Validation of Alternative Methods for Toxicity Testing

Leon H. Bruner,[1] Gregory J. Carr,[2] Rodger D. Curren,[3] and Mark Chamberlain[4]

[1]The Procter & Gamble Company, Health and Beauty Care Europe, Staines, Middlesex, United Kingdom; [2]The Procter & Gamble Company, Miami Valley Laboratories, Cincinnati, Ohio; [3]Microbiological Associates, Incorporated, Institute for In Vitro Sciences, Gaithersburg, Maryland; [4]Unilever Research, Environmental Safety Laboratory, Sharnbrook, Bedford, United Kingdom

Before nonanimal toxicity tests may be officially accepted by regulatory agencies, it is generally agreed that the validity of the new methods must be demonstrated in an independent, scientifically sound validation program. Validation has been defined as the demonstration of the reliability and relevance of a test method for a particular purpose. This paper provides a brief review of the development of the theoretical aspects of the validation process and updates current thinking about objectively testing the performance of an alternative method in a validation study. Validation of alternative methods for eye irritation testing is a specific example illustrating important concepts. Although discussion focuses on the validation of alternative methods intended to replace current *in vivo* toxicity tests, the procedures can be used to assess the performance of alternative methods intended for other uses. — *Environ Health Perspect* 106(Suppl 2):477–484 (1998). *http://ehpnet1.niehs.nih.gov/docs/1998/Suppl-2/477-484bruner/abstract.html*

Key words: validation, alternative methods, prediction model, reliability, relevance, eye irritation testing

Introduction

The use of animals for routine toxicity testing is now questioned by a growing segment of society. The expression of this concern is seen with particular clarity in the 6th Amendment to the European Union Cosmetics Directive (*1*). This directive contains a provision that it will become illegal to market cosmetic products in European Union countries if they contain ingredients or mixtures of ingredients that have been tested in animals (to meet the purposes of the directive) unless there are no valid alternatives to replace the animal tests. New test procedures are now being developed to meet ethical concerns and to provide improved toxicologic information. It is critically important to determine whether such alternative methods are valid for use in the safety assessment process.

If alternative methods are to be successfully incorporated into the safety assessment process, it will be necessary to demonstrate that the new procedures can provide at least an equivalent level of protection to that obtained with current methods (*1*). Additionally, if deadlines imposed by legislation such as the 6th Amendment to the European Union Cosmetics Directive are to be met, it is important that the validation process be conducted in a manner that efficiently and definitively characterizes the performance of the alternative methods.

Important concepts in the theory of alternative method validation outside the area of genotoxicity testing have been discussed extensively since the late 1980s. In 1987, Scala (*2*) reviewed the characteristics of a valid test with particular emphasis on calculating the sensitivity, specificity, and predictive value of new test methods. Shortly thereafter, Frazier defined validation as "the process whereby the reliability and relevance of an alternative method is demonstrated for a particular purpose" (*3*). At approximately the same time, the Amden I Workshop further defined important theoretical aspects related to the validation process (*4*). Five years later the Amden II Workshop (*5*) focused on more practical aspects of validation that had been learned during several large multicenter validation studies that were initiated shortly after the Amden I Workshop. The Center for Alternatives to Animal Testing contributed numerous important documents in this time frame that also developed concepts related to the validation process (*6–20*). The Multicentre Evaluation of *In Vitro* Cytotoxicity program also contributed significantly to the development of theoretical aspects of the validation process (*21,22*).

Validation of alternative methods has also been of considerable interest to regulatory authorities. Consequently, several international organizations, regulatory agencies, and committees have reviewed various aspects of validation and regulatory acceptance of alternative methods. The U.S. Interagency Regulatory Alternatives Group, which comprises scientists from the U.S. Food and Drug Administration, the U.S. Environmental Protection Agency, and the U.S. Consumer Products Safety Commission, has examined the validation and regulatory acceptance process (*23*). This organization gave way to the U.S. National Institute of Environmental Health Sciences Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM), which has completed an extensive review of the

This paper was prepared as background for the 13th Meeting of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Alternative Testing Methodologies held 26–31 January 1997 in Ispra, Italy. Manuscript received at *EHP* 9 May 1998; accepted 9 September 1998.

The authors acknowledge Prof. M. Balls and FRAME, Nottingham, United Kingdom, which supported a meeting between R.D. Bruce, D.P. Lovell, and T. Wiernsberger in August 1992 in which some of the concepts presented in this manuscript were initially discussed. Additionally, it is important to acknowledge the following persons for their useful discussions related to the development of the concepts presented: members of the European Commission/Home Office Validation Study Management Team including M. Balls, P. Botham, and H. Spielmann; members of the COLIPA Eye Irritation Alternatives Validation Study Management Team including P. Brantom, O. DeSilva, L. Earl, D. Lovell, W. Pape, and M. Uttley.

Address correspondence to Dr. L.H. Bruner, The Procter & Gamble Company, Health and Beauty Care Europe, Staines, Middlesex TW18 3AZ, United Kingdom. Telephone: 44 1784 495 059. Fax: 44 1784 495 043. E-mail: brunerlh@pg.com

Abbreviations used: CI, confidence interval; 95%CI_{pred}, 95% confidence interval for the prediction of an *in vivo* test result; CV, coefficient of variation; ECVAM, European Centre for the Validation of Alternative Methods; ICCVAM, Interagency Coordinating Committee for the Validation of Alternative Methods; MAS, maximum average score; OECD, Organisation for Economic Co-operation and Development; SOP, standard operating procedures.

validation and regulatory acceptance of alternative methods (*24*). The European Commission formed the European Centre for the Validation of Alternative Methods (ECVAM), which plays a leading role in facilitating the optimization (prevalidation) and validation of alternative methods. ECVAM has contributed extensively to the field of validation in the form of publications on prevalidation (*25*), and validation (*26–36*), ECVAM workshop reports on the status of method development for a wide range of toxicity end points (*37–61*), and ECVAM task force reports (*25,62*). ECVAM has also coordinated validation studies on eye irritation testing (*63*), photo irritation (*64*), and skin corrosion (*42*). Finally, the Organisation for Economic Co-operation and Development (OECD) test guidelines program has produced an important report on the harmonization and validation of alternative toxicological test methods (*65*).

