ICH Harmonised Tripartite Guideline

Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals

S1A

Current Step 4 version
dated 29 November 1995

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
S1A
Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>30 March 1995</td>
<td>November 2005</td>
</tr>
</tbody>
</table>

**Current Step 4 version**

| S1A | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 29 November 1995 | S1A |
GUIDELINE ON THE NEED FOR CARCINOGENICITY STUDIES OF PHARMACEUTICALS

ICH Harmonised Tripartite Guideline
Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 29 November 1995, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. Introduction ............................................................................................................. 1
2. Historical Background .......................................................................................... 1
3. Objective of the guideline ..................................................................................... 1
4. Factors to consider for carcinogenicity testing .............................................. 2
   4.1 Duration and Exposure .................................................................................. 2
   4.2 Cause for Concern ........................................................................................... 2
   4.3 Genotoxicity .................................................................................................... 2
   4.4 Indication and Patient Population .................................................................. 2
   4.5 Route of Exposure ........................................................................................... 3
   4.6 Extent of Systemic Exposure ......................................................................... 3
   4.7 Endogenous Peptides and Protein Substances or Their Analogs ................. 3
5. Need for Additional Testing ................................................................................. 4
GUIDELINE ON THE NEED FOR CARCINOGENICITY STUDIES OF PHARMACEUTICALS

1. INTRODUCTION
The objectives of carcinogenicity studies are to identify a tumorigenic potential in animals and to assess the relevant risk in humans. Any cause for concern derived from laboratory investigations, animal toxicology studies, and data in humans may lead to a need for carcinogenicity studies. The practice of requiring carcinogenicity studies in rodents was instituted for pharmaceuticals that were expected to be administered regularly over a substantial part of a patient's lifetime. The design and interpretation of the results from these studies preceded much of the available current technology to test for genotoxic potential and the more recent advances in technologies to assess systemic exposure. These studies also preceded our current understanding of tumorigenesis with non-genotoxic agents. Results from genotoxicity studies, toxicokinetics, and mechanistic studies can now be routinely applied in preclinical safety assessment. These additional data are important not only in considering whether to perform carcinogenicity studies but for interpreting study outcomes with respect to relevance for human safety. Since carcinogenicity studies are time consuming and resource intensive they should only be performed when human exposure warrants the need for information from life-time studies in animals in order to assess carcinogenic potential.

2. HISTORICAL BACKGROUND
In Japan, according to the 1990 “Guidelines for Toxicity Studies of Drugs Manual”, carcinogenicity studies were needed if the clinical use was expected to be continuously for 6 months or longer. If there was cause for concern, pharmaceuticals generally used continuously for less than 6 months may have needed carcinogenicity studies. In the United States, most pharmaceuticals were tested in animals for their carcinogenic potential before widespread use in humans. According to the US Food and Drug Administration, pharmaceuticals generally used for 3 months or more required carcinogenicity studies. In Europe, the Rules Governing Medicinal Products in the European Community defined the circumstances when carcinogenicity studies were required. These circumstances included administration over a substantial period of life, i.e., continuously during a minimum period of 6 months or frequently in an intermittent manner so that the total exposure was similar.

3. OBJECTIVE OF THE GUIDELINE
The objective of this guideline is to define the conditions under which carcinogenicity studies should be conducted to avoid the unnecessary use of animals in testing, and to provide consistency in worldwide regulatory assessments of applications. It is expected that these studies will be performed in a manner that reflects currently accepted scientific standards.

The fundamental considerations in assessing the need for carcinogenicity studies are the maximum duration of patient treatment and any perceived cause for concern arising from other investigations. Other factors may also be considered such as the intended patient population, prior assessment of carcinogenic potential, the extent of systemic exposure, the (dis)similarity to endogenous substances, the appropriate study design, or the timing of study performance relative to clinical development.
4. FACTORS TO CONSIDER FOR CARCINOGENICITY TESTING

4.1 Duration and Exposure
Carcinogenicity studies should be performed for any pharmaceutical whose expected clinical use is continuous for at least 6 months (see Note 1).

Certain classes of compounds may not be used continuously over a minimum of 6 months but may be expected to be used repeatedly in an intermittent manner. It is difficult to determine and to justify scientifically what time represents a clinically relevant treatment periods for frequent use with regard to carcinogenic potential, especially for discontinuous treatment periods. For pharmaceuticals used frequently in an intermittent manner in the treatment of chronic or recurrent conditions, carcinogenicity studies are generally needed. Examples of such conditions include allergic rhinitis, depression, and anxiety. Carcinogenicity studies may also need to be considered for certain delivery systems which may result in prolonged exposures. Pharmaceuticals administered infrequently or for short duration of exposure (e.g., anaesthetics and radiolabelled imaging agents) do not need carcinogenicity studies unless there is cause for concern.

4.2 Cause for Concern
Carcinogenicity studies may be recommended for some pharmaceuticals if there is concern about their carcinogenic potential. Criteria for defining these cases should be very carefully considered because this is the most important reason to conduct carcinogenicity studies for most categories of pharmaceuticals. Several factors which could be considered may include: (1) previous demonstration of carcinogenic potential in the product class that is considered relevant to humans; (2) structure-activity relationship suggesting carcinogenic risk; (3) evidence of preneoplastic lesions in repeated dose toxicity studies; and (4) long-term tissue retention of parent compound or metabolite(s) resulting in local tissue reactions or other pathophysiological responses.

4.3 Genotoxicity
Unequivocally genotoxic compounds, in the absence of other data, are presumed to be trans-species carcinogens, implying a hazard to humans. Such compounds need not be subjected to long-term carcinogenicity studies. However, if such a drug is intended to be administered chronically to humans a chronic toxicity study (up to one year) may be necessary to detect early tumorigenic effects.

Assessment of the genotoxic potential of a compound should take into account the totality of the findings and acknowledge the intrinsic value and limitations of both in vitro and in vivo tests. The test battery approach of in vitro and in vivo tests is designed to reduce the risk of false negative results for compounds with genotoxic potential. A single positive result in any assay for genotoxicity does not necessarily mean that the test compound poses a genotoxic hazard to humans (ICH Guidance on Specific Aspects of Regulatory Genotoxicity Tests).

4.4 Indication and Patient Population
When carcinogenicity studies are required they usually need to be completed before application for marketing approval. However, completed rodent carcinogenicity studies are not needed in advance of the conduct of large scale clinical trials, unless there is special concern for the patient population.
For pharmaceuticals developed to treat certain serious diseases, carcinogenicity testing need not be conducted before market approval although these studies should be conducted post-approval. This speeds the availability of pharmaceuticals for life-threatening or severely debilitating diseases, especially where no satisfactory alternative therapy exists.

In instances where the life-expectancy in the indicated population is short (i.e., less than 2 - 3 years) no long-term carcinogenicity studies may be required. For example, oncolytic agents intended for treatment of advanced systemic disease do not generally need carcinogenicity studies. In cases where the therapeutic agent for cancer is generally successful and life is significantly prolonged there may be later concerns regarding secondary cancers. When such pharmaceuticals are intended for adjuvant therapy in tumour free patients or for prolonged use in noncancer indications, carcinogenicity studies are usually needed.

4.5 Route of Exposure
The route of exposure in animals should be the same as the intended clinical route when feasible (ICH Dose Selection for Carcinogenicity Studies of Pharmaceuticals). If similar metabolism and systemic exposure can be demonstrated by differing routes of administration, then carcinogenicity studies should only be conducted by a single route, recognising that it is important that relevant organs for the clinical route (e.g., lung for inhalational agents) be adequately exposed to the test material. Evidence of adequate exposure may be derived from pharmacokinetic data (ICH Guidance on Repeated Dose Tissue Distribution Studies).

4.6 Extent of Systemic Exposure
Pharmaceuticals applied topically (e.g., dermal and ocular routes of administration) may need carcinogenicity studies. Pharmaceuticals showing poor systemic exposure from topical routes in humans may not need studies by the oral route to assess the carcinogenic potential to internal organs. Where there is cause for concern for photocarcinogenic potential carcinogenicity studies by dermal application (generally in mice) may be needed. Pharmaceuticals administered by the ocular route may not require carcinogenicity studies unless there is cause for concern or unless there is significant systemic exposure.

For different salts, acids, or bases of the same therapeutic moiety, where prior carcinogenicity studies are available, evidence should be provided that there are no significant changes in pharmacokinetics, pharmacodynamics, or toxicity. When changes in exposure and consequent toxicity are noted, then additional bridging studies may be used to determine whether additional carcinogenicity studies are needed. For esters and complex derivatives, similar data would be valuable in assessing the need for an additional carcinogenicity study, but this should be considered on a case-by-case basis.

4.7 Endogenous Peptides and Protein Substances or Their Analogs
Endogenous peptides or proteins and their analogs, produced by chemical synthesis, by extraction/purification from an animal/human source or by biotechnological methods such as recombinant DNA technology may require special considerations.

Carcinogenicity studies are not generally needed for endogenous substances given essentially as replacement therapy (i.e., physiological levels), particularly where there is previous clinical experience with similar products (for example, animal insulins, pituitary-derived growth hormone, and calcitonin).
Although not usually necessary, long-term carcinogenicity studies in rodent species should be considered for the other biotechnology products noted above, if indicated by the treatment duration, clinical indication, or patient population (providing neutralising antibodies are not elicited to such an extent in repeated dose studies as to invalidate the results). Conduct of carcinogenicity studies may be important in the following circumstances: (1) For products where there are significant differences in biological effects to the natural counterpart(s); (2) for products where modifications lead to significant changes in structure compared to the natural counterpart; and (3) for products resulting in humans in a significant increase over the existing local or systemic concentration (i.e., pharmacological levels).

5. **Need for Additional Testing**

The relevance of the results obtained from animal carcinogenicity studies for assessment of human safety are often cause for debate. Further research may be needed, investigating the mode of action, which may result in confirming the presence or the lack of carcinogenic potential for humans. Mechanistic studies are useful to evaluate the relevance of tumour findings in animals for human safety.

**Supplementary Note**

*Note 1:* It is expected that most pharmaceuticals indicated for 3 months treatment would also likely be used for 6 months. In an inquiry to a number of pharmaceutical research and regulatory groups no cases were identified in which a pharmaceutical would be used only for 3 months.
This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
## S1B Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1B</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>1 May 1996</td>
<td>November 2005</td>
</tr>
</tbody>
</table>

### Current Step 4 version

| S1B | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 16 July 1997 | S1B |
TESTING FOR CARCINOGENICITY OF PHARMACEUTICALS

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 16 July 1997, this guideline is recommended for adoption to the three regulatory parties to ICH.

TABLE OF CONTENTS

1. OBJECTIVE .............................................................................................................. 1
2. BACKGROUND ........................................................................................................ 1
3. SCOPE OF THE GUIDELINE ................................................................................ 2
4. THE GUIDELINE ..................................................................................................... 2
   4.1 Preamble. .......................................................................................................... 2
   4.2 Experimental approaches to testing for carcinogenic potential .................... 2
      4.2.1 Choice of species for a long-term carcinogenicity study ....................... 2
      4.2.2 Additional in vivo tests for carcinogenicity. ............................................. 2
      4.2.3 Considerations in the choice of short or medium term tests for carcinogenicity. ................................................................. 3
5. MECHANISTIC STUDIES ...................................................................................... 3
   5.1 Cellular changes. .............................................................................................. 3
   5.2. Biochemical measurements ........................................................................... 3
   5.3. Considerations for additional genotoxicity testing ........................................ 3
   5.4. Modified protocols ........................................................................................... 3
6. GENERAL CONSIDERATIONS IN THE CHOICE OF AN APPROPRIATE SPECIES FOR LONG TERM CARCINOGENICITY TESTING ....................... 4
   6.1. Information from surveys on pharmaceuticals .............................................. 4
   6.2. Potential to study mechanisms. ...................................................................... 4
   6.3. Metabolic disposition. ....................................................................................... 5
   6.4. Practicality ....................................................................................................... 5
   6.5. Testing in more than one species. ................................................................... 5
   6.6. Exceptions. ...................................................................................................... 5
7. EVALUATION OF CARCINOGENIC POTENTIAL ......................................... 5
NOTES ............................................................................................................................... 6
ANNEX: Other ICH Guidelines Cited ......................................................................... 7
TESTING FOR CARCINOGENICITY OF PHARMACEUTICALS

1. OBJECTIVE
This document provides guidance on approaches for evaluating the carcinogenic potential of pharmaceuticals.

2. BACKGROUND
Historically, the regulatory requirements for the assessment of the carcinogenic potential of pharmaceuticals in the three regions (E.U., Japan, U.S.) provided for the conduct of long-term carcinogenicity studies in two rodent species, usually the rat and the mouse. Given the cost of these studies and their extensive use of animals, it is in keeping with the mission of ICH to examine whether this practice requiring long term carcinogenicity studies in two species could be reduced without compromising human safety.

This guideline should be read in conjunction with other guidelines (see Annex), especially:

S1.A: Guideline on the Need for Carcinogenicity Studies of Pharmaceuticals.

S1.C: Dose Selection for Carcinogenicity Studies of Pharmaceuticals.

Long-term rodent carcinogenicity studies for assessing the carcinogenic potential of chemicals (including pharmaceuticals) to humans are currently receiving critical examination. Since the early 1970's, many investigations have shown that it is possible to provoke a carcinogenic response in rodents by a diversity of experimental procedures, some of which are now considered to have little or no relevance for human risk assessment. This guideline outlines experimental approaches to the evaluation of carcinogenic potential that may obviate the necessity for the routine conduct of two long-term rodent carcinogenicity studies for those pharmaceuticals that need such evaluation. The relative individual contribution of rat and mouse carcinogenicity studies and whether the use of rats or mice alone would result in a significant loss of information on carcinogenicity relevant to human risk assessment has been addressed by six surveys of the data for human pharmaceuticals. The surveys were those of the International Agency for Research on Cancer (IARC), the U.S. Food and Drug Administration (FDA), the U.S. Physicians' Desk Reference (PDR), the Japanese Pharmaceutical Manufacturers’ Association (JPMA), the EU Committee for Proprietary Medicinal Products (CPMP), and the UK Centre for Medicines Research (CMR). The dimensions of these surveys and the principal conclusions of the analyses can be found in the Proceedings of the Third International Conference (1995) on Harmonization.

