Practical Aspects of the Validation of Toxicity Test Procedures

The Report and Recommendations of ECVAM Workshop 5[1,2]

Michael Balls[3], Bas J. Blaauboer[4], Julia H. Fentem[3], Leon Bruner[5], Robert D. Combes[6], Bjorn Ekwall[7], Robin J. Fielder[8], Andre Guillouzo[9], Richard W. Lewis[10], David P. Lovell[11], Christoph A. Reinhardt[12], Guillermo Repetto[13], Dariusz Sladowski[14], Horst Spielmann[15] and Flavia Zucco[16]

[3]ECVAM, JRC Environment Institute, 21020 Ispra (Va), Italy; [4]RITOX, Utrecht University, P.O. Box 80.176, 3508 TD Utrecht, The Netherlands; [5]Procter & Gamble Health & Beauty Care Europe, Rusham Park, Whitehall Lane, Egham, Surrey TW20 9NW, UK; [6]FRAME, Russell & Burch House, 96-98 North Sherwood Street, Nottingham NG1 4EE, UK; [7]Department of Pharmaceutical Biosciences, Uppsala University, Division of Toxicology, Box 594, S-751 24 Uppsala, Sweden; [8]Department of Health, HEF(M) Division, Skipton House, 80 London Road, London SE1 6LW, UK; [9]INSERM U49, Unite de Recherches Hepatologiques, Hopital Pontchaillou, 35033 Rennes, France; [10]ZENECA Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire SK10 4TJ, UK; [11]BIBRA International, Woodmansterne Road, Carshalton, Surrey SM5 4DS, UK; [12]SIAT — In Vitro Toxicology, Technopark, Pfingstweidstrasse 30, CH-8005 Zurich, Switzerland; [13]National Institute of Toxicology, P.O. Box 863, 41080 Seville, Spain; [14]Department of Transplantology, Institute of Biostructurel Medical School, Chalubinskiego 5, 02-004 Warsaw, Poland; [15]ZEBET, Bundesinstitut fur gesundheitlichen Verbraucherschutz und Veterinarmedizin (BgVV), Diedersdorfer Weg 1, D-12277 Berlin, Germany; [16]Biomedical Technologies Institute — CNR, Via GB Morgagni 301 E, 00161 Rome, Italy

Preface

This is the report of the fifth of a series of workshops organised by the European Centre for the Validation of Alternative Methods (ECVAM). ECVAM's main goal, as defined in 1993 by its Scientific Advisory Committee, is to promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine or replace the use of laboratory animals. One of the first priorities set by ECVAM was the implementation of procedures which would enable it to become well-informed about the state-of-the-art of non-animal test development and validation, and the potential for the possible incorporation of alternative tests into regulatory procedures. It was decided that this would be best achieved by the organisation of ECVAM workshops on specific topics, at which a small group of invited experts would review the current status of various types of *in vitro* tests and their potential uses, and make recommendations about the best ways forward (1).

The workshop on practical aspects of the validation of toxicity test procedures was held in Amden, Switzerland, on 24—28 January 1994, under the co-chairmanship of Michael Balls and Bas Blaauboer. The objective of the workshop was to reconsider the theoretical basis for validation laid down at Amden in

Address for correspondence and reprints: Professor Michael Balls, ECVAM, TP 580, JRC Environment Institute, 21020 Ispra (Va), Italy.

[1]European Centre for the Validation of Alternative Methods. (This workshop was organised for ECVAM by ERGATT [European Research Group for Alternatives in Toxicity Testing]). [2]This document represents the agreed report of the participants as individual scientists.

1990 (2), in light of the practical experience gained since then. In view of recent developments and new ideas concerning the validation of alternative procedures, it was felt that ECVAM needed to have access to expert guidance before it became involved in any further large-scale international validation studies on alternative tests. The European Research Group for Alternatives in Toxicity Testing (ERGATT) was invited to organise the workshop on behalf of ECVAM.

In this ECVAM workshop report, recommendations are made concerning the practical and logistical aspects of validating alternative toxicity testing procedures. These aspects have not really been considered in previous reports on validation.

Introduction

Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose (2). Several approaches to validation may be scientifically acceptable, depending on the particular purpose and goal of the study. Validation studies conducted thus far can be classified on the basis of their apparent objectives (3). These include in-house validation and validation for commercial purposes, in addition to validation which is undertaken to try to secure regulatory acceptance of a new test. Irrespective of the purpose, or the type of validation study, the key factor is that it is of a scientific, rather than a political, nature.

The first Amden report on the validation of toxicity test procedures provided the essential theoretical background to the validation process (2). In the subsequent Vouliagmeni report on promotion of the regulatory acceptance of validated non-animal toxicity test procedures (4), a sound set of principles were proposed for the independent evaluation of properly validated alternatives, and for their incorporation into the regulatory framework if this was considered to be appropriate.

Since their publication in 1990, both sets of recommendations have been widely welcomed. However, the Amden principles have been viewed by some as a set of rules, according to which the validation process must be conducted, rather than as the suggestions of one particular group of individuals (5-8). Experience gained in a variety of validation studies undertaken since 1990 has shown that the Amden principles cannot be applied rigidly. Indeed, it is unlikely that some of them can be applied in practice, due to the constraints faced by the managers of validation studies.

In addition to the first Amden report (2), several other discussions on validation have been published recently. Of particular note are the report of the Validation and Technology Transfer Committee of the Center for Alternatives to Animal Testing (CAAT) at Johns Hopkins University, Baltimore, MD, USA (9), the report of the CAAT/American Tissue Culture Association workshop on the international status of validation of *in vitro* toxicity tests (10), and the report of the Japanese validation study on alternative methods to the Draize eye irritation test (11). Walum *et al.* (8) have also recently discussed various principles for the validation of *in vitro* toxicology tests.

There is still widespread debate about how best to proceed with the validation of alternative methods, in terms of the most time-effective and cost-effective means of assessing their relevance, reliability and predictivity (3, 5, 7, 8, 12). It is hoped that the recommendations made in this ECVAM workshop report, concerning the practical and logistical aspects of conducting validation studies, will enable those involved in validation studies to recognise and learn from the mistakes which have been made in the past, as a basis for devising better validation strategies for the future. Although validation is a continuous process from a scientific perspective, this report concentrates, in particular, on the validation of alternative toxicological testing methods for a defined purpose — their regulatory acceptance. In general, this is the major concern of the sponsors, the participants in the validation study, and the general public.

Problems Encountered in Previous Validation Studies

Numerous problems have been encountered in previous validation studies, most notably the following.

1. The goals of the validation study have not been sufficiently defined, and there have been differences of opinion as to what the objectives should be, and how best to achieve them.
2. Many of the studies have been poorly designed and planned.

3. The responsibilities of the participants (management team, contractors, lead laboratories, and other personnel involved) have not been clearly defined and/or fulfilled.
4. The validation study has been inadequately managed due, for example, to funding and/or- communication problems.
5. Standard protocols defining all aspects of the test procedure have been lacking, or have not been strictly adhered to by some of the participating laboratories.
6. Relevant test chemicals, for which high quality and unambiguous *in vivo* (animal and/or human) toxicity data were available, could not be identified or obtained.
7. Participating laboratories have failed to meet the deadlines set for completion of the testing, submission of data, and other critical stages of the study.
8. Participants in the-validation study have been unable to agree on how to evaluate the data, or on the interpretation of the test results.
9. Laboratories have committed themselves to participating in a validation study, but have subsequently dropped out once the study was under way, for financial or logistical reasons.