This paper summarizes some important aspects of the validation process that have been developed through these efforts. This discussion focuses on the replacement of animal tests that are currently used to determine the hazard of chemicals for regulatory purposes. It provides guidance on the design, execution, and evaluation of validation programs. It also describes how to objectively assess the performance of the alternative methods relative to the *in vivo* test to be replaced and discusses the factors that must be considered when the relevance of an alternative method is assessed. The concepts presented in this review are consistent with and expand on those developed by Frazier (*3*), the Amden Workshops (*4,5*), the ICCVAM (*24*), and OECD (*65*) reviews on validation, and other recent publications on the validation process (*12,13,66*).

Definition of Validation

Validation has been defined as "the establishment of the reliability and relevance of an alternative method for a specific purpose" (*3,4*). To assess the validity of an alternative method, it is important to clearly define the terms reliability and relevance. For a toxicologist to rely on an alternative method, two things must be known about its performance. First, it must be possible to consistently reproduce the results from an alternative method. Second, it must be possible to consistently and correctly convert the results from the alternative method into useful predictions of toxicity so that appropriate safety assessments can be made. Thus, reliability may be defined as the establishment of the reproducibility of the data obtained from a method across different laboratories and the reproducibility of the predictions of toxic hazard after application of a clearly stated prediction model to the alternative method data across appropriately defined sets of test substances (*66*).

Once the reproducibility of an alternative method has been confirmed, then its relevance must be evaluated. Relevance has been defined as establishing the scientific meaningfulness and usefulness of results from an alternative method for a particular purpose (*3,4*). Establishing usefulness and meaningfulness is important because hazard predictions obtained from scientifically credible alternative methods have a higher probability of being correct. To establish relevance, all available information related to the fundamental scientific basis, reliability (as defined above), and practical operation of the alternative method, and to the *in vivo* toxicity test to be replaced must be thoroughly reviewed. Ultimately, a judgment must be made about whether or not a method is relevant for a particular purpose.

Prevalidation of Alternative Methods

A method must be sufficiently developed before it is considered ready for evaluation in a validation study (Figure 1) (*25,66*). First, a test must have been conceived and then developed sufficiently that it can be conducted routinely in an appropriately equipped laboratory by experienced technicians. Second, an adequate prediction model must be available that allows correct interpretation of its results (*66,67*). Third, there should be evidence that an alternative method is relevant for the intended purpose. Fourth, there should be evidence that the method can be reproduced across several laboratories. Finally, adequate protocols and standard operating procedures (SOPs) must be available so that any participating laboratory can conduct the assay. Once it has been confirmed that factors are adequate, a method may then be assessed in a validation study.

Importance of the Prediction Model

For an alternative method to be useful for making safety assessments, it must be possible to translate the results into correct predictions of *in vivo* toxicity. This is usually done by applying algorithm(s) to the alternative method data that convert them into toxicity predictions. Because such algorithms constitute models that allow the prediction of toxicity, they have been called prediction models (*66*). If an alternative method does not have an adequate prediction model, it cannot be used in the safety assessment process. It is therefore essential that validation programs test the utility of the prediction model associated with each alternative method evaluated. In fact, if the prediction model is not defined prior to the start of a study, its validity cannot be assessed (*67*).

An adequate prediction model must have at least four components (*66*). First, there must be a clear definition of every type of data available from the alternative method. Second, the prediction model must provide an algorithm that allows an individual to convert each data type into a prediction of the *in vivo* end point of interest. Third, the prediction model should provide an indication of the accuracy and the precision of the predictions. For example, the 95% confidence intervals (CI) for a given prediction (95% CI_{pred}) may be provided. Finally, the prediction model must define the test substances for which the method is valid, limitations on the use of the method, and the specific purpose for which the test is to be used.

Practical experience has shown the benefits in having a prediction model clearly defined before a validation study starts. First, it provides a clear picture of the results that should be expected at the end of a validation study if the method is valid. This allows the reviewers of a validation study to objectively assess the performance of the method at the end of the study. Second, if the prediction model is defined at the start, it is possible to work with statisticians to design the validation study in an efficient and proper way. Data-based methods can be used to determine the appropriate number of test substances and laboratories to include in a study in order to adequately assess a method's performance.

Assessing the Reliability of an Alternative Method in a Validation Study

The first step in assessing the validity of an alternative method is to conduct a study designed to measure reliability. To conduct a validation study, there are several important steps that must be completed (Figure 1). The study must be designed, the participating laboratories must be identified and recruited, a reference set of test substances must be assembled and distributed under code, the quality of the *in vivo* data must be assessed, and each test substance must be

