

Summary of a Workshop on Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints

Joanna S. Jaworska,¹ M. Comber,² C. Auer,³ and C.J. Van Leeuwen⁴

¹Procter & Gamble Eurocor, Strombeek-Bever, Belgium; ²ExxonMobil Biomedical Sciences Inc., Annandale, New Jersey, USA;

³U.S. Environmental Protection Agency, Washington, DC, USA; ⁴European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Ispra, Italy

The “Workshop on Regulatory Use of (Q)SARs for Human Health and Environmental Endpoints,” organized by the European Chemical Industry Council and the International Council of Chemical Associations, gathered more than 60 human health and environmental experts from industry, academia, and regulatory agencies from around the world. They agreed, especially industry and regulatory authorities, that the workshop initiated great potential for the further development and use of predictive models, that is, quantitative structure–activity relationships [(Q)SARs], for chemicals management in a much broader scope than is currently the case. To increase confidence in (Q)SAR predictions and minimization of their misuse, the workshop aimed to develop proposals for guidance and acceptability criteria. The workshop also described the broad outline of a system that would apply that guidance and acceptability criteria to a (Q)SAR when used for chemical management purposes, including priority setting, risk assessment, and classification and labeling. *Key words:* quantitative structure–activity relationships, regulatory acceptance. *Environ Health Perspect* 111:1358–1360 (2003). doi:10.1289/ehp.5757 available via <http://dx.doi.org/> [Online 5 February 2003]

A 3-day scientific workshop titled “Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints,” hosted by the European Centre for Ecotoxicology and Toxicology of Chemicals and organized by the International Council of Chemical Associations (ICCA) and the European Chemical Industry Council (CEFIC; as part of their long-range research initiative) was held 4–6 March 2002 in Setubal, near Lisbon, Portugal. Participants of the Setubal workshop had a diverse background both in human and environmental safety and in associations with academic institutions, government bodies, or industry from Europe, North America, and Japan. Participants agreed that the workshop initiated great potential for the further development of predictive models and their application for chemicals management, including priority setting, risk assessment, and classification and labeling.

One of the key messages during the workshop was that both industry and regulatory authorities share the same goal, that is, to use quantitative structure–activity relationships [(Q)SARs] in a much broader scope than currently practiced for safety evaluation and chemicals management. Consequently, there was a clear agreement on the need to continue dialogue and cooperation.

(Q)SARs are simplified mathematical representations of complex chemical–biological interactions. They can be divided into two major types, QSARs and SARs. QSARs are all quantitative models yielding a continuous or categorical result. The most common techniques for developing QSARs are regression analysis, neural nets, and classification methods. Examples of regression models include

ordinary least squares and partial least squares, whereas for neural nets back-propagation methods would be commonly used. Examples of classification methods are discriminant analysis, decision trees, and distance-based similarity analysis. SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity.

(Q)SAR predictions have the potential to save time and money and minimize the use of animal testing. However, to fulfill this potential, the predictions, especially those considered for regulatory decision making, need to be scientifically valid, appropriate for the purpose intended, reliable, and accepted by decision makers. Approaches to determine the acceptability of (Q)SARs have been developed in the past [e.g., guidance from the Organisation for Economic Development (OECD)], but because of their breadth and generality, they have not been widely applied or respected by either (Q)SAR users or developers. As a consequence, decision making with the help of existing models must be done with care and considerable knowledge. The workshop in Setubal aimed at reopening the debate to develop more specific guidance and acceptability criteria and a system that would support the use of (Q)SARs such that the guidance and acceptability criteria were actually applied when a (Q)SAR was used for chemicals management purposes.

Acceptability Criteria for (Q)SARs

(Q)SAR predictions are derived from simplified mathematical representations of complex chemical–biological interactions and,

consequently, are potentially more uncertain than the underlying test data. This imposes limitations on the acceptable scope of a (Q)SAR use in chemicals management and decision making. The general acceptability criteria developed for alternative methods to animal testing by the European Center for the Validation of Alternative Methods were discussed in the context of (Q)SARs and were fully accepted. These criteria indicate that an alternative model should *a*) be associated with a defined end point that it serves to predict; *b*) take the form of an unambiguous and easily applicable algorithm for predicting a pharmacotoxicologic end point; *c*) ideally have a clear mechanistic basis; *d*) be accompanied by a definition of the domain of its applicability, for example, the physicochemical classes of chemicals for which it is applicable; *e*) be associated with a measure of its goodness of fit and internal goodness of prediction estimated with cross-validation or similar method to a training set of data; and *f*) be assessed in terms of its predictive power by using data that were not used in the development of the model (external validation).

The workshop participants agreed that (Q)SARs are one of the alternative methods to animal testing, and therefore, these generic criteria can and should be further refined specifically for (Q)SARs. The acceptability criteria were divided into two components, statistical and nonstatistical. The discussions on statistical criteria centered on proposals circulated before the meeting, which focused on the regression and two-way classification models. As a result of the quantitative nature of these criteria, it was possible to make them very specific.