Satisfactory solutions to these and similar problems must be found as a matter of urgency, if potentially useful alternative methods are to be properly validated.

Validation of Alternative Procedures Used for Different Purposes

Before any general guidance on the nature of validation studies can be given, the purpose of such studies must be considered. Experience gained since the first Amden workshop on validation (2) indicates that a pragmatic division into the following three main types of validation studies can be made, according to their intended purpose:

a) validation of alternative procedures for use in non-regulatory studies;
b) validation of alternative tests for inclusion as part of hierarchical approaches in regulatory guidelines; and
c) validation of alternative procedures for the replacement of existing regulatory guidelines.

In addition, a fourth type of validation study is possible, in which the alternative test is intended to provide part of the information which is required by a regulatory testing guideline. By conducting the alternative test and the regulatory method in parallel, sufficient data may eventually accrue to support the case for replacing the animal test. These various types of studies are considered in the following sections.

Non-regulatory studies

These studies differ from the others in that the w.ork does not form part of a regulatory submission. They are undertaken to provide sufficient supporting data to give reassurance that the information obtained using the alternative test is adequate for the purpose for which it is intended. For example, these studies may be carried out for selection or priority-setting within a range of developmental substances or preparations, or for the safety assessment of a range of similar products.

A second kind of non-regulatory study involves using *in vitro* tests to provide mechanistic information relevant to the type and extent of toxic effects which might be caused by a particular chemical or formulation. Such studies may eventually lead to the refinement or replacement of existing animal tests. The role of mechanistic toxicology in test method validation has recently been discussed by Frazier (12).

There is not necessarily a need to provide convincing evidence to support the formal and wider acceptance of the methods used in non-regulatory studies, except as part of the normal evolution of scientific methods, although they will need to be validated within the laboratory where they are being used.

Inclusion as part of a hierarchical approach in regulatory guidelines

The use of *in vitro* tests, or of more humane *in vivo* methods, may have a role in a hierarchical, or step-wise, approach within a given regulatory guideline. In such cases, the available data are not adequate to allow these tests to be a complete replacement for the existing procedure, but they do allow "positive" compounds to be identified in a more humane way, thus obviating the need for the

full animal test with such compounds. The Organisation for Economic Cooperation and Development (OECD) would appear to support this type of approach (see the later section on Progression Toward Regulatory Acceptance).

Examples of *in vivo* methods which are used in this way are the mouse local lymph node test and the mouse ear swelling test in the OECD procedure for skin sensitisation testing. Positive compounds in these preliminary tests do not need to be investigated in the Magnusson & Kligman test (the method preferred by the regulators) or the Buehler test, which involve more animals, and subject these animals to more distress, than the preliminary tests.

Examples of *in vitro* tests that may be used in this way are methods for identifying substances which are corrosive to the skin (for example, electrical conductivity measurements on skin slices), or are severe eye irritants (for example, enucleated eye tests). The OECD is now discussing the possibility of incorporating *in vitro* screens for the identification of positive compounds into test guidelines, in a similar manner to the way in which the use of more humane *in vivo* methods has been introduced. Such an approach may enable a number of currently available *in vitro* tests to be included in OECD test guidelines, for example, the proposed incorporation of screening methods into a test guideline on reproductive toxicity. Such *in vitro* screens would be used as the initial stage in a hierarchical approach to the safety testing of chemicals.

The extent of validation necessary for these screening tests is less than that required for an alternative procedure which is intended to be a complete replacement, since "negative" compounds will subsequently be investigated in the animal test. The use of a hierarchical approach will thus not result in any increase in the incidence of false negatives relative to the use of the existing method alone. However, it is recognised that adequate validation is needed to ensure that there is not an unacceptable false positive rate, since potentially useful chemicals may be lost and because over-classification should be avoided. As an example of the possible extent of validation required (and which was acceptable to a regulatory authority), the use of the local lymph node test in a hierarchical approach for investigating skin sensitisation, as detailed in the OECD test guideline, was accepted essentially on the basis of a validation study in which 25 compounds were tested in four laboratories (13).

Replacement of an existing regulatory guideline

At present, the only successful validation studies which have led to the replacement of an existing regulatory guideline relate to *in vivo* methods that are more humane (with a reduction in the numbers of animals required and a refinement of the techniques applied to them). In the case of the acceptance of the Fixed Dose Procedure as an alternative to the LD50 test for investigating acute oral toxicity, a validation exercise in the UK, involving five laboratories testing a total of 41 substances between them, and using the new method and the existing method in parallel, resulted in some refinement of the initial approach.

The performance of the "optimised" method, including its interlaboratory variability, was then investigated in a large international study which involved 20 compounds being tested blind in 31 laboratories in 11 countries (14). This was considered to provide adequate validation by both the Member States of the EU and by the member countries of the OECD. In the design of this validation study, a large number of laboratories were believed to be necessary, because of concerns about the subjective nature of the main endpoint measurement (evident toxicity) compared with that used in the existing method (death). The involvement of this number of laboratories would clearly be excessive for validation in most other contexts.

Provision of partial information

In many areas of toxicity testing, it is unlikely that sufficient data will be provided, at least in the short-term, to enable the existing method to be replaced. In such situations, it may be appropriate to build up a database by undertaking *in vitro* methods and the existing regulatory guideline procedure concurrently, using the same samples of test materials. Provision of data on male fertility, which is currently obtained by undertaking a reproductive toxicity study, can be considered as an example. A battery of *in vitro* tests could be devised to provide information on various aspects of male reproductive function (for example, sperm maturation and sperm transport). In due course, enough tests and suitable

data may accumulate for such a battery of tests to provide adequate information for the replacement of animal procedures in this area. In the meantime, such methods may be of value in mode-of-action studies.

The concept and value of parallel *in vivo/in vitro* testing were discussed in detail in the Vouliagmeni report on regulatory acceptance (4), and have been formulated as a practical proposal in the ERGATT/CFN (Swedish Board for Laboratory Animals) Integrated Toxicity Testing Scheme, ECITTS (15, 16).

Stages in the Evolution of New Tests

Practical experience gained since the publication of the first Amden report (2) has shown that more emphasis should be placed on *test development* than on what was defined at that time as *intralaboratory assessment.* Also, based on the experience gained during the validation of short-term genotoxicity tests, *test database development was* previously seen as an essential stage in the validation process, but it has proved unrealistic to insist that groups of 200-250 carefully selected chemicals should have been tested before a new method could be considered to be validated. In very few, if any, areas of toxicology would adequate *in vivo* data be available on such large numbers of chemicals.

Meanwhile, the Vouliagmeni report (4) introduced the concept *of independent assessment* of the outcome of a validation study, before a proposal for modification of test guidelines is submitted to the regulatory authorities. It is therefore proposed that the progression of new tests from conception to regulatory acceptance should be considered to consist of five main stages: test development, prevalidation, validation, independent assessment, and progression toward regulatory acceptance (Table I).

Test development

Before a test can be considered for inclusion in a validation study, it must have been properly developed. The development of a test involves a description of its basis and a definition of its scientific purpose, as well as an explanation of the need for it in relation to the type and extent of toxic effect (for example, skin corrosive, mild ocular irritant), the type of assessment (i.e. of toxic potential, potency, hazard or risk), the chemical spectrum to which the test can be applied (2), and the availability of other tests. A case for support of the test, in terms of its potential relevance and reliability, should be established. Finally, a detailed and comprehensive protocol suitable for the preliminary evaluation of its interlaboratory transferability should be produced.

