Development of a framework for risk tradeoff analysis of chemical substance substitution

— An approach to risk assessment using relative comparison—

Masashi Gamo* **and Jun-ichi Takeshita**

A chemical substance is often substituted for another to reduce risks associated with use of the original substance. However, the replacement may be associated with new risks, and this introduces a risk tradeoff problem. Although the concept of risk tradeoff analysis has been discussed in this context, no feasible method has yet been developed. In this study, a novel assessment method was proposed, based on relative risk comparison among substances, through an examination of some possible approaches. Case studies were also conducted to assess the efficacy of the method.

Keywords : Chemical substances, risk, tradeoff, relative comparison, assessment methodology

1 Introduction

While the use of chemical substances is necessary for achieving a prosperous and sustainable society, there is concern about related risks against humans and ecosystems. The production, use, and emission of chemical substances determined to pose significant risks are regulated by laws and through self-management. For example, such substances as polychlorinated biphenyls (PCBs), some chloride agrichemicals, brominated flame retardants, and heavy metals such as mercury, cadmium, and lead are strictly regulated, while the emissions of substances like air pollutants such as toluene are declining every year due to voluntary emission control.[1]

Although the risk associated with a particular substance can be reduced by regulating the use and emission of that substance, related measures may be costly; and replacement by different, less hazardous substances is often made, while maintaining the functionality of the product in which the original substance was used. For example, the brominated flame retardant, decabromodiphenyl ether (decaBDE), has been replaced by substances such as bisphenol-A bis(diphenyl phosphate) (BDP);[2] lead solder alloys by 'lead-free solder alloys,' such as tin-silver-copper alloys, which do not contain lead;[3] and chlorinated solvents by carbohydrate or aqueous varieties, for use as industrial cleansers.[4]

The occurrence of new risk when reducing a certain risk is called 'risk trade off.' To assess whether the replacement of a given substance constitutes appropriate risk management, it is insufficient merely to demonstrate reduction in the risks associated with the original substance. It is necessary also to consider the potential risks associated with the replacement substance. It is necessary to determine whether the overall risk of the replacement substance is less than that of the original substance; and in addition, by comparing the risks before and after substitution, to assess whether the given substitution produces a risk reduction effect that exceeds the cost of countermeasures.

Kishimoto[5] argued for the necessity of developing an assessment method for human health risks associated with chemical substance use, which would reflect social requirements, including the new social demand for "comparing risks of different types of chemical substances, and the assessment of cost-effectiveness of emission reduction measures;" and presented a case study involving toluene risk assessment. Here, he proposed a method for quantifying human health risks, using quality of life (QOL) as the risk index, by backcasting from the demand. The method enabled the comparison of the cost-effectiveness of risk reduction measures for toluene with that of risk reduction measures for other chemical substances, infectious diseases, accidents, disasters, etc.

However, Kishimoto's proposal is, in practice, applicable only to substances such as toluene, for which abundant toxicity information is available. In risk tradeoff analysis associated with substance substitution, the substitution often involves replacing a substance for which there is relatively abundant information, with one for which there is insufficient information. For such risk tradeoff analysis to become a

Research Institute of Science for Safety and Sustainability, AIST   Tsukuba West, 16-1 Onogawa, Tsukuba 305-8569, Japan
* E-mail: masashi-gamo@aist.go.jp

realistic assessment method, then, it is necessary to develop a technique whereby Kishimoto's risk comparison conception is executable.

In this paper, we discuss the approaches taken in this regard, as well as the development of basic techniques enabling risk tradeoff analysis of chemical substance substitution, and the results of related case studies. We also discuss future prospects for this analytical method.

2 Investigation and discussion of approaches

Here, we consider the merits and demerits of current approaches to risk tradeoff analysis, in terms of the comparative risk before and after substance substitution. Figure 1 shows the relationship between the basic techniques and the complex of analytical methods considered below.

2.1 General risk analysis method

The first approach involves the application of a general assessment method for chemical substance risks. In this method, the nature and level of risk is assessed through comparison of levels of tolerance to those of exposure to a given chemical substance. Tolerance level, the exposure level at which the manifestation of toxicity is not a concern, is often calculated by applying an uncertainty factor (safety factor) to the results of animal tests and human epidemiological studies; whereas, exposure level is determined by multiplying the concentration in environmental media (air, water, food, etc.) by the amount of intake of the media. The concentration in environmental media is obtained by actual measurement or by simulation-based prediction. There is considered to be no risk if the exposure level is less than the tolerance level, and risk if the former is greater the latter.

However, since the comparison of exposure and tolerance levels is conducted separately for each individual substance, if the respective ratios of exposure to tolerance level for two

substances are calculated to be 0.1 and 0.5, for example, neither is considered to involve risk, and no relation is established between them. This lack of relation becomes especially problematic when the types of toxicity differ between the two substances. With this approach, then, any difference in the level of risk before and after substance substitution can be considered only as difference in the presence (or absence) of risk.

In addition, calculation of the tolerance level requires the results of animal tests and human epidemiological studies; and since sufficient information on toxicity may not be available to enable accurate tolerance level calculation for a given substance after substitution, it must be concluded that risk tradeoff analysis based on this method is difficult.

2.2 Method using a common risk index
To solve the problem faced by such general risk analysis methods, wherein the relative risks of multiple and varied substances cannot be compared, Kishimoto proposed an approach involving quantitative expression of the magnitude of health risks, based on a common index.[5] As common indices, lifespan reduction due to adverse health effects, or similar reduction but with adjustment for quality of life, are both widely used. For example, Gamo *et al.*[6] used an index called lost life expectancy (lifespan reduction) to evaluate replacement of a termite control agent; and Gamo *et al.*[7] used this same index to rank major environmental pollutants. Kishimoto[5] used an index based on loss of quality-adjusted life-years (QALYs: lifespan adjusted for quality of life) in the detailed risk evaluation of toluene; and Cohen *et al.*[8] used this index to compare the benefit of polysaturated fatty acid, with the risk of methyl mercury, in fish consumption. Disability-adjusted life expectancy years (lifespan with adjustment for disabilities caused by disease) is used as an index in calculations of the worldwide burden of disease by the World Health Organization (WHO);[9] and Havelaar *et al.*[10] used this index to compare the risk of bromate
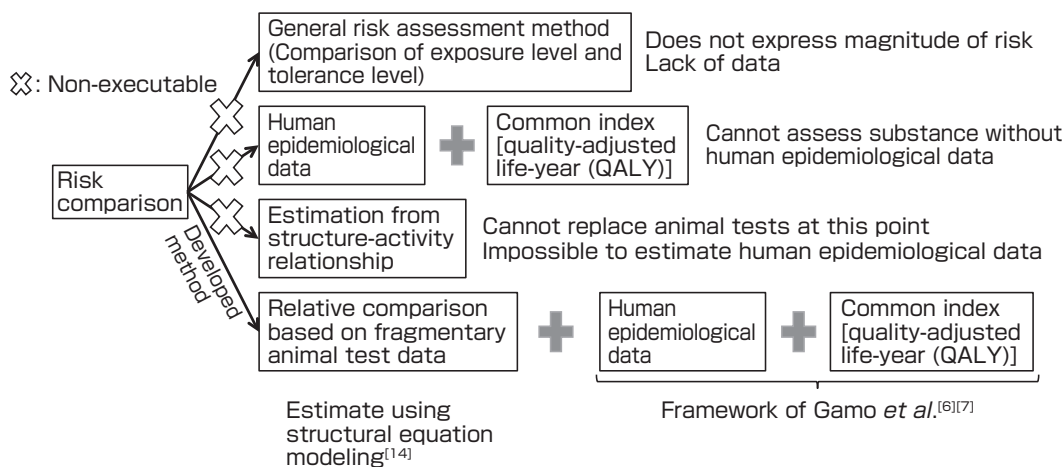

