboratories were 94.3 ± 1.2%, 92.5 ± 1.4%, 99.0 ± 4.4%, and 42.2 ± 5.4%, respectively (Table 3). These SD were relatively small.

The inter-laboratory SD estimated by REML method using the random-effects model for physiological saline, physiological saline with 5% (w/w) DMSO, mineral oil, and 0.01% SLS are shown in Table 4 and the SD values were from 0.0% to 5.2%. The inter-laboratory SD values for cell viability turned out to be smaller or almost similar with the intra-laboratory SD values by the REML method (Table 4).

3.4. Predictive capacity – correspondence of STE irritation category (NI or I) to GHS irritation category (NI or I) for 25 blinded chemicals

The code number, chemical name, GHS category, and GHS rank are shown in Table 5. In addition, the relative viability at 5% and STE irritation category for each test chemical based on cell viability at 5% obtained from each laboratory are shown in Table 5. All laboratories selected same solvent for each of the chemicals tested: solvents used are shown in Table 5. Nineteen chemicals were identified correctly. The exceptions were chemical E (ethanol), J (methyl amyl ketone), R (2-ethyl-1-hexanol), S (acetone), W (gluconolactone), and X (methyl ethyl ketone). There were only two test chemicals for which classifications did not agree among all laboratories (i.e., chemical E (ethanol) and S (acetone)). Table 6 shows the corresponding relationships between the irritation category classifications as I and NI, according to cell viability of the 5% concentration in the STE test for each laboratory; also shown are two categories of I or NI, according to GHS. In the table 7, the sensitivity, specificity, positive predictivity, negative predictivity, and accuracy are shown for each laboratory. Sensitivity of each laboratory was from 75.0% to 83.3%. The combine sensitivity for all five laboratories was 78.3%. The individual specificity for each laboratory was from 92.3% to 100%, and the overall specificity was 96.9%. The positive predictivity/negative predicitivity for each laboratory was 100/81.3%, 100/81.3%, 90.9/85.7%, 100/86.7%, and 90.0/80.0%. The accuracy ranged from 84.0% to 92.0% for each laboratory and had a mean value of 88.0%. These findings provide data sufficient support of the STE as alternative method for eye irritation.

3.5. Predictive capacity – correspondence of STE irritation rank to GHS irritation Rank for 25 blinded chemicals

The STE rank, calculated from the relative viabilities at 5% and 0.05%, and the GHS rank of each test chemical are shown in Table 5. Fig. 4 shows a scatter plot of relative viability with the 5% concentration as the vertical axis and the 0.05% concentration as the horizontal axis for each test chemical. The reference line is drawn at 70%, which is the standard for classification of STE rank. Thus, among the 25 test chemicals used in the validation study, 18 chemicals were in-line with the GHS irritation rank and were identified correctly in all laboratories (Table 5). Three test chemicals (chemical E (ethanol), S (acetone), and T (cyclohexanol)) did not match correctly with GHS irritation rank in all laboratories. The other four test chemicals exhibiting discrepancy in only one or two laboratories were chemical J (methyl amyl ketone), R (2-ethyl-1-hexanol), W (gluconolactone), and X (methyl ethyl ketone) (Table 5). All these chemicals except chemical J (methyl amyl ketone) showed one rank discrepancy. Table 8 shows the corresponding relationships between STE rank (1, 2, and 3) and GHS rank (NI, Category 2 and Category 1). The accuracy between STE rank and GHS rank in each laboratory ranged from 80.0% to 88.0% (Table 8). These findings provide data sufficient support of the STE as alternative method for eye irritation.

3.6. Inter-laboratory reproducibility for 25 blinded chemicals

The SD of relative viability for the 25 test chemicals in five laboratories was basically not large (Table 5); indicating that the inter-laboratory reproducibility of relative viability was good. The chemicals with SD greater than 8.7% were six test chemicals, i.e., chemical D (Tween 20) at 5% concentration, J (methyl amyl ketone) at 5% and 0.05% concentrations, K (2-methyl-1-pentanol) at 0.05% concentration, R (2-ethyl-1-hexanol) at 5% concentration, V (2-eth-

Table 5
Summary data of the STE test results and GHS data for 25 chemicals. Code number, test chemical name, GHS category (bold and upper), and GHS rank (normal and lower), solvent used, relative viability for 5% and 0.05% test concentration (mean value of three experiments), and STE irritation category (NI or I)/rank (1, 2, and 3) of 25 test chemicals are shown. NI: non-irritant; cat 2 (category 2): irritating to the eye; cat 1 (category 1): irreversible effects on the eye. *Evaluation using mineral oil as solvent.