Positive results in long-term carcinogenicity studies that are not relevant to the therapeutic use of a pharmaceutical present a dilemma to all parties: regulatory reviewers, companies developing drugs and the public at large. The conduct of one long-term carcinogenicity study (rather than two long term studies) would, in part, allow resources to be diverted to other approaches to uncover potential carcinogenicity relevant to humans. A “weight of evidence” approach, that is use of scientific judgment in evaluation of the totality of the data derived from one long-term carcinogenicity study along with other appropriate experimental investigations, enhances the assessment of carcinogenic risk to humans.
3. SCOPE OF THE GUIDELINE
The guideline embraces all pharmaceutical agents that need carcinogenicity testing as indicated in Guideline S1A. For biotechnology-derived pharmaceuticals refer to Guideline S6.

4. THE GUIDELINE

4.1 Preamble.
The strategy for testing the carcinogenic potential of a pharmaceutical is developed only after the acquisition of certain key units of information, including the results of genetic toxicology (Guidelines S2A and S2B), intended patient population, clinical dosage regimen (Guideline S1A), pharmacodynamics in animals and in humans (selectivity, dose-response) (Guideline S1C), and repeated-dose toxicology studies. Repeated-dose toxicology studies in any species (including nonrodents) may indicate that the test compound possesses immunosuppressant properties, hormonal activity, or other activity considered to be a risk factor for humans, and this information should be considered in the design of any further studies for the assessment of carcinogenic potential (see also Note 1).

4.2 Experimental approaches to testing for carcinogenic potential.
Flexibility and judgment should be exercised in the choice of an approach which should be influenced by the information cited in the above preamble. Given the complexity of the process of carcinogenesis, no single experimental approach can be expected to predict the carcinogenic potential of all pharmaceuticals for humans.

The basic principle:
The basic scheme comprises one long-term rodent carcinogenicity study, plus one other study of the type mentioned in §4.2.2 that supplements the long term carcinogenicity study and provides additional information that is not readily available from the long term assay.

4.2.1 Choice of species for a long-term carcinogenicity study.
The species selected should be appropriate, based on considerations that include the following:
(a) Pharmacology.
(b) Repeated-dose toxicology.
(c) Metabolism (see also Guidelines S1C and S3A).
(d) Toxicokinetics (see also Guidelines S1C, S3A, and S3B).
(e) Route of administration (e.g., less common routes such as dermal and inhalation).

In the absence of clear evidence favoring one species, it is recommended that the rat be selected. This view is based on the factors discussed in §6.

4.2.2 Additional in vivo tests for carcinogenicity.
Additional tests may be either (a) or (b) (see Note 2).
(a) Short or medium-term in vivo rodent test systems.
Possibilities should focus on the use of in vivo models providing insight into carcinogenic endpoints. These may include models of initiation-promotion in rodents, or models of carcinogenesis using transgenic or neonatal rodents (Note 3).
(b) A long-term carcinogenicity study in a second rodent species is still considered acceptable (see § 4.2.1 for considerations).

4.2.3 Considerations in the choice of short or medium term tests for carcinogenicity.

Emphasis should be placed on selection of a test method that can contribute information valuable to the overall “weight of evidence” for the assessment of carcinogenic potential. The rationale for this choice should be documented and based on information available at the time of method selection about the pharmaceutical such as pharmacodynamics and exposure compared to human or any other information that may be relevant. This rationale should include a scientific discussion of the strengths and weaknesses of the method selected for the pharmaceutical(see Note 4).

5. MECHANISTIC STUDIES

Mechanistic studies are often useful for the interpretation of tumor findings in a carcinogenicity study and can provide a perspective on their relevance to human risk assessment. The need for or the design of an investigative study will be dictated by the particular properties of the drug and/or the specific results from the carcinogenicity testing. Dose dependency and the relationship to carcinogenicity study conditions should be evaluated in these investigational studies. Suggestions include:

5.1 Cellular changes.

Relevant tissues may be examined for changes at the cellular level using morphological, histochemical, or functional criteria. As appropriate, attention may be directed to such changes as the dose-relationships for apoptosis, cell proliferation, liver foci of cellular alteration, or changes in intercellular communication.

5.2. Biochemical measurements.

Depending on the putative mode of tumorigenic action, investigations could involve measurements of:

- plasma hormone levels, e.g. T3/T4, TSH, prolactin
- growth factors
- binding to proteins such as $\alpha_2\mu$-globulin
- tissue enzyme activity, etc.

In some situations, it may be possible to test a hypothesis of, for example, a hormone imbalance with another study in which the imbalance has been, at least in part, compensated.

5.3. Considerations for additional genotoxicity testing

(see Guidelines S2A and S2B).

Additional genotoxicity testing in appropriate models may be invoked for compounds that were negative in the standard test battery but which have shown effects in a carcinogenicity test with no clear evidence for an epigenetic mechanism. Additional testing can include modified conditions for metabolic activation in in vitro tests or can include in vivo tests measuring genotoxic damage in target organs of tumor induction (e.g., DNA damage and repair tests, 32P-postlabeling, mutation induction in transgenes).

5.4. Modified protocols.
Modified protocols may be helpful to clarify the mode of tumorigenic action of the test substance. Such protocols might include groups of animals to explore, for example, the consequence of interrupted dosage regimens, or the reversibility of cellular changes after cessation of dosing.

6. GENERAL CONSIDERATIONS IN THE CHOICE OF AN APPROPRIATE SPECIES FOR LONG TERM CARCINOGENICITY TESTING

There are several general considerations which, in the absence of other clear indications, suggest that the rat will normally be the species of choice for a long term carcinogenicity study.

6.1. Information from surveys on pharmaceuticals.

In the six analyses, attention was given to data on genetic toxicology, tumor incidence, strain of animal, route and dosage regimen, pharmacological or therapeutic activity, development and/or regulatory status, and, if relevant, reason for termination of development. Inevitably, there was considerable overlap of the data, but that is not necessarily an impediment to drawing valid conclusions.

The main overall conclusions from the analysis were:

a. Although very few instances have been identified of mouse tumors being the sole reason for regulatory action concerning a pharmaceutical, data from this species may have contributed to a “weight of evidence” decision and in identifying agents that caused tumors in two rodent species.

b. Of the compounds displaying carcinogenic activity in only one species, the number of "rat-only" compounds was about double the number of "mouse-only" compounds, implying in a simplistic sense that the rat is more "sensitive" than the mouse.

c. As with other surveys accessible in the literature, the data for pharmaceuticals were dominated by the high incidence of rodent liver tumors. The high susceptibility of mouse liver to nongenotoxic chemicals has been the subject of many symposia and workshops. These have concluded that these tumors may not always have relevance to carcinogenic risk in humans and can potentially be misleading.

6.2. Potential to study mechanisms.

The carcinogenic activity of nongenotoxic chemicals in rodents is characterized by a high degree of species, strain, and target organ specificity and by the existence of thresholds in the dose-response relationship. Mechanistic studies in recent years have permitted the distinction between effects that are specific to the rodent model and those that are likely to have relevance for humans. Progress has often been associated with increased understanding of species and tissue specificity. For example, receptor-mediated carcinogenesis is being recognized as of growing importance. Most of these advances are being made in the rat, and only rarely in the mouse.
6.3. Metabolic disposition.
Neither rats nor mice would seem, on metabolic grounds, to be a priori generally more suitable for the conduct of long term carcinogenicity studies. However, much attention is now being given to pharmacokinetic-pharmacodynamic relationships and rapid progress is occurring in knowledge of the P-450 isozymes that mediate the biotransformation of drugs. Most of this research activity is confined to rats and humans. Therefore, in the near future at least, where specific information on the P-450 isozymes involved in biotransformation is critical for the evaluation it appears that mice would be less likely to provide this mechanistic information.

6.4. Practicality.
Pertinent to the above two topics is the question of feasibility of investigative studies. Size considerations alone put the mouse at a severe disadvantage when it comes to the taking of serial blood samples, microsurgery/catheterization, and the weighing of organs. Blood sampling often requires the sacrifice of the animals, with the result that many extra animals may be needed when mice are subject to such investigations.

6.5. Testing in more than one species.
Most of the currently available short and medium term in vivo models for carcinogenicity testing involve the use of mice. In order to allow testing in more than one species for carcinogenic potential, when this is considered important and appropriate, the rat will often be used in the long term carcinogenicity study.

6.6. Exceptions.
Despite the above considerations, there may be circumstances under which the mouse or another rodent species could be justified on mechanistic, metabolic, or other grounds as being a more appropriate species for the long term carcinogenicity study for human risk assessment (c.f. §4.2.1). Under such circumstances it may still be acceptable to use the mouse as the short term or medium term model.

7. EVALUATION OF CARCINOGENIC POTENTIAL.
Evidence of tumorigenic effects of the drug in rodent models should be evaluated in light of the tumor incidence and latency, the pharmacokinetics of the drug in the rodent models as compared to humans, and data from any ancillary or mechanistic studies that are informative with respect to the relevance of the observed effects to humans.

The results from any tests cited above should be considered as part of the overall “weight of evidence” taking into account the scientific status of the test systems.
NOTES

Note 1. Data from in vitro assays, such as a cell transformation assay, can be useful at the compound selection stage.

Note 2. If the findings of a short or long-term carcinogenicity study and of genotoxicity tests and other data indicate that a pharmaceutical clearly poses a carcinogenic hazard to humans, a second carcinogenicity study would not usually be useful.

Note 3. Several experimental methods are under investigation to assess their utility in carcinogenicity assessment. Generally, the methods should be based on mechanisms of carcinogenisis that are believed relevant to humans and applicable to human risk assessment. Such studies should supplement the long term carcinogenicity study and provide additional information that is not readily available from the long term assay. There should also be consideration given animal numbers, welfare and the overall economy of the carcinogenic evaluation process. The following is a representative list of some approaches that may meet these criteria and is likely to be revised in the light of further information.

(a) The initiation-promotion model in rodent. One initiation-promotion model for the detection of hepatocarcinogens (and modifiers of hepatocarcinogenicity) employs an initiator, followed by several weeks of exposure to the test substance. Another multi-organ carcinogenesis model employs up to five initiators followed by several months of exposure to the test substance.

(b) Several transgenic mouse assays including the p53+/− deficient model, the Tg.AC model, the TgHras2 model, the XPA deficient model, etc.

(c) The neonatal rodent tumorigenicity model.

Note 4. While there may be a number of approaches that will in general meet the criteria described in Note 3 for use as the additional in vivo study, not all may be equally suitable for a particular pharmaceutical. The following are examples of factors that should be considered and addressed in the rationale:

1. Can results from the model provide new information not expected to be available from the long-term study that is informative with respect to hazard identification and/or risk assessment?

2. Can results from the model address concerns related to the carcinogenic process arising from prior knowledge of the pharmaceutical or compounds with similar structures and/or mechanisms of action? These concerns may include genotoxic, mitogenic, promotional, or receptor-mediated effects, etc.

3. Does the metabolism of the pharmaceutical shown in the animal model affect the evaluation of carcinogenic risk for humans?

4. Is adequate systemic or local exposure attained in relation to human exposure?

5. How extensively has the model been evaluated for its intended use? Prior to using any new in vivo methods in testing the carcinogenic potential of pharmaceuticals for humans, it is critical that the method be evaluated for its ability to contribute to the weight of evidence assessment. Many experimental studies are in progress (1997) to evaluate the new short or medium tests for carcinogenic potential. These include selected
pharmaceuticals with known potencies and known mechanism of carcinogenic activity in rodents, and also putative human non-carcinogens. When the results of these studies become available, it may be possible to offer clearer guidance on which of these tests have the most relevance for cancer assessment in humans.

ANNEX: Other ICH Guidelines Cited


Guideline S3B: Guidance on Repeat-Dose Tissue Distribution Studies.