Fig. 1 Scenario for developing the methodology for risk tradeoff analysis, and the basic techniques

byproducts, with risk reduction in infectious disease by the disinfection of drinking water.

The problem with this approach is that its estimations require the results of human epidemiological studies, and many potential replacement substances do not even have sufficient animal test data, let alone human epidemiological study results. In sum, this approach is less executable than the general risk analysis method described in Subchapter 2.1.

2.3 Estimation based on structure-activity relationships and/or cell-based assays

In light of the problems faced by the foregoing methods, one way to counter the lack of information on the potential toxicity of substances after substitution is a risk estimation approach based on structure-activity relationships and/or cell-based assays. The structure-activity relationship is an expression relating the structure of a given chemical substance to its activity (in this case, toxicity). By constructing the expression, based on data for multiple chemical substances, it is possible to estimate the potential toxicity of substances without animal or human study data. Several related studies have been conducted, and recently a prediction system based on categorization for repeated dose toxicity, called the Hazard Evaluation Support System (HESS), was developed in Japan.[11]

In a similar manner, the approach based on cell-based assays involves conducting tests using cells instead of animals, and assessing the presence and degree of toxicity of the target substance based on these tests. The Ames test (detection of mutagenicity using bacteria), which evaluates the mutagenicity of chemical substances, is well known, and several cell-test methods are described in the guidelines of the Organisation for Economic Co-operation and Development (OECD). However, the general consensus is that it is difficult to correlate the effect on cells with the effect on individual organisms.

While use of these methods is expected to increase in the future, due to demands for cost reduction in animal testing and from the perspective of animal rights, they cannot replace animal tests at this point; and it is technically impossible to estimate information equivalent to human epidemiological studies, based on the results of these methods. Moreover, even if technological development progresses to the point where it becomes possible to determine the nature and level of risk for each substance, this alone would not enable multiple-substance risk comparison, as discussed above with regard to the general risk evaluation method (Subchapter 2.1).

2.4 Method of multiple-substance relative comparison

To enable comparative risk tradeoff analysis of substance substitution, we have proposed an approach involving the relative comparison of substance risks, based on a quantitative risk evaluation method (Subchapter 2.2), to

enable relevant estimation when human epidemiological data is unavailable. In this approach, the substance for which the risk can be evaluated using a common scale based on human epidemiological data is established as a reference substance, and hazard assessment of the target substance is conducted by relative comparison with the reference substance. This avoids the difficulty of estimating the relevant human epidemiological data based on animal test results (or on cell-based assays or structure-activity relationships).

Several reports have been published on the approach of assessing substance risk without sufficient test data, based on such comparison. For example, Maier[12] proposed the "parallelogram approach," where workplace tolerance concentrations are established by comparing the activities of pharmaceutical intermediates without hazard data, with those of substances with sufficient animal test data, based on cell-based assays. This is similar to our proposal in the sense that the results of cell-based assays are not directly used for estimating individual effects, but instead as a tool for relative comparison. Nakanishi *et al.*[13] proposed a "two-axis approach" for conducting risk evaluation of carbon nanotubes (CNT). To evaluate the risk to workers of inhalation exposure to various CNTs, the inhalation exposure test, which is standard but costly, is conducted only for representative CNTs, while the simpler method of an intertracheal administration test (where a CNT suspension is applied to an animal's trachea) is conducted for other CNTs, in order to establish a wide range of comparison.

In the case of the method discussed here, the established method[6][7] was chosen as a basis for quantifying the risk of reference substances. However, as there is no existing research on a method for relative risk comparison of reference and target substances, it was necessary to develop an original method for conducting such comparison based on limited information.[14] The details of this method will be presented in the next chapter.

3 Framework of assessment and basic techniques

3.1 Overall structure

It is assumed that, for the target substance, there is only fragmentary animal test data, and no human epidemiological data, to enable the sort of quantitative risk assessment as is described in Subchapter 2.2. Here, fragmentary animal test data denotes cases where, for example, there exist data or papers characterizing the specific effects of the substance on the liver, but none describing all the effects on all organs; and human epidemiological data refers specifically to the dose-response relationship (relationship between the exposure level and occurrence rate of the effect) in humans.

In reality, it is impossible to directly estimate such target substance human epidemiological data based solely on

fragmentary animal test data, given the species difference between humans and animals. Therefore, we devised a framework for conducting such an estimation with the following procedure (Fig. 2 shows a conceptual diagram of the procedure).

1) The substance with human epidemiological data, and for which quantitative risk assessment is possible, is established as the reference substance.
2) A relative comparison is made between the reference substance and the target substance, at the level of animal test data, to calculate the relative toxicity value, which is the ratio of the exposure level of the target substance, to that of the reference substance at which both substances present the same level of toxicity.
3) The dose-response relationship of the target substance is estimated by multiplying the relative toxicity value calculated in (2) with the known dose-response relationship of the reference substance (Fig. 3). In this case, it is

desirable to establish a confidence interval for the estimated dose-response relationship equation.
4) The quantitative risk assessment is conducted by combining the dose-response relationship of the target substance and the value of severity of the expected health effect as expressed in terms of a common index.

Typically, chemical substances affect multiple organs, but the level of exposure required to trigger an adverse effect varies with the given organ. Therefore, to reflect the diversity in substance toxicity, it is desirable to conduct this four-step estimation for each organ. Thus, we decided to treat the liver and kidney independently, as they were major organs in which adverse effects occurred. The reference substances were established and relative toxicity values were calculated for each organ.

