Editorial

No Prediction Model, No Validation Study

There has recently been increasing interest in the development and validation of alternative methods that could be used in the place of *in vivo* toxicity tests. The goal is that a toxicologist will be able to test a substance by an alternative method, convert the results obtained into correct predictions of toxic hazard and, ultimately, use the predictions for making decisions about the safety of a test substance. If a toxicologist can be assured that the predictions obtained from an alternative method will lead to correct risk assessment decisions, the method may replace the *in vivo* test.

Theoretical discussions on the assessment of alternative method validity have emphasised the importance of confirming the technical performance of an alternative method, but have been less clear about how to confirm that a method correctly predicts a toxic endpoint.[1-3] The discussions commonly recommend that the data obtained from a validation study be thoroughly searched *after* the study has been completed, in order to determine whether the alternative method might be useful for making predictions. In this situation, the data from the validation study are used to construct the models to be used for making toxicity predictions.

When Prediction Models are constructed after a study is finished, reviewers of validation studies are left wondering whether the new models are specific only to the reference set of test substances used in the study, whether the variability is generally representative of variability in each test system (alternative method and *in vivo* test), and whether the models will have general applicability for the majority of substances to be tested in the future. This is because a *post hoc* assessment *develops* Prediction Models. However, the real purpose of a validation study is to *confirm* Prediction Models, not *develop* them. In order to make validation the confirmatory process it has to be, the models used to convert the results from an alternative method into predictions of toxic hazard must be defined *before* the study starts.

What is a Prediction Model?

A Prediction Model is the tool that is used to convert the results from an alternative method into a prediction of toxicity *in vivo*. It is created from the data generated during the thorough evaluation of results obtained from an optimised method.[4] A Prediction Model is essential for any test, because it defines exactly how to use the test results to predict a desired toxicity endpoint. Whether we have realised it or not, toxicologists have always used Prediction Models when making safety assessments. For example, linking hazard descriptions to results from the Draize eye irritation scoring scheme is a type of Prediction Model. So are classification schemes used by regulators for classifying the toxic hazard of test substances. The ability to make correct predictions is extremely important, since a toxicologist uses the predictions to make decisions during a safety assessment. If a test method does not have an adequate Prediction Model, there is simply no way to use it.

A Prediction Model associated with an alternative method can be considered adequate when it consists of four elements. These elements include a definition of the specific purpose(s) for which the alternative method is to be used, a definition of all the possible results that may be obtained from an alternative method (inputs), an algorithm that defines how to convert each alternative method result into a prediction of the *in vivo* toxicity endpoint (outputs), and an indication of the accuracy and precision of outputs obtained from

the model. Although the important factors associated with these elements have been described in detail elsewhere,[5] it is useful to review them briefly.

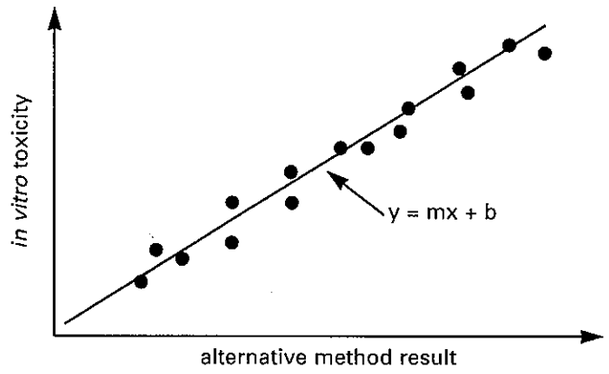
Firstly, a Prediction Model must define the specific purposes for which the alternative method is to be used. This must include a clear description of the endpoint the alternative method is used to predict. This may be a prediction of an endpoint as general as the maximum average score from the eye irritation test, or as specific as measurement of toxicity in a single cell in a specific organ (for example, renal proximal tubule cells). The Prediction Model must also define the chemical classes, product categories and physical forms of test substances for which it can be applied.

Secondly, a Prediction Model must define all of the possible types of results that may be obtained from the alternative method. Experience suggests that there are several types of results obtainable from alternative methods, including quantitative data, censored data, qualitative data, and non-qualified data.