Code	Chemical name	GHS	Solvent used	Concentration	Lab. 1		Lab. 2		Lab. 3		Lab. 4		Lab. 5		Mean (%)	SD(%)
					Viability (%)	NI/I	Viability (%)	NI/I	Viability(%)	NI/I	Viability(%)	NI/I	Viability(%)	NI/I		
A	3-Methoxy-1,2-propanediol	**NI**	Saline	5%	95.7	NI	96.6	NI	101.0	NI	88.7	NI	91.1	NI	94.6	4.8
		NI		0.05%	104.7		98.1		99.1		94.2		93.6		98.0	4.5
				Rank	1		1		1		1					
B	Polyethyleneglycol 400	**NI**	Saline	5%	92.1	NI	98.0	NI	108.1	NI	99.1	NI	96.7	NI	98.8	5.8
		NI		0.05%	100.2		101.1		100.9		101.2		100.5		100.8	0.4
				Rank	1		1		1		1					
C	Glycerol	**NI**	Saline	5%	90.3	NI	103.4	NI	104.7	NI	97.2	NI	99.0	NI	98.9	5.7
		NI		0.05%	100.8		101.7		104.2		104.4		101.6		102.6	1.6
				Rank	1		1		1		1					
D	Tween 20	**NI**	Saline	5%	77.0	NI	101.4	NI	82.0	NI	101.3	NI	94.0	NI	91.1	11.2
		NI		0.05%	91.6		94.5		89.9		92.8		87.6		91.3	2.7
				Rank	1		1		1		1					
E	Ethanol	**I**	Saline	5%	99.8	NI	95.8	NI	99.9	NI	92.7	NI	101.3	NI	97.9	3.5
		cat 2		0.05%	101.7		101.5		98.5		103.5		97.3		100.5	2.5
				Rank	1		1		1		1					
F	Sodium hydroxide	**I**	Saline	5%	4.0	I	0.9	I	0.6	I	0.3	I	0.2	I	1.2	1.6
		cat 1		0.05%	3.7		0.5		1.0		0.5		0.0		1.1	1.5
				Rank	3		3		3		3					
G	Triton X-100	**I**	Saline	5%	3.9	I	2.7	I	0.4	I	0.7	I	0.4	I	1.6	1.6
		cat 1		0.05%	6.5		3.8		1.1		2.2		0.1		2.7	2.5
				Rank	3		3		3		3					
H	Cetylpyridinium bromide	**I**	Saline	5%	1.6	I	1.5	I	0.0	I	0.9	I	0.0	I	0.8	0.8
		cat 1		0.05%	12.7		5.9		2.8		5.3		2.7		5.9	4.1
				Rank	3		3		3		3					
I	Benzalkonium chloride	**I**	Saline	5%	3.9	I	2.8	I	1.4	I	1.7	I	0.6	I	2.1	1.3
		cat 1		0.05%	6.5		3.2		1.6		1.7		1.5		2.9	2.1
				Rank	3		3		3		3					
J	Methyl amyl ketone*	**NI**	Mineral oil	5%	85.1	NI	93.0	NI	63.0	I	73.3	NI	77.8	NI	78.4	11.4
		NI		0.05%	95.3		100.0		68.9		89.7		92.7		89.3	12.0
				Rank	1		3		1		1					
K	2-Methyl-1-pentanol*	**I**	Mineral oil	5%	10.6	I	2.1	I	3.3	I	9.1	I	3.7	I	5.8	3.8
		cat 2		0.05%	85.0		100.6		73.0		90.2		99.3		89.6	11.3
				Rank	2		2		2		2					
L	*n*-Hexanol	**I**	Mineral oil	5%	11.1	I	1.0	I	2.1	I	9.0	I	2.0	I	5.0	4.7
		cat 2		0.05%	99.3	103.2		92.9		96.9		104.2		99.3	4.6	
				Rank	2		2		2		2					
M	3,3-Dimethyl pentane*	**NI**	Mineral oil	5%	101.2	NI	89.3	NI	95.7	NI	98.5	NI	93.9	NI	95.7	4.5
		NI		0.05%	97.0		96.5		91.4		96.8		92.7		94.9	2.6
				Rank	1		1		1		1		1			
P	Toluene*	**NI**	Mineral oil	5%	89.3	NI	103.1	NI	83.0	NI	92.8	NI	96.1	NI	92.9	7.5
		NI		0.05%	93.5		99.8		83.3		90.8		100.8		93.6	7.2
				Rank	1		1		1		1					
Q	1-Octanol*	**I**	Mineral oil	5%	7.7	I	44	I	3.0	I	8.3	I	2.4	I	5.2	2.7
		cat 2		0.05%	104.7		111.9		89.9		94.9		103.1		100.9	8.6
				Rank	2		2		2		2					
R	2-Ethyl-1-hexanol*	**I**	Mineral oil	5%	64.3	I	70.2	NI	27.4	I	56.6	I	80.4	NI	59.8	20.1
		cat 2		0.05%	103.0		112.2		97.5		109.1		102.4		104.8	5.8

(*continued on next page*)

Table 5 (*continued*)

Code	Chemical name	GHS	Solvent used	Concentration	Lab. 1		Lab. 2		Lab. 3		Lab. 4		Lab. 5		Mean (%)	SD(%)
					Viability (%)	NI/I	Viability (%)	NI/I	Viability(%)	NI/I	Viability(%)	NI/I	Viability(%)	NI/I		
				Rank	2		1		2		2		1			
S	Acetone	**I**	Saline	5%	80.8	NI	96.9	NI	91.8	NI	84.6	NI	98.6	NI	90.6	7.7
		cat 2		0.05%	95.2		106.6		91.6		93.1		100.3		97.4	6.2
				Rank	1		1		1		1					
T	Cyclohexanol*	**I**	Mineral oil	5%		I	8.8	I	3.5	I	6.1	I	4.0	I	5.5	2.1
		cat 1		0.05%	97.9		101.5		100.7		99.7		109.6		101.9	4.5
				Rank	2		2		2		2					
U	*n,n*-Dimethylguanidine sulfate	**NI**	Saline	5%	83.4	NI	99.2	NI	92.8	NI	81.9	NI	92.8	NI	90.0	7.2
		NI		0.05%	100.5		110.2		102.4		98.9		104.5		103.3	4.4
				Rank	1		1		1		1					
V	2-Ethylhexyl p-dimethyl-amino benzoate*	**NI**	Mineral oil	5%	84.5	NI	88.8	NI	85.1	NI	101.2	NI	88.7	NI	89.6	6.7
		NI		0.05%	92.9		93.2		87.1		112.3		94.9		96.1	9.5
				Rank	1		1		1		1					
W	Gluconolactone	**NI**	Saline	5%	79.1	NI	78.7	NI	73.0	NI	74.8	NI	65.8	I	74.3	5.4
		NI		0.05%	90.7		83.4		93.5		90.9		83.9		88.5	4.5
				Rank	1		1		1		2					
X	Methyl ethyl ketone	**I**	Saline	5%	71.0	NI	36.0	I	55.8	I	43.2	I	51.8	I	51.6	13.3
		2		0.05%	100.2		103.4		98.9		93.0		102.0		99.5	4.0
				Rank	1		2		2		2		2			
Y	Propylene glycol	**NI**	Saline	5%	94.4	NI	96.6	NI	96.2	NI	96.0	NI	100.0	NI	96.6	2.1
		NI		0.05%	95.2		101.1		96.4		97.2		99.2		97.8	2.3
				Rank	1		1		1		1		1			

Table 6
Correlation of STE test and GHS eye irritation category in each laboratory. Each value represents the number of chemicals designated as NI or I by STE test and GHS classification.

	STE (5%)		Sum
	NI (CV > 70)	I (CV < 70)	
Lab. 1 GHS			
NI	13	0	13
I (cat. 1 and 2)	3	9	12
Sum	16	9	25
Lab. 2 GHS			
NI	13	0	13
I (cat. 1 and 2)	3	9	12
Sum	16	9	25
Lab. 3 GHS			
NI	12	1	13
I (cat. 1 and 2)	2	10	12
Sum	14	11	25
Lab. 4 GHS			
NI	13	0	13
I (cat. 1 and 2)	2	10	12
Sum	15	10	25
Lab. 5 GHS			
NI	12	1	13
I (cat. 1 and 2)	3	9	12
Sum	15	10	25

Table 7
Overall performance of the STE test using GHS classification in each laboratory for 25 test chemicals. Each value represents the percentage of each performance item in laboratory 1–5.