ICH Harmonised Tripartite Guideline

Dose Selection for Carcinogenicity Studies of Pharmaceuticals
S1C(R2)

Current Step 4 version
Parent Guideline dated 27 October 1994
(Addendum on a Limit Dose dated 17 July 1997 and incorporated in November 2005)
Revised on 11 March 2008

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
### S1C(R2)

#### Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1C(R2)</td>
<td></td>
<td></td>
<td>November 2005</td>
</tr>
</tbody>
</table>

#### Parent Guideline: Dose Selection for Carcinogenicity Studies of Pharmaceuticals

<table>
<thead>
<tr>
<th>S1</th>
<th>Approval by the Steering Committee under <em>Step 2</em> and release for public consultation.</th>
<th>27 October 1993</th>
<th>S1</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1C(R1)</td>
<td>Approval by the Steering Committee under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>27 October 1994</td>
<td>S1C(R1)</td>
</tr>
</tbody>
</table>

#### Addendum to the Parent Guideline: Addition of a Limit Dose and Related Notes

<table>
<thead>
<tr>
<th>S1C(R)</th>
<th>Approval of the Addendum by the Steering Committee under <em>Step 2</em> and release for public consultation.</th>
<th>6 November 1996</th>
<th>in S1C(R1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1C(R)</td>
<td>Approval of the Addendum by the Steering Committee under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>17 July 1997</td>
<td>in S1C(R1)</td>
</tr>
</tbody>
</table>

#### First Revision (R1)

| S1C and S1C(R) | The parent guideline is now renamed S1C(R1) as the Addendum has been incorporated to the parent guideline. The new title is: “Dose Selection for Carcinogenicity Studies of Pharmaceuticals & Limit Dose”. | November 2005 | S1C(R1) |

#### Current *Step 4* version

| S1C(R2) | Approval by the Steering Committee of the second revision directly under *Step 4* without further public consultation, to delete Note 2 of the parent guideline and revise the text referring to the Notes. The title has been changed by deleting “& Limit Dose”. The Addendum has been integrated in the text. In this revision the pharmacokinetic endpoint of 25 is declared to be applicable also for pharmaceuticals with positive genotoxicity signals. This change has implications on “Refinement” (one of the 3R’s) in enhancing the welfare, i.e., reducing the pain or discomfort of the animals at the MTD. | 11 March 2008 | S1C(R2) |
DOSE SELECTION FOR CARCINOGENICITY STUDIES OF PHARMACEUTICALS

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 11 March 2008, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. INTRODUCTION............................................................................................................ 1
   1.1 General Considerations for the Conduct of Dose-Ranging Studies......................... 2
   1.2 Toxicity Endpoints in High Dose Selection .............................................................. 2
   1.3 Pharmacokinetic Endpoints in High Dose Selection ................................................. 3
   1.4 Criteria for Comparisons of AUC in Animals and Man for use in High Dose Selection ......................................................................................................................... 3
   1.5 Saturation of Absorption in High Dose Selection....................................................... 4
   1.6 Pharmacodynamic Endpoints in High Dose Selection ............................................. 4
   1.7 Maximum Feasible Dose ............................................................................................ 4
   1.8 Limit Dose ................................................................................................................. 4
   1.9 Additional Endpoints in High Dose Selection .......................................................... 5
   1.10 Selection of Middle and Low Doses in Carcinogenicity Studies ............................. 5
2. SUMMARY..................................................................................................................... 5
3. NOTES.......................................................................................................................... 5
DOSE SELECTION FOR CARCINOGENICITY STUDIES
OF PHARMACEUTICALS

1. INTRODUCTION

Traditionally, carcinogenicity studies for chemical agents have relied upon the maximally tolerated dose (MTD) as the standard method for high dose selection (Note 1). The MTD is generally chosen based on data derived from toxicity studies of 3 months' duration.

In the past, the criteria for high dose selection for carcinogenicity studies of human pharmaceuticals have not been uniform among international regulatory agencies. In Europe and Japan, dose selection based on toxicity endpoints or attaining high multiples of the maximum recommended human daily dose (>100x on a mg/kg basis) has been accepted. However, in the United States, dose selection based on the MTD has traditionally been considered the only appropriate practice. All regions have used a maximum feasible dose as an appropriate endpoint.

For pharmaceuticals with low rodent toxicity, use of the MTD can result in the administration of very large doses in carcinogenicity studies, often representing high multiples of the clinical dose. This has led to the concern that exposures in rodents greatly in excess of the intended human exposures might not be relevant to human risk; because they so greatly alter the physiology of the test species, the findings might not reflect what would occur following human exposure.

Ideally, the doses selected for rodent bioassays for pharmaceuticals should provide an exposure to the agent that (1) allows an adequate margin of safety over the human therapeutic exposure, (2) is tolerated without significant chronic physiological dysfunction and is compatible with good survival, (3) is guided by a comprehensive set of animal and human data that focus broadly on the properties of the agent and the suitability of the animal, and (4) permits data interpretation in the context of clinical use.

In order to achieve international harmonisation of requirements for high dose selection for carcinogenicity studies of pharmaceuticals, and to establish a rational basis for high dose selection, the ICH Expert Working Group on Safety initiated a process to arrive at common, scientifically based criteria for high dose selection. Several features of pharmaceutical agents distinguish them from other environmental chemicals and can justify a guideline which might differ in some respects from other guidelines. This should enhance the relevance of the carcinogenicity study for pharmaceuticals. Thus much knowledge might be available on the pharmacology, pharmacokinetics, and metabolic disposition in humans. In addition, there will usually be information on the patient population, the expected use pattern, the range of exposure, and the toxicity and/or side effects that cannot be tolerated in humans. Diversity of the chemical and pharmacological nature of the substances developed as pharmaceuticals, and the diversity of mechanisms of carcinogenesis, call for a flexible approach to dose selection. This document proposes that any one of several approaches could be useful for dose selection, and should provide for a more rational approach to dose selection for carcinogenicity studies for pharmaceuticals. These include: 1) toxicity-based endpoints; 2) pharmacokinetic endpoints; 3) saturation of absorption; 4) pharmacodynamic endpoints; 5) maximum feasible dose; 6) limit dose; and 7) additional endpoints.
Consideration of all relevant animal data and integration with available human data is paramount in determining the most appropriate endpoint for selecting the high dose for the carcinogenicity study. Relevant pharmacokinetic, pharmacodynamic, and toxicity data should be considered in the selection of doses for the carcinogenicity study, regardless of the primary endpoint used for high dose selection.

In the process of defining such a flexible approach, it is recognised that the fundamental mechanisms of carcinogenesis are only poorly understood at the present time. Further, it is also recognised that the use of the rodent to predict human carcinogenic risk has inherent limitations, although this approach is the best available option at this time. Thus, while the use of plasma levels of drug-derived substances represents an important attempt at improving the design of the rodent bioassay, progress in this field calls for continuing examination of the best method to detect human risk. This guideline is therefore intended to serve as guidance in this difficult and complex area, recognising the importance of updating the specific provisions outlined below as new data become available.

1.1 General Considerations for the Conduct of Dose-Ranging Studies
The considerations involved when undertaking dose-ranging studies to select the high dose for carcinogenicity studies are the same regardless of the final endpoint utilised.

1. In practice, carcinogenicity studies are carried out in a limited number of rat and mouse strains for which there is reasonable information on spontaneous tumour incidence. Ideally, rodent species/strains with metabolic profiles as similar as possible to humans should be studied (Note 2);
2. Dose-ranging studies should be conducted for both males and females for all strains and species to be tested in the carcinogenicity bioassay;
3. Dose selection is generally determined from 90-day studies using the route and method of administration that will be used in the bioassay;
4. Selection of an appropriate dosing schedule and regimen should be based on clinical use and exposure patterns, pharmacokinetics, and practical considerations;
5. Ideally, both the toxicity profile and any dose-limiting toxicity should be characterised. Consideration should also be given to general toxicity, the occurrence of preneoplastic lesions and/or tissue-specific proliferative effects, and disturbances in endocrine homeostasis;
6. Changes in metabolite profile or alterations in metabolising enzyme activities (induction or inhibition) over time should be understood to allow for appropriate interpretation of studies.

1.2 Toxicity Endpoints in High Dose Selection
ICH 1 agreed to evaluate endpoints other than the MTD for the selection of the high dose in carcinogenicity studies. These were to be based on the pharmacological properties and toxicological profile of the test compound. There is no scientific consensus on the use of toxicity endpoints other than the MTD. Therefore, the ICH Expert Working Group on Safety has agreed to continue use of the MTD as a useful toxicity-based endpoint for high dose selection for carcinogenicity studies.

The following definition of the MTD is considered consistent with those published previously by international regulatory authorities (Note 1): The top dose or maximum
tolerated dose is that which is predicted to produce a minimum toxic effect over the course of the carcinogenicity study. Such an effect can be predicted from a 90-day dose range-finding study in which minimal toxicity is observed. Factors to consider are alterations in physiological function which would be predicted to alter the animal's normal life span or interfere with interpretation of the study. Such factors include: no more than 10% decrease in body weight gain relative to controls; target organ toxicity; significant alterations in clinical pathological parameters.

1.3 Pharmacokinetic Endpoints in High Dose Selection
A systemic exposure representing a large multiple of the human area under the exposure curve (AUC) (at the maximum recommended daily dose) can be an appropriate endpoint for dose selection for carcinogenicity studies for pharmaceuticals which have similar metabolic profiles in humans and rodents and low organ toxicity in rodents (i.e., high doses are well tolerated in rodents). The level of animal systemic exposure should be sufficiently great compared to exposure to provide reassurance of an adequate test of carcinogenicity.

It is recognised that the doses administered to different species might not correspond to tissue concentrations because of different metabolic and excretory patterns. Comparability of systemic exposure is better assessed by blood concentrations of parent drug and metabolites than by administered dose. The unbound drug in plasma is thought to be the most relevant indirect measure of tissue concentrations of unbound drug. The AUC is considered the most comprehensive pharmacokinetic endpoint since it takes into account the plasma concentration of the compound and residence time in vivo.

There is, as yet, no validated scientific basis for use of comparative drug plasma concentrations in animals and humans for the assessment of carcinogenic risk to humans. However, for the present, and based on an analysis of a database of carcinogenicity studies performed at the MTD, the selection of a high dose for carcinogenicity studies which represents a 25- to-1 exposure ratio of rodent to human plasma AUC of parent compound and/or metabolites is considered pragmatic (Note 3).

1.4 Criteria for Comparisons of AUC in Animals and Man for use in High Dose Selection
The following criteria are especially applicable for use in determining a pharmacokinetically-defined exposure for high dose selection.

1. Rodent pharmacokinetic data are derived from the strains used for the carcinogenicity studies using the route of compound administration and dose ranges planned for the carcinogenicity study (Notes 4, 5 and 6);

2. Pharmacokinetic data are derived from studies of sufficient duration to take into account potential time-dependent changes in pharmacokinetic parameters which might occur during the dose ranging studies;

3. Documentation is provided on the similarity of metabolism between rodents and humans (Note 7);

4. In assessing exposure, scientific judgement is used to determine whether the AUC comparison is based on data for the parent, parent and metabolite(s), or metabolite(s). The justification for this decision is provided;
5. Interspecies differences in protein binding are taken into consideration when estimating relative exposure (Note 8);

6. Human pharmacokinetic data are derived from studies encompassing the maximum recommended human daily dose (Note 9).

1.5 Saturation of Absorption in High Dose Selection
High dose selection based on saturation of absorption measured by systemic availability of drug-related substances can be considered. The mid and low doses selected for the carcinogenicity study should take into account saturation of metabolic and elimination pathways.

1.6 Pharmacodynamic Endpoints in High Dose Selection
The utility and safety of many pharmaceuticals depend on their pharmacodynamic receptor selectivity. Pharmacodynamic endpoints for high dose selection will be highly compound-specific and can be considered for individual study designs based on scientific merits. The high dose selected should produce a pharmacodynamic response in dosed animals of such magnitude as would preclude further dose escalation. However, the dose should not produce disturbances of physiology or homeostasis which would compromise the validity of the study. Examples include hypotension and inhibition of blood clotting (because of the risk of spontaneous bleeding).

1.7 Maximum Feasible Dose
Currently, the maximum feasible dose by dietary administration is considered to be 5% of diet. International regulatory authorities are re-evaluating this standard. It is believed that the use of pharmacokinetic endpoints (AUC ratio) for dose selection of low toxicity pharmaceuticals, discussed in this guideline, should significantly reduce the number of instances where high doses are selected based on feasibility criteria.

When routes other than dietary administration are appropriate, the high dose can be limited based on considerations including practicality and local tolerance.

1.8 Limit Dose
In determining the high dose for carcinogenicity studies using the approaches outlined in this guideline it is appropriate to limit this dose to 1500 mg/kg/day (Note 10). This limit dose applies where the maximum recommended human dose does not exceed 500 mg/day (Note 11).

Data should be provided comparing exposure of rodents and humans to drug and metabolites primarily to support dose selection for and interpretation of the carcinogenicity study. Based on such information, there might be cases where the limit of 1500 mg/kg/day is not applicable because it cannot be assured that animal exposure after 1500 mg/kg/day is sufficiently high compared to the exposure achieved in humans. The rodent systemic exposure at 1500 mg/kg/day should be greater by at least an order of magnitude than human exposure measured at the intended human therapeutic dose. (If this is not the case, efforts should be made to increase the rodent exposure or to reconsider the animal model in a case-by-case approach.) If the human dose exceeds 500 mg/day the high dose can be increased up to the maximum feasible dose.
1.9 Additional Endpoints in High Dose Selection

It is recognised that there might be merit in the use of alternative endpoints not specifically defined in this guidance on high dose selection for rodent carcinogenicity studies. Use of these additional endpoints in individual study designs should be based on scientific rationale. Such designs should be evaluated based on their individual merits (Note 12).

1.10 Selection of Middle and Low Doses in Carcinogenicity Studies

Regardless of the method used for the selection of the high dose, the selection of the mid and low doses for the carcinogenicity study should provide information to aid in assessing the relevance of study findings to humans. The doses should be selected following integration of rodent and human pharmacokinetic, pharmacodynamic, and toxicity data. The rationale for the selection of these doses should be provided. While not all-encompassing, the following points should be considered in selection of the middle and low doses for rodent carcinogenicity studies:

1. Linearity of pharmacokinetics and saturation of metabolic pathways;
2. Human exposure and therapeutic dose;
3. Pharmacodynamic response in rodents;
4. Alterations in normal rodent physiology;
5. Mechanistic information and potential for threshold effects;
6. The unpredictability of the progression of toxicity observed in short-term studies.

2. SUMMARY

This guidance outlines six criteria for selection of the high dose for carcinogenicity studies of therapeutics: the maximum tolerated dose, 25-fold AUC ratio (rodent:human), dose-limiting pharmacodynamic effects, saturation of absorption, maximum feasible dose, and limit dose. The use of other pharmacodynamic-, pharmacokinetic-, or toxicity-based endpoints in study design should be considered based on scientific rationale and individual merits. In all cases, appropriate dose ranging studies should be conducted. All relevant information should be considered for dose and species/strain selection for the carcinogenicity study. This information should include knowledge of human use, exposure patterns, and metabolism. The availability of multiple criteria for dose selection will provide greater flexibility in optimising the design of carcinogenicity studies for therapeutic agents.