Prevalidation

Experience indicates that a prevalidation step is needed following test development, prior to the possible inclusion of a test in a large-scale, formal validation study. The prevalidation stage should involve any optimisation and standardisation of the protocol which may be considered necessary, the identification of any unexpected problems with the test design and procedure, including those relating to analysis of the resulting data, and an initial assessment of the interlaboratory transferability of the method. Prevalidation need not involve the blind testing of coded chemicals.

Validation

The main purpose of a validation study is to conduct an interlaboratory blind trial, as a basis for assessing whether one or more tests, test batteries or testing strategies can be shown to be relevant and reliable for one or more specific purposes, according to predefined performance criteria. Formal validation studies should comprise a preliminary phase (in which a small number of coded chemicals, a "training set", are tested), and a definitive phase. This should be followed by data analysis and an evaluation of the outcome of the study.

Independent assessment

Before any regulatory authorities are asked to consider the formal acceptance of any satisfactorily validated alternative procedure for incorporation into the regulatory framework, the published results of a validation study should be considered by one or more independent assessment panels, under the auspices of appropriate national or international organisations, as outlined in the Vouliagmeni report (4). Such organisations need not necessarily be governmental bodies, but could be industry associations or learned societies. The membership of such panels should be independent of any particular validation study under consideration, and

Table I: Stages in the evolution of new tests

1.	**Test development** (laboratory of origin)	
	Purpose of the test	
	Need for the test	
	Derivation of the method	
	Application to appropriate chemicals	
	Case for inclusion in a validation study	
	Production of a protocol	
2.	**Prevalidation** (informal interlaboratory study)	
	Assessment of the interlaboratory transferability	
	Optimisation of the protocol	
3.	**Validation** (formal interlaboratory study, including a blind trial)	
	Two phases:	preliminary phase (training set of chemicals)
		definitive phase
	Main stages:	study design
		selection of tests
		selection of laboratories
		selection and distribution of chemicals
		data collection and analysis
		assessment of outcome
4.	**Independent assessment of study and proposals**	
5.	**Progression toward regulatory acceptance**	

should be representative of the scientific, toxicological, industrial, regulatory and animal welfare communities (4).

Criteria for the evaluation of validation studies were also given in the first Amden report (2). These encompass an assessment of the performance of each test in terms of its reproducibility and relevance. The tests should be evaluated according to the purpose of the validation study and, in particular, should take into account the assessment of the outcome by the management team and by the participating laboratories.

Assessment of the outcome of a validation study should involve close examination of every aspect of the programme, including all definitive statements made about the validity of each alternative procedure included in the study. The first main task of the independent assessment panel should be to consider the purpose and objectives of the validation study and whether these have been met. The independent assessment panel should then assess the value of the scientifically validated alternative procedure in competition with other tests, including those already validated or known to be in the course of development or validation (4). They should also evaluate the need for the procedure and the practicability of its use as part of the regulatory process.

Progression toward regulatory acceptance

It is essential that any new method that is considered to be adequately validated as a replacement for an existing method receives as widespread international recognition as possible. For example, the OECD test guidelines are particularly important in this respect, since they are used for tests conducted in member countries in Europe and North America, and in Japan, Australia and New Zealand. Furthermore, under the OECD Mutual Acceptance of Data Agreement, member countries have agreed to accept data

from tests performed according to OECD test guidelines, provided that the principles of Good Laboratory Practice (GLP) are observed.

The OECD has established a procedure for updating test guidelines and for the introduction of new test methods (17). This takes into account both advances in science and proposals that are based on animal welfare considerations. The latter point is particularly important, since many OECD member countries have regulations (for example, *Directive 86/609/EEC)* which require that, when a reasonably practical and adequately validated test is available that does not entail the use of animals protected by laboratory animal welfare legislation, it must be used. In addition, if it is unavoidable to use animals, the regulations may require that a method using the minimum number of animals and the most humane techniques, consistent with obtaining acceptable data, is used.

The OECD, through its Test Guidelines Programme, is currently involved in discussing how it could play a more active role in the development of alternative methods, and whether there is a need for a more formal mechanism for the introduction of alternative tests. In particular, the OECD is concentrating on the possibility of incorporating *in vitro* screens, for the identification of positive compounds, in test guidelines. The advantages of this would be that the alternative tests would be conducted in compliance with GLP, and the results would be reported as part of the notification/registration procedures, thereby contributing to the development of a database.

Once the management team, steering committee and sponsors of a validation study, together with the independent assessment panel(s), consider that an alternative procedure has been shown to be valid for a particular purpose, then, in the case of the OECD, a proposal should be submitted (via a national coordinator of one member country or, possibly, directly via the OECD Secretariat) for entry of the test into the procedure established for the updating of test guidelines.

Practical Validation Studies

Study design

The design of a validation study should reflect the objectives of the study and should take into account various other essential considerations. The objective should be clearly stated and, most importantly, it should represent a realistic goal. Laying down objectives which are too vague or over-ambitious, or defining the specific use of the results too precisely, should be avoided. The definition of the objectives of a study is the responsibility of its sponsors, and the steering committee and the management team which they appoint. Attention must be paid to: the selection of tests; the selection of participating laboratories; the selection of test chemicals, and their coding and distribution; data collection and analysis; the procedures for the management of the study, including contractual arrangements among the various parties involved and an agreed time-scale for each stage; and the procedures for the assessment and reporting of the outcome.

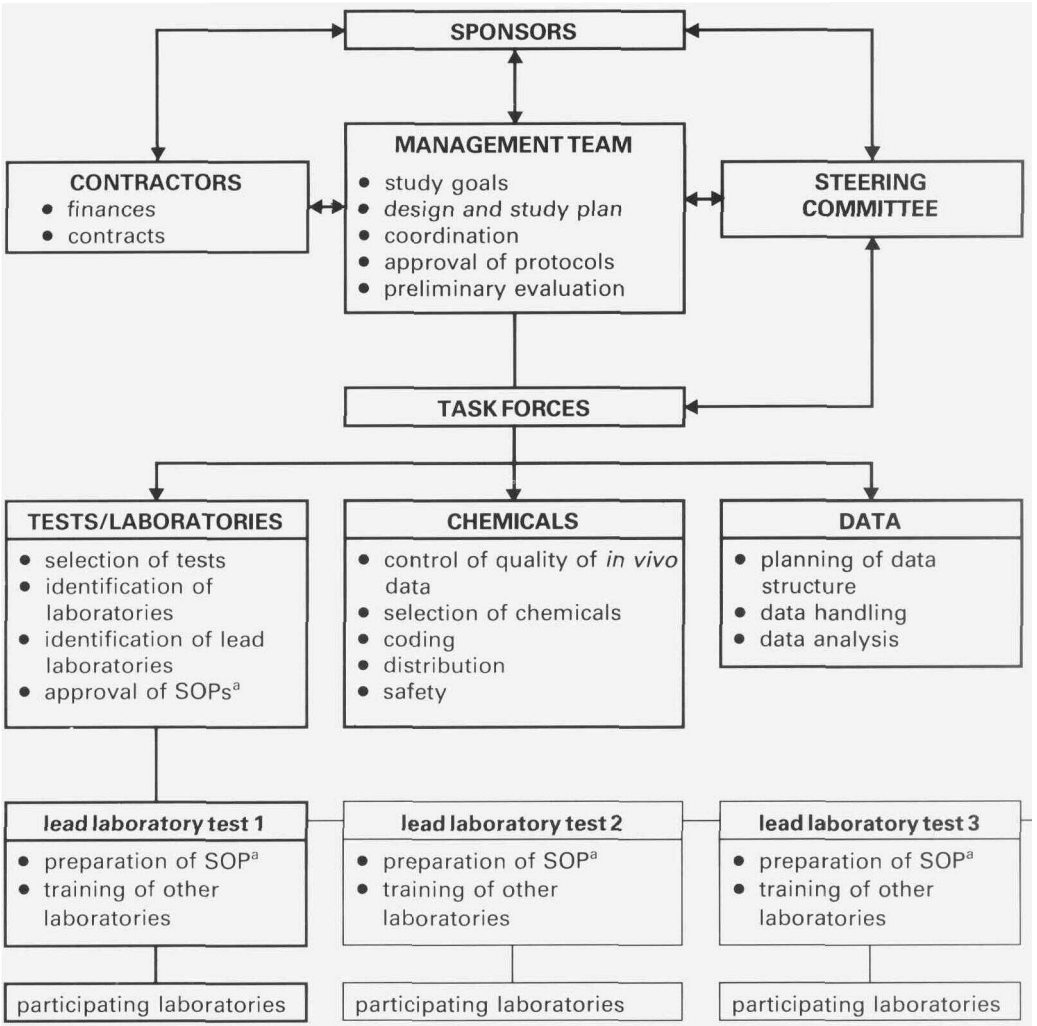
Expert statistical guidance should be sought, and it is desirable that a statistician should form part of the team designing the study. It should be recognised during the planning stage that most studies reflect a compromise between what is ideally required and what may be practically possible. It is important to ensure that the design is appropriate.