	Lab. 1	Lab. 2	Lab. 3	Lab. 4	Lab. 5	Mean
Sensitivity (%)	75.0	75.0	83.3	83.3	75.0	78.3
Specificity (%)	100	100	92.3	100	92.3	96.9
Positive predictivity (%)	100	100	90.9	100	90.0	96.2
Negative predictivity (%)	81.3	81.3	85.7	86.7	80.0	83.0
Accuracy (%)	88.0	88.0	88.0	92.0	84.0	88.0

ylhexyl *p*-dimethyl-amino benzoate) at 0.05% concentration, and X (methyl ethyl ketone) at 5% concentration.

Four of 25 test chemicals did not have the same category in one or two laboratories. These are chemical J (methyl amyl ketone (Lab. 3)), R (2-ethyl-1-hexanol (Labs. 2 and 5)), W (gluconolactone (Lab. 5)), and X (methyl ethyl ketone (Lab. 1)). The κ-values between laboratories were calculated (Table 9) and these values are from 0.75 ± 0.20 to 0.92 ± 0.20.

In the rank classification, test chemical J (methyl amyl ketone (Lab. 3)), R (2-ethyl-1-hexanol (Labs. 2 and 5)), W (gluconolactone (Lab. 5)), and X (methyl ethyl ketone (Lab. 1)) had variable results among the laboratories (Table 5 and Fig. 4). For test chemical J (methyl amyl ketone (Lab. 3)) showed different irritation scores at both 5% and 0.05% concentrations with other laboratories. This resulted in a predicted rank classification of two rankings lower compared to the other laboratories (Table 5). The remaining three chemicals for this group had different irritation scores at the 5% concentration in one or two laboratories, which resulted a predicted rank classification of one ranking lower or higher compared to the other laboratories (Table 5). From Fig. 4, the cell viabilities of most chemicals are concentrated in a similar area for each laboratory.

4. Discussion

By now, a lot of eye irritation alternative methods are being developed around the world (Balls et al., 1999; Ohno et al., 1999;

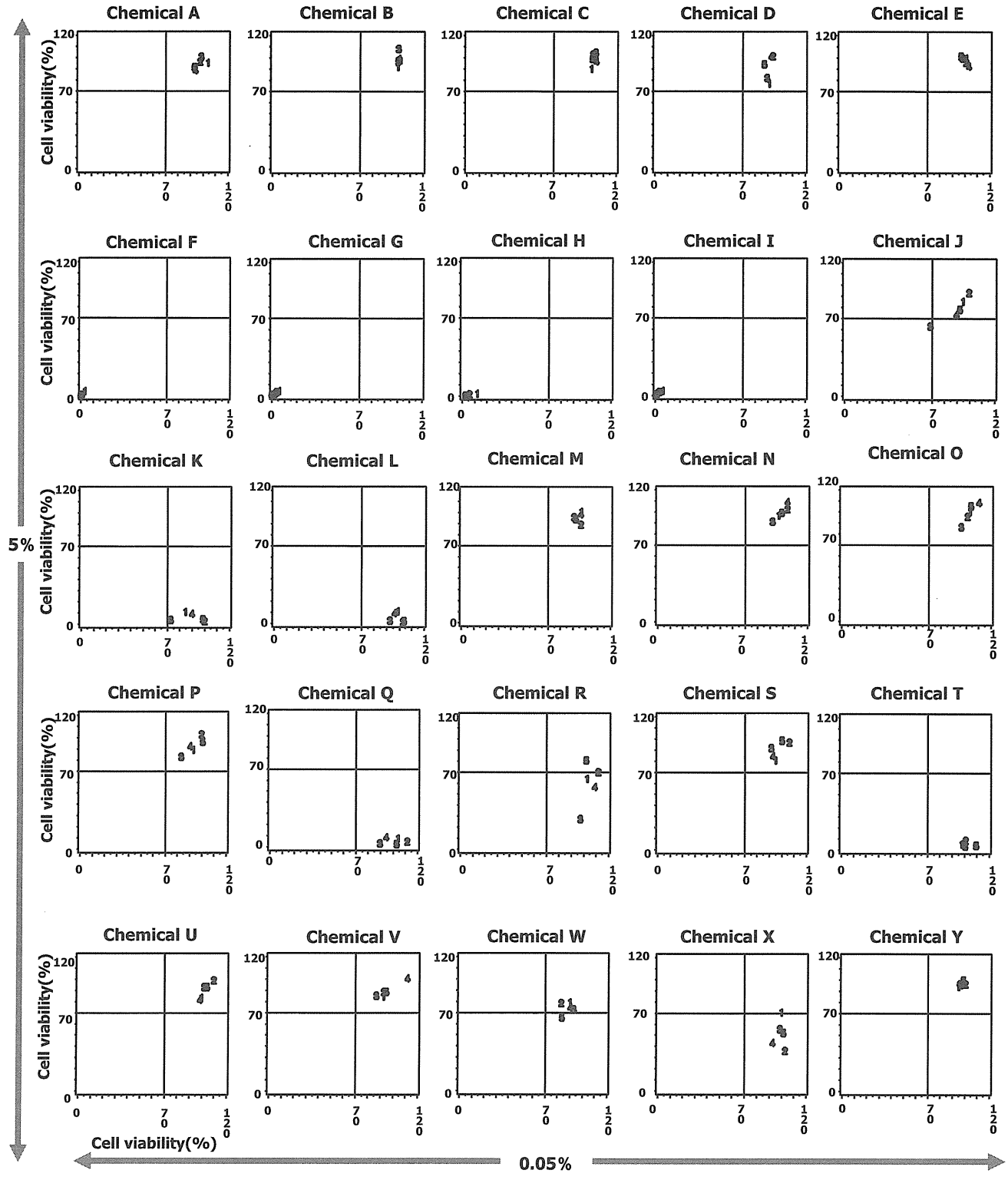

Fig. 4. Scatter charts of cell viability for 25 test chemicals. The scatter plots show mean relative viability with the 5% concentration as the vertical axis and the 0.05% concentration as the horizontal axis for each test chemical. The reference line is drawn at 70%, which is the standard for classification of STE rank.