3.2 Reference substances and dose-response relationships
In establishing the reference substances for liver and kidney
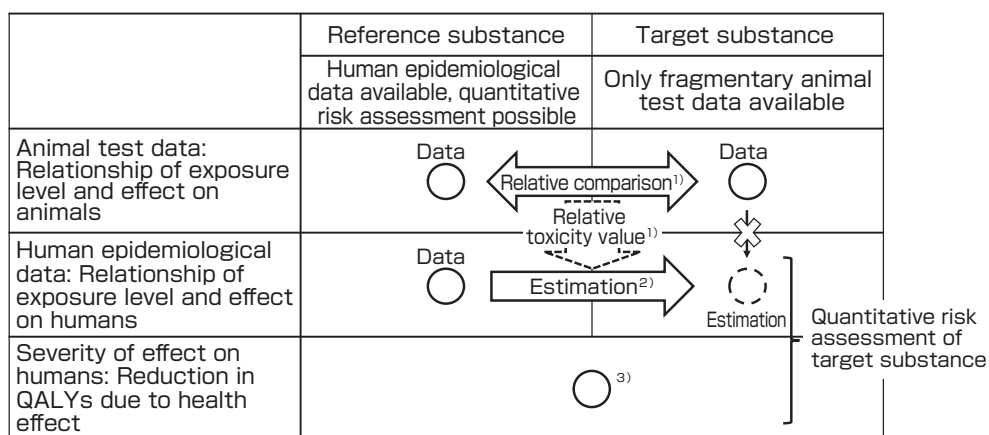
1) Relative comparison between the reference substance and the target substance, calculation of relative toxicity value from Takeshita *et al.*[14] (Subchapter 3.2)
2) The conceptual diagram for estimating the human epidemiological data for the target substance, based on the human epidemiological data and relative toxicity values of the reference substance, is shown in Fig. 3.
3) These values are general values for types of effects, and it was assumed that they are not dependent on types of exposed substance.

Fig. 2 Framework of the risk tradeoff analysis based on relative assessment of substances
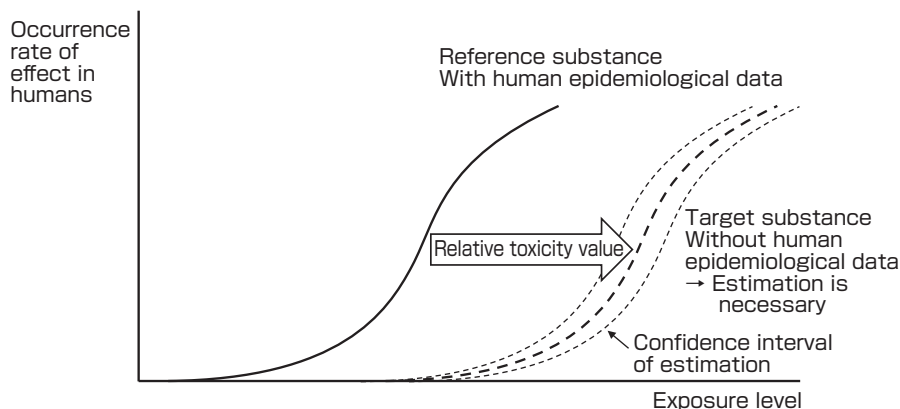

Fig. 3 Estimation of the dose-response relationship of the target substance, based on the dose-response relationship of the reference substance and the relative toxicity value

effects, the CHE Toxicant and Disease Database[15] of the Collaborative on Health and the Environment (CHE), an international environmental group, was used to search for chemical substances that were known to cause adverse effects on organs. This database, which is based on three famous textbooks on toxicology, enables searches for chemical substances that may be the cause of disease.

We selected substances for which the evidence was "strong" for liver and kidney effects, and then narrowed the search to substances with published human epidemiological data, by referring to existing documents on hazard assessment. As a result, we established as reference substances vinyl chloride monomer (chloroethene) for liver, and cadmium for kidney effects. The human dose-response relationship information for each substance is as follows.

3.2.1 Liver effect
Ho *et al.*[16] studied workers who were exposed to vinyl chloride monomer in the air, and reported that of 271 subjects exposed to 1–20 ppm (equivalent to 2.5–50 mg/m^3), 12 showed liver dysfunction (4 had liver enlargement, 4 had liver and spleen enlargement, and 4 had spleen enlargement). As this was occurrence rate data for a specific concentration range, we decided to apply the distribution of sensitivities to noncarcinogenic effects, as proposed by Huijbregts *et al.*;[17] that is, a geometric standard deviation of 1.82 when the lognormal distribution is assumed for individual differences. As a result, for inhalation exposure to vinyl chloride monomer, the dose-response relationship of liver effect was set as the lognormal distribution (with a geometric mean value of 31 mg/m^3, and a geometric standard deviation of 1.82). In using this value for the risk assessment of oral exposure, the concentration in air was converted to a daily intake amount (unit: mg/kg/day) by using assumed values for respiratory volume and weight.

3.2.2 Kidney effect
The effects of cadmium on humans have been studied in detail, and renal tubular disorder is known as a highly sensitive adverse effect. In this case, then, the established dose-response relationship value was used. Renal tubular disorder was defined as the situation where the urine β_2-microglobulin concentration exceeded 1000 μg/g creatinine. The sensitivity of this tubular disorder is age dependent, and Gamo *et al.*[18] and Nakanishi *et al.*[19] summarize the dose-response relationship parameters for each age.

3.3 Derivation method for the relative toxicity value[14][20]
The following two factors were considered requisite for a relative toxicity derivation method to be developed.
• It is possible to complement missing data: On the assumption that the available animal test data is fragmentary, it may be the case that either or both the data for liver and kidney are lacking. In such a case, it is difficult to use the data

in actual risk tradeoff analysis unless the relative toxicity values can be estimated for both organs. The method in which the significance of effects in multiple organs are mutually estimated is called the quantitative activity-activity relationship (QAAR).
• It is possible to estimate the confidence interval: Estimation is inherently uncertain, yet the estimation of uncertainty is essential in considering the results of risk tradeoff analysis. For example, in cases where it seems that the risk has decreased with the replacement of a given substance, the estimation of a confidence interval is necessary to determine whether the decrease is dubious or not.

With these factors and the methodological aim in mind, we decided to apply structural equation modeling (SEM), which is a statistical analysis method encompassing linear regression analysis and factor analysis. In SEM, the statistical model parameters used in establishing relationships between the variables are determined such that the variance and covariance calculated from the model will best correspond to the variance and covariance calculated from the actual data. Figure 4 shows an abstract diagram of this procedure.

To construct the training data set, existing test data were obtained from literature on roughly 165 substances listed in the Japanese Pollutant Release and Transfer Register as Class I Designated Chemical Substances. There are 45 toxicity endpoint items in total, for the combination of target organs (liver, kidney, blood, urine, body weight, death, spleen, digestive tract, respiratory organ, brain, etc.), test animal species (rat, mouse), and administration methods (oral exposure, inhalation exposure).

Using the constructed model, the no-observed-effect level (NOEL) is estimated for each item and substance, and based on this result, the relative toxicity value of the two substances is estimated. Let Substance A be the reference, Substance B be the target, and $\hat{a}$, $\widehat{\sigma_a}$, $\hat{b}$, $\widehat{\sigma_b}$ be, respectively, the logarithmic NOEL value for Substance A, its standard deviation, the logarithmic NOEL value for Substance B, and its standard deviation. Then, the relative toxicity value of Substance B versus Substance A (the reference) and its 95 % prediction interval are calculated by the following equations.