Selection of tests

The tests which are selected for validation should meet previously defined scientific criteria for inclusion. The steering committee and the management team appointed by the sponsors of the study should oversee the selection of the tests, or the batteries of tests, to be included. It is recommended that a task force is established (Figure 1), specifically to select the tests (and the laboratories). The tests should be chosen according to the following considerations.

1. The defined aims of the validation study.
2. The type of assessment involved (for example, hazard identification or safety assessment; screening of chemicals for classification of positives).
3. The types of chemicals to be tested, according to chemical class, physical form, and type and degree of toxicity.
4. The state of development of the test, including a defined purpose, evidence of its relevance and reliability, and the existence of an adequate protocol.
5. Evidence of the need for the test in relation to other tests which have been developed.

Figure 1: Management and organisation of validation studies



[a] *SOP = standard operating procedure*

The following criteria of readiness should be applied to a test, before it is accepted for inclusion in a validation study.

1. The purpose of the test and its mechanistic basis, if any, should have been defined, and satisfactory evidence of its potential usefulness for a specific purpose should have been provided.
2. An adequate protocol, including detailed standard operating procedures (SOPs), should have been published, or deposited at an independent institution which will protect it until its release is authorised (for example, *INVITTOX* [18], other data banks, or a patent office).
3. The area of applicability and limitations

of the test should have been clearly defined, including the classes and types of test materials which can and cannot be tested.

4. The relevance of the test to its proposed area of applicability should have been established in studies involving appropriate test chemicals, with suitable positive and negative controls.
5. The intralaboratory reproducibility of the test should have been demonstrated. Ideally, an initial assessment of its interlaboratory transferability should also have been undertaken.
6. The measurements which are made to provide the raw data should be clearly defined, as should the way in which these data are subsequently used to calculate the test results.
7. The way in which results are analysed and interpreted, in line with the stated purpose of the test, should be fully described.
8. The advantages of the test in comparison with other tests, and how it could be used in a tiered testing approach, or to complement or confirm the results of other tests in a test battery approach, should have been addressed.

Study plans, protocols and standard operating procedures

When a study is being designed, a study (or "project") plan should be produced to define the scientific conduct of the work and the procedures to be applied. The study plan is the basis of an agreement between the sponsor, the management team and the contractors to undertake the study. It is this plan, not the protocols or SOPs, which is exchanged, discussed and agreed upon. All study plans issued prior to agreement should be referred to as "draft study plans". Any alterations to the study plan, made subsequent to the agreement, should be issued in the form of study plan amendments. They must be authorised by the relevant principal investigator and agreed with the management team representing the sponsor. Unless self-evident, the reasons for making amendments must be provided, and they should be made in accordance with GLP requirements. The study plan, protocols and SOPs should be issued, or be available, to all staff involved in the study, as well as to any quality assurance (QA) staff assigned to the study.

The term "protocol" refers to the precise step-by-step description of a test. Where a single protocol covers a range of test substances and/or procedures, it needs to distinguish clearly between the different substances and tests or testing strategies. According to GLP requirements, a protocol should define all principal stages of the investigation, including a descriptive title, the purpose of the procedure, and details of experimental design, data acquisition, presentation and analysis.

SOPs provide written definitions of all the necessary technical and logistical steps to be used in the laboratory during the conduct of experimental work. They define the performance of routine activities, not normally specified in detail in a protocol. In other words, they provide information on exactly how to undertake each step of a particular procedure. SOPs should exist for all routine operations and for the use and maintenance of all equipment. Where a method is not defined in an SOP, it must be detailed in the protocol or documented, in detail, in the original plan for the study. If any method deviates from that detailed in the SOP, this deviation should be documented and requires prior approval by the principal investigator for that study. SOPs should be reviewed regularly and amended as required. Thus, it should be possible to define all activities relating to a validation study by reference only to the study plan, protocols, SOPs and original raw data.

Selection of laboratories

Laboratories should be chosen primarily on the basis of demonstrable competence in the test undergoing validation. The selection of the participating laboratories should be overseen by the steering committee and management team appointed by the sponsors of the study. It is recommended that a task force is established to select the laboratories (and tests) to be included (Figure 1).

Four laboratories per test is an adequate number for interlaboratory assessment. An increase above this number increases costs and logistical problems, without necessarily improving the scientific quality of the outcome. A lead laboratory should be appointed for each group of laboratories performing a particular test. The lead laboratory should be very experienced in conducting the test, and should be involved in preparing the protocol, and in guiding and training personnel from

the participating laboratories. A contact person should be appointed in each laboratory, to ensure ease of communication with the management team and, where appropriate, with other laboratories participating in the validation study.

Participating laboratories should meet minimum standards in terms of the availability of competent staff, laboratory facilities, safety and QA procedures. The laboratories chosen should usually include academic, government and industrial laboratories. A formal contract should be drawn up between the sponsors and the laboratories participating in the study, perhaps via the management team.

Good Laboratory Practice

GLP "...is intended to promote the quality and validity of test data. It is a managerial concept covering the organisational process and the conditions under which laboratory studies are planned, performed, monitored, recorded and reported" (19). Common basic principles and procedures for GLP were developed by the OECD, to facilitate the mutual acceptance of test data generated in OECD member countries in accordance with OECD test guidelines (20, 21).

While not all the laboratories participating in validation studies may have formally implemented GLP, or be subject to national or international GLP monitoring programmes, it is considered that the following requirements are essential for the mutual acceptance of information produced in the validation process.