Thirdly, a Prediction Model must adequately define the conversion algorithms that translate each alternative method result into a prediction of the *in vivo* toxicity endpoint. An example of such an algorithm is illustrated in Figure 1. This plot shows that the results from a hypothetical alternative method are directly related to the level of toxicity measured *in vivo*. In this case, the relationship can be described in terms of the standard equation for a line, $y = mx + b$, where m is the slope of the regression line, and b represents the value of the y intercept of the regression line. If this algorithm is true for all test materials, then any result, x', from this alternative method could be input into the algorithm, $y = mx + b$, to obtain an output, y', which represents the prediction of toxicity *in vivo*. Even though this example illustrates a linear model, non-linear models can also be employed. Figure 2 illustrates a non-linear model and the prevalidation data set used to construct it.

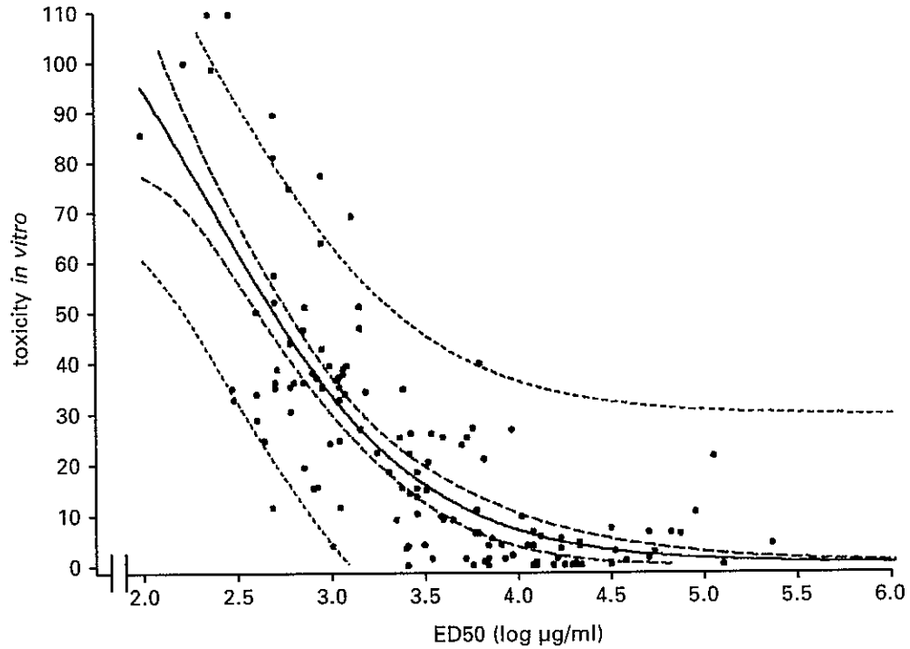
It is important to note that the conversion algorithms do not necessarily need to be mathematical equations. For example, algorithms may describe how to convert the alternative method data into classifications that fit a particular *in vivo* toxicity test

Figure 1: Algorithm prediction of *in vivo* toxicity endpoint using a linear model



In order for an alternative method to be useful, there must be a consistent and definable relationship between toxicity measured in vivo and corresponding results in the alternative method. In this case, the relationship is described in terms of the mathematical algorithm, $y = mx + b$. If this algorithm were true for all test materials, then any result, x', obtained from the alternative method could be input into the algorithm $y = mx + b$, to obtain an output, y', which would represent the prediction of toxicity in vivo. Such algorithms can be incorporated into Prediction Models that translate the results from an alternative method into a prediction of toxicity in vivo.

Figure 2: Algorithm prediction of *in vivo* toxicity endpoint using a non-linear model



This figure shows a non-linear model (and the prevalidation data used to generate it) that can be used for converting results from an alternative method into a prediction of toxicity in vivo. The solid line is the best fit curve for a three parameter logistic model. The broken lines closest to the regression line indicate the confidence interval for the regression line. The dotted lines furthest from the regression line indicate the 95% confidence intervals for the prediction of an in vivo result.

classification scheme. No matter what approach is used, each algorithm must provide an unambiguous description of how to arrive at a prediction of *in vivo* toxicity, given any possible set of results obtained from one or more alternative methods. The definitions of the algorithms should be clear enough to permit a reasonably well-trained individual to perform this translation.