Eskes et al., 2005). However, it may not be likely to completely replace the Draize test by a single in vitro test. The tiered approach of several in vitro assays combined was proposed in order to estimate the irritation potential for a wide range of chemical classes (Hagino et al., 2008; McNamee et al., 2009; Scott et al., 2010). Many eye irritation alternative methods can be classified into five main method types (Eskes et al., 2005). The 1st type is isolated organs (e.g., BCOP); the 2nd is organotypic methods such as the Chorio-Allantoic Membrane methods (e.g., Hens Egg Test on the Chorio-Allantoic Membrane assay); the 3rd one is reconstituted human tissue models (e.g., EpiOcular™ assay); the 4th one is cell-based cytotoxicity methods (e.g., Neutral Red Uptake assay); and the last

Table 8
Correlation of STE and GHS eye irritation ranking in each laboratory. Each value represents the number of chemicals classified as STE irritation rank 1, 2, or 3, and GHS rank NI, Category 2 or Category 1.

	STE rank			Sum
	1	2	3	
Lab. 1 GHS				
NI	13	0	0	13
Category 2	3	4	0	7
Category 1	0	1	4	5
Sum	16	5	4	25
Accuracy		84.0%		
Lab. 2 GHS				
NI	13	0	0	13
Category 2	3	4	0	7
Category 1	0	1	4	5
Sum	16	5	4	25
Accuracy		84.0%		
Lab. 3 GHS				
NI	12	0	1	13
Category 2	2	5	0	7
Category 1	0	1	4	5
Sum	14	6	5	25
Accuracy		84.0%		
Lab. 4 GHS				
NI	13	0	0	13
Category 2	2	5	0	7
Category 1	0	1	4	5
Sum	15	6	4	25
Accuracy		88.0%		
Lab. 5 GHS				
NI	12	1	0	13
Category 2	3	4	0	7
Category 1	0	1	4	5
Sum	15	6	4	25
Accuracy		80.0%		

Table 9
The κ-value between laboratories for category classification. The κ-value are presented with their 95% confidence intervals.

	Lab. 1	Lab. 2	Lab. 3	Lab. 4	Lab. 5
Lab. 1	–	0.83 ± 0.20	0.83 ± 0.20	0.92 ± 0.20	0.75 ± 0.20
Lab. 2	–	–	0.83 ± 0.20	0.92 ± 0.20	0.92 ± 0.20
Lab. 3	–	–	–	0.92 ± 0.20	0.75 ± 0.20
Lab. 4	–	–	–	–	0.83 ± 0.20
Lab. 5	–	–	–	–	–

5th type is cell function-based assays (e.g., Fluorescein Leakage test). As we described in Section 1, the BCOP and ICE, which are the 1st type alternative eye irritation test, were adopted by OECD and a corresponding test guideline was generated. But the other alternative type methods have yet to be accepted by OECD. At this time, the validation of SkinEthic HCE™ and EpiOcular™, the 3rd alternative method type, are ongoing by European Centre for the Validation of Alternative Methods (ECVAM) (ECVAM home page). The STE test, one of the cytotoxicity method types, addresses several disadvantages of the conventional cytotoxicity test (Takahashi et al., 2008). The issues associated with the conventional cytotoxicity tests (Bagley et al., 1994; Harbell et al., 1997; Ohno et al., 1999) that the STE test improves are the evaluation of water insoluble test chemicals, susceptibility to the effect of medium, and inappropriateness of evaluating acidic and alkaline raw chemicals. To date, good transferability and inter-laboratory reproducibility have been confirmed by inter-laboratory study with two or three laboratories (Takahashi et al., 2009, 2010). The purpose of this validation study was to confirm transferability (by using three standard chemicals), intra- and inter-laboratory reproducibility, and predictive capacity by evaluating four assay controls and 25 blind test chemicals in five laboratories.

4.1. Transferability

To confirm the transferability of the STE method in five laboratories, three experiments for each chemical (SLS, CT, and TW80) were evaluated using 5%, 0.5%, and 0.05% test concentrations in physiological saline. The relative viability of each chemical was similar to the background data of the test method developing laboratory (Table 2). Basically, the greatest amount of difference was in the middle range of cell viability (about 20–85%). The exception though was the relative viabilities of 0.5% CT: relative viability for each laboratory of the five validation laboratories were from 64.3% to 76.3%, and for the developing laboratory was 64.7. Also, the irritation rank based on the STE prediction model for these chemicals (SLS, CT, and TW80) were 3, 2, and 1, respectively, in all five laboratories and the test method developing laboratory. From these data, a good transferability of the STE test was obtained. Takahashi et al. (2009) reported that the test method developing laboratory and two naive laboratories had good transferability data using the same three standard chemicals. Combining the findings from Takahashi et al. (2009) and the present study, a good transferability was obtained for the above three standard chemicals in the STE test.

4.2. Intra-laboratory reproducibility

The SD of mean cell viability for the physiological saline, physiological saline with 5% (w/w) DMSO, and mineral oil – where the mean cell viability was near 100% – was small. The reason for the relatively small variation in SD was due to similar ranges. The SD for the physiological saline was in the range of 3.0–4.6%, 2.4–4.4% for the 5% (w/w) DMSO in physiological saline, and 3.1–5.7% for the mineral oil (Table 3). For the 0.01% SLS, mean cell viability was in the range of 34.2–47.6% (overall mean was 42.2%). However, SD was in the range of 2.9–7.3%, also relatively small, and mean was 5.4% (Table 3). As described in Section 2.12, the range of permissible SD for analyses of intra-laboratory reproducibility was difficult to determine. To address this task, we compared using the value obtained by dividing 15% by the square root of 3 (8.7%). The SD of four assay controls (physiological saline, 5% (w/w) DMSO in physiological saline, mineral oil, and 0.01% SLS) were smaller than above reference values, 8.7%. For the intra-laboratory SD estimated by the REML method for the four same assay controls, the SD range was 3.2–4.8% and are shown in Table 4. Goethem et al. (2006) reported that mean cell viability ± SD of 1.0% sodium dodecyl sulfate (same chemical to SLS) for the HCE model from the SkinEthic laboratory were 61.6 ± 13.3%, 56.8 ± 9.3%, 67.7 ± 1.8%, and 51.6 ± 6.6% for four runs (overall mean of four runs = 59.4 ± 6.8%). The data for the four runs was not significantly different between runs. Although the test methods and concentrations differed, both mean cell viability and SD of positive control for evaluating intra-laboratory reproducibility were similar for both tests.

From these data, intra-laboratory reproducibility was achieved for four assay controls.