1. Qualified personnel, and appropriate facilities, equipment and materials should be available.
2. Records of the qualifications, training and experience, and a job description for each professional and technical individual, should be maintained.
3. SOPs should be established and followed.
4. A study plan should be provided, and any amendments to this should be documented.
5. For each investigation, an individual with appropriate qualifications, training and experience should be appointed to be responsible for its overall conduct and for any report issued.
6. Space should be provided for the confidential storage and retrieval of raw data, reports, samples and specimens.
7. Apparatus used for the generation of data should be inspected regularly, cleaned, maintained and calibrated according to SOPs. Records of these processes should be kept, and made available for inspection on request.
8. Reagents should be labelled, as appropriate, to indicate their source, identity, concentration and stability. The labelling should include the preparation and expiry dates, and specific storage conditions.
9. Proper conditions should be established and maintained, not only for the housing, husbandry and care of animals, plants and microbial organisms, but also for other cellular and subcellular systems, in order to ensure the quality of the data. The origin of the biological system should be well defined, and its homogeneity and stability should be assured.
10. All data generated during a study should be recorded directly, promptly and legibly by the individual(s) responsible. These entries should be signed and dated.
11. Data generated as direct computer input should be identified at the time of input by the individual(s) responsible.
12. All changes to data should be identified with the date and the identity of the individual responsible, and a reason for the change should be documented at the time.
13. A final report should be provided and signed by the designated responsible individual.

Studies may be audited by the management team or by a group appointed by them, the steering committee or the sponsors of the validation study.

Selection and distribution of chemicals

The overall responsibility for choosing the number and type of test chemicals lies with the sponsors, steering committee and management team of a validation study. It is recommended that they establish a task force to deal with all aspects relating to the selection and distribution of chemicals (Figure

1). The management team will appoint an independent contractor to purchase, code and supply the test chemicals to the participating laboratories.

The criteria for inclusion of chemicals in a validation study are dependent on, and should be guided by, the needs and purpose of the particular 'study. The major factor governing chemical selection is the availability of appropriate and reliable *in vivo* data, against which to judge the performance of the alternative test. Incorrect selection of test chemicals will seriously compromise the outcome of the validation study. Therefore, it is imperative that selection of the chemicals to be tested is consistent with the defined aims of the validation study. In addition, it is necessary to take into account the experience gained during test development when selecting the set of test chemicals.

Test chemicals data

Chemicals must be chosen according to the availability of relevant and reliable *in vivo* data of high quality, which provide an unequivocal assessment of their toxic potencies. Any final toxicological classification should be supported by raw data, to permit such reassessment as may be deemed necessary in the future. Preferably, high quality human data should be used for comparative purposes where man is the target species of interest (2). This will not usually be possible, so reliable data derived from experimental animals will have to be used. Such animal data should be derived from studies conducted according to international testing guidelines (for example, OECD guidelines), and in compliance with GLP.

Several other criteria for the selection of test chemicals are outlined in Table II. The chemicals included should be single chemical entities (preferably), or formulations, of known high and consistent purity. All the laboratories in the validation study must, wherever possible, use the same batch of each chemical or formulation, which should be as close as possible to the material tested *in vivo*. It is recognised that only infrequently (for example, in parallel testing) will it be possible to use the same batch as that tested *in vivo*. The chemicals must be readily available (preferably from commercial sources) in sufficient quantity. They must be stable under the defined conditions of storage for at least the duration of the validation study.

Assessment of in vivo toxicology data

The approach to the assessment of *in vivo* data will depend entirely upon the source and nature of the data and the needs and goals of the validation study. For example, for a validation study designed to assess the utility of a test, or a battery of tests, as a replacement for a particular animal test undertaken for regulatory purposes, the toxicological information should be of sufficient quality for the chemicals to be classified unambiguously, according to the appropriate regulatory guideline(s). However, for a validation study designed for other purposes, for example, to explore structure-activity relationships, a set of chemicals with *in vivo* activities not classified according to any recognised regulatory scheme would be acceptable. In such cases, toxicity assessments of the chemicals selected should be based on a well-defined and informed peer-review process.

Table II: Selection of test chemicals

Essential criteria:

1. must be supported by relevant and reliable *in vivo* data of high quality
2. must be single chemical entities (preferably) or formulations, of known, high and consistent purity
3. must be readily available, preferably from commercial sources
4. must be stable under defined conditions for at least the duration of the validation study

Additional criteria dependent on the specific purpose of the validation study:

5. must cover the desired range of toxic effects/potencies (for example, non-toxic-highly toxic)
6. must include relevant classes of chemicals (for example, surfactants, acids)
7. must include relevant physical states (i.e. solids, liquids)

Initiatives taken by organisations such as the European Centre for the Ecotoxicology and Toxicology of Chemicals (ECETOC) and the European Cosmetic Toiletry and Perfumery Association (COLIPA) to establish task forces to select sets of chemicals appropriate for use in specific validation studies (for example, on eye irritancy [22], skin irritancy and phototoxicity [23]) are most welcome. They are in line with the call for reference chemical data banks, made in the Amden (2) and Vouliagmeni (4) reports, as well as by Purchase (24). Such sets of chemicals are useful in test development, as well as in validation studies. The magnitude of this task should not be underestimated, since it is proving difficult to find adequate numbers of chemicals, representative of the full spectrum of compound groups and of different levels of toxicity, which are backed by *in vivo* data which meet the criteria set by the task forces.

Number of test chemicals

It is impossible to be precise as to the optimum number of test chemicals, since this can depend upon the number of participating laboratories, the number of tests included in the validation study, the range of toxic effects and potencies being investigated, the diversity of chemical classes and structural relationships between the chemicals, and the range of physical states to be considered, in addition to the availability of relevant and reliable *in vivo* data for the chemicals.

Sets of test chemicals

It is suggested that a subset of the chemicals selected for a validation study is employed initially as a training set (for the preliminary phase), to ensure that all the participating laboratories are complying with the agreed protocols. The composition of this training set should be representative of the chemistry, range of toxic effects and physical states of the main set of test chemicals.

Supply and distribution of test chemicals

Chemicals should be dispatched in suitable containers, from a central repository, and should be labelled in accordance with relevant transport regulations. An adequate amount of chemical for conducting the whole of the validation of the test should be supplied, together with a physicochemical data sheet and a safety information sheet. The test chemicals should be coded uniquely and dispatched simultaneously to all participating laboratories. Advance notification of shipping should be provided. To ensure against unforeseen circumstances, it is suggested that the identities of the coded test chemicals be maintained confidentially at two separate sites. In addition, a decision on when to break the test codes will need to be made. These are the responsibilities of the management team.

Physicochemical data and description of chemicals

Laboratories should be provided with the following essential information about the test chemicals:

a) visual appearance;
b) physical state;
c) weight or volume of the sample of test chemical dispatched;
d) physicochemical data on the pH, solubility (including suggested solvents), volatility and stability; and
e) storage instructions.

Safety information

Participating laboratories should be provided with a sealed package, containing necessary information about the hazards of the test chemicals. This should include clear instructions for action in the case of accidents. Such safety information should be lodged with a named person within the testing laboratory or organisation, who is not involved with the conduct of the actual validation study (for example, a safety officer). This information should satisfy local legal requirements (for example, OSHA for the USA and COSHH for the UK). At the end of the validation study, the package should be returned unopened, at the time of data submission. In the event that an accident occurs, however, and the package is opened, this must be notified as soon as possible to the management team for the validation study.

Alternatively, instead of supplying the laboratories with sealed packages, it could be arranged for the chemical codes and accompanying health and safety information to be deposited with an independent organisation offering a 24-hour emergency service, such as a national poisons information centre. In some previous validation studies, a sealed envelope containing the codes for all the chemicals has been lodged with the West Midlands Poisons Unit (Birmingham, UK), and all the participating laboratories have been provided with telephone numbers for the poisons centre. If the code has to be broken in the event of an emergency, the poisons

centre would notify the management team for the validation study.