3. NOTES

Note 1

The following are considered equivalent definitions of the toxicity-based endpoint describing the maximum tolerated dose:

The US Interagency Staff Group on Carcinogens has defined the MTD as follows: "The highest dose currently recommended is that which, when given for the duration of the chronic study, is just high enough to elicit signs of minimal toxicity without significantly altering the animal's normal lifespan due to effects other than carcinogenicity. This dose, sometimes called the maximum tolerated dose (MTD), is determined in a subchronic study (usually 90 days duration) primarily on the basis of
mortality, toxicity and pathology criteria. The MTD should not produce morphologic evidence of toxicity of a severity that would interfere with the interpretation of the study. Nor should it comprise so large a fraction of the animal's diet that the nutritional composition of the diet is altered, leading to nutritional imbalance.

"The MTD was initially based on a weight gain decrement observed in the subchronic study; i.e., the highest dose that caused no more than a 10% weight gain decrement. More recent studies and the evaluation of many more bioassays indicate refinement of MTD selection on the basis of a broader range of biological information. Alterations in body and organ weight and clinically significant changes in haematologic, urinary, and clinical chemistry measurements can be useful in conjunction with the usually more definitive toxic, pathologic or histopathologic endpoints." (Environmental Health Perspectives, Vol. 67, pp. 201-281, 1986)

The Ministry of Health and Welfare in Japan prescribes the following: "The dose in the preliminary carcinogenicity study that inhibits body weight gain by less than 10% in comparison with the control and causes neither death due to toxic effects nor remarkable changes in the general signs and laboratory examination findings of the animals is the highest dose to be used in the full-scale carcinogenicity study." (Toxicity test guideline for pharmaceuticals, Chapter 5, pg. 127, 1985)

The Committee on Proprietary Medicinal Products of the European Community prescribes the following: "The top dose should produce a minimum toxic effect, for example a 10% weight loss or failure of growth, or minimal target organ toxicity. Target organ toxicity will be demonstrated by failure of physiological functions and ultimately by pathological changes." (Rules Governing Medicinal Products in the European Community, Vol. III, 1987)

**Note 2**
This does not imply that all possible rodent strains should be surveyed for metabolic profile, but rather that standard strains used in carcinogenicity studies should be examined.

**Note 3**
In order to select a multiple of the human AUC that would serve as an useful endpoint for dose selection for carcinogenicity studies, a retrospective analysis was performed on data from carcinogenicity studies of therapeutics conducted at the MTD for which there was sufficient human and rodent pharmacokinetic data for comparison of AUC values.

In 35 drug carcinogenicity studies carried out at the MTD for which there were adequate pharmacokinetic data in rats and humans, approximately 1/3 had a relative systemic exposure ratio less than or equal to 1, and another 1/3 had ratios between 1 and 10.

An analysis of the correlation between the relative systemic exposure ratio, the relative dose ratio (rat mg/kg:human mg/kg MRD), and the dose ratio adjusted for body surface area (rat mg/M2 MTD:human mg/M2 MRD), performed in conjunction with the above-described database analysis, indicates that the relative systemic exposure corresponds better with dose ratios expressed in terms of body surface area rather than body weight. When 123 compounds in the expanded FDA database were analysed by this approach, a similar distribution of relative systemic exposures was observed. In the selection of a relative systemic exposure ratio (AUC ratio) to apply in
high dose selection, consideration was given to a ratio value that would represent an adequate margin of safety, would detect known or probable human carcinogens, and could be attained by a reasonable proportion of compounds.

To address the issue of detection of known or probable human carcinogenic pharmaceuticals, an analysis of exposure and/or dose ratios was performed on IARC class 1 and 2A pharmaceuticals with positive rat findings. For phenacetin, sufficient rat and human pharmacokinetic data are available to estimate that a relative systemic exposure ratio of at least 15 was found to produce positive findings in a rat carcinogenicity study. For most of 14 IARC 1 and 2A drugs evaluated with positive carcinogenicity findings in rats, there is a lack of adequate pharmacokinetic data for analysis. For these compounds, the body surface area adjusted dose ratio was employed as a surrogate for the relative systemic exposure ratio. The results of this analysis indicated that using doses in the rodent corresponding to body surface area ratios of 10 or more would identify the carcinogenic potential of these pharmaceuticals.

As a result of the evaluations described above, a minimum systemic exposure ratio of 25 can be considered as a useful pharmacokinetic endpoint for high dose selection. This value was attained by approximately 25% of compounds tested in the FDA database (see Note 10), is high enough to detect known or probable (IARC 1, 2A) human carcinogenic drugs, and represents an adequate margin of safety. Those pharmaceuticals tested using a 25 fold or greater AUC ratio for the high dose will have exposure ratios greater than do 75% of pharmaceuticals tested previously in carcinogenicity studies performed at the MTD.

**Note 4**
The rodent AUCs and metabolite profiles can be determined from separate steady-state kinetic studies, as part of the subchronic toxicity studies, or dose-ranging studies.

**Note 5**
AUC values in rodents are usually obtainable using a small number of animals, depending on the route of administration and the availability of data on the pharmacokinetic characteristics of the test compound.

**Note 6**
Equivalent analytical methods of adequate sensitivity and precision should be used to determine plasma concentrations of pharmaceuticals in rodents and humans.

**Note 7**
It is recommended that *in vivo* metabolism be characterised in humans and rodents, if possible. However, in the absence of appropriate *in vivo* metabolism data, *in vitro* metabolism data (e.g., from liver slices, uninduced microsomal preparations) can provide appropriate support for the similarity of metabolism across species.

**Note 8**
While *in vivo* determinations of unbound drug might be the best approach, *in vitro* determinations of protein binding using parent and/or metabolites as appropriate (over the range of concentrations achieved *in vivo* in rodents and humans) might be used in the estimation of AUC unbound. When protein binding is low in both humans
and rodents, or when protein binding is high and the unbound fraction of drug is greater in rodents than in humans, the comparison of total plasma concentration of drug is appropriate. When protein binding is high and the unbound fraction is greater in humans than in rodents, the ratio of the unbound concentrations should be used.

**Note 9**

Human systemic exposure data can be derived from pharmacokinetic monitoring in normal volunteers and/or patients. The possibility of extensive inter-individual variation in exposure should be taken into consideration. In the absence of knowledge of the maximum recommended human daily dose, at a minimum, doses producing the desired pharmacodynamic effect in humans should be used to derive the pharmacokinetic data.

**Note 10**

Review of the FDA carcinogenicity database of nearly 900 carcinogenicity tests indicated that about 20 tests had been conducted that used doses of 1000 mg/kg or greater as the highest dose tested. About 10 of these tests were considered as having demonstrated a carcinogenic response. Seven of these were positive only at or above 1000 mg/kg. Regulatory action has resulted from some of these cases. Based on these results, the limit dose for carcinogenicity testing should be 1500 mg/kg rather than 1000 mg/kg to eliminate the risk that a carcinogen will not be able to be identified as a result of adoption of a limit dose of 1000 mg/kg.

**Note 11**

It has been agreed that if a drug is only positive in rodents at doses above those producing a 25-fold exposure over exposure in humans, such a finding would not be considered likely to reflect a relevant risk to humans.

It has been shown that systemic exposure comparisons between rodents and humans are better estimated by a dose using mg/m² than using mg/kg (see Note 3 above). Therefore, the human dose should be at least 25-fold lower on a mg/m² basis than the high dose in the carcinogenicity study. The factor 6-7 (6.5) is used to convert rat doses from mg/kg to mg/m² and the factor 40 is used to convert human doses from mg/kg to mg/m². Thus, the estimated systemic exposure ratio of 25-fold rodent/human is equal to about a 25-fold mg/m² ratio or a 150-fold mg/kg ratio (150 ≈ 25 x 40/6.5). Therefore a human dose below 10 mg/kg/day (about 500 mg/day or less) could be tested in rats at 1500 mg/kg as the high dose.

**Note 12**

Additional pharmaceutical-specific endpoints to select an appropriate high dose are currently under discussion (e.g., additional pharmacodynamic, pharmacokinetic and toxicity endpoints as well as alternatives to a maximum feasible dose).
This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
S2(R1)
Document History

<table>
<thead>
<tr>
<th>Code</th>
<th>History</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2A</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>10 March 1994</td>
</tr>
<tr>
<td>S2A</td>
<td>Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>19 July 1995</td>
</tr>
</tbody>
</table>

S2B:
**Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals**

| S2B   | Approval by the Steering Committee under Step 2 and release for public consultation. | 2 October 1996 |
| S2B   | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 16 July 1997 |

S2(R1):
**Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals**

Revision of the S2A and S2B Guidelines which have been merged as part of the revision

| S2(R1) | Approval by the Steering Committee of S2(R1) under Step 2 and release for public consultation. | 6 March 2008 |

Current Step 4 version

| S2(R1) | Approval by the Steering Committee of S2(R1) under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 9 November 2011 |
GUIDANCE ON GENOTOXICITY TESTING AND DATA INTERPRETATION FOR PHARMACEUTICALS INTENDED FOR HUMAN USE

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 9 November 2011, this Guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. INTRODUCTION ............................................................................................................. 1
1.1 OBJECTIVES OF THE GUIDELINE ........................................................................... 1
1.2. Background .....................................................................................................................1
1.3. Scope of the Guideline ................................................................................................. 1
1.4 General Principles ...................................................................................................... 1
2 THE STANDARD TEST BATTERY FOR GENOTOXICITY .................................... 2
2.1 Rationale .........................................................................................................................2
2.2 Description of the Two Options for the Standard Battery ...........................................2
2.3 Modifications to the Test Battery ..................................................................................4
  2.3.1 Exploratory Clinical Studies ..................................................................................4
  2.3.2 Testing Compounds that are Toxic to Bacteria ......................................................4
  2.3.3 Compounds Bearing Structural Alerts for Genotoxic Activity ..............................4
  2.3.4 Limitations to the Use of In Vivo Tests ..................................................................4
2.4 Detection of Germ Cell Mutagens ..............................................................................4
3. RECOMMENDATIONS FOR IN VITRO TESTS .................................................. 5
3.1 Test Repetition and Interpretation ..............................................................................5
3.2 Recommended Protocol for the Bacterial Mutation Assay ........................................5
   3.2.1 Selection of Top Dose Level ................................................................................5
   3.2.2 Study Design/Test Protocol ...............................................................................5
3.3 Recommended Protocols for the Mammalian Cell Assays ...........................................6
   3.3.1 Selection of Top Concentration ...........................................................................6
   3.3.2 Study Design/Test Protocols .............................................................................6
   3.3.3 Positive Controls .................................................................................................7
4. RECOMMENDATIONS FOR IN VIVO TESTS .................................................. 7
4.1 Tests for the Detection of Chromosome Damage In Vivo ..........................................7
4.2 Other In Vivo Genotoxicity Tests ..............................................................................7
4.3 Dose Selection for In Vivo Assays .............................................................................8
1. INTRODUCTION

1.1 Objectives of the Guideline

This guidance replaces and combines the ICH S2A and S2B Guidelines. The purpose of the revision is to optimize the standard genetic toxicology battery for prediction of potential human risks, and to provide guidance on interpretation of results, with the ultimate goal of improving risk characterization for carcinogenic effects that have their basis in changes in the genetic material. The revised guidance describes internationally agreed upon standards for follow-up testing and interpretation of positive results \textit{in vitro} and \textit{in vivo} in the standard genetic toxicology battery, including assessment of non-relevant findings. This guidance is intended to apply only to products being developed as human pharmaceuticals.

1.2. Background

The recommendations from the latest Organization for Economic Co-operation and Development (OECD) guidelines and the reports from the International Workshops on Genotoxicity Testing (IWGT) have been considered where relevant. In certain cases, there are differences from the OECD or IWGT recommendations, which are noted in the text. The following notes for guidance should be applied in conjunction with other ICH guidances.

1.3. Scope of the Guideline

The focus of this guidance is testing of new “small molecule” drug substances, and the guidance does not apply to biologics. Advice on the timing of the studies relative to clinical development is provided in the ICH M3 (R2) guidance.

1.4 General Principles

Genotoxicity tests can be defined as \textit{in vitro} and \textit{in vivo} tests designed to detect compounds that induce genetic damage by various mechanisms. These tests enable hazard identification with respect to damage to DNA and its fixation. Fixation of damage to DNA in the form of gene mutations, larger scale chromosomal damage or recombination is generally considered to be essential for heritable effects and in the multi-step process of malignancy, a complex process in which genetic changes might possibly play only a part. Numerical chromosome changes have also been associated with tumorigenesis and can indicate a potential for aneuploidy in germ cells. Compounds that are positive in tests that detect such kinds of damage have the potential to be human carcinogens and/or mutagens. Because the relationship between exposure to particular chemicals and carcinogenesis is established for humans, whilst a similar relationship has been difficult to prove for heritable diseases, genotoxicity tests have been used mainly for the prediction of carcinogenicity. Nevertheless, because germ line mutations are clearly associated with human disease, the suspicion that a compound might induce heritable effects is considered to be just as serious as the suspicion that a compound might induce cancer. In addition, the outcome of genotoxicity tests can be valuable for the interpretation of carcinogenicity studies.
2 THE STANDARD TEST BATTERY FOR GENOTOXICITY

2.1 Rationale

Registration of pharmaceuticals requires a comprehensive assessment of their genotoxic potential. Extensive reviews have shown that many compounds that are mutagenic in the bacterial reverse mutation (Ames) test are rodent carcinogens. Addition of \textit{in vitro} mammalian tests increases sensitivity for detection of rodent carcinogens and broadens the spectrum of genetic events detected, but also decreases the specificity of prediction; i.e., increases the incidence of positive results that do not correlate with rodent carcinogenicity. Nevertheless, a battery approach is still reasonable because no single test is capable of detecting all genotoxic mechanisms relevant in tumorigenesis.