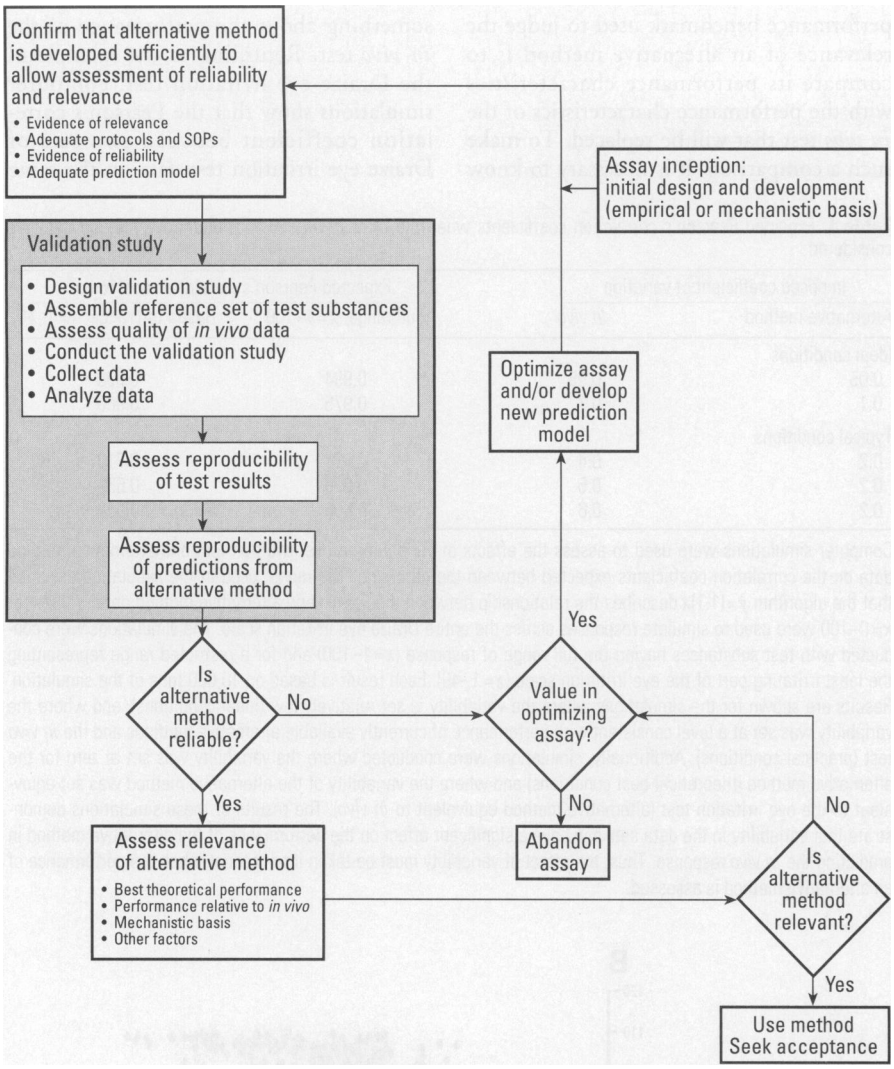

Figure 1. The validation of an alternative method. The flow chart depicts one of several possible approaches that may be used as guides to design and conduct a validation program. The steps on the left side of the chart represent the validation process. The pathway within the shaded box represents the validation study process. The right side of the chart depicts the steps associated with improving the performance of the alternative method and defining another prediction model prior to inclusion of the method in a subsequent validation study. Any new method, whether based on a fundamental understanding of toxic mechanisms or on empirical correlations, may be assessed for validity using this approach. From Bruner et al. (*66*).

evaluated in the alternative method (*66*). Ideally, all data supporting the validity of the test method should be obtained and reported in accordance with Good Laboratory Practices (*24,65*). Then, when the alternative method data are available, the prediction model defined before the start of the study must be used to predict the *in vivo* toxicity of each test substance. If the toxicity predictions are similar to the actual toxicity of the test substances, and if the same results were obtained across all of the participating laboratories, it would provide evidence that the method is reliable. If, however, the toxicity is not predicted correctly, or if the results are not similar across the participating laboratories, it would not be possible to consider the method is reliable. If the alternative method is found not reliable, it may be optimized, a new prediction model developed, and then the new method tested in a subsequent validation study. Alternatively, the method may be abandoned if additional work is unlikely to be fruitful (Figure 1).

Assessing the Relevance of an Alternative Method

As noted above, an alternative method may be considered relevant when it is shown that the predictions of toxicity obtained are meaningful and useful for a specific purpose. Establishing relevance is a judgmental process requiring evaluation of all available supporting data and scientific evidence supporting the use of an alternative method. This involves evaluation of key performance benchmarks that provide a useful context for interpreting the results obtained from a validation study, the mechanistic basis for the test, and other factors related to the performance of the method (*66*).

Establishment of Key Performance Benchmarks

Theoretical Best Performance. One criterion used for assessing relevance is to estimate the theoretical best performance expected from the alternative method. Ideally, there should be a high correlation coefficient in the relationship between the *in vivo* and alternative method data, and a narrow 95% CI_{pred}. However, there are certain technical limitations associated with alternative methods that decrease the likelihood that such performance will be observed. If this is true, the question that must be asked is, What level of performance is possible and reasonable? Computer simulations may be used to provide guidance on answering this question (*66*). The results from a hypothetical eye irritation alternative method validation study provide an example of how this may be accomplished. Let us assume that the relationship between the maximum average score (MAS) from the Draize eye irritation test, y, and an alternative method result, x, is defined by the following equation:

$$y = 1.1(x),$$

where the alternative method scores, x, range between 0 and 100. In this case the predicted MAS will range between 0 and 110, which is consistent with the Draize eye irritation test scoring scheme (*68*). The simulation may be run many times (10,000 in this example). Each run of the simulation produces a corresponding value of y, which is a predicted MAS. Simulations may also be conducted with scores for x restricted to a range between 0 and 40, which is about half of the eye irritation scale. This can be done to simulate expected results from studies that use a reference set of test substances in the least irritating half of the Draize eye irritation scale as would occur with more mild test substances.

If a low level of variability is assumed in both *in vivo* and alternative method data,

the Pearson's correlation coefficient is high for both the full range of eye irritancy and for the least irritating part of the eye irritation scale (Figure 2, Table 1). However, *in vivo* and alternative method data can show considerable variability. Experience shows that the coefficient of variation (CV) is approximately 10 to 30% for typical alternative methods (*66*). The CV for the Draize eye irritation test ranges between 40 and 60%. Accordingly, computer simulations conducted with the CV for the alternative method and the *in vivo* data set at 20 and 40%, respectively, show that the expected correlation coefficients will be approximately 0.86 for the full set of test materials, and approximately 0.7 for the least irritating portion or the Draize eye irritation scale (Figure 2, Table 1). The results of these simulations also indicate that the 95% CI_{pred} for a predicted MAS of 55 is relatively wide at ±35.2 (Table 2). Thus, if an alternative method using the algorithm, $y = 1.1x$, produces a correlation between the alternative method data and *in vivo* data of approximately 0.7 to 0.8 with a 95% CI_{pred} in the range of ±35, it would provide evidence supporting its relevance as a replacement for the *in vivo* test.