Finally, the Prediction Model should indicate the level of uncertainty associated with the predictions obtained from the alternative method. This can be estimated with computer simulations based on data that define the accuracy and precision of results from *both* the alternative method undergoing validation and the *in vivo* method to be replaced. The level of uncertainty in predictions from the alternative method can be stated in terms such as the half-width of the 95% confidence interval for the prediction of the *in vivo* result from a given alternative method score. This interval indicates the range of *in vivo* scores that are likely to occur with a given *in vitro* result.

The Prediction Model must also be relevant for its intended purpose. The algorithms used in the Prediction Model must convert alternative method data into toxicity predictions that are sufficiently accurate and precise for toxicologists to make correct safety decisions.

The relevance of the Prediction Model should be assessed before a validation study starts. This assessment requires knowledge of both the alternative method and the *in vivo* test it is intended to replace. Ultimately, a scientific judgement must be made regarding the relevance of a Prediction Model. The factors that must be considered in order to judge the relevance of an alternative method are reviewed elsewhere.[5]

The Advantages of Defining a Prediction Model at the Beginning of a Validation Study

There are two important reasons why a Prediction Model should be defined prior to the start of a validation study. Firstly, if an adequate Prediction Model is defined in advance, it allows those evaluating an alternative method to construct a clear picture of what the results from a valid assay will look like *before* the analysis begins. Objective comparisons can then be made between the predefined picture provided by the Prediction Model and the actual study results. Such an approach has the advantage that it makes validation a confirmatory process and minimises post-study data fitting that does not provide definitive answers on alternative method performance.[4]

Secondly, the Prediction Model is a tool that can be used to guide the design of a validation study. When the models used for making the predictions are stated at the beginning, statisticians can use the information to provide data-based advice on such things as the number of test substances to be included in the reference set of test substances, the number of participating laboratories needed, and the range of toxicity needed in order to adequately assess alternative method performance. Thus, the incorporation of the Prediction Models into the validation process at the beginning has the potential to decrease the cost and time required to validate an alternative method, by facilitating better study design. This is particularly important, given the high costs of large multicentre validation studies.[4]

Ultimately, the testing of Prediction Models along with test methods (as defined by protocols and Standard Operating Procedures) adds solid scientific method to the validation process. In effect, the hypothesis tested in a validation study is whether the stated Prediction Model is true for the reference set of test materials evaluated. A clearly stated hypothesis provides guidance to the assessors of a validation study on how to objectively evaluate the performance of the alternative method. If the data obtained from the study fit the predefined Prediction Model (i.e. the hypothesis is supported), and if the Prediction Model and test method are judged adequate for the defined purpose (i.e. it is relevant), it would provide strong evidence that the new test is valid. Each alternative method evaluated in a validation study must have a clearly defined Prediction Model. In fact, if a Prediction Model is not tested in a validation programme, it is not possible to demonstrate the validity of an alternative method.

Leon Bruner,[a] Gregory Carr,[b] Mark Chamberlain[c] and Rodger Curren[d]

[1] Frazier, J.M. (1990). *OECD Environment Monograph 36, Scientific Criteria for Validation of In Vitro Toxicity Tests*. Paris: OECD.
[2] Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stammati, A-L. & Walum, E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA* **18**, 313–337.
[3] Balls, M., Blaauboer, B., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. & Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM Workshop 5. *ATLA* **23**, 129–147.
[4] Curren, R.D., Southee, J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the deveopment, validation and acceptance of alternative methods. *ATLA* **23**, 211–217.
[5] Bruner, L.H., Carr, G.J., Chamberlain, M. & Curren, R.D. (1996). Validation of alternative methods for toxicity testing. *Toxicology in Vitro*, in press.

[a]The Proctor & Gamble Company, Health & Beauty Care Europe, Egham, Surrey TW20 9NW, UK; [b]The Proctor & Gamble Company, Miami Valley Laboratories, Cincinnati, OH 45239, USA; [c]Unilever Research, Environmental Safety Laboratory, Bedford MK44 1LQ, UK; [d]Microbiological Associates Inc., Rockville, MD 20850, USA.