Relative toxicity value: exp $(\hat{b}-\hat{a})$
95 % prediction interval:
[exp $(\hat{b}-\hat{a}-2\sqrt{\widehat{\sigma_a^2}+\widehat{\sigma_b^2}})$, exp $(\hat{b}-\hat{a}+2\sqrt{\widehat{\sigma_a^2}+\widehat{\sigma_b^2}})$]

The estimation accuracy of the respective NOELs for each substance was quantified according to the OECD principles of verification, by applying the leave-one-out cross-validation method, which dictates that each observed value is removed and then estimated by the remaining observed values. This is repeated for all values, and then the estimated values are compared with the true values. In this case, the

– 189 –

correlation coefficient between the observed and estimated values was 0.89; and in terms of estimation accuracy, 93 % of the observed values were included in the 95 % prediction interval, and 97 % of the ratios of estimated and observed values were less than 10.

3.4 Common index of the effects

As noted in Subchapter 2.2, among the common indices of human health effect are lost life expectancy, quality-adjusted life-years (QALYs), and disability-adjusted life-years. Here, we used the QALY index, which incorporates both lifespan and QOL reduction. In the case of QOL reduction, various values are typically reported, according to the disease and its state, even if the same organ is affected; but here we set a general value, without assuming any specific disease state. As the liver and kidney effect information obtainable for the reference substances differed, the estimation method differed for the two organs.

3.4.1 Liver effect

The liver dysfunctions reported by Ho *et al.*,[16] as discussed in Section 3.2.1, were based on the observation of workers, and were considered to reflect a relatively light, chronic effect due to long-term exposure to chemical substances. In the risk ranking of environmental pollutants, Gamo *et al.*[7] conducted an assessment by specifying as one year the reduction in lifespan due to health conditions accompanied by some expression of subjective symptoms, and our present evaluation adopted this assessment. For QOL, the data compiled and organized by Tengs *et al.*[21] was reviewed, and the QOL value of 0.01 was used, in the QALY calculation, for the lightest liver disorder among the liver diseases. This value means that, for example, when an individual lives (for about 80 years) in this state of health, the disorder will be considered to have the same degree of severity as a lifespan reduction of 80 yr × 0.01 = 0.8 yr.

3.4.2 Kidney effect

Renal tubular disorder by cadmium exposure is thought to occur in those over 50 years old, and it has been reported that the mortality of people in this health state is, respectively, 1.57 times (males) and 1.81 times (females) the mortality rate of healthy individuals.[22] Based on this information, lifespan reduction was calculated based on the life table (a table of mortality by age, which enables calculation of average lifespan and other values). For the QOL reduction due to kidney disease, the data compiled and organized by Tengs *et al.*[21] was similarly reviewed, and the QOL value of 0.01 was used for the lightest kidney disorder.

4 Case study

The above method was used for the three risk tradeoff analyses[2]–[4] conducted by the Research Institute of Science

Step 1: Preliminary analysis of the data set
Checking correlation among the toxicity endpoint items (NOEL for each animal species, administration method, and organ)

	Rat, oral, liver	Rat, inhalation, liver	Mouse, oral, liver	Mouse, inhalation, liver	Rat, oral, kidney
Rat, oral, liver	—	0.635 (34)	0.607 (41)	0.385 (16)	0.890 (84)
Rat, inhalation, liver		—	0.745 (15)	0.830 (25)	0.466 (33)
Mouse, oral, liver			—	0.828 (8)	0.598 (35)
Mouse, inhalation, liver				—	0.639 (15)
Rat, oral, kidney					—

Part of scatter diagram matrix
Bottom-left: scatter plots (both axes represent logarithmic NOEL values)
Top-right: Pearson's correlation coefficients (parentheses: number of samples)

Step 2: Structural equation modeling

Determination of relationships among all the items based on the strength of correlation among the toxicity endpoint items

⬇

Derivation of the (implied) covariance structure of the data set

Step 3: Determination of the best prediction equation

$$\begin{cases} \hat{y} = \hat{\mu}_y + \Sigma_{xy}^t \Sigma_{xx}^{-1}(x - \hat{\mu}_x), \\ \mathrm{Cov}(e) = \Sigma_{yy} - \Sigma_{xy}^t \Sigma_{xx}^{-1} \Sigma_{xy}. \end{cases}$$

⟲ Existing animal test data of the substance to be estimated

○ The vectors and matrices that quantify the relationships among the toxicity endpoint items obtained in Step 2

Equation 1: Equation for estimating the mean values
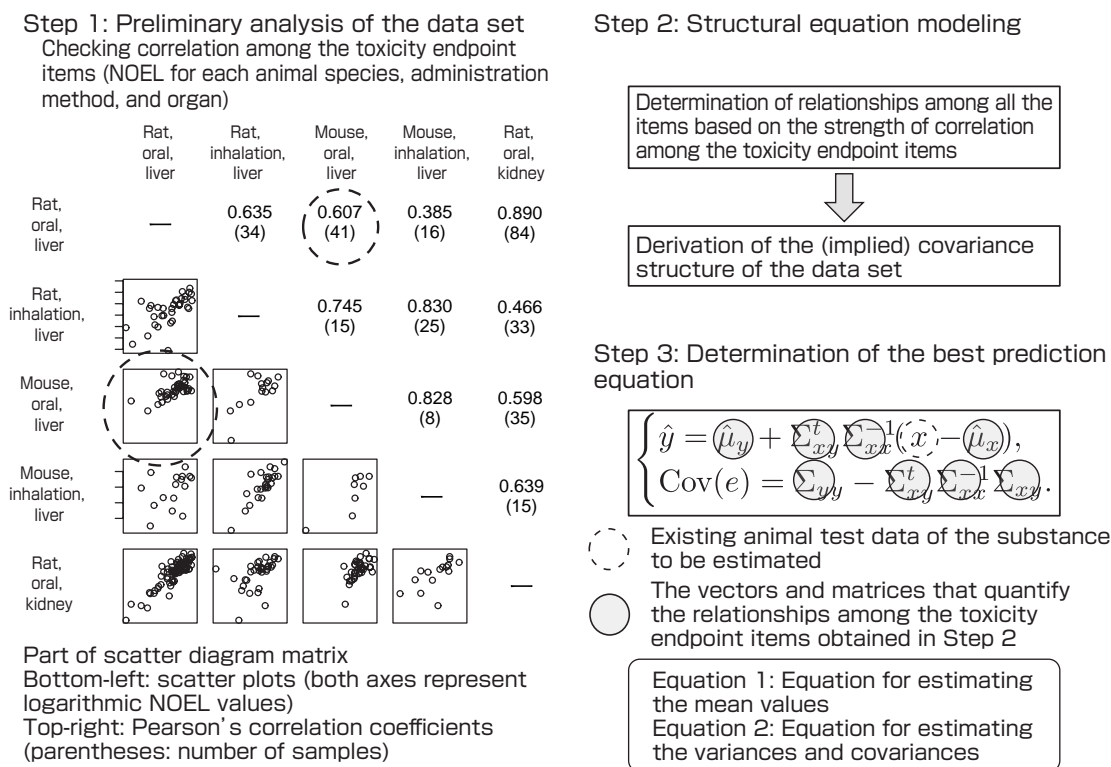Equation 2: Equation for estimating the variances and covariances

Fig. 4 Procedure for developing the QAAR method using structural equation modeling
The dashed circle in the scatter diagram matrix in Step 1 shows the Pearson's correlation coefficient and scatter diagram, when the NOEL of the liver effect due to oral exposure is compared for rats and mice (horizontal axis: rat value; vertical axis: mouse value).