Comparison with the Performance of the* in Vivo *Test to Be Replaced. A second performance benchmark used to judge the relevance of an alternative method is to compare its performance characteristics with the performance characteristics of the *in vivo* test that will be replaced. To make such a comparison, it is necessary to know something about the performance of the *in vivo* test. Returning to the example of the Draize eye irritation test, computer simulations show that the Pearson's correlation coefficient between two sets of Draize eye irritation test data on the same

Table 1. Expected Pearson's correlation coefficients when the error in *in vivo* and alternative method data are considered.

Imposed coefficient of variation		Expected Pearson's correlation coefficient	
Alternative method	*In vivo*	Full range, x=1–100	Restricted range, x=1–40
Ideal conditions			
0.05	0.05	0.994	0.990
0.1	0.1	0.975	0.960
Typical conditions			
0.2	0.4	0.860	0.719
0.2	0.5	0.828	0.652
0.2	0.6	0.803	0.608

Computer simulations were used to assess the effects of variability in eye irritation test and alternative method data on the correlation coefficients expected between the data sets. The model used in the simulation assumed that the algorithm $y=(1.1)x$ describes the relationship between the *in vivo* and alternative method data. Values for x=0–100 were used to simulate responses across the entire Draize eye irritation scale. The simulations were conducted with test substances having the full range of response (x=1–100) and for a restricted range representing the least irritating part of the eye irritation scale (x=1–40). Each result is based on 10,000 runs of the simulation. Results are shown for the simulations where the variability is set relatively low (ideal conditions), and where the variability was set at a level consistent with performance of currently available alternative methods and the *in vivo* test (practical conditions). Additionally, simulations were conducted where the variability was set at zero for the alternative method (theoretical best conditions) and where the variability of the alternative method was set equivalent to the eye irritation test (alternative method equivalent to *in vivo*). The results of these simulations demonstrate that variability in the data sets can have a significant effect on the performance of the alternative method in predicting the *in vivo* response. Thus, the effect of variability must be taken into account when the performance of an alternative method is assessed.

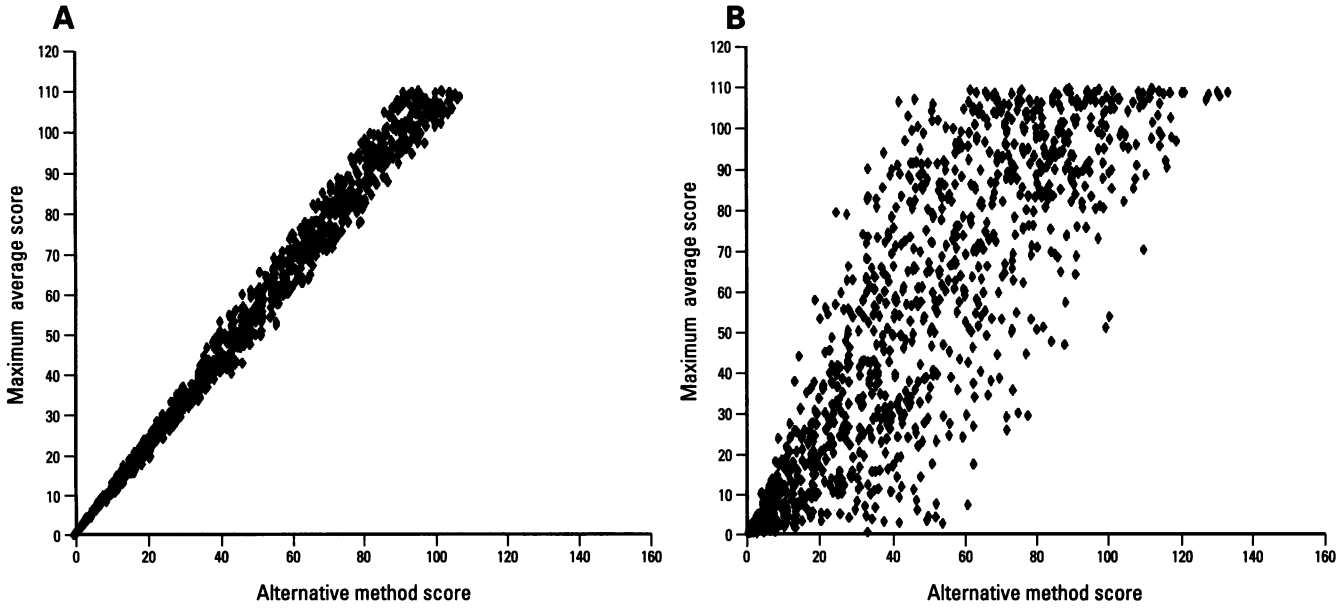

Figure 2. Effects of variability in the Draize test and alternative methods on correlation. Computer simulations were used to assess the effects of variability in the eye irritation test; alternative method data on the relationship between the two data sets is illustrated. The model used in the simulation assumed that the algorith, $y=(1.1)x$ describes the relationship between the *in vivo* (MAS) and alternative method data. Values for x=0–100 were used to simulate responses across the entire Draize eye irritation scale. Different levels of variability were added to the alternative method and *in vivo* scores in each run of the simulation. The x and y values generated in 1000 runs of the simulation are plotted on the figures. (*A*): The expected relationship between the MAS and the alternative method results when the variability is relatively low. In this case, the CVs applied to both the *in vivo* and alternative method data were 5%. (*B*): The expected relationship between the maximum average scores and alternative method results under typical conditions. The CVs applied to the *in vivo* and alternative method data were 50 and 20%, respectively.

test substances will be approximately 0.87 if the *in vivo* CV = 40% (Figure 3). The 95% CI_{pred} in this case will be approximately ±35 for a predicted MAS of 55 (Table 2).