The general features of a standard test battery are as follows:

i. Assessment of mutagenicity in a bacterial reverse gene mutation test. This test has been shown to detect relevant genetic changes and the majority of genotoxic rodent and human carcinogens.

ii. Genotoxicity should also be evaluated in mammalian cells \textit{in vitro} and/or \textit{in vivo} as follows.

Several \textit{in vitro} mammalian cell systems are widely used and can be considered sufficiently validated: The \textit{in vitro} metaphase chromosome aberration assay, the \textit{in vitro} micronucleus assay (Note 1) and the mouse lymphoma L5178Y cell Tk (thymidine kinase) gene mutation assay (MLA). These three assays are currently considered equally appropriate and therefore interchangeable for measurement of chromosomal damage when used together with other genotoxicity tests in a standard battery for testing of pharmaceuticals, if the test protocols recommended in this Guideline are used.

\textit{In vivo} test(s) are included in the test battery because some agents are mutagenic \textit{in vivo} but not \textit{in vitro} (Note 2) and because it is desirable to include assays that account for such factors as absorption, distribution, metabolism and excretion. The choice of an analysis either of micronuclei in erythrocytes (in blood or bone marrow), or of chromosome aberrations in metaphase cells in bone marrow, is currently included for this reason (Note 3). Lymphocytes cultured from treated animals can also be used for cytogenetic analysis, although experience with such analyses is less widespread.

\textit{In vitro} and \textit{in vivo} tests that measure chromosomal aberrations in metaphase cells can detect a wide spectrum of changes in chromosomal integrity. Breakage of chromatids or chromosomes can result in micronucleus formation if anacentric fragment is produced; therefore assays that detect either chromosomal aberrations or micronuclei are considered appropriate for detecting clastogens. Micronuclei can also result from lagging of one or more whole chromosome(s) at anaphase and thus micronucleus tests have the potential to detect some aneuploidy inducers. The MLA detects mutations in the \textit{Tk} gene that result from both gene mutations and chromosome damage. There is some evidence that MLA can also detect chromosome loss.

There are several additional \textit{in vivo} assays that can be used in the battery or as follow-up tests to develop weight of evidence in assessing results of \textit{in vitro} or \textit{in vivo} assays (see below). Negative results in appropriate \textit{in vivo} assays (usually two), with adequate justification for the endpoints measured, and demonstration of exposure (see Section 4.4) are generally considered sufficient to demonstrate absence of significant genotoxic risk.

2.2 Description of the Two Options for the Standard Battery

The following two options for the standard battery are considered equally suitable (Note 4):
Option 1
i. A test for gene mutation in bacteria.

ii. A cytogenetic test for chromosomal damage (the in vitro metaphase chromosome aberration test or in vitro micronucleus test), or an in vitro mouse lymphoma Tk gene mutation assay.

iii. An in vivo test for genotoxicity, generally a test for chromosomal damage using rodent hematopoietic cells, either for micronuclei or for chromosomal aberrations in metaphase cells.

Option 2
i. A test for gene mutation in bacteria.

ii. An in vivo assessment of genotoxicity with two different tissues, usually an assay for micronuclei using rodent hematopoietic cells and a second in vivo assay. Typically this would be a DNA strand breakage assay in liver, unless otherwise justified (see below; also Section 4.2 and Note 12).

There is more historical experience with Option 1, partly because it is based on S2A and B. Nevertheless, the reasoning behind considering Options 1 and 2 equally acceptable is as follows: When a positive result occurs in an in vitro mammalian cell assay, clearly negative results in two well conducted in vivo assays, in appropriate tissues and with demonstrated adequate exposure, are considered sufficient evidence for lack of genotoxic potential in vivo (see Section 5.4.1.1 below). Thus a test strategy in which two in vivo assays are conducted is the same strategy that would be used to follow up a positive result in vitro (Note 4).

Under both standard battery options, either acute or repeat dose study designs in vivo can be used. In case of repeated administrations, attempts should be made to incorporate the genotoxicity endpoints into toxicity studies, if scientifically justified. When more than one endpoint is evaluated in vivo it is preferable that they are incorporated into a single study. Often sufficient information on the likely suitability of the doses for the repeat-dose toxicology study is available before the study begins and can be used to determine whether an acute or an integrated test will be suitable.

For compounds that give negative results, the completion of either option of the standard test battery, performed and evaluated in accordance with current recommendations, will usually provide sufficient assurance of the absence of genotoxic activity and no additional tests are warranted. Compounds that give positive results in the standard test battery might, depending on their therapeutic use, need to be tested more extensively (see Section 5).

There are several in vivo assays that can be used as the second part of the in vivo assessment under Option 2 (see Section 4.2), some of which can be integrated into repeat-dose toxicology studies. The liver is typically the preferred tissue because of exposure and metabolizing capacity, but choice of in vivo tissue and assay should be based on factors such as any knowledge of the potential mechanism, of the metabolism in vivo, or of the exposed tissues thought to be relevant.

Information on numerical changes can be derived from the mammalian cell assays in vitro and from the micronucleus assays in vitro or in vivo. Elements of the standard protocols that can indicate such potential are elevations in the mitotic index, polyploidy induction and micronucleus evaluation. There is also experimental evidence that spindle poisons can be detected in MLA. The preferred in vivo cytogenetic test under Option 2 is the micronucleus assay, not a chromosome aberration assay, to include more
direct capability for detection of chromosome loss (potential for aneuploidy).

The suggested standard set of tests does not imply that other genotoxicity tests are generally considered inadequate or inappropriate. Additional tests can be used for further investigation of genotoxicity test results obtained in the standard battery (see Sections 4.2 and 5). Alternative species, including non-rodents, can also be used if indicated, and if sufficiently validated.

Under conditions in which one or more tests in the standard battery cannot be employed for technical reasons, alternative validated tests can serve as substitutes provided sufficient scientific justification is given.

2.3 Modifications to the Test Battery

The following sections describe situations where modification of the standard test battery might be advisable.

2.3.1 Exploratory Clinical Studies

For certain exploratory clinical studies, fewer genotoxicity assays or different criteria for justification of the maximum dose in vivo might apply (see ICH M3(R2) guidance).

2.3.2 Testing Compounds that are Toxic to Bacteria

In cases where compounds are highly toxic to bacteria (e.g., some antibiotics), the bacterial reverse mutation (Ames) test should still be carried out, just as cytotoxic compounds are tested in mammalian cells, because mutagenicity can occur at lower, less toxic concentrations. In such cases, any one of the in vitro mammalian cell assays should also be done, i.e., Option 1 should be followed.

2.3.3 Compounds Bearing Structural Alerts for Genotoxic Activity

Structurally alerting compounds (Note 5) are usually detectable in the standard test battery since the majority of “structural alerts” are defined in relation to bacterial mutagenicity. A few chemical classes are known to be more easily detected in mammalian cell chromosome damage assays than bacterial mutation assays. Thus negative results in either test battery with a compound that has a structural alert is usually considered sufficient assurance of a lack of genotoxicity. However, for compounds bearing certain specific structural alerts, modification to standard protocols can be appropriate (Note 5). The choice of additional test(s) or protocol modification(s) depends on the chemical nature, the known reactivity and any metabolism data on the structurally alerting compound in question.

2.3.4 Limitations to the Use of In Vivo Tests

There are compounds for which many in vivo tests (typically in bone marrow, blood or liver) do not provide additional useful information. These include compounds for which data on toxicokinetics or pharmacokinetics indicate that they are not systemically absorbed and therefore are not available to the target tissues. Examples of such compounds are some radioimaging agents, aluminum based antacids, some compounds given by inhalation, and some dermally or other topically applied pharmaceuticals. In cases where a modification of the route of administration does not provide sufficient target tissue exposure, and no suitable genotoxicity assay is available in the most exposed tissue, it might be appropriate to base the evaluation only on in vitro testing. In some cases evaluation of genotoxic effects at the site of contact can be warranted, although such assays have not yet been widely used (Note 6).

2.4 Detection of Germ Cell Mutagens

Results of comparative studies have shown that, in a qualitative sense, most germ cell
mutagens are likely to be detected as genotoxic in somatic cell tests so that negative results of in vivo somatic cell genotoxicity tests generally indicate the absence of germ cell effects.

3. **RECOMMENDATIONS FOR IN VITRO TESTS**

3.1 **Test Repetition and Interpretation**

Reproducibility of experimental results is an essential component of research involving novel methods or unexpected findings; however, the routine testing of drugs with standard, widely used genotoxicity tests often does not call for replication. These tests are sufficiently well characterized and have sufficient internal controls that repetition of a clearly positive or negative assay is not usually warranted. Ideally it should be possible to declare test results clearly negative or clearly positive. However, test results sometimes do not fit the predetermined criteria for a positive or negative call and therefore are declared “equivocal”. The application of statistical methods can aid in data interpretation; however, adequate biological interpretation is of critical importance. An equivocal test that is repeated might result in (i) a clearly positive outcome, and thus an overall positive result; (ii) a negative outcome, so that the result is not reproducible and overall negative, or (iii) another equivocal result, with a final conclusion that remains equivocal.

3.2 **Recommended Protocol for the Bacterial Mutation Assay**

Advice on the protocols is given in the OECD guideline (1997) and the IWGT report (Gatehouse et al., 1994).

3.2.1 **Selection of Top Dose Level**

**Maximum dose level**

The maximum dose level recommended is 5000 µg/plate (or 5 µL/plate for liquid test substance) when not limited by solubility or cytotoxicity.

**Limit of solubility**

For bacterial cultures, precipitating doses are scored provided precipitate does not interfere with scoring, toxicity is not limiting, and the top concentration does not exceed 5000 µg/plate (or 5 µL/plate for liquid test substance). If no cytotoxicity is observed, then the lowest precipitating dose should be used as the top dose scored. If dose related cytotoxicity or mutagenicity is noted, irrespective of solubility, the top dose scored should be based on cytotoxicity as described below.

**Limit of cytotoxicity**

In the Ames test, the doses scored should show evidence of significant toxicity, but without exceeding a top dose of 5000 µg/plate. Toxicity might be detected by a reduction in the number of revertants, and/or clearing or diminution of the background lawn.

3.2.2 **Study Design/Test Protocol**

The recommended set of bacterial strains (OECD) includes those that detect base substitution and frameshift mutations as follows:

- *Salmonella typhimurium* TA98;
- *Salmonella typhimurium* TA100;
- *Salmonella typhimurium* TA1535;
- either *Salmonella typhimurium* TA1537 or TA97 or TA97a;
and either \textit{Salmonella typhimurium} TA102 or \textit{Escherichia coli} WP2 \textit{uvrA} or \textit{Escherichia coli} WP2 \textit{uvrA} (pKM101).

One difference from the OECD and IWGT recommendations is that, based on experience with testing pharmaceuticals, a single bacterial mutation (Ames) test is considered sufficient when it is clearly negative or positive, and carried out with a fully adequate protocol including all strains with and without metabolic activation, a suitable dose range that fulfills criteria for top dose selection, and appropriate positive and negative controls. Also, for testing pharmaceuticals, either the plate incorporation or the pre-incubation method is considered appropriate for this single experiment (Note 7). Equivocal or weak positive results might indicate that it would be appropriate to repeat the test, possibly with a modified protocol such as appropriate spacing of dose levels.

3.3 \textbf{Recommended Protocols for the Mammalian Cell Assays}

Advice on the protocols is given in the OECD guidelines (1997) and the IWGT publications (e.g., Kirsch-Volders et al., 2003; Moore et al., 2006). Advice on interpretation of MLA results is also given (Moore et al., 2006), including use of a global evaluation factor. Several differences from these recommendations are noted here for testing pharmaceuticals, notably for selection of the top concentration. (See details below.)

3.3.1 \textbf{Selection of Top Concentration}

\textbf{Maximum concentration}

The maximum top concentration recommended is 1 mM or 0.5 mg/ml, whichever is lower, when not limited by solubility in solvent or culture medium or by cytotoxicity (Note 8).

\textbf{Limit of solubility}

When solubility is limiting, the maximum concentration, if not limited by cytotoxicity, should be the lowest concentration at which minimal precipitate is visible in cultures, provided there is no interference with scoring. Evaluation of precipitation can be done by naked eye or by methods such as light microscopy, noting precipitate that persists or appears during culture (by the end of treatment).

\textbf{Cytotoxicity}

For \textit{in vitro} cytogenetic assays for metaphase chromosome aberrations or for micronuclei, cytotoxicity should not exceed a reduction of about 50\% in cell growth (Notes 9 and 10). For the MLA, at the top dose there should be 80-90\% cytotoxicity as measured by an RTG between 20-10\% (Note 9).

3.3.2 \textbf{Study Design/Test Protocols}

For the cytogenetic evaluation of chromosomal damage in metaphase cells \textit{in vitro}, the test protocol should include the conduct of tests with and without metabolic activation, with appropriate positive and negative controls. Treatment with the test articles should be for 3 to 6 hours with a sampling time approximately 1.5 normal cell cycles from the beginning of the treatment. A continuous treatment without metabolic activation up to the sampling time of approximately 1.5 normal cell cycles should be conducted in case of negative or equivocal results for both short treatments, with and without metabolic activation. The same principles apply to the \textit{in vitro} micronucleus assay, except that the sampling time is typically 1.5 to 2 normal cell cycles from the beginning of treatment to allow cells to complete mitosis and enter the next interphase. For both \textit{in vitro} cytogenetic assays, there might be a need to modify the protocol for certain types of chemicals that could be more readily detected by longer treatment, delayed sampling times or recovery periods, e.g., some nucleoside analogues and some
Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use

nitrosamines. In the metaphase aberration assay, information on the ploidy status should be obtained by recording the incidence of polyploid (including endoreduplicated) metaphases as a percentage of the number of metaphase cells. For MLA, the test protocol should include the conduct of tests with and without metabolic activation, with appropriate positive and negative controls, where the treatment with the test article is for 3 to 4 hours. A continuous treatment without metabolic activation for approximately 24 hours should be conducted in case of a negative or equivocal result for both short treatments, with and without metabolic activation. A standard MLA should include (i) the incorporation of positive controls that induce mainly small colonies, and (ii) colony sizing for positive controls, solvent controls and at least one positive test compound concentration (should any exist), including the culture that gave the greatest mutant frequency.