Table 1. Change in estimated exposure level (decaBDE equivalent) with the substitution of flame retardant: Liver effect

		Estimated exposure level (mg/kg/day: decaBDE equivalent)			
		decaBDE	BDP	TPP	Total
Scenario with substitution	When mean value of relative toxicity value is used	2.1×10^{-4}	8.8×10^{-5}	4.0×10^{-7}	2.9×10^{-4}
	When 95 % upper limit of relative toxicity value is used	2.1×10^{-4}	4.1×10^{-4}	7.9×10^{-6}	6.3×10^{-4}
Scenario without substitution		3.3×10^{-4}	–	–	3.3×10^{-4}

Table 2. Change in estimated exposure level (decaBDE equivalent) with the substitution of flame retardant: Kidney effect

		Estimated exposure level (mg/kg/day: decaBDE equivalent)			
		decaBDE	BDP	TPP	Total
Scenario with substitution	When mean value of relative toxicity value is used	2.1×10^{-4}	6.9×10^{-5}	5.1×10^{-8}	2.7×10^{-4}
	When 95 % upper limit of relative toxicity value is used	2.1×10^{-4}	4.2×10^{-4}	3.1×10^{-7}	6.2×10^{-4}
Scenario without substitution		3.3×10^{-4}	–	–	3.3×10^{-4}

Table 3. Risk associated with the flame retardant in each substitution scenario[2] (day : quality-adjusted life-year : value per person for lifetime exposure)

	With substitution (current situation)		Without substitution (imaginary situation)
	When mean value of relative toxicity value is used	When 95 % upper limit of relative toxicity value is used	
Liver effect	<< 0.001 (2.8×10^{-57})	<< 0.001 (2.0×10^{-53})	<< 0.001 (9.5×10^{-57})
Kidney effect	<< 0.001 (1.4×10^{-140})	<< 0.001 (1.0×10^{-122})	<< 0.001 (8.8×10^{-137})
Total	<< 0.001 (2.8×10^{-57})	<< 0.001 (2.0×10^{-53})	<< 0.001 (9.5×10^{-57})

for Safety and Sustainability at AIST. Here, we present a summary of the replacement of a flame retardant used as a plastic additive.[2] A study was done on a scenario where decabromodiphenyl ether (decaBDE) was partially replaced by bisphenol-A bis(diphenyl phosphate) (BDP) [which includes triphenyl phosphate (TPP) as impurity], and an imaginary scenario where a replacement for decaDBE was not sought. For these scenarios, we conducted a material flow analysis consisting of the assessment of the demand volume, community-acquired stock volume, and waste volume of each substance. Then the respective exposure levels of each substance, through indoor air, environment, and food, was estimated. The risk tradeoff analysis was conducted based on the estimated exposure levels for each substance in each scenario.

Before conducting the quantitative risk analysis based on common indices, we first investigated the qualitative difference in risk due to substitution. DecaBDE, the original substance, was set as the reference substance, and the relative toxicity values for BDP and TPP were calculated; after which, the estimated exposure levels in each scenario were calculated and totaled. The overall exposure level was expressed as an equivalent amount of decaBDE (mg/kg/day). Table 1 shows the liver effects, and Table 2 the kidney effects. In the scenario with substitution, the estimated exposure level of decaBDE equivalent was less than in the scenario without substitution. Even with the addition of BDP and TPP, it could be determined, using the estimated mean value of relative toxicity, that the exposure level of decaBDE equivalent decreased slightly in the scenario with substitution. However, when the estimated 95 % upper limit of the relative toxicity value was used, the exposure level of decaBDE equivalent increased with substance substitution. This means that, when the uncertainty involved in the estimation of the replacement substance's toxicity was considered, it could not be determined whether the substance substitution would necessarily contribute to risk reduction. Here, the ratios of the 95 % upper limit and the mean of the relative toxicity value for the liver effect, for BDP and TPP, were 4.7 and 20, respectively; and for the kidney effect, 6.0 and 6.0, respectively. The difference in the ratio values for the different substances and organs reflects the fact that the availability and/or reported values of NOEL varied.

Table 3 shows the results of the quantitative risk comparison

by common indices, based on the method described in Chapter 3.[2] The magnitude of risk is expressed as the QALYs (days) per person, through lifetime exposure. In both scenarios, extremely small values, both for liver and for kidney effect, were found. A lifetime probability of cancer occurrence of 10^{-5}, often used as the upper limit of the tolerable risk level, is equivalent to a lifespan reduction of about 0.04 d.[23] Therefore, when the loss in QALYs is less than 0.001 d, it can be concluded that the effectiveness of the substance substitution, in terms of risk reduction, cannot be determined.

5 Discussion

The substance substitution risk tradeoff analyzed in the case studies of the flame retardant[2] (Chapter 4), lead solder alloy,[3] and industrial cleanser,[4] arose from concern for the risk presented by the original substances. To determine whether such substitutions are appropriate from the perspective of risk reduction, quantitative risk analysis based on scientific evidence is essential, and the aforementioned studies performed at the Research Institute of Science for Safety and Sustainability were the first instances of such analysis. This was primarily made possible by the development of the assessment method described in this paper. Kishimoto[5] developed the concept, and we engaged in developmental research on basic techniques to realize it.

Our risk tradeoff analysis involved chemical substance substitution aimed at preservation of the original substance's functionality; for example, preserving the flame-retardant property in plastic. However, related measures may involve simply reducing the amount used of a certain substance, to reduce the associated risks. In such cases, a tradeoff occurs which may involve risks other than chemical substance toxicity alone. To take two examples, the reduction of product functionality may increase the risk of fires or accidents, and reducing energy efficiency may increase the risk of global warming. While we have constructed a provisional framework for comparing the risks associated with different chemical substances, there remains the question of how to compare risk types other than toxicity, with the attendant questions of how to express such different risk types in terms of common indices, and how society should understand tradeoff involving risks with greatly varying properties. Such questions must be addressed, for example, in the risk analysis of substances (such as nanomaterials) used in future technologies, for which, as practical application is lacking, there is, as yet, no risk tradeoff concern, though strict regulation may be applied. However, if such regulation should limit the possibilities for future technology, there is a chance of forsaking future risk reduction in some fields, and this too must be considered.