Table 2. 95% Confidence interval for predicting an *in vivo* eye irritation score (95% CI_{pred}) when the predicted maximum average score is 55.

Imposed coefficient of variation		
Alternative method	*In vivo*	95% CI_{pred}
Alternative method predicting *In vivo*		
0.2	0.4	±35.2
0.2	0.5	±40.2
0.2	0.6	±45.6
Draize predicting Draize		
—	0.4	±34.8
—	0.5	±43.2
—	0.6	±50.6

Computer simulations were used to assess the effects of variability in eye irritation test and alternative method data on the 95% CI_{pred}. For predictions of *in vivo* scores from an alternative method result, the model used in the simulation assumed that the algorithm, $y=(1.1)x$, describes the relationship between the *in vivo* and alternative method data. Values for $x=0$–100 were used to simulate responses across the entire Draize eye irritation scale. For predictions of *in vivo* scores from the *in vivo* result, the model used in the simulation assumed that the algorithm, $y=x$, describes the relationship between the two sets of data. Each result is based on 10,000 runs of the simulation. The coefficient of variation (CV) for the alternative method was set at 20%. The CV used for the *in vivo* data ranged from 40–60% which is consistent with reports in the scientific literature (*69*).

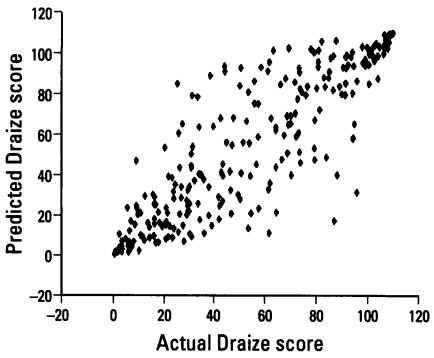

Figure 3. Effects of variability on the capacity of the Draize test to predict its own result. Computer simulations were used to assess the effects of variability in the capacity of the Draize eye irritation test to predict its own result. The model used in the simulation assumed that the algorithm $y=x$ describes the relationship between the *in vivo* data sets (MAS). Values for $x=0$–110 were used to simulate responses across the entire Draize eye irritation scale. The x and y values generated in 1000 runs of the simulation are plotted on the figure. The expected relationship between an actual MAS and a predicted MAS when the CV applied is 40% as illustrated.

If the predictive capacity of the alternative method is similar to or better than these values, it would support the relevance of the alternative method.

Other Factors Supporting the Relevance of an Alternative Method. In addition to these performance benchmarks, it is important to consider other factors supporting the relevance of an alternative method (*66*). First, the mechanistic basis of the new assay should be understood (*9*). A stronger mechanistic understanding increases confidence that the predictions from the alternative method will be correct. Second, it is important to define the known limitations in the use of an alternative method. For example, a new procedure may be valid for only a small number of substances relative to the universe of materials that must be tested. If the method is limited in its application, it may not be very relevant for general use in the safety assessment process. Third, the technical limitations of an alternative method must be known. An assay that can handle all types of test substances may ultimately be more relevant for general use than one restricted to only one type (e.g., water-soluble test materials). Finally, performance of the alternative method reported in the scientific literature should also be considered.

Once this information has been assembled and evaluated, the overall relevance of the method for its defined purpose must be assessed. If the conclusion is that the alternative method is not relevant, the test cannot be considered valid, and it is necessary to consider whether there is value in optimizing the assay, developing a new prediction model, and assessing it in a subsequent validation study (Figure 1). Conversely, if the data support its relevance, that would suggest the alternative method may be used in the safety assessment process and should be considered for official acceptance by regulatory authorities (Figure 1). To gain regulatory acceptance, regulatory authorities and independent reviewers should receive all data supporting the conclusions obatined from the program so that the results and conclusions can be given a complete peer review (*24*). Publication of results in a high-quality, peer-reviewed journal provides additional credibility to the conclusions obtained from a study (*24*).

Conclusion

The reliability and relevance of an alternative method for a specific purpose are established during the validation process. The validation study, a part of the overall validation process, should be considered a confirmation step that provides quantitative evidence that an alternative method is reliable. To efficiently assess the reliability of an alternative method, a prediction model must be defined before the commencement of the study. The assessment of relevance requires a thorough review of all the performance data and other supporting information related to both the alternative method and the *in vivo* test it will replace. Ultimately, those participating in the validation process must integrate this information and render a judgment on whether the method, when used for a specifically defined purpose, is useful and meaningful.

The importance of the prediction model has been stressed because the primary purpose of an alternative method is to provide predictions of toxicity that will be used by toxicologists to make decisions during the safety assessment process. Because the prediction of toxicity is the critical piece of information needed from an alternative method, it is important that the procedures used to arrive at these predictions be validated during the validation process. Previous discussions of validation have indirectly addressed the need for the prediction model, but have focused on the identification of such models after the validation study is completed (*4,5*) However, if an adequate prediction model is defined at the beginning of a validation study, 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 study begins. When the results from the validation study become available, objective comparisons can be made between the predefined picture provided in the prediction model and the actual study results. Such an approach has an advantage in that it makes validation a confirmatory process and minimizes post hoc data fitting that does not provide definitive answers on alternative method performance. The value of defining the prediction model prior to the start of a validation study has been demonstrated in a recently completed eye irritation test method validation program (*70*).

In addition to facilitating objective assessment of the predictive capacity of an alternative method, the prediction model is also an important 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 of a validation study, statisticians can use the information to provide data-based advice on

such things as the numbers 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 to adequately assess alternative method performance. Thus, the incorporation of the prediction models into the validation process at the beginning not only improves a reviewer's ability to assess the validity of an alternative method, but also 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, multicenter validation studies.