For mammalian cell assays in vitro, built-in confirmatory elements, such as those outlined above (e.g., different treatment lengths, tests with and without metabolic activation), should be used. Following such testing, further confirmatory testing in the case of clearly negative or positive test results is not usually warranted. Equivocal or weak positive results might call for repeating tests, possibly with a modified protocol such as appropriate spacing of the test concentrations.

3.3.3 Positive Controls
Concurrent positive controls are important, but in vitro mammalian cell tests for genetic toxicity are sufficiently standardized that use of positive controls can generally be confined to a positive control with metabolic activation (when it is done concurrently with the non-activated test) to demonstrate the activity of the metabolic activation system and the responsiveness of the test system.

4. RECOMMENDATIONS FOR IN VIVO TESTS

4.1 Tests for the Detection of Chromosome Damage In Vivo
Either the analysis of chromosomal aberrations or the measurement of micronucleated polychromatic erythrocytes in bone marrow cells in vivo is considered appropriate for the detection of clastogens. Both rats and mice are considered appropriate for use in the bone marrow micronucleus test. Micronuclei can also be measured in immature (e.g., polychromatic) erythrocytes in peripheral blood in the mouse, or in the newly formed reticulocytes in rat blood (Note 3). Likewise, immature erythrocytes can be used from any other species which has shown an adequate sensitivity to detect clastogens/aneuploidy inducers in bone marrow or peripheral blood (Note 3). Systems for automated analysis (image analysis and flow cytometry) can be used if appropriately validated (OECD, 1997; Hayashi et al., 2000; 2007). Chromosomal aberrations can also be analyzed in peripheral lymphocytes cultured from treated rodents (Note 11).

4.2 Other In Vivo Genotoxicity Tests
The same in vivo tests described as the second test in the standard battery (Option 2) can be used as follow-up tests to develop weight of evidence in assessing results of in vitro or in vivo assays (Notes 11 and 12). While the type of effect seen in vitro and any knowledge of the mechanism can help guide the choice of in vivo assay, investigation of chromosomal aberrations or of gene mutations in endogenous genes is not feasible with standard methods in most tissues. Although mutation can be measured in transgenes in rodents, this entails prolonged treatment (e.g., 28 days) to allow for mutation expression, fixation and accumulation, especially in tissues with little cell division (Note 12). Thus the second in vivo assay will often evaluate a DNA damage endpoint as a surrogate. Assays with the most published experience and advice on protocols include
the DNA strand break assays such as the single cell gel electrophoresis (“Comet”) assay and alkaline elution assay, the \textit{in vivo} transgenic mouse mutation assays and DNA covalent binding assays, (all of which may be applied in many tissues, Note 12), and the liver unscheduled DNA synthesis (UDS) assay.

4.3 \textbf{Dose Selection for In Vivo Assays}

Typically three dose levels are analyzed (Hayashi et al., 2005).

\subsection*{4.3.1 Short-Term Studies}

For short-term (usually 1 to 3 administrations) studies, the top dose recommended for genotoxicity assays is a limit dose of 2000 mg/kg, if this is tolerated, or a maximum tolerated dose defined (for example for the micronucleus assay (OECD)) as the dose producing signs of toxicity such that higher dose levels, based on the same dosing regimen, would be expected to produce lethality. Similar recommendations have been made for the Comet assay (Hartmann et al., 2003) and transgenic mutation assay (Heddle et al., 2000). Suppression of bone marrow red blood cell production should also be taken into account in dose selection. Lower doses are generally spaced at approximately two to three fold intervals below this.

\subsection*{4.3.2 Multiple Administration Studies}

\textbf{Option 1 Battery}

When the \textit{in vivo} genotoxicity test is integrated into a multiple administration toxicology study, the doses are generally considered appropriate when the toxicology study meets the criteria for an adequate study to support human clinical trials; this can differ from dose selection criteria in the OECD guideline for the \textit{in vivo} micronucleus assay. This applies when the \textit{in vitro} mammalian cell test is negative (or “non-relevant positive”; see Section 5).

\textbf{Follow-up studies or Option 2 Battery}

When carrying out follow-up studies to address any indication of genotoxicity, or when using Option 2 with no \textit{in vitro} mammalian cell assay, several factors should be evaluated to determine whether the top dose is appropriate for genotoxicity evaluation. Any one of the criteria listed below is considered sufficient to demonstrate that the top dose in a toxicology study (typically in rats) is appropriate for micronucleus analysis and for other genotoxicity evaluation:

\begin{enumerate}
\item Maximum Feasible Dose (MFD) based on physico-chemical properties of the drug in the vehicle (provided the MFD in that vehicle is similar to that achievable with acute administration; Note 13).
\item Limit dose of 1000 mg/kg for studies of 14 days or longer, if this is tolerated.
\item Maximal possible exposure demonstrated either by reaching a plateau/saturation in exposure or by compound accumulation. In contrast, substantial reduction in exposure to parent drug with time (e.g., \text{≥} 50\% reduction from initial exposure) can disqualify the study (unless a blood sample taken in the first few days is available). If this is seen in one sex, generally the sex with reduced exposure would not be scored at the end of the study, unless there is enhanced exposure to a metabolite of interest.
\item Top dose is \text{≥} 50\% of the top dose that would be used for acute administration, i.e., close to the minimum lethal dose, if such acute data are available for other reasons. (The top dose for acute administration micronucleus tests is currently described in
OECD guidance as the dose above which lethality would be expected; similar guidance is given (e.g., Hartmann et al., 2003) for other in vivo assays.)

Selection of a top dose based only on an exposure margin (multiple over clinical exposure) without toxicity is not considered sufficient justification.

4.3.3 Testing Compounds that are Toxic for Blood or Bone Marrow

Many compounds that induce aneuploidy, such as potent spindle poisons, are detectable in in vivo micronucleus assays in bone marrow or blood only within a narrow range of doses approaching toxic doses. This is also true for some clastogens. If toxicological data indicate severe toxicity to the red blood cell lineage (e.g., marked suppression of Polychromatic Erythrocytes (PCEs) or reticulocytes), doses scored should be spaced not more than about 2 fold below the top, cytotoxic dose. If suitable doses are not included in a multi-week study, additional data that could contribute to the detection of aneugens and some toxic clastogens could be derived from any one of the following:

i. Early blood sampling (at 3-4 days) is advisable when there are marked increases in toxicity with increasing treatment time. For example, when blood or bone marrow is used for micronucleus measurement in a multiweek study (e.g., 28 days), and reticulocytes are scored, marked hematotoxicity can affect the ability to detect micronuclei; i.e., a dose that induces detectable increases in micronuclei after acute treatment might be too toxic to analyze after multiple treatments (Hamada et al., 2001). The early sample can be used to provide assurance that clastogens and potential aneugens are detected (but see Notes 14 and 15).

ii. An in vitro mammalian cell micronucleus assay.

iii. An acute bone marrow micronucleus assay.

4.4 Demonstration of Target Tissue Exposure for Negative In Vivo Test Results

In vivo tests have an important role in genotoxicity test strategies. The value of in vivo results is directly related to the demonstration of adequate exposure of the target tissue to the test compound. This is especially true for negative in vivo test results when in vitro test(s) have shown convincing evidence of genotoxicity, or when no in vitro mammalian cell assay is used. Evidence of adequate exposure could include toxicity in the tissue in question, or toxicokinetic data as described in the following section.

4.4.1 When an In Vitro Genotoxicity Test is Positive (or not done)

Assessments of in vivo exposure should be made at the top dose or other relevant doses using the same species, strain and dosing route used in the genotoxicity assay. When genotoxicity is measured in toxicology assays, exposure information is generally available as part of the toxicology assessment.

Demonstration of in vivo exposure should be made by any of the following measurements:

i. Cytotoxicity:
   a. For cytogenetic assays: By obtaining a significant change in the proportion of immature erythrocytes among total erythrocytes in the tissue used (bone marrow or blood) at the doses and sampling times used in the micronucleus test or by measuring a significant reduction in mitotic index for the chromosomal aberration assay.
   b. For other in vivo genotoxicity assays: Toxicity in the liver or tissue being assessed, e.g., by histopathological evaluation or blood biochemistry toxicity
indicators.

ii. Exposure:

a. Measurement of drug related material either in blood or plasma. The bone marrow is a well perfused tissue and levels of drug related materials in blood or plasma are generally similar to those observed in bone marrow. The liver is expected to be exposed for drugs with systemic exposure regardless of the route of administration.

b. Direct measurement of drug-related material in target tissue, or autoradiographic assessment of tissue exposure.

If systemic exposure is similar to or lower than expected clinical exposure, alternative strategies might be called for such as:

i. Use of a different route of administration;

ii. Use of a different species with higher exposure;

iii. Use of a different tissue or assay (see Section 2.3.4, “Limitations to the use of standard in vivo tests”).

When adequate exposure cannot be achieved (e.g., with compounds showing very poor target tissue availability) conventional in vivo genotoxicity tests have little value.

4.4.2 When In Vitro Genotoxicity Tests are Negative

If in vitro tests do not show genotoxic potential, in vivo (systemic) exposure can be assessed by any of the methods above, or can be assumed from the results of standard Absorption, Distribution, Metabolism and Excretion (ADME) studies in rodents done for other purposes.

4.5 Sampling Times for In Vivo Assays

Selection of the sampling time in the in vivo Micronucleus (MN), chromosomal aberration and UDS test should follow OECD (1997).

When micronucleus analysis is integrated into multi-week studies, sampling of blood or bone marrow can be done the day after the final administration (see recommendation for additional blood sampling time above).

For other genotoxicity assays, sampling time should be selected as appropriate for the endpoint measured; for example, DNA damage/strand break measurements are usually made a few (e.g., 2-6) hours after the last administration for the multiple daily administration. In the case of single administration, two sampling times should be used: a few hours and 24 hours after the treatment.

In principle, studies of any length can be considered appropriate, provided the top dose/exposure is adequate.

4.6 Number of Animals Analyzed

The number of animals analyzed is determined by current recommendations for the micronucleus assay (OECD) or other genotoxicity assays and generally does not include all the animals treated for a toxicology study. Animals used for genotoxicity analyses should be randomly selected from the group used for the toxicology study.

4.7 Use of Male/Female Rodents in In Vivo Genotoxicity Tests

If sex-specific drugs are to be tested, then the assay can be done in the appropriate sex. In vivo tests with the acute protocol can generally be carried out in only one sex. For
acute tests, both sexes should be considered only if any existing toxicity, metabolism or exposure (Cmax or AUC) data indicate a toxicologically meaningful sex difference in the species being used. Otherwise, the use of males alone is considered appropriate for acute genotoxicity tests. When the genotoxicity test is integrated into a repeat-dose toxicology study in two sexes, samples can be collected from both sexes, but a single sex can be scored if there is no substantial sex difference evident in toxicity/metabolism. The dose levels for the sex(es) scored should meet the criteria for appropriate dose levels (see Sections 4.3.2 and 4.3.3).

Similar principles can be applied for other established in vivo genotoxicity tests.

4.8 Route of Administration

The route of administration is generally the expected clinical route, e.g., oral, intravenous or subcutaneous, but can be modified if appropriate in order to obtain systemic exposure, e.g., for topically applied compounds (see Section 2.3.4).

4.9 Use of Positive Controls for In Vivo Studies

For in vivo studies, it is considered sufficient to treat animals with a positive control only periodically, and not concurrently with every assay, after a laboratory has established competence in the use of the assay (Note 16).

5. GUIDANCE ON EVALUATION OF TEST RESULTS AND ON FOLLOW-UP TEST STRATEGIES

Comparative trials have shown conclusively that each in vitro test system generates both false negative and false positive results in relation to predicting rodent carcinogenicity. Genotoxicity test batteries (of in vitro and in vivo tests) detect carcinogens that are thought to act primarily via a mechanism involving direct genetic damage, such as the majority of known human carcinogens. Therefore, these batteries are not expected to detect non-genotoxic carcinogens. Experimental conditions, such as the limited capability of the in vitro metabolic activation systems, can lead to false negative results in in vitro tests. The test battery approach is designed to reduce the risk of false negative results for compounds with genotoxic potential. On the other hand a positive result in any assay for genotoxicity does not always mean that the test compound poses a genotoxic/carcinogenic hazard to humans.

Although positive in vitro data could indicate intrinsic genotoxic properties of a drug, appropriate in vivo data determine the biological significance of these in vitro signals in most cases. Also, because there are several indirect mechanisms of genotoxicity that operate only above certain concentrations, it is possible to establish a safe level (threshold) for classes of drugs with evidence for such mechanisms (see 5.2. below, Müller and Kasper, 2000; Scott et al., 1991; Thybaud et al., 2007).

5.1 Assessment of Biological Relevance

The recommendations below assume that the test has been conducted using appropriate spacing of doses, levels of toxicity etc.

Small increases in apparent genotoxicity in vitro or in vivo should first be assessed for reproducibility and biological significance. Examples of results that are not considered biologically meaningful include:

i. Small increases that are statistically significant compared with the negative or solvent control values but are within the confidence intervals of the appropriate historical control values for the testing facility.
ii. Weak/equivocal responses that are not reproducible.