The method of chemical substance risk tradeoff analysis

presented in this paper must also be improved in the following respects. The toxicity data that formed the basis for the QAAR model constructed for the relative comparison of toxicity consists of animal test results published in general toxicological journals, and the reliability of the data has not been carefully investigated. Currently, government agencies tend to utilize highly reliable test data, collected within the framework of the Law Concerning the Examination and Regulation of Manufacture, etc. of Chemical Substances, for the purpose of creating a structure-activity relationship model for toxicity. The rebuilding of models, based on such data, will increase the reliability of the assessment results. Also, this paper's QAAR model is based solely on the correlation among the toxicity endpoint items observed in animals; however, to enable the assessment of substances for which no animal test results are available, NOEL estimation values based on structure-activity relationships, as well as the results of cell-based assays, and chemical substance structure descriptors, should be used as variables in structural equation modeling. In addition, to increase the model's reliability, both variable selection and establishment of cause-effect relationships can be performed based on the information on the mode of action of chemical substances. Finally, in this study, the focus was on the effects on major organs (here, the liver and kidney). However, chemical substances may also have neurotoxicity and/or sensitizing effects, and future research topics must include the selection of reference substances, the understanding of dose-response relationships, and the estimation of QALYs, for these other effects.

Improvement is also needed in the treatment of uncertainty. In this paper, we drew attention to the importance of the estimation of uncertainty in the tradeoff analysis, and quantified the uncertainty in the estimation of relative toxicity values, based on the correlation among toxicity endpoint items. However, other relevant uncertainties were not treated here explicitly. Notable among these are the uncertainties involved in the selection of reference substances (even if the focus is on 'liver effect,' it may be necessary to establish different reference substances for different substances), the estimation of dose-response relationships for reference substances (uncertainty in the reliability of the human epidemiological data on the given reference substance, and/or uncertainty accompanying the derivation method for obtaining the given dose-response relationship), and the determination of QALYs for a given effect (even if the focus is on 'liver effect,' the degree of severity may differ among substances). Also, in terms of exposure assessment, there are uncertainties in the establishment of the substitution scenarios and estimation of exposure levels. We must, to some extent, accept the fact that uncertainty is an inherent feature of assessment. For example, though in the research and development of assessment methods, the main aim is to reduce the uncertainty in factors with a large degree of uncertainty, and much R&D work is focused on this aim,

there are cases where there is a limit on reducing uncertainty, and/or there is little or no effect on decision-making despite the presence of large uncertainty (for example, in the context of the present study, when the superiority of Substance A over Substance B is very clear, even with uncertainty considered). Nonetheless, to develop a risk tradeoff analysis method that contributes to effective decision making in future, the quantitative estimation of uncertainty, and the formation of consensus regarding the magnitude of uncertainty, will become increasingly important.

In presenting our novel method for risk tradeoff analysis, and the related case study, our hope is that consideration of the risk tradeoffs involved chemical substance substitution will ultimately extend to society at large; and in this way, it may be possible to avoid unhelpful regulations or the substitution of substances based on poor scientific evidence. As noted earlier, substance substitution incurs costs, and the risk may even increase in some cases. Effective risk tradeoff analysis is necessary both for the companies that manufacture chemical substances and for the general public that uses them. While there is no move for incorporating such assessment in specific regulations at this point, the OECD, for example, has established an ad hoc group focused on the substitution of hazardous chemical substances, and is discussing the development of methods to support safe assessment of substance substitution.[24] We, for our part, are making timely presentations of our assessment method and case studies to such bodies.

It is expected that the relative comparison-based approach to risk assessment, as described here, will be applied to risk assessments in diverse fields, not only to chemical substances. Particularly in the case of new technologies that may replace conventional ones, rather than simply estimating the risks that may be introduced by such new technologies, it will be more realistic and convincing to society at large to perform relative comparisons of the two, based on a clear understanding of the current risks posed by the conventional technologies. Future topics to be studied, then, include the opportunities and requirements for applying the relative comparison approach, the identification and organization of uncertainties inherent in such assessment, and the relationship between such uncertainties and acceptability of the assessment by society.

Acknowledgments

References

[1] National Institute of Technology and Evaluation (NITE): PRTR todokede data no nendo hikaku (Comparison by fiscal year of PRTR notification data), http://www.nite.go.jp/chem/prtr/prepdata.html, accessed 2016-03-31 (in Japanese).

[2] Research Institute of Science for Safety and Sustainability, AIST (2012): Risk tradeoff hyokasho plastic tenkazai—Nannenzai (Risk tradeoff analysis of plastic additives—Flame retardant), https://www.aist-riss.jp/assessment/12151/, accessed 2016-03-31 (in Japanese).

[3] Research Institute of Science for Safety and Sustainability, AIST (2012): Risk tradeoff hyokasho kinzoku—Namari handa (Risk tradeoff analysis of metals—Lead solder alloy), https://www.aist-riss.jp/assessment/12151/, accessed 2016-03-31 (in Japanese).

[4] Research Institute of Science for Safety and Sustainability, AIST (2012): Risk tradeoff hyokasho kogyoyo senjo (Risk tradeoff of industrial cleanser), https://www.aist-riss.jp/assessment/12151/, accessed 2016-03-31 (in Japanese).

[5] A. Kishimoto: A strategic approach for comparing different types of health risks—Risk assessment of toluene exposure using quality-adjusted life years, *Synthesiology*, 1 (1), 31–37 (2008) (in Japanese) [*Synthesiology English edition*, 1 (1), 31–37].

[6] M. Gamo, T. Oka and J. Nakanishi: A method evaluating population risks from chemical exposure: A case study concerning prohibition of chlordane use in Japan, *Regul. Toxicol. Pharm.*, 21 (1), 151–157 (1995).

[7] M. Gamo, T. Oka and J. Nakanishi: Ranking the risks of 12 major environmental pollutants that occur in Japan, *Chemosphere*, 53 (4), 277–284 (2003).

[8] J. T. Cohen, D. C. Bellinger, W. E. Connor, P. M. Kris-Etherton, R. S. Lawrence, D. A. Savitz, B. A. Shaywitz, S. T. Teutsch and G. M. Gray: A quantitative risk-benefit analysis of changes in population fish consumption, *Am. J. Prev. Med.*, 29 (4), 325–334 (2005).

[9] World Health Organization (WHO): Global burden of disease, http://www.who.int/topics/global_burden_of_disease/en/, accessed 2016-03-31.

[10] A. H. Havelaar, A. E. M. De Hollander, P. F. M. Teunis, E. G. Evers, H. J. Van Kranen, J. F. M. Versteegh, J. E. M. Van Koten and W. Slob: Balancing the risks and benefits of drinking water disinfection: Disability adjusted life-years on the scale, *Environ. Health Perspect.*, 108 (4), 315–321 (2000).

[11] Y. Sakuratani, H. Q. Zhang, S. Nishikawa, K. Yamazaki, T. Yamada, J. Yamada, K. Gerova, G. Chankov, O. Mekenyan and M. Hayashi: Hazard evaluation support system (HESS) for predicting repeated dose toxicity using toxicological categories, *SAR QSAR Environ. Res.*, 24 (5), 351–363 (2013).