The computer simulations on the Draize eye irritation test provide a striking view of the results that can be expected from a validation study if the level of uncertainty in the data from the reference test to be replaced is high. In such cases, it will not be possible to demonstrate that alternative methods provide predictions that have high levels of certainty. As noted earlier, one of the most important factors to consider in the design of a validation study is to assure that the quality of data used for comparisons against the alternative method results are as high as possible. It has become apparent that obtaining test substances with high quality *in vivo* data is a difficult problem that must be overcome if rapid progress in the development and validation of alternatives is to be made. The simulations also demonstrate why it is important to establish objective criteria to be used as the basis for judging alternative method performance. The establishment of data-based performance benchmarks will better guide reviewers of a validation study in setting realistic performance expectations given the real-world technical limitations characteristic of the current state of the art (*69,71,72*).

REFERENCES AND NOTES

1. EEC. Council Directive 93/35/EEC of 14 June, 1993 amending for the 6th time Directive76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. Off J Eur Comm L151:32–36 (1994).
2. Scala RA. Theoretical approaches to validation. In: Approaches to Validation. Alternative Methods in Toxicology. Vol 5 (Goldberg AM, ed). New York:Mary Ann Liebert, 1987;1–10.
3. Frazier JM. OECD Environment Monograph No. 36: Scientific Criteria for Validation of *In Vitro* Toxicology Tests, 1990. Paris:Organisation for Economic Co-operation and Development, 1990.
4. Balls M, Blaauboer B, Brusick D, Frazier J, Lamb D, Pemberton M, Reinhardt C, Roberfroid M, Rosenkranz H, Schmid B, et al. Report and recommendations of the CAAT/ERGATT Workshop on the Validation of Toxicity Test Procedures. ATLA 18:313–337 (1990).
5. Balls M, Blaauboer B, Fentem JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, et al. Practical aspects of the Validation of Toxicity Test Procedures. The report and recommendations of ECVAM Workshop 5. ATLA 23:129–147 (1995).
6. Frazier JM. Validation of *in vitro* models. J Am Coll Toxicol 9: 355–359 (1990).
7. Frazier JM. Validation of *in vitro* toxicity tests. In: *In Vitro* Toxicity Testing: Applications to Safety Evaluations (Frazier JM, ed). New York:Marcel Dekker, 1992;245–252.
8. Frazier JM. *In vitro* models for toxicological research and testing. Toxicol Lett 68:73–90 (1993).
9. Frazier JM. The role of mechanistic toxicology in test methods validation. Toxicol In Vitro 8:787–791 (1994).
10. Frazier JM. Interdisciplinary approach to toxicity test development and validation. Toxicol in Vitro 9:8925–8949 (1995)
11. Goldberg AM, ed. Alternative Methods in Toxicology. Vols 1–11. New York:Mary Ann Liebert, 1983–1995.
12. Goldberg AM, Frazier JM, Brusick D, Dickens MS, Flint O, Gettings SD, Hill RN, Lipnick RL, Renskers KJ, Bradlaw JA, et al. Framework for validation and implementation of *in vitro* toxicity tests. In Vitro Cell Dev Biol 29A:688–692 (1993).
13. Goldberg AM, Epstein LD, Zurlo J. A modular approach to validation: a work in progress. In Vitro Toxicol 8:431–435 (1995).
14. Goldberg AM, Frazier JM. Alternatives to animals in toxicity testing. Sci Am 261:24–30 (1989).
15. Frazier, JM, Bradlaw J. Technical Report No 1: Technical Problems Associated with *In Vitro* Toxicity Testing Systems. Baltimore:The Johns Hopkins Center for Alternatives to Animal Testing, 1989.
16. Sehnert SS. Technical Report No 2: Structure–Activity Relationships in Predictive Toxicology. Baltimore:The Johns Hopkins Center for Alternatives to Animal Testing, 1990.
17. CAAT. Technical Report No 4: Cell Culture Systems and *In Vitro* Toxicity Testing: Baltimore:The Johns Hopkins Center for Alternatives to Animal Testing, 1992.
18. Frazier JF. Technical Report No 5: The International Status of Validation of *In Vitro* Toxicity Tests. Baltimore The Johns Hopkins Center for Alternatives to Animal Testing, 1993.
19. CAAT. Technical Report No 6: Final Scientific Reports of CAAT Grantees. Baltimore:The Johns Hopkins Center for Alternatives to Animal Testing, 1994.
20. CAAT. Technical Report No 7: Molecular and Cellular Approaches to Extrapolation for Risk Assessment. Baltimore: The Johns Hopkins Center for Alternatives to Animal Testing, 1994.
21. Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, Clothier R, Cottin M, Curren R, Daniel-Szolgay E, et al. MEIC evaluation of acute systemic toxicity. Part 1: Methodology of 68 *in vitro* toxicity assays used to test the first 30 reference chemicals. ATLA 24:251–272 (1996).
22. Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, Clothier R, Cottin M, Curren R, Dierickx P, et al. MEIC evaluation of acute systemic toxicity. Part II: *In vitro* results from 68 toxicity assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. ATLA 24:273–311 (1996).
23. Bradlaw JA, Wilcox NL. Workshop on eye irritation testing: Practical applications of non-whole animal alternatives. Fd Chem Toxicol 35:1–11 (1997).
24. NIEHS. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publ No 97-3981. Research Triangle Park, NC:National Institute of Environmental Health Sciences, 1997.
25. Curren RD, Southee JA, Spielmann H, Liebsch M, Fentem JH, Balls M. The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23: 211–217 (1995).
26. Balls, M. Replacement of animal procedures: alternatives in research, education and testing. Lab Anim 28:193–211 (1994).
27. Marafante E, Smyrniotis T, Balls M. ECVAM: The European Centre for the Validation of Alternative Methods. Toxicol in Vitro 8:803–805 (1994).
28. Balls M. *In vitro* methods in regulatory toxicology: the crucial significance of validation. In: Toxicology in Transition (Degen GH, Seiler JP, Bentley P, eds). Archives of Toxicology, Suppl 17. Berlin/Heidelberg:Springer Verlag, 1995;155–162.
29. Balls M. Scientific validation: a crucial and unavoidable prerequisite to the acceptance of new tests and testing strategies. ATLA 23:474–479 (1995).