If either of the above conditions applies, the weight of evidence indicates a lack of genotoxic potential, the test is considered negative or the findings not biologically relevant, and no further testing is called for.

## 5.2 Evaluation of Results Obtained in In Vitro Tests

In evaluating positive results, especially for the microbial mutagenicity test, the purity of the test compound should be considered, to determine whether the positive result could be attributable to a contaminant.

### 5.2.1 Evaluation of Positive Results Obtained In Vitro in a Bacterial Mutation Assay

Since positive results in the Ames test are thought to indicate DNA reactivity, extensive follow-up testing to assess the in vivo mutagenic and carcinogenic potential would be warranted to assess the potential risk for treatment of patients, unless justified by appropriate risk-benefit analysis.

There are some well characterized examples of artifactual increases in colonies that are not truly revertants. These can occur due to contamination with amino acids (i.e., providing histidine for *Salmonella typhimurium* strains or tryptophan for *Escherichia coli* strains), so that the bacterial reversion assay is not suitable for testing a peptide that is likely to degrade. Certain cases exist where positive results in bacterial mutation assays might be shown not to indicate genotoxic potential in vivo in humans, for example when bacterial-specific metabolism occurs, such as activation by bacterial nitroreductases.

### 5.2.2 Evaluation of Positive Results Obtained In Vitro in Mammalian Cell Assays

Recommendations for assessing weight of evidence and follow-up testing for positive genotoxicity results are discussed in IWGT reports (e.g., Thybaud et al., 2007). In addition, the scientific literature gives a number of conditions that can lead to a positive in vitro result of questionable relevance. Therefore, any in vitro positive test result should be evaluated based on an assessment of the weight of evidence as indicated below. This list is not exhaustive, but is given as an aid to decision-making.

i. The conditions do not occur in vivo (pH; osmolality; precipitates).

(Note that the 1 mM limit avoids increases in osmolality, and that if the test compound alters pH it is advisable to adjust pH to the normal pH of untreated cultures at the time of treatment).

ii. The effect occurs only at the most toxic concentrations.

For in vitro cytogenetics assays when growth is suppressed by ≥50%.

If any of the above conditions apply the weight of evidence indicates a lack of genotoxic potential; the standard battery (Option 1) can be followed. Thus, a single in vivo test is considered sufficient.

### 5.2.3 Evaluation of In Vitro Negative Results

For in vitro negative results further testing should be considered in special cases, such as (the examples given are not exhaustive, but are given as an aid to decision-making): The structure or known metabolism of the compound indicates that standard techniques for in vitro metabolic activation (e.g., rodent liver S9) might be inadequate; the structure or
known activity of the compound indicates that the use of other test methods/systems might be appropriate.

5.3 Evaluation of Results Obtained from In Vivo Tests

In vivo tests have the advantage of taking into account absorption, distribution and excretion, which are not factors in in vitro tests, but are potentially relevant to human use. In addition metabolism is likely to be more relevant in vivo compared to the systems normally used in vitro. If the in vivo and in vitro results do not agree, then the difference should be considered/explained on a case-by-case basis, e.g., a difference in metabolism; rapid and efficient excretion of a compound in vivo.

In vivo genotoxicity tests also have the potential to give misleading positive results that do not indicate true genotoxicity. As examples:

i. Increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoiesis (Tweets et al., 2007, I).

ii. DNA adduct data should be interpreted in the light of the known background level of endogenous adducts.

iii. Indirect, toxicity-related effects could influence the results of the DNA strand break assays (e.g., alkaline elution and Comet assays).

Thus it is important to take into account all the toxicological and hematological findings when evaluating the genotoxicity data (Note 15). Indirect effects related to toxicological changes could have a safety margin and might not be clinically relevant.

5.4 Follow-up Strategies for Positive Results

5.4.1 Follow-up to Findings in In Vitro in Mammalian Cell Tests

The following discussion assumes negative results in the Ames bacterial mutation assay.

5.4.1.1 Mechanistic/In Vivo Follow-up

When there is insufficient weight of evidence to indicate lack of relevance, recommended follow-up for positive mammalian cell assays would be to provide experimental evidence, either by additional in vitro studies (i, below) or by carrying out two appropriate in vivo assays (ii, below), as follows:

i. Mechanistic information that contributes to a weight of evidence for a lack of relevant genotoxicity is often generated in vitro, for example evidence that a test compound that induces chromosome aberrations or mutations in the MLA is not a DNA damaging agent (e.g., other negative mutation/DNA damage tests in addition to the Ames test; structural considerations), or evidence for an indirect mechanism that might not be relevant in vivo or might have a threshold (e.g., inhibition of DNA synthesis, reactive oxygen species produced only at high concentrations) (Galloway et al., 1998; Scott et al., 1991; Müller and Kasper, 2000). Similar studies can be used to follow up a positive result in the in vitro micronucleus assay, or in this case evidence can include a known mechanism that indicates chromosome loss/aneuploidy, or centromere staining experiments (Note 17) that indicate chromosome loss. Polyploidy is a common finding in chromosome aberration assays in vitro. While aneugens can induce polyploidy, polyploidy alone does not indicate aneugenic potential and can simply indicate cell cycle perturbation; it is also commonly associated with increasing cytotoxicity. If polyploidy, but no structural chromosome breakage, is seen in an in vitro assay, generally a negative in vivo micronucleus assay with assurance of appropriate exposure would provide
sufficient assurance of lack of potential for aneuploidy induction.

If the above mechanistic information and weight of evidence supports the lack of relevant genotoxicity, only a single in vivo test with appropriate evidence of exposure is called for in order to establish the lack of genotoxic activity. This is typically a cytogenetic assay, and the micronucleus assay in vivo is called for when following up potential for chromosome loss.

If there is not sufficient weight of evidence or mechanistic information to rule out relevant genotoxic potential, two in vivo tests are generally called for, with appropriate endpoints and in appropriate tissues (usually two different tissues), and with an emphasis on obtaining sufficient exposure in the in vivo models.

Or

ii. Two appropriate in vivo assays are done, usually with different tissues, and with supporting demonstration of exposure.

In summary, negative results in appropriate in vivo assays, with adequate justification for the endpoints measured and demonstration of exposure (see Section 4.4.1) are considered sufficient to demonstrate absence of significant genotoxic risk.

5.4.1.2 Follow-up to an In Vitro Positive Result That is Dependent upon S9 Activation

When positive results are seen only in the presence of the S9 activation system, it should first be verified that metabolic activation is responsible and not some other difference in conditions (e.g., low or no serum in the S9 mix, compared with ≥10% serum in the non-activated incubations). The follow-up strategy is then aimed at determining the relevance of the results in vitro to conditions in vivo, and will generally focus on in vivo studies in liver (Note 18).

5.4.2 Follow-up to a Positive In Vivo Micronucleus Assay

If there is an increase in micronuclei in vivo, all the toxicological data should be evaluated to determine whether a non-genotoxic effect could be the cause or a contributing factor (Note 15). If non-specific effects of disturbed erythropoiesis or physiology (such as hypo/hyperthermia) are suspected, an in vivo assay for chromosome aberrations might be more appropriate. If a “real” increase is suspected, strategies should be used to demonstrate whether the increase is due to chromosome loss or chromosome breakage (Note 17). There is evidence that aneuploidy induction, e.g., with spindle poisons, follows a non-linear dose response. Thus, it might be possible to determine that there is a threshold exposure below which chromosome loss is not expected and to determine whether an appropriate safety margin exists compared with clinical exposure.

In conclusion, the assessment of the genotoxic potential of a compound should take into account the totality of the findings and acknowledge the intrinsic values and limitations of both in vitro and in vivo tests.

5.5 Follow-up Genotoxicity Testing in Relation to Tumor Findings in a Carcinogenicity Bioassay

Additional genotoxicity testing in appropriate models can be conducted for compounds that were negative in the standard test battery but which have shown increases in tumors in carcinogenicity bioassay(s) with insufficient evidence to establish a non-genotoxic mechanism. To help understand the mode of action, additional testing can include modified conditions for metabolic activation in in vitro tests or can include in vivo tests measuring genetic damage in target organs of tumor induction, such as DNA
strand break assays (e.g., comet or alkaline elution assays), liver UDS test, DNA covalent binding (e.g., by $^{32}$P-postlabelling), mutation induction in transgenes, or molecular characterization of genetic changes in tumor-related genes (Kasper et al., 2007).

6. NOTES

1. The in vitro micronucleus assay has been widely evaluated in international collaborative studies (Kirsch-Volders et al., 2003), is validated by ECVAM (Corvi et al., 2008), and is the subject of an OECD guideline 487 (2010).

2. There is a small but significant number of genotoxic carcinogens that are reliably detected by the bone marrow tests for chromosomal damage but have yielded negative/weak/conflicting results in the in vitro tests outlined in the standard battery options. Carcinogens such as procarbazine, hydroquinone, urethane and benzene fall into this category. Some other examples from a survey of companies are described by Tweats et al., 2007, II.

3. In principle, micronuclei in hematopoietic cells can be evaluated in bone marrow from any species, and in blood from species that do not filter out circulating micronucleated erythrocytes in the spleen. In laboratory mice, micronuclei can be measured in polychromatic erythrocytes in blood, and mature (normochromatic) erythrocytes can be used when mice are treated continuously for about 4 weeks or more. Although rats rapidly remove micronucleated erythrocytes from the circulation, it has been established that micronucleus induction by a range of clastogens and aneugens can be detected in rat blood reticulocytes (Wakata et al., 1998; Hamada et al., 2001). Rat blood can be used for micronucleus analysis provided methods are used to ensure analysis of the newly formed reticulocytes (Hayashi et al., 2007; MacGregor et al., 2006), and the sample size is sufficiently large to provide appropriate statistical sensitivity given the lower micronucleus levels in rat blood than in bone marrow (Kissling et al., 2007). Whichever method is chosen, bone marrow or blood, automated or manual analysis, each laboratory should determine the appropriate minimum sample size to ensure that scoring error is maintained below the level of animal-to-animal variation.

Some experience is now available for micronucleus induction in the dog and rhesus monkey (Harper et al., 2007; Hotchkiss et al., 2008). One example where such alternative species might be useful would be in evaluation of a human metabolite that was not sufficiently represented in rodents but was formed in the dog or monkey.

4. While the two options in the battery are equally suitable, specific knowledge about an individual test compound can indicate that one option is preferable. For example, if systemic exposure in animal models is equal to or less than anticipated clinical exposure, in vitro assays should be employed: Option 1 (see also Sections 2.3.4 and 4.4.1). On the other hand Option 2, including a test in liver, is recommended in cases where short-lived reactive metabolites are expected to be generated in the liver.

5. Certain structurally alerting molecular entities are recognized as being causally related to the carcinogenic and/or mutagenic potential of chemicals. Examples of structural alerts include alkylating electrophilic centers, unstable epoxides, aromatic amines, azo-structures, N-nitroso groups, and aromatic nitro-groups (Ashby and Paton, 1994). For some classes of compounds with specific structural alerts, it is established that specific protocol modifications/additional tests are important for optimum detection of genotoxicity (e.g., molecules containing an azo-group, glycosides, compounds such as nitroimidazoles requiring nitroreduction for activation,
compounds such as phenacetin requiring a different rodent S9 for metabolic activation).

6. There is some experience with *in vivo* assays for micronucleus induction in skin and colon (Hayashi et al., 2007), and DNA damage assays in these tissues can also be an appropriate substitute.

7. A few chemicals are more easily detectable either with plate-incorporation or with pre-incubation methods, though differences are typically quantitative rather than qualitative (Gatehouse et al., 1994). Experience in the pharmaceutical industry where drugs have been tested in both protocols has not resulted in different results for the two methods, and, in the IWGT report (Gatehouse et al., 1994), the examples of chemical classes listed as more easily detectable in the pre-incubation protocol are generally not pharmaceuticals and are positive in *in vivo* genotoxicity tests in liver. These include short chain aliphatic nitrosamines; divalent metals; aldehydes (e.g., formaldehyde, crotonaldehyde); azo dyes (e.g., butter yellow); pyrrolizidine alkaloids; allyl compounds (allylisothiocyanate, allyl chloride), and nitro (aromatic, aliphatic) compounds.

8. The rationale for a maximum concentration of 1 mM for *in vitro* mammalian cell assays includes the following: The test battery includes the Ames test and an *in vivo* assay. This battery optimizes the detection of genotoxic carcinogens without relying on any individual assay alone. There is a very low likelihood of compounds of concern (DNA damaging carcinogens) that are not detected in Ames test or *in vivo* genotoxicity assay, but are detectable in an *in vitro* mammalian assay only above 1 mM. Second, a limit of 1 mM maintains the element of hazard identification, being higher than clinical exposures to known pharmaceuticals, including those that concentrate in tissues (Goodman & Gilman, 2001), and is also higher than the levels generally achievable in preclinical studies *in vivo*. Certain drugs are known to require quite high clinical exposures for therapeutic effect, e.g., nucleoside analogs and some antibiotics. While comparison of potency with existing drugs can be of interest to sponsors, perhaps even above the 1 mM limit, it is ultimately the *in vivo* tests that determine relevance for human safety. For pharmaceuticals with unusually low molecular weight (e.g., less than 200) higher test concentrations should be considered.

9. Although some genotoxic carcinogens are not detectable in *in vitro* genotoxicity assays unless the concentrations tested induce some degree of cytotoxicity, DNA damaging agents are generally detectable with only moderate levels of toxicity (Greenwood et al., 2004). As cytotoxicity increases, mechanisms other than direct DNA damage by a compound or its metabolites can lead to ‘positive’ results that are related to cytotoxicity and not genotoxicity. Such indirect induction of DNA damage secondary to damage to non-DNA targets is more likely to occur above a certain concentration threshold. The disruption of cellular processes is not expected to occur at lower, pharmacologically relevant concentrations.