[12] M. S. V. Maier: Setting occupational exposure limits for unstudied pharmaceutical intermediates using in vitro parallelogram approach, *Toxicol. Mech. Methods*, 21 (2), 76–85 (2011).

[13] J. Nakanishi, Y. Morimoto, I. Ogura, N. Kobayashi, M. Naya, M. Ema, S. Endoh, M. Shimada, A. Ogami. T. Myojyo, T. Oyabu, M. Gamo, A. Kishimoto, T. Igarashi and S. Hanai: Risk assessment of the carbon nanotube group, *Risk Anal.*, 35 (10), 1940–1956 (2015).

[14] J. Takeshita, M. Gamo, K. Kanefuji and H. Tsubaki: A quantitative activity-activity relationship model based on covariance structure analysis, and its use to infer the NOEL values of chemical substances, *J. Math-for-Ind.*, 5B, 151–159 (2013).

[15] The Collaborative on Health and the Environment

(CHE): CHE Toxicant and Disease Database, http://www.healthandenvironment.org/tddb, accessed 2016-03-31.

[16] S. F. Ho, W. H. Phoon, S. L. Gan and Y. K. Chan: Persistent liver dysfunction among workers at a vinyl chloride monomer polymerization plant, *Occ. Med.*, 41 (1), 10–16 (1991).

[17] M. A. Huijbregts, L. J. Rombouts, A. M. Ragas and D. van de Meent: Human-toxicological effect and damage factors of carcinogenic and noncarcinogenic chemicals for life cycle impact assessment, *Integr. Environ. Assess. Manag.*, 1 (3), 181–244 (2005).

[18] M. Gamo, K. Ono and J. Nakanishi: Meta-analysis for deriving age and gender-specific dose-response relationship between urinary cadmium concentration and β_2-microglobulinuria under environmental exposure, *Environ. Res.*, 101 (1), 104–112 (2006).

[19] J. Nakanishi, K. Ono, M. Gamo and K. Miyamoto: *Shosai Risk Hyokasho Series 13 Cadmium* (Detailed Risk Analysis Series No. 13—Cadmium), Maruzen (2008) (in Japanese).

[20] Research Institute of Science for Safety and Sustainability, AIST (2012): Kagaku busshitsu no daitai ni tomonau risk tradeoff hyoka no tameno guidance (hito kenko) yobiteki kokaiban [Guidance for risk tradeoff analysis in the use of substitution chemical substances (Human health) Preliminary disclosure version], https://www.aist-riss.jp/assessment/12131/, accessed 2016-03-31 (in Japanese).

[21] T. O. Tengs and A. Wallace: One thousand health-related quality-of-life estimates, *Med. Care*, 38 (6), 583–637 (2000).

[22] H. Nakagawa: Cadmium osen chiiki jumin no kenko shogai ni kansuru kenkyu: Jin-nyosaikan shogai teido oyobi cadmium bakuroryo to seimei yogo—15 nennkan no tsuiseki chosa (Research on the health damage to residents of areas polluted by cadmium: Levels of renal tubular disorder, cadmium exposure level, and life expectancy—15 years of follow-up survey), *Kankyo Hoken Report*, 65, 76–79 (1999) (in Japanese).

[23] M. Gamo, T. Oka and J. Nakanishi: Estimation of the loss of life expectancy from cancer risk due to exposure to carcinogens using life table, *Environmental Science*, 9 (1), 1–8 (1996) (in Japanese).

[24] Organisation for Economic Co-operation and Development (OECD): Substitution of hazardous chemicals, http://www.oecd.org/env/ehs/risk-management/substitutionofhazardous chemicals.htm, accessed 2016-03-31.

Authors

Masashi GAMO
Graduated from the Department of Urban Engineering at the University of Tokyo, in 1991. Completed the doctoral program of the Department of Urban Engineering, in the Graduate School of Engineering at the University of Tokyo, in 1996. PhD (Engineering). Joined the National Institute for Resources and Environment of the Agency of Industrial Science and Technology, in 1996. Joined the Research Center for Chemical Risk Management at AIST, in 2001 (became Team Leader, Risk Management Strategy Team, as of 2002). Group Leader, Risk Assessment Strategy Group, Research Institute of Science for Safety and Sustainability, in 2008. Engages in the development of risk assessment methods for chemical substances and nanomaterials, and conducts cumulative case studies. In this paper, was in charge of investigating the framework of the risk tradeoff analysis method, and write-up of relevant sections of the paper.

Jun-ichi TAKESHITA
Graduated from the Department of Applied Physics, in the School of Science and Engineering at Waseda University, in 2004. Completed the doctoral program in the Department of Mathematics of the Graduate School of Mathematics at Kyushu University, in 2009. PhD (Functional Mathematics). Joined AIST in 2011, assigned to the Risk Assessment Strategy Group of the Research Institute of Science for Safety and Sustainability. Specialty is applied mathematics. Engages in the research of mathematical analysis methods for risk assessment and management of chemical substances, as well as conducting case studies. In this paper, was in charge of building the model for the relative toxicity value calculation for substances, case studies, and write-up of related sections of the paper.

Discussions with Reviewers

1 Overall
Comment (Akira Ono, AIST)

This paper proposes a novel, practical risk assessment methodology, based on the technique of introducing an index that can be broadly employed to quantitatively compare the magnitude of risks with differing factors. It should be noted that this method was applied to existing cases to verify its effectiveness.

Some different scenarios for developing the practical risk assessment methodology were compared. The processes of scenario selection are described, and this paper is appropriate as a research paper for *Synthesiology*.

2 Uncertainty in the risk estimation
Question (Akira Ono)

I value highly the fact that you considered and executed the method for estimating the uncertainty of risk (95 % confidence interval: 2σ equivalent). Uncertainty is an important concept, and I think it will be at the core of future research development. Based on this way of thinking, I pose the following question, as well as the subsequent questions and comments.

I would like to understand the specific magnitude of uncertainty. In Tables 1, 2, and 3, how many times greater is the uncertainty (2σ) than the mean value?

You estimate the relative toxicity value of the target substance without human epidemiological data using a statistical method, but what are the factors of uncertainty in this case? Please give examples of factors that have large effects (for example, the quality and amount of animal test data is insufficient, the correlation among toxicity endpoint items is low, or the similarity between the selected reference substance and the target substance is low).

I think one of the key challenges in this research is how to decrease the uncertainty involved in estimating the relative toxicity value of the target substance. To do so, I think it is necessary to clarify the factors that enhance the uncertainty, and make efforts to reduce them, beginning with the largest contributor. What do you think?

Answer (Masashi Gamo)

We did not show the relative toxicity value itself; but the ratios between its mean and 95 % upper limit values for the liver effect, for BDP and TPP, were 4.7 and 20, respectively; and for the kidney effect, 6.0 and 6.0, respectively. The differences in the ratio values for different substances and organs reflect the fact that there are differences in the availability and/or reported values of NOEL.