4.3. Inter-laboratory reproducibility

The mean cell viability was calculated for the four assay controls (physiological saline, 5% (w/w) DMSO in physiological saline, mineral oil, and 0.01% SLS) for each laboratory and the SD obtained was 1.2%, 1.4%, 4.4%, and 5.4%, respectively (Table 3). As described in Section 2.12, the value obtained by dividing 15% by the square root of 3 (8.7%) was used as reference values. The SDs of four assay

controls are smaller than above reference values. Goethem et al. (2006) reported that the inter-laboratory mean value ± SD for 1.0% sodium dodecyl sulfate (same as SLS) in four laboratories was 43.6 ± 5.3% for the HCE model. The data from these four laboratories did not significantly differ between laboratories.

The inter-laboratory SD estimated by the REML method for four assay controls are shown in Table 4 and range from 0.0% to 5.2%. This inter-laboratory SD for the physiological saline and 5% (w/w) DMSO in physiological saline were smaller than the intra-laboratory SD obtained with the REML method. The inter-laboratory SD for mineral oil was 4.2%, which is almost the same as the intra-laboratory SD (Table 4). The mean cell viability for 0.01% SLS was 34.2–47.6%, depending on the laboratory (Table 3). Since cell viability can range from 0% to around 100% (in this study, cell viability of SLS needs from 21.1% to 62.3% by Criterion 3), any variation in the mean cell viability for SLS 0.01% would be expected to be larger than that for physiological saline, 5% (w/w) DMSO in physiological saline, or mineral oil. The SD of mean cell viability for the SLS 0.01% was in the range of 2.9–7.3% in each laboratory, and the inter-laboratory SD by the REML method was 5.2%. Although this inter-laboratory SD was slightly larger than that of any of the three solvents and intra-laboratory SD of 0.01% SLS, the SD did not appear to be large enough to pose a problem (Table 4). From analysis of our own data and comparison of other test method data, the four assay controls provided a good inter-laboratory reproducibility.

Among the 25 blinded test chemicals, the inter-laboratory reproducibility of relative viability for most of the chemicals was good and the SD was not large. The chemicals with SD greater than 8.7% were test chemical D (Tween 20) at 5% concentration (11.2%), J (methyl amyl ketone) at 5% and 0.05% concentrations (11.4% and 12.0%, respectively), K (2-methyl-1-pentanol) at 0.05% concentration (11.3%), R (2-ethyl-1-hexanol) at 5% concentration (20.1%), V (2-ethylhexyl *p*-dimethyl-amino benzoate) at 0.05% concentration (9.5%), and X (methyl ethyl ketone) at 5% concentration (13.3%). For these chemicals, Lab. 3 tended to have slightly lower relative viability values than the other laboratories. Test chemical R (2-ethyl-1-hexanol) at 5% concentration had the largest SD, and the mean relative viability value at this concentration was 59.8% (range: 27.4–80.4%). The larger SD may be expected for chemicals having a medium level of cell viability (approximately 20–85%). This tendency was observed for test chemical J (methyl amyl ketone) at 5% concentration and X (methyl ethyl ketone) at 5% concentration.

In classifying chemicals by category, the results varied between one or two laboratories among the laboratories only for four test chemicals. Test chemical J (methyl amyl ketone (Lab. 3)) and W (gluconolactone (Lab. 5)) were predicted as I and were false positive. Test chemical R (2-ethyl-1-hexanol (Labs. 2 and 5)), and X (methyl ethyl ketone) were predicted as NI and were false negative. Compared to other chemicals, these four chemicals have the cell viabilities closer to the cut-off value of 70%; perhaps one of the causes is the relatively larger amount of variation in mean cell viability compared to other chemicals whose values are in the vicinity of 0% or 100% (Table 5 and Fig. 4).

For this validation study, the κ-value between laboratories for category classification was from 0.75 to 0.92. Brantom et al. (1997) reported the κ-values for 6 laboratories between 0.4333 and 0.949 in red blood cells. Goethem et al. (2006) reported the κ-values of four laboratories for HCE model to be from 0.58 to 1.00. A κ-value of 0.75 or higher suggests strong agreement above chance (Fleiss, 1981). The κ-values of this validation study were slightly better than those observed with other method and higher than Fleiss (1981) proposed value. The κ-values suggest a good inter-laboratory reproducibility was obtained for category classification in this validation study.

In the rank classification, test chemicals J (methyl amyl ketone (Lab. 3)), R (2-ethyl-1-hexanol (Labs. 2 and 5)), W (gluconolactone (Lab. 5)), and X (methyl ethyl ketone) had variable results among the laboratories. For test chemical J (methyl amyl ketone), Lab. 3 showed 63.0% and 68.9% relative viability at the 5% and 0.05% concentrations, respectively. This resulted in a predicted rank classification two rankings lower compared to the other laboratories (Table 5). Also the relative viabilities for chemical R (2-ethyl-1-hexanol) at 5% in Labs. 2 and 5; chemical W at 5% in Lab. 5; and chemical X at 5% in Lab. 1 are 70.2%, 80.4%, 65.8%, and 71.0%, respectively. These findings (i.e., the relative viabilities) suggest the likelihood that for test chemicals around the cut-off value of 70%, different rankings can be obtained due to larger differences among laboratories. Thus, when interpreting the rank classification results, relative viabilities at two concentrations need to be carefully considered in order to make appropriate rankings.

Overall, these findings indicate that the inter-laboratory reproducibility is good.

4.4. Predictive capacity – category classification

Of the 25 test chemicals, 19 chemicals were identified correctly by STE category classification when these chemicals were compared with GHS category (NI or I). Chemical E (ethanol) and S (acetone) were the only two test chemicals for which classifications did not agree among all laboratories. Also, four chemicals (J, methyl amyl ketone (Lab. 3); R, 2-ethyl-1-hexanol (Labs. 2 and 5); W, gluconolactone (Lab. 5); and X, methyl ethyl ketone (Lab. 1)) were not identified correctly by 1 or 2 laboratories. In this validation study, the five laboratories had a mean value of 78.3% (range: 75.0–83.3%) for sensitivity, 96.6% (range: 92.3–100.0%) for specificity, and 88.0% (range: 84.0–92.0%) for accuracy. All values were high. Twenty of the 25 chemicals that we tested were evaluated in the HCE model by SkinEthic Laboratories in a pre-validation study using four laboratories (Goethem et al., 2006). The data reported was as follows; sensitivity was 100% (range: 100%); specificity was 59% (range: 56–67%); and accuracy was 81% (range: 80–85%). Comparing the STE test data to the HCE model data, the sensitivity of the STE is lower than HCE model. However, false positive data in the STE test was only for two data points (J, methyl amyl ketone (Lab. 3) and W, gluconolactone (Lab. 5)) of 125 possible data points (five laboratories × 25 test chemicals). Therefore, the specificity of STE test was better than HCE model. In addition, accuracy of this STE test was better than HCE model. The good predictive parameters indicate that the STE is a promising assay. Once the applicability domain of the STE test is clarified, the STE test will become even a more promising alternative method for eye irritation test if used in battery system with, for example, the HCE model.