Specific criteria for continuous QSAR models. One of the clear outcomes was that for regulatory purposes a QSAR’s predictive power and the prediction uncertainty must be reported along with the goodness of fit. Typical values were recommended, and it was agreed these should be used in the subsequent testing of the criteria.

This article is part of the mini-monograph “Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints.”

Address correspondence to J.S. Jaworska, Procter & Gamble Eurocor, Temselaan 100, 1853 Strombeek-Bever, Belgium. Telephone: 32-2-456-20-76. Fax: 32-2-568-30-98. E-mail: jaworska.j@pg.com

The authors declare they have no conflict of interest.

Received 2 May 2002; accepted 20 August 2002.

Specific criteria for classification QSARs. Again, values describing goodness of fit, including specificity, sensitivity, and negative and positive predictive power, were proposed and accepted. Another factor that needs to be evaluated is the minimization of false positive/negatives by sequential use of models.

Specific criteria for SARs. The issues arising from the need to assess such models were felt to be model specific, but clearly included similarity analysis. It was recognized that further research was needed to address SARs and the application of expert knowledge models.

Nonstatistical criteria for (Q)SARs. The nonstatistical criteria discussed were associated with the endpoint, chemical descriptors, mechanism, domain, and transparency. In particular, it was agreed that predictive models should be transparent. Transparency in this context means that there should be access to the training and validation data sets as well as to the methods used for the development and validation of the model. Thus, an informed user with the correct tools would be able to re-create the model using the same data and techniques as the original developer.

Another outcome of the discussion about nonstatistical criteria was realization of the difference between human health and environmental predictive models because of differences in the nature of the endpoints studied. In general, QSARs for environmental endpoints are founded on relatively large quantitative databases with sufficient mechanistic understanding to enable the model to have useful predictive capability. Furthermore, it is relatively easy to support the prediction with subsequent testing. In contrast, the ability to predict local effects in humans is currently limited by a lack of good quality data and, consequently, has limited regulatory use. For systemic human health endpoints, the models are poor because the traditional endpoints (e.g., LD₅₀, no-observable-adverse-effect level, lowest-observable-adverse-effect level), although suitable for current methods of chemicals management, are not suitable for (quantitative) modeling. These complex human health endpoints are expressed through many different mechanisms, are often receptor mediated, and are multistage processes comprising absorption, distribution, metabolism, and excretion (ADME), frequently with site-specific interactions. Furthermore, it was concluded that often the endpoints were not defined by a clear dose response and that steady-state concentrations in animals were often not achieved. In the light of these discussions, the workshop participants felt that further work was needed to increase the availability of good quality data where possible. It was also recognized that before attempting improvements to the predictive models for complex *in vivo* human health endpoints, the

existing methods needed to be evaluated for their potential to generate additional data more useful for modeling purposes. It should be noted that, because of lack of appropriate experts, the workshop participants did not extensively discuss reproductive toxicity and repeated dose effects.

Future QSAR Applications

By developing the acceptability criteria, the workshop participants agreed that considerable progress had been made in refining when and which (Q)SARs can support chemicals management decisions. This was recognized as a convergence of industry and regulatory agency positions regarding the scope of QSAR use and an acceptance that both positive and negative assignment of a chemical can be achieved with (Q)SAR models. Participants agreed on the following points:

- Acceptable levels of uncertainty in a prediction will depend upon the chemical management decision being made; that is, the model should be fit for purpose.
- The smaller the change in a prediction that would affect a decision, the more certain that prediction should be. This means that there will need to be a balance between the accuracy of a (Q)SAR prediction and the (Q)SAR applicability domain, depending upon the decision to be made.
- Uncertainty in a prediction should always be considered in the light of the underlying variability of the experimental data. If the prediction uncertainty matches the inherent variability associated with the endpoint, then animal testing should be avoided.

It was thus concluded that if the uncertainty of the prediction were such that a “wrong” decision might be made, targeted testing could be conducted to confirm the data point. In other circumstances, the (Q)SAR prediction would be judged acceptable and no testing would be warranted.

The workshop participants assumed a scenario with a (Q)SAR that met all the validation and acceptability requirements previously agreed. It was agreed that such a (Q)SAR could potentially be used for prioritization, risk assessment, and/or classification and labeling. It was felt that (Q)SARs appear to be most useful for risk-based priority setting, for risk assessment at the lower tiers of the assessment, and for rational prioritization for testing and test design. In this way, it is possible to evaluate whether the substance would trigger specific concerns on the basis of structural alerts, analog information, or (Q)SAR. The tiered approach in risk assessment allows for the collection of more accurate effects data and, if warranted, guides further and higher tier animal testing. Such a tiered approach will also help in decisions on test species selection. For example, a confirmatory test with one of the

most sensitive species could be made instead of simultaneously generating data on all species of a regulatory scheme. If a prediction and an experimental value agree, then further testing could be derogated and QSAR results for the other species could be used.