In this paper, we highlighted the importance of the estimation of uncertainty in the tradeoff analysis, and quantified the uncertainty involved in the estimation of relative toxicity values based on the correlation among toxicity endpoint items. On the other hand, as you indicated, there are several uncertainties that are not treated explicitly. Notable among these are the uncertainties involved in the selection of reference substances (even if the focus is on 'liver effect,' it may be necessary to establish different reference substances for different substances), the estimation of dose-response relationships for reference substances (uncertainty in the reliability of the human epidemiological data on the given reference substance, and/or uncertainty accompanying the derivation method for obtaining the given dose-response relationship), and the determination of QALYs for a given effect (even if the focus is on 'liver effect,' the degree of severity may differ among substances). Also, in terms of exposure assessment, there are uncertainties in the establishment of the substitution scenarios and estimation of exposure levels. We must, to some extent, accept the fact that uncertainty is an inherent feature of assessment. For example, though in the research and development of assessment methods, the main aim is to reduce the uncertainty in factors with a large degree of uncertainty, and much R&D work is focused on this aim, there are cases where there is a limit on reducing uncertainty, and/or there is little or no effect on decision-making despite the presence of large uncertainty (for example, in the context of the present study, when the superiority of Substance A over Substance B is very clear, even with uncertainty considered). Nonetheless, to develop a risk tradeoff analysis method that contributes to effective decision making in future, the quantitative estimation of uncertainty, and the formation of consensus regarding the magnitude of uncertainty, will become increasingly important.

I have added these comments to the text.

3 Verification of the adequacy of the estimation method for uncertainty
Question (Akira Ono)

You estimate the uncertainty of risk estimate values using statistical methods. Is there any way to verify this method's adequacy?

For example, let's say you select several reference substances for a certain target substance, and see how much variation there is in the risk estimate values of the target substance with respect to each reference substance; can you use this as an index of the adequacy of the method?

Or, for example, say you select two substances that have sufficient human epidemiological data, position one of these as the reference substance and the other as the target substance, and then estimate the target substance risk using the method described in this paper. Here, you do not use the human epidemiological data for the target substance, but only the animal test data. I think you could do the same thing by interchanging the reference substance and the target substance. Might not the difference between the highly reliable estimate value based on the human epidemiological data, and the estimate value based only on animal test data, function as an index of the adequacy of the method?
Answer (Masashi Gamo)

Of the methods presented in the paper, we reported the results of the leave-one-out cross-validation method concerning

the adequacy of the estimation of uncertainty involved in the estimation of relative toxicity value, as shown in Subchapter 3.3. We found that 93 % of the observed values were included in the 95 % prediction interval, and we believe the estimation of uncertainty was adequate. This method is the most widely accepted, by various bodies (including the OECD), as verification of statistical estimation.

On the other hand, for the method of estimating the uncertainty involved in the establishment of reference substances, I think it is effective, as you suggested, to specify several such substances and compare the estimate results based on these, or to mutually estimate and verify the NOEL of reference substances; although there are not many substances with human epidemiological data, which means that this is one of the topics to be investigated in the future.

4 Use of data from the scenario that was not selected
Comment (Akira Ono)

In Fig. 1, you compare four scenarios and select the fourth scenario for this research, regarding the three non-selected scenarios as non-executable. However, isn't it true that the exposure/tolerance concentration data, structure-activity relationships, and cell-based assay data in the scenarios that you deemed non-executable contain a certain amount of useful information on toxicity, although it is impossible to individually execute the related risk tradeoff analysis? And if you could incorporate these data into this research in some form, could you not thereby reduce even more the uncertainty involved in the risk assessment?

As a future direction in Paragraph 3 of Chapter 5 (Discussion), you note that the point in increasing the reliability of estimation results is to incorporate and integrate such data into the methodology of this study. Please explain, even if it is just your current thoughts, by what mechanism you can reduce the uncertainty by integrating this data.
Answer (Masashi Gamo)

As you indicated, we think that the three non-selected scenarios may contain data and methods that can be utilized effectively in the risk tradeoff analysis.

For substances that have absolutely no animal test data, it is important, from the perspective of animal rights, to use the structure-activity relationships and cell-based assay results. One possible approach is to use the NOEL estimation values based on structure-activity relationships, cell-based assay results, and the chemical substance structure descriptors, as observation variables in the structural equation modeling. On the other hand, in the method presented in this paper, the characterization of the correlation among the variables in the structural equation modeling depended only on the correlation among toxicity endpoint items. However, we can also select variables, and establish the cause-effect relationships among them, based on the information on the mode of action of chemical substances, which is also discussed in the treatment of the general risk assessment method.

Paragraph 3 of Chapter5 (Discussion) was modified to reflect these concerns.

5 Reason for replacing the flame retardant
Question (Akira Ono)

Table 3 shows the results of risk assessment, taking the example of a conventional flame retardant. I understand that the right column of the table (with no substitution, the imaginary situation) shows the reduction in QALYs if the conventional flame retardant is used, compared to when we stop using that flame retardant.

The conventional flame retardant and replacement substance

both have very small risk estimation values, in fact dramatically small compared to the reduction in QALYs of 0.04 days, which is the upper limit of the tolerable risk level. What is the reason for the extremely low risk values obtained in the assessment? Does this show that the risk has been overrated, and is in fact very small? Or is it due to your introduction of the common index of QALYs? Please explain.

And another question along the same lines. According to this assessment result, the original-substance reduction in QALYs is negligibly small, and you conclude that the substance substitution was scientifically meaningless. I think this conclusion is correct; but then, why was this flame retardant substituted for the original? What was the reason? The reason might not have been scientifically sound, but it must have been convincing enough at the time. Please provide the authors' view, to the extent of your knowledge, on what the reason was.

Answer (Masashi Gamo)

In Table 3, the magnitude relationships between risk estimation values (QALYs) in the scenarios with and without substitution are reversed, depending on whether one uses the mean or the 95 % upper limit of the relative toxicity value. This is true also in Tables 1 and 2, but it indicates that one cannot determine whether the substitution of a given substance necessarily contributes to risk reduction, considering the uncertainty involved in the assessment.

In this specific case study, it was calculated that the risk estimation value was extremely small. I think we would have reached the conclusion that there was no risk, using the general risk assessment method (comparison of exposure and tolerance levels) as well. However, as explained in Subchapter 2.1, the general risk assessment method could only provide the conclusion that 'there is no risk,' whereas in this study, the magnitude of risk could be specifically presented in terms of QALYs. By doing so, it was clarified that the risk was not at a level that necessitated discussion of risk reduction measures.

The reason for the flame retardant substitution was probably that people were interested only in the toxicity of the conventional flame retardant, and there was widespread desire to avoid this substance. I do not think even a general risk assessment was done at the time, and comparison of the hazards or risks of the original and replacement substances was not done in detail, unlike the careful analysis performed in this paper.