4.5. Predictive capacity – rank classification

The mean accuracy value for STE rank classification to the GHS classification ranks (i.e., NI, Category 2 and Category 1) for all laboratories was 84.0% (range: 80.0–88.0%). Adriaens et al. (2008) reported that the eye irritancy potency could classify chemicals to the EU eye irritation categories (NI, R36, and R41) in a pre-validation study with the slug mucosal irritation test. Also their data indicated the possibility of ranking 20 chemicals to the GHS categories (NI, Category 2 and Category 1) in their test model. Of the 20 chemicals in the Adriaens et al. (2008) study, there were 12 chemicals evaluated in the current STE validation study. Two chemicals (R, 2-ethyl-1-hexanol and S, acetone), where the STE test did not correctly classify the GHS ranking in one or all laboratories, were classified correctly by the slug mucosal irritation test. But chemical H (cetylpyridinium bromide) that was classified correctly (NI) by STE test was not predicted correctly (i.e., Category 1 (GHS)/

R41 (EU)) in the slug mucosal irritation test. These data suggest that multiple methods may be needed to correctly classify GHS rankings (NI, Category 2 and Category 1). However, the STE test had good results in classifying 25 chemicals and the surprise is that the test is a very simple to conduct. The possibility of using the STE to rank classify chemicals in the future is good since the accuracy showed 84% even though the number of evaluated chemicals was only 25.

4.6. Chemicals judged to be false-positive or false-negative

Several chemicals had false-negative results when evaluating the correspondence between the STE category (NI or I) and the published GHS category (NI and I (category 1 and category 2)). Of the 25 chemicals evaluated, four chemicals (E, ethanol; R, 2-ethyl-1-hexanol; S, acetone; and X, methyl ethyl ketone) had false-negative results.

Results for test chemicals E (ethanol) and S (acetone) were false-negative in all laboratories. Test chemical E (ethanol) is thought to cause cytotoxicity by creating a wide hydrophobic surface in the solution that then affects the cell membrane (lipid layer). We speculate that at the 5% concentration used in the test, the chemical may be too dilute to cause an effect on cell membrane. Cell injury was not observed and viability was high, which resulted in false-negative results.

In the case of test chemical S (acetone), the high volatility (vapor pressure: 24.7 kPa (20 °C)) of acetone may have caused cells to be exposed to a concentration lower than the test concentration of 5%. The reason we speculate the lower concentration is that evaporation of acetone can occur between the time of sample preparation and cell exposure, which will lead to a lower test concentration than 5%. In addition, the partition coefficient of octanol/water (log Pow) is low at −0.24, which may drive the acetone to the physiological saline and possibly not released. As a result, cells would not have been sufficiently exposed to the test chemical. When mineral oil was used in place of physiological saline for evaluation of acetone, cell viability for the 5% concentration was 9.6% and acetone was classified as I (Takahashi et al., 2008). This may be one possibility. However, we also need to consider other possibilities in order to clarify the reason of the false negative of acetone.

In addition, findings for test chemicals R (2-ethyl-1-hexanol) and X (methyl ethyl ketone) were considered false negative results in two and one laboratory, respectively. The viabilities in the individual three STE tests using a 5% solution were as follows: 63.6%, 74.6%, and 72.4% (mean viability was 70.2%) for chemical R in Lab. 2; 79.1%, 78.7%, and 83.6% (mean viability was 80.4%) for test chemical R in Lab. 5; and 73.6%, 69.5%, and 69.9% (mean viability was 71.0%) for test chemical X in Lab. 1. Thus, the cell viabilities observed were near the cut-off value of 70%. It is possible that chemicals with a viability of around 70% may yield discrepancies in findings, depending on the laboratory. Goethem et al. (2006) also reported that especially chemicals which induced an intermediate viability (between 20% and 70%) showed higher variability in reconstituted human cornea model as measured with MTT assay.

On the other hand, false positive results were obtained for test chemical J (methyl amyl ketone in one laboratory (Lab. 3) and for test chemical W (gluconolactone) in Lab. 5 alone. In Lab. 3, test chemical J (methyl amyl ketone) was the first chemical evaluated using mineral oil as a solvent. Since mineral oil has a higher viscosity than physiological saline, the cells could have been physically scraped off in the process of performing plate-washing by higher power with the lack of experience, and that viability was consequently lower. Following the completion of this STE validation study, chemical J (methyl amyl ketone) was re-evaluated by Lab. 3 by appropriate power in the process of performing plate-washing as mineral oil and was identified as NI (data not shown). For test chemical W (gluconolactone), the viabilities in Lab. 5 were 67.6%, 66.6%, and 63.1%. Like the false-negative chemicals (test chemical R (2-ethyl-1-hexanol) and X (methyl ethyl ketone)), chemicals with a viability of around 70% may have resulted in false-positives, depending on the laboratory. Again, the viability around 70% may be one reason for the false positive.

Discrepancies were also noted in all laboratories when STE rank and GHS rank were compared: Three chemicals – namely, test chemicals E (ethanol), S (acetone), and T (cyclohexanol) – demonstrated discrepancies at all testing laboratories. For test chemicals E (ethanol) and S (acetone), the discrepancy from GHS classification has already been discussed above and is associated with high viability at 5%. For test chemical T (cyclohexanol), the viability at 0.05% was greater than 70%, which resulted in a STE rank of 2 in all the laboratories. Since this chemical is an alcohol, it is our thought that its toxicity cannot be detected at 0.05% in the STE test with the same reason of chemical E (ethanol). On the other hand, this chemical exhibits very strong eye irritation in the Draize test score, with a score of 79.8 at 100%; however, at 10%, it exhibits almost no toxicity (i.e., Draize score of 4.0). Takahashi et al. (2008) reported that STE test with 0.05% test solution could be predicted eye irritation of a 10% solution in the Draize test. Therefore, the result that the cell viability of 0.05% test solution of chemical T (cyclohexanol) was over 70%, was also reasonable.

4.7. Limitations of the STE validation study

In the present study, a total of 28 chemicals, including the three standard chemicals in the preliminary investigation, were evaluated. The results found that the STE test provided an excellent predictive ability. There were a few chemicals that exhibited cell viability around 70% in the STE test. Those that did have viability around 70% seemed also to have higher variability in classification between laboratories. In addition, since scattering in the intermediate range (around 20–85%) of mean cell viability was relatively high, the interpretation of rank-classification results must be performed carefully. Hence, many more chemicals with cell viability near the 70% cut-off point need to be evaluated and added to the databank for future analysis.