If chemicals pose a particular disposal problem, information relating to this needs to be included with the physicochemical information sheet. Otherwise, it is recommended that test chemicals be disposed of in accordance with 'local regulations, once clearance has been given by the management team.

Data collection and analysis

It is recommended that a task force is appointed, by the steering committee and management team, to be responsible for all aspects of the study relating to data collection and analysis (Figure 1). The management team would appoint an independent contractor to receive, process and analyse the data.

Reporting of results

A test protocol should state explicitly the type and amount of data to be collected. Laboratories participating in the validation study should agree on exactly what form the results should take, i.e. quantitative or qualitative. The protocol should define the experimental design, in terms of the number of repeat experiments each laboratory should perform and the number of independent replicate measures of the endpoint in each repeat experiment.

The protocol should define a standard form for data entry, to be used by all participating laboratories. Considerable care is needed to ensure that such forms are not ambiguous, and clear directions should be given on how to complete the forms. For example, a statement on whether to use a full stop or a comma to represent decimal points should be given. In international studies, it is essential to remember that many participants will not be using their first language. Some method for allowing extra information to be provided, but in a way which does not interfere with the actual reporting of the data, is needed. Problems associated with reporting the results, such as the need to distinguish between "zero", "not determined", "could not be tested" and "no effect", should be anticipated and explicitly addressed in the instructions. It is advisable to try out the data entry form in a prevalidation study, before beginning the formal validation study.

The experimental protocol must state explicitly how the data specifying the endpoint are to be determined, and the calculations used to derive these values must be specified. The original data obtained and any values derived from them should be subjected to QA and quality control (QC).

Results must be submitted to the data analyst in the agreed format. Specific computer-based systems may be used to collect, analyse and report the data, but this must be agreed upon as part of the project at the study design stage. There are advantages and disadvantages to the submission of data in either computer-based electronic format or on paper (whether handwritten or printed). The latter has the potential disadvantages of transcription errors and illegibility of the data. The former may raise technical and logistical issues for some studies. If data are submitted in electronic format, a written copy should be provided as well.

The data requested on the standard form should be relevant to the purpose of the study. Any extra data may form part of the report of the study. It is extremely important that the nature of the data required is clearly identified at the study design stage. Potential problems, such as possible inconsistencies in the nature of the data reported for different chemicals or between laboratories (such as qualitative results from one laboratory and quantitative data from another), should be addressed at the protocol agreement stage.

Data analysis

The data analysis should be carried out independently, in order to reduce the possibility of bias in analysis and interpretation. This has potential advantages in terms of the future acceptance of the results of the study. A disadvantage is that the independent analyst may be unaware of specific aspects of the experimental system, and this may complicate the analysis. This should be avoided if biostatisticians and laboratory scientists are both involved at the planning stage. Those responsible for undertaking the data analysis should be directly responsible to the management team.

In general, two types of information may be required from the study. Firstly, an assessment of the amount of variability within and between the participating laboratories (i.e. the intralaboratory and interlaboratory variability) for a particular test is necessary. Secondly, for those tests where it is relevant, a direct comparison between the results obtained using an alternative test and those from an *in vivo* test carried out with the same chemical will be needed. The statistical meth-

ods used for assessing the two different aspects of the experimental design will differ.

Statistical methods can be used both to test a hypothesis, such as whether significant differences exist in the performance of a test between laboratories, and for estimation, such as in providing an estimate of how great the variability is. In practice, a complete statistical analysis involves the use of both approaches to provide an exploration of a complex set of data. It is important to remember that a distinction may need to be made between the statistical significance of a finding and its biological importance.

The investigation of the variability in a test can be carried out by the use of a hierarchical analysis. This permits estimates of the size of the variation at different levels in the design, and tests of the statistical significance using analysis of variance (ANOVA) techniques (25). A typical design might be the testing of a chemical by a series of laboratories, each conducting a set of repeat experiments, each making a number of independent replicate measures of a particular endpoint. The precise design may vary from test to test. The terms "repeat" and "replicate" are used specifically here to designate the different levels in the design; however, these terms tend to be used interchangeably in the statistical literature.

An experimental design explicitly incorporating a comparison between alternative and *in vivo* data permits tests to be made for the agreement between the two sets of data. Several different methods are used. These methods depend, in part, on the nature of the data. Qualitative or categorical data have often been described in terms of the sensitivity and specificity of the test results, often in conjunction with estimates of the positive prediction value (PPV) and the negative prediction value (NPV) of the test. These statistics may be a useful description of the results, but sometimes have little predictive value because of the dependence of the PPV and NPV on the prevalence (the proportion of the chemicals showing the characteristics of interest in the population of chemicals being studied). An initial appropriate selection of test chemicals in terms of prevalence would avoid the problem of the limited predictive value of the PPV and NPV.

Quantitative data have been compared using correlation and regression models. Correlation should be used for determining the degree of association between two measures, while regression is used for predicting one measure from another. The difference between these two measures is often not appreciated. Neither method is ideal for the measurement of agreement between two parameters, and more suitable statistical techniques exist for determining such agreement. Criteria for determining acceptable levels of agreement need to be established on a case-by-case basis.

Statistical tests for variability among the levels in a hierarchical design can be carried out, for quantitative data, by using ANOVA procedures in a step-wise manner (25). The statistical significance of the degrees of variability between different levels can be tested and estimates given of the variability. It may be appropriate to carry out a transformation of the data to permit a parametric approach. Qualitative or categorical data may be difficult to analyse in a hierarchical design, and specialised statistical advice may need to be sought in such cases.

In the case of explicit comparisons between *in vivo* and alternative test data for a series of chemicals, the set of chemicals can be divided into two subsets, one being considered to be a "training" set and the other a "test" set. Such a division allows *post hoc* comparisons (for example, specific cut-off values) to be developed on the data from the training set and then to be tested on the test set. These are common cross-validation procedures. This division of chemicals into subsets has no implications for the actual experimental procedures, and the division should not be known by the laboratories participating in the study.

Assessment of outcome

An important stage following the definitive validation study and data analysis is the critical assessment of the outcome by the management team and steering committee. This assessment should be an orderly process which reviews every aspect of the study. Ultimately, a definitive statement on the validity of the tests included should be given.

Firstly, the goal statement of the validation study must be reviewed. This statement should have clearly indicated the way in which a particular test would be used in the safety assessment process. The next step should be an examination of the overall study design. Factors to be considered are: whether the test chemicals and reference compounds

were commensurate with the study goals; whether there were common, clearly written protocols and SOPs, which were strictly adhered to by the participating laboratories; whether all the necessary data have been provided; and whether the testing was performed according to critical GLP procedures.

The quality of the *in vivo* data on the test chemicals must also be evaluated. It is important that the toxicity of these chemicals is defined as accurately as possible. If the *in vivo* data are of questionable accuracy, it will be of little value using them as a reference for investigating the validity of the alternative tests. The quality of the data obtained using the alternative methods must also be assured; this should include an assessment of the intra-experimental, intralaboratory and interlaboratory variability in the data. An alternative method which does not produce consistently reproducible results will be of limited value.

Once all these factors have been considered, the relationship between the *in vivo* and the alternative test data must be assessed, to determine whether the proposed method is valid for its stated purpose. The statistical analysis should have provided a set of quantitative measures of the relationship between the *in vivo* and the alternative test data. A decision about the validity of the proposed methods should then be based on these results. If it is concluded that the method is valid, it could then be used in the safety assessment process and, if appropriate, may be proposed for regulatory acceptance.