The workshop participants recognized that, until recently, there had been limited regulatory uses of (Q)SARs. The major use had been in the support of chemical assessment and notification in the United States. In the last few years, this had begun to change, as both within the European Union (Denmark, the Netherlands) and in other countries (e.g., Canada and Japan), programs are being developed that will considerably increase the use of (Q)SARs.

Management of Accepted (Q)SAR Models

It was acknowledged that there is no rigorous framework for the use of (Q)SAR. The workshop participants agreed that such a framework is needed because this will support the users, both regulatory and industry, in their decision making. The system discussed had the following key elements: *a*) transparent databases with flexible search engines, *b*) validated (Q)SAR models that meet agreed acceptability criteria, *c*) a biotransformation and metabolic simulation model

It was also recognized that such a system, regardless of its complexity, should be user friendly, incorporate tools to aid the selection of appropriate (Q)SARs, and be generally available via the Internet. The workshop agreed that such a decision support framework should attempt to help non-QSAR experts choose the most appropriate models, thus aiding them in the decision-making process.

The decision support system should be dynamic, that is, should allow for the continuous refinement of existing (Q)SARs. It is expected that, when the acceptability criteria are properly applied to a (Q)SAR, there will not be a major impact on chemical-specific predictions. With consequent model improvements, the predictions for “old” structures would not be significantly altered. Rather, the models would improve through expansion of the applicability domain and thus predictions for structures not previously covered by the (Q)SAR. It was recognized, however, that this was an area where further work is needed to help build confidence by all users in the approaches being advocated.

The decision support system should be supported and maintained by an independent organization that can hold data (experimental or models) and/or validate data and models. This organization might also be a potential holder of proprietary data or provide a mechanism for better sharing of data between model developers.

Actions and Recommendations

Three review papers were circulated among participants before the workshop and provided a common background and helped in the discussions (Cronin et al. 2003a, 2003b; Eriksson et al. 2003).

These papers have been invited for publication in *Environmental Health Perspectives* together with this summary of the workshop.

The workshop urged industry to take the lead to further develop the acceptability criteria and test existing (Q)SARs against the proposed criteria for use with the various applications, that is, priority setting, risk assessment, and classification. This exercise is felt to be necessary to quickly build general confidence in QSAR use for these purposes and especially to address the acceptability of both negative and positive classifications by all interested parties. This work would then be used to further the development of an appropriate decision support system.

This would have the additional benefit of enhancing research within industry to focus on development and selection of alternative chemicals, which was a key component of the proposed U.S. Sustainable Futures program (<http://www.epa.gov/chemrtk.volchall.htm>), and the European Union White Paper on New

Chemical Policy (<http://europa.eu.int/comm/environment/chemicals/whitepaper.htm>).

Although it was recognized that the U.S. Environmental Protection Agency's High Production Volume Chemicals Challenge program (<http://www.epa.gov/chemrtk/volchall.htm>) was unlikely to generate large volumes of new data, the data it will generate should be used to validate existing (Q)SARs, to add to or to develop new (Q)SARs, and to further test the proposed criteria.

The recommendations from the workshop regarding the need for acceptability criteria, validation of (Q)SAR models for the purpose of regulatory applications were submitted to OECD. In response, in November 2002 OECD initiated a (Q)SAR program by forming an ad-hoc expert group that developed a 2-year work plan. The work plan was approved in June 2003 by the OECD joint meeting and contains the following items:

- Apply the principles agreed upon at the ICCA Workshop on Regulatory Acceptance of (Q)SARs and the general OECD validation principles for new and updated test methods to selected (Q)SARs for regulatory use.
- Develop guidance documents for development, validation, and regulatory acceptance of (Q)SARs.

- Identify practical approaches to enable (Q)SARs to be readily available and accessible, including the development of a database of accepted (Q)SARs.

The work will be undertaken within the Test Guidelines Programme. Involvement of OECD member countries and other stakeholders in revising and expanding the current OECD guidance on (Q)SARs along the lines described by the workshop will be a very important driver in expanding the role for, and reliance on, (Q)SARs in regulatory decision making.

REFERENCES

- Cronin MTD, Jaworska JS, Walker JD, Comber MHI, Watts CD, Worth A. 2003a. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ Health Perspect* 111:1391–1401.
- Cronin MTD, Walker JD, Jaworska JS, Comber MHI, Watts CD, Worth AP. 2003b. Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environ Health Perspect* 111:1376–1390.
- Eriksson, L, Jaworska J, Worth AP, Cronin MTD, McDowell, RM, Gramatica Paola. 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 111:1361–1375.