Among the 25 chemicals evaluated in the present validation study, several chemicals, alcohol test chemicals E (ethanol) and T (cyclohexanol) had predictive rankings in the STE test in all laboratories that differed from the rankings of the GHS classification. On the other hand, the alcohol test chemicals L (*n*-hexanol), Q (1-octanol), and R (2-ethyl-1-hexanol), which all have the same GHS classification (Category 2) as test chemical E (ethanol), were evaluated correctly by the STE test. Although alcohols can be evaluated in the STE test, alcohols should be considered chemicals that could generate false negative results or be predicted to have a weaker toxicity potential. In the future, data on many more alcohols should be obtained as well as other chemicals with mean cell viabilities near the 70% cut-off value in order to evaluate the predictive ability of the STE test for alcohols and the assay. But until then, we will continue to use the 70% cut-off value.

5. Conclusion

On the basis of a unified test protocol, a validation study was performed in five laboratories. Three standard chemicals were first used to confirm transferability of method, and then four assay controls and 25 blinded test sample chemicals with GHS classification categories were used to confirm intra-/inter-laboratory reproducibility and predictive capacity. The results show that the STE test not only provided good transferability, but it also had good intra-

and inter-laboratory reproducibility, and a high predictive capacity for predicting GHS classification of various chemicals. In conclusion, the STE test is considered a good alternative test method to predict eye irritation.

Acknowledgments

The Validation Executive Committee was organized under the JSAAE and performed the present study. The JSAAE, through its Validation Committee, provided both laboratory personnel and financial support for completion of this study. The authors are very grateful for this support. In addition, in performing the validation study, the authors received considerable support and advice from the Japanese Center for the Validation of Alternative Methods. The authors offer sincere thanks to them as well.

The present study was supported by many people. Without their support, the present study could not have been possible. Especially we offer heartfelt thanks to Ms. Mirei Koike (Kao Corporation, Safety Science Research Laboratories).

References

Adriaens, E., Bytheway, H., De Wever, B., Eschrich, D., Guest, R., Hansen, E., Vanparys, P., Schoeters, G., Warren, N., Weltens, R., Whittingham, A., Remon, J.P., 2008. Successful prevalidation of the slug mucosal irritation test to assess the eye irritation potency of chemicals. Toxicol. in Vitro 22, 1285–1296.

Bagley, D., Booman, K.A., Bruner, L.H., Casterton, P.L., Demetrulias, J., Heinze, J., Innis, J.D., Mccormick, W.C., Neun, D.J., Rothenstein, A.S., Sedlak, R.I., 1994. The SDA alternatives program phase III: comparison of in vitro data with animal eye irritation data on solvents, surfactants, oxidizing agents, and prototype cleaning products. J. Toxicol. Cut. Ocul. Toxicol. 13, 127–155.

Balls, M., Berg, N., Bruner, L.H., Curren, R.D., De Silva, O., Earl, L.K., Esdaile, D.J., Fentem, J.H., Liebsch, M., Ohno, Y., Prinsen, M.K., Spielmann, H., Worth, A.P., 1999. Eye irritation testing: the way forward, The report and recommendations of ECVAM workshop 34. Altern. Lab. Anim. 27, 53–77.

Brantom, P.G., Bruner, L.H., Chamberlain, M., De Silva, O., Dupuis, J., Earl, L.K., Lovell, D.P., Pape, W.J., Uttley, M., Bagley, D.M., Baker, F.W., Bracher, M., Courtellemont, P., Declercq, L., Freeman, S., Steiling, W., Walker, A.P., Carr, G.J., Dami, N., Thomas, G., Harbell, J., Jones, P.A., Pfannenbecker, U., Southee, J.A., Tcheng, M., Argembeaux, H., Castelli, D., Clothier, R., Esdaile, D.J., Itigaki, H., Jung, K., Kasai, Y., Kojima, H., Kristen, U., Larnicol, M., Lewis, R.W., Marenus, K., Moreno, O., Peterson, A., Rasmussen, E.S., Robles, C., Stern, M., 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. Toxicol. in Vitro 11, 141–179.

Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003. Official Journal of the European Union, L66, pp. 26–35.

ECETOC, 1998. Eye Irritation Reference Chemicals Data Bank, second ed. Technical Report, No. 48(2).

ECVAM homepage. Avaliable from: <http://ecvam.jrc.ec.europa.eu/page_pdf.cfm?voce=s&idvoce=120/>.

Eskes, C., Bessou, S., Bruner, L., Curren, R., Harbell, J., Jones, P., Kreiling, R., Liebsch, M., McNamee, P., Pape, W., Prinsen, M.K., Seidle, T., Vanparys, P., Worth, A., Zuang, V., 2005. Subchapter 3.3. Eye Irritation. In: Eskes, C., Zuang, V. (Eds.), Alternative (non-animal) methods for cosmetics testing: Current status and future prospects, A report prepared in the context of the 7th amendment to the cosmetics directive for establishing the timetable for phasing out animal testing. Alternatives to Laboratory Animals, vol. 33, pp. 47–81.

Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions. John Wiley, New York. p. 321.

Gautheron, P., Dukic, M., Alix, D., Sina, J.F., 1992. Bovine corneal opacity and permeability test: an in vitro assay of ocular irritancy. Fundam. Appl. Toxicol. 18, 442–449.

Goethem, F.V., Adriaens, E., Alepee, N., Straube, F., Wever, B.D., Cappadoro, M., Catoire, S., Hansen, E., Wolf, A., Vanparys, P., 2006. Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. Toxicol. in Vitro 20, 1–17.

Hagino, S., Okazaki, Y., Itagaki, H., 2008. An in vitro tier evaluation for the identification of cosmetic ingredients which are not ocular irritants. Alternat. Lab. Anim. 36, 641–652.

Harbell, J.W., Koontz, S.W., Lewis, R.W., Lovell, D., Acosta, D., 1997. IRAG working group 4: cell cytotoxicity assays. Food Chem. Toxicol. 35, 79–126.

McNamee, P., Hibatallah, J., Costabel-Farkas, M., Goebel, C., Araki, D., Dufour, E., Hewitt, N.J., Jones, P., Kirst, A., Le Varlet, B., Macfarlane, M., Marrec-Fairley, M., Rowland, J., Schellauf, F., Scheel, J., 2009. A tired approach to the use of alternative to animal testing for the safety assessment of cosmetics: eye irritation. Regul. Toxicol. Pharmacol. 54, 197–209.