Management

Validation studies will normally be conducted under the auspices of sponsors, such as international bodies (for example, COLIPA), government agencies (for example, CFN, ECVAM, the National Institute of Environmental Health Sciences [NIEHS], ZEBET), national organisations (for example, the Cosmetic Toiletry and Fragrance Association [CTFA]), or other independent organisations (for example, FRAME). The sponsors of a study will need to appoint a steering committee composed of independent experts, which in turn will appoint a small management team. The management team will coordinate the study and, in consultation with the steering committee and the task forces which have been appointed, will be responsible for defining the purpose of the study, for the selection of tests, participating laboratories and test chemicals, and for the procedures for data collection and analysis (Figure 1).

Contractual aspects

The sponsors of a validation study will need to appoint a main contractor to be responsible for the issuing of contracts, as appropriate, and for the financial management of the study. An international study can involve many languages, many legal systems, and many different financial practices and contractual systems, all of which can greatly complicate the work of a management team. There is an imperative need for a very clear understanding by all individuals involved in a validation study of their precise duties and rights, for example in relation to payment for work done, requirement for strict adherence to the protocol, the time-table for submission of data, ownership of results, the right to publish jointly or independently, and the right to comment on the data and the outcome.

It is a great advantage to have these issues clearly addressed in a formal contract, or contracts. This can become quite complicated and can be a source of difficulties. For example, the sponsors of a study may have expectations with regard to timing which are not reasonable in light of what has to be done at the laboratory bench or in the data analysis phase. This can cause problems for a management team caught in the middle between the sponsors and the participating laboratories. At worst, a sponsor may threaten to withdraw financial support, while some laboratories and other sub-contractors will expect to be paid for work they have done.

Dissemination of information

The Vouliagmeni report emphasised the need for the early, continuous and comprehensive dissemination of information about validation studies (4), which should include the following.

1. An early announcement of the existence of the study.
2. Early reporting on the aims, design and time-table for the study.
3. Reporting on the progress and eventual outcome of the study.
4. Publication of a report of the study in the peer-review literature (although not nee-

essarily with peer-review in the usual "submitted manuscript" sense).

5. Depositing of all raw data which are not in the published report in a place where they will be freely available.
6. Encouragement of independent published review of the entire validation study, somewhat like a book review.

Nevertheless, at present, not all the managers of validation studies are living up to these ideals. It is very difficult to find out what studies are in progress, or even what studies have been completed, let alone what studies are at the planning stage. Test protocols are rarely available. In addition, the reports of some studies are not in the peer-review literature (for example, the EC pilot studies on alternatives for skin and eye irritation testing). In other cases, reports only become available long after studies have been completed (for example, the three phases of the CTFA programme). It is hoped that ECVAM will be able to help here, for example by providing a repository for raw data as part of the information services required of ECVAM by the EU, and by insisting on the publication by *INVITTOX* of protocols for tests involved in validation studies conducted under the auspices of ECVAM.

Summary of Conclusions and Recommendations

The following recommendations are made concerning the practical and logistical aspects of validating alternative toxicity testing procedures.

1. A pragmatic division into three main types of validation studies can be made, according to their intended purpose: validation of alternative procedures for use in non-regulatory studies, for inclusion as part of hierarchical approaches in regulatory guidelines, and for the replacement of existing regulatory guidelines.
2. Validation studies will normally be conducted under the auspices of national or international sponsors, who will need to appoint a steering committee of independent experts, which in turn will appoint a small management team.
3. It is recommended that the management team and steering committee establish task forces, which are given responsibilities for the selection of the tests to be included and the participating laboratories, the selection of the test chemicals, and the development of the procedures for data collection and analysis.
4. The management team will coordinate the study and, in consultation with the steering committee and the task forces, will be responsible for defining the purpose of the study, for the selection of tests, laboratories and test chemicals, and for implementing the procedures for data collection and analysis.
5. Independent contractors will need to be appointed to purchase, code and distribute the chemicals, to receive, process and analyse the data, and to issue contracts and manage the financial aspects of the study.
6. It is important that all the participants in a validation study understand precisely what is required and expected of them. These issues should be clearly addressed in formal contracts.
7. Early, continuous and comprehensive dissemination of information about all validation studies is essential. It is hoped that ECVAM, with the cooperation of all other interested parties, will be able to ensure that this is the case.
8. The design of a validation study should reflect the objectives of the study, which must be clearly stated and must represent a realistic goal. Attention must be paid to the selection of tests, participating laboratories and test chemicals, data collection and analysis, and the procedures for the management of the study, and for the assessment and reporting of the outcome. A biostatistician should be included in the team responsible for designing the study.
9. A study plan should be produced which defines the scientific conduct of the work and the procedures to be applied. The study plan should be the basis of an agreement between the sponsor, the management team and the contractors to undertake the validation study.
10. It is proposed that the progression of new tests should be considered to consist of five main stages: test development,

prevalidation, validation, independent assessment, and progression toward regulatory acceptance.

11. More emphasis should be placed on test development, than on intralaboratory assessment as defined in the first Amden report (2).

12. Before a test can be considered for inclusion in a validation study, it must have been properly developed. A case for support of the test, in terms of its potential relevance and reliability, should be established, and a detailed protocol suitable for the preliminary evaluation of its interlaboratory transferability should have been produced.

13. The prevalidation stage should involve optimisation and standardisation of the protocol, the identification of any unexpected problems with respect to the test design, procedure and data analysis, and an initial assessment of the interlaboratory transferability of the method. Prevalidation need not involve the blind testing of coded chemicals.

14. Formal validation studies, involving an interlaboratory blind trial, should comprise a preliminary phase (in which a small number of coded chemicals, a "training set", are tested) and a definitive phase.

15. Four laboratories per test is an adequate number for interlaboratory assessment. The laboratories should be chosen primarily on the basis of demonstrable competence in the test undergoing validation. They should meet minimum GLP standards.

16. A lead laboratory should be appointed for each group of laboratories performing a particular test.

17. The selection of the chemicals to be tested must be consistent with the defined aims of the validation study. The chemicals must be chosen according to the availability of relevant and reliable *In vivo* data of high quality, which provide an unequivocal assessment of their toxic potencies.

18. The chemicals included should be single chemical entities (preferably), or formulations, of known high and consistent purity.

19. A subset of chemicals, representative of the chemistry, range of toxic effects and physical states of the main set of test chemicals, should be used as a training set for the preliminary phase of the validation stage.

20. Although test database development was previously considered to be an essential stage in the validation process, it is unrealistic to insist that groups of 200-250 carefully selected chemicals should have been tested before a new method can be considered to be validated. In very few, if any, areas of toxicology would adequate *in vivo* data be available on such large numbers of chemicals.

21. Data analysis should be carried out independently, to reduce the possibility of bias in analysis and interpretation.

22. A test protocol should state explicitly the form and amount of data to be collected. This protocol should define a standard data entry sheet, to be used by all participating laboratories. The protocol must also state how the data specifying the endpoint are to be determined.

23. The assessment of the outcome by the management team and the steering committee should include a review of every aspect of the study. Ultimately, a definitive statement on the validity of the tests included should be given.

24. Before any regulatory authorities are asked to consider the formal acceptance of validated alternative procedures, the published results of a validation study should be evaluated by one or more independent assessment panels, under the auspices of appropriate national or international organisations.

25. Once the management team, steering committee and sponsors of a validation study, together with the independent assessment panel(s), consider that an alternative procedure has been shown to be valid for a particular purpose, a proposal should be submitted for entry of the test into the appropriate procedure established for the updating of test guidelines.