30. Balls M. Defining the role of ECVAM in the development, validation and acceptance of alternative tests and testing strategies. Toxicol in Vitro 9:863–869 (1995).
31. Balls M, Blaauboer BJ, eds. The validation of replacement alternative methods. In: Proceedings of the ECVAM Opening Symposium, 18 October 1994, Ispra, Italy. Toxicol in Vitro 9:789–869 (1995).
32. Balls M, Karcher W. The validation of alternative test methods. ATLA 23:884–886(1995).
33. Fentem JH, Balls M. Alternative methods in toxicological research and testing. Comments Toxicol 5:199–202 (1995).
34. Fentem JH, Prinsen MK, Spielmann H, Walum E, Botham PA. Validation—lessons learned from practical experience. Toxicol in Vitro 9:857–862 (1995).
35. Balls M, Fentem JH. Progress toward the validation o alternative tests. ATLA 25:33–43 (1997).
36. Fentem JH. ECVAM workshops: summary of conclusions and recommendations. Toxicol Ecotoxicol News 3:69–78 (1996).
37. Blaauboer BJ, Boobis AR, Castell JV, Coecke S, Groothuis GMM, Guillouzo A, Hall TJ, Hawksworth GM, Lorenzon G, Miltenburger HG, et al. The practical applicability of hepatocyte cultures in routine testing. The report and recommendations of ECVAM Workshop 1. ATLA 22:231–241 (1994).
38. Spielman, H, Lovell WW, Hölzle E, Johnson BE, Maurer T, Miranda MA, Pape WJW, Sapora O, Sladowski D. *In vitro* phototoxicity testing. The report and recommendations of ECVAM Workshop 2. ATLA 22:314–348 (1994).
39. Atterwill CK, Bruinink A, Drejer J, Duarte E, McFarlane Abdulla E, Meredith C, Nicotera P, Regan C, Rodriguez-Farre E, et al. *In vitro* neurotoxicity testing. The report and recommendations of ECVAM Workshop 3. ATLA 22:350–362 (1994).
40. Hendriksen CFM, Garthoff B, Aggerbeck H, Bruckner L, Castle P, Cussler K, Dobbelaer R, van de Donk H, van der Gun J, Lefrancois S, et al. Alternatives to animal testing in the quality control of immunobiologicals: current status and future prospects. The report and recommendations of ECVAM Workshop 4. ATLA 22:420–434 (1994).
41. Balls M, Blaauboer BJ, Fentem, JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, et al. Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM Workshop 5. ATLA 23:129–147 (1995).
42. Botham PA, Chamberlain M, Barratt MD, Curren RD, Esdaile DJ, Gardner JR, Gordon VC, Hildebrand B, Lewis,RW, Liebsch M, et al. A prevalidation study on *in vitro* skin corrosivity testing. The report and recommendations of ECVAM Workshop 6. ATLA 23:219–255 (1995)
43. Balls M, De Klerck W, Baker F, van Beek M, Bouillon C, Bruner L, Carstensen J, Chamberlain M, Cottin M, Curren R., et al. Development and validation of non-animal tests and testing strategies: the identification of a coordinated response to the challenge and the opportunity presented by the 6th Amendment to the Cosmetics Directive (76/768/EEC). The report and recommendations of an ECVAM/CPS Workshop (ECVAM Workshop 7. ATLA 23:398–409 (1995).
44. Barratt MD, Castell JV, Chamberlain M, Combes RD, Dearden JC, Fenten JH, Gerner I, Giuliani, A, Gray TJB, Livingstone DK, et al. The integrated use of alternative approaches for predicting toxic hazard. The report and recommendations of ECVAM Workshop 8. ATLA 23:410–429 (1995).
45. Garthoff B, Hendriksen C, Bayol A, Goncalves D, Grauer A, de Leeuw R, van Noordwijk J, Pares M, Pirovano R, Rieth M, et al. Safety and efficacy testing of hormones and related products. The report and recommendations of ECVAM Workshop 9. ATLA 23:699–712 (1995).
46. Hawksworth GM, Bach PH, Dekant W, Diezi JE, Harpur E, Lock EA, MacDonald C, Morin J-P, Nagelkerke JF, Pfaller, W, et al. Nephrotoxicity testing *in vitro*. The report and recommendations of ECVAM Workshop 10. ATLA 23: 713–727 (1995).
47. Balls M, Goldberg AM, Fentem JH, Broadhead CL, Burch RL, Festing, MFW, Frazier JM, Hendriksen CFM, Jennings M, van der Kamp MDO, et al. The three Rs: the way forward. The report and recommendations of ECVAM Workshop 11. ATLA 23:838–866 (1995).
48. Brown NA, Spielmann H, Bechter R, Flint, OP, Freeman SJ, Jelínek RJ, Koch E, Nau H, Newall DR, Palmer AK, et al. Screening chemicals for reproductive toxicity: the current alternatives. The report and recommendations of an ECVAM/ETS workshop (ECVAM Workshop 12). ATLA 23:868–882 (1995).
49. Howes D, Guy R, Hadgraft J, Heylings J, Hoeck U, Kemper, F, Mailbach H, Marty J-P, Merk H, Parra J, et al. Methods for assessing percutaneous absorption. The report and recommendations of ECVAM Workshop 13. ATLA 24:81–106 (1996).
50. Gribaldo L, Bueren J, Deldar A, Hokland P, Meredith C, Moneta, D, Mosesso P, Parchment R., Parent-Massin D, Pessina A, et al. The use of *in vitro* systems for evaluating haematotoxicity. The report and recommendations of ECVAM Workshop 14. ATLA 24:211–231 (1996).
51. Blaauboer BJ, Bayliss MK, Castell JV, Evelo CTA. Frazier JM, Groen K, Gülden M, Guillouzo A, Hissink AM, Houston JB, et al. The use of biokinetics and *in vitro* methods in toxicological risk evaluation. The report and recommendations of ECVAM Workshop 15. ATLA 24:473–497 (1996).
52. Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, Ekwall B, Garle MJ, GA, Hissink AM, Houston J, et al. The *in vitro* and the classification and labelling of chemicals. The report and recommendations of ECVAM Workshop 16. ATLA 24:499–510 (1996).
53. Svendsen O, Garthoff B, Spielmann H, Hensten-Pettersen A, Jensen JC, Kuijpers MR, Leimgruber R, Liebsch M, Müller-Lierheim WGK, Rydhög G, et al. Alternatives to the animal testing of medical devices. The report and recommendations of ECVAM Workshop 17. ATLA 24:659–669 (1996).
54. Lambré CR, Aufderheid, M, Bolton RE, Fubini B, Haagsman HP, Hext PM, Jorissen M, Landry Y, Morin J-P, Nemery B, et al. *In vitro* tests for respiratory toxicity. The report and recommendations of ECVAM Workshop 18. ATLA 24:671–681 (1996).
55. de Silva O, Basketter DA, Barratt MD, Corsini E, Cronin MTD, Das PK, Degwert J, Enk A, Garrigue JL, Hauser C, et al. Alternative methods for skin sensitisation testing. The report and recommendations of ECVAM Workshop 19. ATLA 24:683–705 (1996).
56. Bach PH, Vickers AEM, Fisher R, Baumann A, Brittebo E, Carlile DJ, Koster HJ, Lake BG, Salmon F, Sawyer TW, et al. The use of tissue slices for pharmacotoxicology studies. The report and recommendations of ECVAM Workshop 20. ATLA 24:893–923 (1996).
57. Schade R, Staak C, Hendriksen C, Erhard M, Hugl H, Koch G, Larsson A, Pollmann W, van Regenmortel M, Rijke E, et al. The production of avian (egg yolk) antibodies: IgY. The report and recommendations of ECVAM Workshop 21. ATLA 24:925–934 (1996).
58. Leahy DE, Duncan R, Ahr HJ, Bayliss MK, de Boer AG, Darvas F, Fentem JH, Fry JR, Hopkins R, Houston JB, Pharmacokinetics in early drug research. The report and recommendations of ECVAM Workshop 22. ATLA 25:17–31 (1997).
59. Marx U, Embleton MJ, Fischer R., Gruber FP, Hansson U, Heuer J, de Leeuw WA, Logetenberg T, Merx W, Portetelle D, et al. Monoclonal antibody production. The report and recommendations of ECVAM Workshop 23. ATLA 25:121–137 (1997) .
60. Dearden JC, Barratt MD, Benigni, R, Bristol DW, Combes RD Cronin, MTD, Judson PN, Payne MP, Richard AM, Tichy M, et al. The development and validation of expert systems for predicting toxicity. The report and recommendations of an ECVAM/ECB workshop (ECVAM Workshop 24). ATLA 25:223–252 (1997).
61. Janusch A, van der Kamp MDO, Bottrill K, Grune B,