OECD Guideline for the Testing of Chemicals. Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants. Paris, 2009. Available from: <http://titania.sourceoecd.org/vl=2663626/cl=26/nw=1/rpsv/cgi-bin/fulltextew.pl?prpsv=/ij/oecdjournals/1607310x/v1n4/s54/p1.idx/>.

OECD Guideline for the Testing of Chemicals. Test No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants. Paris, 2009. Available from: <http://titania.sourceoecd.org/vl=2663626/cl=26/nw=1/rpsv/cgi-bin/fulltextew.pl?prpsv=/ij/oecdjournals/1607310x/v1n4/s55/p1.idx/>.

Ohno, Y., Kaneko, T., Inoue, T., Morikawa, Y., Yoshida, T., Fujii, A., Masuda, M., Ohno, T., Hayashi, M., Monma, J., Uchiyama, T., Chiba, K., Ikeda, N., Imanishi, Y., Itagaki, H., Kakishima, H., Kasai, Y., Kurishita, A., Kojima, H., Matsukawa, K., Nakamura, T., Ohkoshi, K., Okumura, H., Saijyo, K., Sakamoto, K., Suzuki, T., Takano, K., Tatsumi, H., Tani, N., Usami, M., Watanabe, R., 1999. Interlaboratory validation of the in vitro eye irritation test for cosmetic ingredients. (1) Overview of the validation study and Draize scores for the evaluation of the tests. Toxicol. in Vitro 13, 73–98.

Scott, L., Eskes, C., Hoffmann, S., Adriaens, E., Alepee, N., Bufo, M., Clothier, R., Facchini, D., Faller, C., Guest, R., Harbell, J., Hartung, T., Kamp, H., Le Varlet, B., Meloni, M., McNamee, P., Osborne, R., Pape, W., Pfannenbecker, U., Prinsen, M., Seaman, C., Spielman, H., Stokes, W., Trouba, K., Berghe, C., Goethem, F., Vassallo, M., Vinardell, P., Zuang, V., 2010. A proposed eye irritation testing strategy to reduce and replace in vivo studies using Bottom-Up and Top-Down approaches. Toxicol. in Vitro 24, 1–9.

Takahashi, Y., Koike, M., Honda, H., Ito, Y., Sakaguchi, H., Suzuki, H., Nishiyama, N., 2008. Development of the short time exposure (STE) test: an in vitro eye irritation test using SIRC cells. Toxicol. in Vitro 22, 760–770.

Takahashi, Y., Hayashi, T., Watanabe, S., Hayashi, K., Koike, M., Aisawa, N., Ebata, S., Sakaguchi, H., Nakamura, T., Kuwahara, H., Nishiyama, N., 2009. Interlaboratory study of short time exposure (STE) test for predicting eye irritation potential of chemicals and correspondence to globally harmonized system (GHS) classification. J. Toxicol. Sci. 34, 611–626.

Takahashi, Y., Hayashi, T., Koike, M., Sakaguchi, H., Kuwahara, H., Nishiyama, N., 2010. An interlaboratory study of the short time exposure (STE) test using SIRC cells for predicting eye irritation potential. Cutan. Ocul. Toxicol. 29, 77–90.

United Nations, 2003. Globally Harmonised System of Classification and Labeling of Chemicals (GHS), New York and Geneva (ST/SG/AC.10/30).

原著

高齢者におけるベンゾジアゼピン系薬の服用量変更と転倒との関連：急性期病院入院患者を対象にした解析

高橋佳苗[*1]，長尾能雅[*2]，足立由起[*3]，森本　剛[*4]
市橋則明[*5]，坪山直生[*5]，大森　崇[*6]，佐藤俊哉[*7]

Changes in the Dose of Benzodiazepines and Falls in Elderly Inpatients in an Acute-care Hospital

Kanae TAKAHASHI[*1], Yoshimasa NAGAO[*2], Yuki ADACHI[*3], Takeshi MORIMOTO[*4], Noriaki ICHIHASHI[*5], Tadao TSUBOYAMA[*5], Takashi OMORI[*6], Tosiya SATO[*7]

[*1] Research and Development Initiative Center, National Cerebral and Cardiovascular Center, Japan
[*2] Department of Quality and Patient Safety, Nagoya University Hospital, Japan
[*3] Department of Nursing, Kyoto University Hospital, Japan
[*4] Center for Medical Education, Kyoto University Graduate School of Medicine, Japan
[*5] Human Health Sciences, Graduate School of Medicine, Kyoto University, Japan
[*6] Faculty of Culture and Information Science, Doshisha University, Japan
[*7] Department of Biostatistics, Kyoto University School of Public Health, Japan

〈Abstract〉
Objective：It is well known that the use of benzodiazepines is associated with falling in elderly people, but there have been few researches focused on changes in the dose of benzodiazepines and falls. If the association between changes in the dose of benzodiazepines and falling becomes clear, we may take an action to prevent falling. In this study, we investigated the association between changes in the dose of benzodiazepines and falling among elderly inpatients in an acute-care hospital.
Design：Falling generally results from an interaction of multiple and diverse risk factors and situations, and medication history of each subject must be considered in this study. We conducted a case-crossover study in which a case was used as his/her own control at different time periods. Therefore covariates that were not time-dependent were automatically adjusted in this study.
Methods：Subjects were patients who had falling at one hospital between April 1, 2008 and November 30, 2009. Data were collected from incident report forms and medical records. Odds ratio for changes in the dose of benzodiazepines were calculated using conditional logistic regression analyses.
Results：A total of 422 falling by elderly people were eligible for this study. The odds ratio for increased amounts of benzodiazepines was 2.02 (95% Confidence Interval (CI)：1.15, 3.56). On the other hand, the

[*1] 国立循環器病研究センター研究開発基盤センター　[*2] 名古屋大学医学部附属病院医療の質・安全管理部
[*3] 京都大学医学部附属病院看護部　[*4] 京都大学大学院医学研究科医学教育推進センター
[*5] 京都大学大学院医学研究科人間健康科学系専攻　[*6] 同志社大学文化情報学部
[*7] 京都大学大学院医学研究科社会健康医学系専攻医療統計学分野
別刷請求先：〒565-8565 大阪府吹田市藤白台 5-7-1　国立循環器病研究センター研究開発基盤センター先進医療・治験推進部 DM/統計室　高橋佳苗