Acknowledgements

The authors would like to thank Dr Manfred Liebsch (ZEBET, BgVV, Germany) for his help with Figure 1, and Dr Philip Botham (ZENECA, UK) and Dr Erik Walum (Pharmacia, Sweden) for commenting on the first draft of the report.

References

1. Anon. (1994). ECVAM News & Views. *ATLA* **22,** 7-11.
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. (1992). *In vitro* test validation: high hurdling but not pole vaulting. *A TLA* 20,355-357.
4. Balls, M., Botham, P., Cordier, A., Fumero, S., Kayser, D., Koeter, H., Koundakjian, P., Lindquist, N.G., Meyer, O., Pioda, L., Reinhardt, C., Rozemond, H., Smyrniotis, T., Spielmann, H., Van Looy, H., van der Venne, M.T. & Walum, E. (1990). Report and recommendations of an international workshop on promotion of the regulatory acceptance of validated non-animal toxicity test procedures. *ATLA* 18, 339-344.
5. Ekwall, B. (1992). Validation of *in vitro* cytotoxicity tests. In *In Vitro Alternatives to Animal Pharmaco-toxicology* (ed. J.V. Castell & M.J. Gomez-Lechon), pp. 361-390. Madrid, Spain: Farmaindustria.
6. Ekwall, B., Bondesson, I., Hellberg, S., Hogberg, J., Romert, L., Stenberg, K. & Walum, E. (1991). Validation of *in vitro* cytotoxicity tests — past and present strategies. *ATLA* 19, 226-233.
7. Flint, O. (1992). *In vitro* test validation: a house built on sand. *ATLA* **20,** 196-198.
8. Walum, E., Clemedson, C. & Ekwall, B. (1994). Principles for the validation of *in vitro* toxicology test methods. *Toxicology in Vitro* 8, 807-812.
9. Goldberg, A.M., Frazier, J.M., Brusick, D., Dickens, M.S., Flint, O., Gettings, S.D., Hill, R.N., Lipnick, R.L., Renskers, K.J., Bradlaw, J.A., Scala, R.A., Veronesi, B., Green, S., Wilcox, N.L. £ Curren, R.D. (1993). Report of the Validation and Technology Transfer Committee of the Johns Hopkins Center for Alternatives to Animal Testing. Framework for validation and implementation of *in vitro* toxicity tests. *Xenobiotica* 23, 563-572.
10. Frazier, J.F., ed. (1993). *CAAT Technical Report No. 5: Report on the CAAT/TCA Workshop on the International Status of Validation of In Vitro* Toxicity Tests. Baltimore, USA: CAAT.
11. Ohno, Y., Kaneko, T., Kobayashi, T., Inoue, T., Kuroiwa, Y., Yoshida, T., Momma, J., Hayashi, M., Akiyama, J., Atsumi, T., Chiba, K., Endo, T., Fujii, A., Kakishima, H., Kojima, H., Masamoto, K., Masuda, M., Matsukawa, S., Ohkoshi, K., Okada, J., Sakamoto, K., Takano, K. & Takanaka, A. (1994). First-phase validation of the *in vitro* eye irritation tests for cosmetic ingredients. *In Vitro Toxicology* 7, 89-101.
12. Frazier, J.F. (1994). The role of mechanistic toxicology in test method validation. *Toxicology in Vitro 8,* 787-791.
13. Basketter, D.A., Scholes, E.W., Kimber, L, Botham, P.A., Hilton, J., Miller, K., Robbins, M.C., Harrison, P.T.C. & Waite, S.J. (1991). Interlaboratory evaluation of the local lymph node assay with 25 chemicals and comparison with guinea-pig test data. *Toxicology Methods* 1, 30^3.
14. van den Heuvel, M.J., Clark, D.C., Fielder, R.J., Koundakjian, P.P., Oliver, G.J.A., Felling, D., Tomlinson, N.J. & Walker, A.P. (1990). The international validation of a fixed dose procedure as an alternative to the classical LD50 test. *Food and Chemical Toxicology* 28, 469-482.
15. Blaauboer, B.J., Balls, M., Bianchi, V., Bolcsfoldi, G., Guillouzo, A., Moore, G.A., Odland, L., Reinhardt, C.A., Spielmann, H. & Walum, E. (1994). The ECITTS integrated toxicity testing scheme: the application of *in vitro* test systems to the hazard assessment of chemicals. *Toxicology in Vitro* 8, 845-846.
16. Walum, E., Balls, M., Bianchi, V., Blaauboer, B., Bolcsfoldi, G., Guillouzo, A., Moore, G.A., Odland, L., Reinhardt, C. & Spielmann, H. (1992). ECITTS: an integrated approach to the application of *in vitro* test systems to the hazard assessment of chemicals. *ATLA* 20, 406-428.
17. Koeter, H.B.W.M. (1994). Principles for a pragmatic approach to the regulatory acceptance of alternative tests. *Toxicology in Vitro* 8, 925-930.
18. Ungar, K. (1993). *INVITTOX:* promoting *in vitro* methods for toxicity testing. *AATEX* 2, 37-43.
19. Anon. (1992). *Environment Monograph No. 45: The OECD Principles of Good Laboratory Practice 1,* 29pp. Paris, France: Organisation for Economic Cooperation and Development.
20. Anon. (1981). The OECD Principles of Good Laboratory Practice: Annex II of the decision of the Council concerning the mutual acceptance of data in the assessment of chemicals [C(81) 30 (Final)], 25 pp. Paris, France: Organisation for Economic Cooperation and Development.
21. Anon. (1981). *OECD Guidelines for the Testing of Chemicals.* Paris, France: Organisation for Economic Cooperation and Development.
22. Bagley, D.M., Botham, P.A., Gardner, J.R., Holland, G., Kreiling, R., Lewis, R.W., Stringer, D.A. & Walker, A.P. (1992). Eye irritation: reference chemicals data bank. *Toxicology in Vitro* 6, 487-491.
23. Spielmann, H., Lovell, W.W., Holzle, E., Johnson, B.E., Maurer, T., Miranda, M.A., Pape, W.J.W., Sapora, O. & Sladowski, D. (1994). *In vitro* phototoxicity testing. The report and recommendations of ECVAM workshop 2. *ATLA* 22, 314-348.
24. Purchase, I.F.H. (1990). An international reference chemical data bank would accelerate the development, validation and regulatory acceptance of alternative toxicology tests. *ATLA* 18, 345-348.
25. Snedecor, G.W. & Cochran, W.G. (1980). *Statistical Methods,* 7th edn. Iowa, USA: Iowa State University Press.

Appendix A: Terminology

Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose.

Reliability of a procedure describes whether it can be performed reproducibly within and among laboratories and over time.

Relevance of a procedure describes whether it is meaningful and useful for a particular purpose.

Test development is the process by which the components of a protocol (for example, experimental system, exposure conditions, endpoint, endpoint measurement and data analysis procedures) are defined for a specific purpose, and is normally carried out in the laboratory of origin.

Interlaboratory assessment establishes whether or not a test can be successfully transferred from one laboratory to another. It should include two phases: prevalidation, which results in fine-tuning of the study design and final optimisation and agreement of the standard protocol to be used, and formal validation, in which a set of test chemicals is evaluated by the standard protocol.

Procedure refers to a test, test battery or testing strategy.

Test refers to the combination of the experi mental system used, exposure conditions, endpoint, endpoint measurement and data analysis method.

Endpoint refers to the processes, responses or effects assessed.

Endpoint measurement refers to the techniques used to assess endpoints.

Protocol refers to the precise step-by-step description of a test.