Anderson DC, Ekwall B, Howald M, Kolar R, Kuiper HJD, Larson J, et al. Current status and future developments of databases on alternative methods. The report and recommendations of ECVAM Workshop 25. ATLA 25:411–422 (1997).
62. Holzhütter H-G, Archer G, Dam, N, Lovell DP, Saltelli A, Sjöström M. . Recommendations for the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA 24:511–530 (1996).
63. Balls M, Botham P, Bruner LH, Spielmann H. The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test. Toxicology in Vitro 9:871–929 (1995).
64. Spielmann H, Balls M, Brand M, Döring B, Holzhütter HG, Kalweit S, Klecak G, L'Eplattenier H, Liebsch M, Lovell WW, et al. EEC/COLIPA project on *in vitro* phototoxicity testing: first results obtained with a Balb/c 3T3 cell phototoxicity assay. Toxicol in Vitro 8:793–796 (1995).
65. OECD. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. ENV/MC/CHEM/TG(96)9. Paris:Organisation for Economic Co-operation and Development, 1996.
66. Bruner LH, Carr GJ, Chamberlain M, Curren RD. Validation of alternative methods for toxicity testing. Toxicol in Vitro 10:479–501 (1996).
67. Bruner L, Carr GJ, Chamberlain M, Curren RD. No prediction model, no validation study ATLA 24:139–142 (1996).
68. Draize JH, Woodard G, Calvery HO. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J Pharmacol Exp Ther 82:377–390 (1944).
69. Weil CA, Scala A. Study of intralaboratory and inter-laboratory variability in the results of rabbit eye and skin irritation tests. Toxicol Appl Pharmacol 19:276–360(1971).
70. Brantom PG, Bruner LH, Chamberlain M, de Silva O, Dupuis J, Earl LK, Lovell DP, Pape WJW, Uttley M, et al. A summary report of the COLIPA international validation study on the alternatives to the Draize rabbit eye irritation test. Toxicol in Vitro 11:141–179 (1997).
71. Hunter WJ, Lingk W, Recht RJ. Intercomparson study on the determination of single administration toxicity in rats. J Assoc Off Analyt Chem 62:864–873 (1979).
72. van den Heuvel MJ, Clark DG, Fielder RJ, Koundakjian PP, Oliver GJA, Pelling D, Tomlinson N, Walker AP. The international validation of a fixed-dose procedure an alternative to the classical LD_{50} test. Food Chem Toxicol 28:469–482 (1990).