In cytogenetic assays, even weak clastogens that are known to be carcinogens are positive without exceeding a 50% reduction in cell counts. On the other hand, compounds that are not DNA damaging, mutagenic or carcinogenic can induce chromosome breakage at toxic concentrations. For both *in vitro* cytogenetic assays, the chromosome aberration assay and the *in vitro* micronucleus assay, a limit of about 50% growth reduction is considered appropriate.

For cytogenetic assays in cell lines, measurement of cell population growth over time (by measuring the change in cell number during culture relative to control, e.g., by
the method referred to as Population Doubling (PD; Note 10), has been shown to be a useful measure of cytotoxicity, as it is known that cell numbers can underestimate toxicity. For lymphocyte cultures, an inhibition of proliferation not exceeding about 50% is considered sufficient; this can be measured by Mitotic Index (MI) for metaphase aberration assays and by an index based on cytokinesis block for in vitro micronucleus assays. In addition, for the in vitro micronucleus assay, since micronuclei are scored in the interphase subsequent to a mitotic division, it is important to verify that cells have progressed through the cell cycle. This can be done by use of cytochalasin B to allow nuclear division but not cell division, so that micronuclei can be scored in binucleate cells (the preferred method for lymphocytes). For cell lines, other methods to demonstrate cell proliferation, including cell population growth over time (PD) as described above, can be used (Kirsch-Volders et al., 2003).

For MLA, appropriate sensitivity is achieved by limiting the top concentration to one with close to 20% Relative Total Growth (RTG) (10-20%) both for soft agar and for microwell methods (Moore et al., 2002). Reviews of published data using the current criteria found very few chemicals that were positive in MLA only at concentrations with less than 20% RTG and that were rodent carcinogens, and convincing evidence of genotoxic carcinogenesis for this category is lacking. The consensus is that caution is appropriate in interpreting results when increases in mutation are seen only below 20% RTG, and a result would not be considered positive if the increase in mutant fraction occurred only at \( \leq 10\% \) RTG.

In conclusion, caution is appropriate in interpreting positive results obtained as reduction in growth/survival approaches or exceeds 50% for cytogenetics assays or 80% for MLA. It is acknowledged that the evaluation of cells treated at these levels of cytotoxicity/clonal survival can result in greater sensitivity but bears an increased risk of non-relevant positive results. The battery approach for genotoxicity is designed to ensure appropriate sensitivity without relying on single in vitro mammalian cell tests at high cytotoxicity.

To obtain an appropriate toxicity range, a preliminary range-finding assay over a broad range of concentrations is useful, but in the genotoxicity assay it is often critical to use multiple concentrations that are spaced quite closely (less than two-fold dilutions). Extra concentrations can be tested but not all concentrations need be evaluated for genotoxicity. It is not intended that multiple experiments be carried out to reach exactly 50% reduction in growth, for example, or exactly 80% reduction in RTG.

10. For in vitro cytogenetic assays it is appropriate to use a measure of relative cell growth to assess toxicity, because cell counts can underestimate toxicity (Greenwood et al., 2004). Using calculated population doublings (see Glossary) to estimate the 50% growth reduction level, it was demonstrated that the frequency of positive results with compounds that are not mutagenic or carcinogenic is reduced, while agents that act via direct interaction with DNA are reliably positive.

11. In certain cases it can be useful to examine chromosome aberrations at metaphase in lymphocytes cultured from test animals after one or more administrations of test compound, just as bone marrow metaphase cells can be used. Since circulating lymphocytes are not replicating, agents that require replication for their genotoxic effect (e.g., some nucleoside analogs) are not expected to be detected in this cell type. Because some lymphocytes are relatively long-lived, in principle there is the potential for accumulation of un-repaired DNA damage in vivo that would give rise to aberrations when the cells are stimulated to divide in vitro. The in vivo lymphocyte
Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use

assay can be useful in following up indications of clastogenicity, but in general another tissue such as liver is a more informative supplement to the micronucleus assay in hematopoietic cells because exposure to drug and metabolite(s) is often higher in liver.

12. The inclusion of a second in vivo assay in the battery is to provide assurance of lack of genotoxicity by use of a tissue that is well exposed to a drug and/or its metabolites; a small number of carcinogens that are considered genotoxic gave positive results in a test in liver but were negative in a cytogenetics assay in vivo in bone marrow. These examples likely reflect a lack of appropriate metabolic activity or lack of reactive intermediates delivered to the hematopoietic cells of the bone marrow.

Assays for DNA strand breaks, DNA adducts, and mutations in transgenes have the advantage that they can be applied in many tissues. Internationally agreed protocols are not yet in place for all the in vivo assays, although considerable experience and published data and protocol recommendations exist for DNA strand break assays (Comet and alkaline elution assays), DNA adduct (covalent binding) measurements, and transgenic rodent mutation assays, in addition to the UDS assay. For a compound that is positive in vitro in the MLA and induces predominantly large colonies, and is also shown not to induce chromosome breakage in an in vitro metaphase assay, an in vivo assay for mutation, such as a transgenic mouse mutation assay, should be considered in preference to a DNA strand break assay. The UDS assay is considered useful mainly for compounds that induce bulky DNA adducts or are positive in the Ames test. Because cytotoxicity induces DNA strand breakage, careful cytotoxicity assessment is needed to avoid confounding the results of DNA strand break assays. This has been well characterized for the in vitro alkaline elution test (Storer et al., 1996) but not yet fully validated for the Comet assay. In principle the DNA strand break assays can be used in repeat-dose toxicology assays with appropriate dose levels and sampling times.

Since liver of mature animals is not a highly mitotic tissue, often a non-cytogenetic endpoint is used for the second assay, but when dividing hepatocytes are present, such as after partial hepatectomy, or in young rats (Hayashi et al., 2007), micronucleus analysis in liver is possible, and detects known genotoxic compounds.

13. For common vehicles like aqueous methyl cellulose this would usually be appropriate, but for vehicles such as Tween 80, the volume that can be administered could be as much as 30 fold lower than that given acutely.

14. Caution is appropriate if the toxicological study design includes additional blood sampling, e.g., for measurement of exposure. Such bleeding could perturb the results of micronucleus analysis since erythropoiesis stimulated by bleeding can lead to increases in micronucleated erythrocytes.

15. Increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoiesis (such as regenerative anemia; extramedullary hematopoiesis), stress, and hypo- and hyperthermia (reviewed by Tweats et al., 2007, I). In blood, changes in spleen function that affect clearance of micronucleated cells from the blood could lead to small increases in circulating micronucleated red blood cells.

16. Positive controls for either short-term or repeat dose genotoxicity studies: For micronucleus (and other cytogenetic) assays, the purpose of the positive control is to verify that the individuals scoring the slides can reliably detect increases in micronuclei. This can be accomplished by use of samples from periodic studies (every few months) of small groups of animals (one sex) given acute treatment with a
positive control. For manual scoring such slides can be included in coded slides scored from each study. Positive control slides should not be obvious to readers based on their staining properties or micronucleus frequency. For automated scoring, appropriate quality control samples should be used with each assay.

For other in vivo genotoxicity assays, the purpose of positive controls is to demonstrate reliable detection of an increase in DNA damage/mutagenicity using the assay in the chosen species, tissue and protocol. After a laboratory has demonstrated that it can consistently detect appropriate positive control compounds in multiple independent experiments, carrying out positive control experiments periodically is generally sufficient provided experimental conditions are not changed. However, currently it is considered that for the Comet assay concurrent positive controls are advisable.

17. Determination of whether micronucleus induction is due primarily to chromosome loss or to chromosome breakage could include staining micronuclei in vitro or in vivo to determine whether centromeres are present, e.g., using Fluorescent in situ Hybridization (FISH) with probes for DNA sequences in the centromeric region, or a labeled antibody to kinetochore proteins. If the majority of induced micronuclei are centromere positive, this suggests chromosome loss. (Note that even potent tubule poisons like colchicine and vinblastine do not produce 100% kinetochore positive micronuclei, but more typically 70 to 80%, and are accepted as primarily aneugens for assessing risk). An alternative approach is to carry out an in vitro or in vivo assay for metaphase structural aberrations; if negative this would imply that micronucleus induction is related to chromosome loss.

18. Standard induced S9 mix has higher activation capacity than human S9, and lacks phase two detoxification capability unless specific cofactors are supplied. Also, non-specific activation can occur in vitro with high test substrate concentrations (see Kirkland et al., 2007). Genotoxicity testing with human S9 or other human-relevant activation systems can be helpful. Analysis of the metabolite profile in the genotoxicity test incubations for comparison with known metabolite profiles in preclinical species (in uninduced microsomes or hepatocytes, or in vivo) or in preparations from humans can also help determine the relevance of test results (Ku et al., 2007), and follow-up studies will usually focus on in vivo testing in liver. A compound that gives positive results in vitro with S9 might not induce genotoxicity in vivo because the metabolite is not formed, is formed in very small quantities, or is metabolically detoxified or rapidly excreted, indicating a lack of risk in vivo.
7. **Glossary**

**Alkaline elution assay:**
See DNA strand break assay.

**Aneuploidy:**
Numerical deviation of the modal number of chromosomes in a cell or organism.

**Base substitution:**
The substitution of one or more base(s) for another in the nucleotide sequence. This can lead to an altered protein.

**Cell proliferation:**
The ability of cells to divide and to form daughter cells.

**Centromere/kinetochore:**
Structures in chromosomes essential for association of sister chromatids and for attachment of spindle fibers that move daughter chromosomes to the poles and ensure inclusion in daughter nuclei.

**Clastogen:**
An agent that produces structural breakage of chromosomes, usually detectable by light microscopy.

**Cloning efficiency:**
The efficiency of single cells to form clones. It is usually measured after seeding low numbers of cells in a suitable environment.

**Comet assay:**
See DNA strand break assay.

**Culture confluency:**
A quantification of the cell density in a culture by visual inspection.

**Cytogenetic evaluation:**
Chromosome structure analysis in mitosis or meiosis by light microscopy or micronucleus analysis.

**DNA adduct:**
Product of covalent binding of a chemical to DNA.

**DNA repair:**
Reconstitution of the original DNA sequence after DNA damage.

**DNA strand breaks:**
Single or double strand scissions in the DNA.

**DNA strand break test:**
Alkaline treatment that converts certain types of DNA lesions into strand breaks that can be detected by the alkaline elution technique, measuring migration rate through a filter, or by the single cell gel electrophoresis or Comet test (in which cells embedded in a thin layer of gel on a microscope slides are subjected to electric current, causing shorter pieces of DNA to migrate out of the nucleus into a “Comet tail”). The extent of DNA migration is measured visually under the microscope on stained cells.

**Frameshift mutation:**
A mutation (change in the genetic code) in which one base or two adjacent bases are
added to (inserted in) or deleted from the nucleotide sequence of a gene. This can lead to an altered or truncated protein.

**Gene mutation:**
A detectable permanent change within a single gene or its regulating sequences. The changes can be point mutations, insertions, or deletions.

**Genetic endpoint:**
The precise type or class of genetic change investigated (e.g., gene mutations, chromosomal aberrations, DNA strand breaks, DNA repair, DNA adduct formation, etc).

**Genotoxicity:**
A broad term that refers to any deleterious change in the genetic material regardless of the mechanism by which the change is induced.

**Micronucleus:**
Particle in a cell that contains nuclear DNA; it might contain a whole chromosome(s) or a broken centric or acentric part(s) of chromosome(s).

**Mitotic index:**
Percentage of cells in the different stages of mitosis amongst the cells not in mitosis (interphase) in a preparation (slide).

**Numerical chromosome changes:**
Chromosome numbers different from the original haploid or diploid set of chromosomes; for cell lines, chromosome numbers different from the modal chromosome set.

**Plasmid:**
Genetic element additional to the normal bacterial genome. A plasmid might be inserted into the host chromosome or form an extra-chromosomal element.

**Point mutations:**
Changes in the genetic codes, usually confined to a single DNA base pair.

**Polychromatic erythrocyte:**
An immature erythrocyte in an intermediate stage of development that still contains ribosomes and, as such, can be distinguished from mature normochromatic erythrocytes (lacking ribosomes) by stains selective for RNA.

**Polyploidy:**
Numerical deviation of the modal number of chromosomes in a cell, with approximately whole multiples of the haploid number. Endoreduplication is a morphological form of polyploidy in which chromosome pairs are associated at metaphase as “diplochromosomes”.

**Population doubling or culture growth:**
This can be calculated in different ways; one example of an appropriate formula is: Population doublings (PDs) = the log of the ratio of the final count (N) to the starting (baseline) count (X₀), divided by the log of 2. That is: PD = \( \frac{\log(N \div X₀)}{\log 2} \).

**Recombination:**
Breakage and balanced or unbalanced rejoining of DNA.

**RTG (relative total growth):**
This measure of cytotoxicity takes the relative suspension growth (based on cell loss and cell growth from the beginning of treatment to the second day post-treatment) and multiplies it by the relative plating efficiency at the time of cloning for mutant
quantization.

**Single Cell Gel Electrophoresis assay:**
Comet assay. See **DNA strand break assay**.

**Survival (in the context of mutagenicity testing):**
Proportion of living cells among dead cells, usually determined by staining or colony counting methods after a certain treatment interval.

**Transgene:**
An exogenous or foreign gene inserted into the host genome, either into somatic cells or germ line cells.

**Unscheduled DNA synthesis (UDS):**
DNA synthesis that occurs at some stage in the cell cycle other than S-phase in response to DNA damage. It is usually associated with DNA excision repair.
8. REFERENCES


Kasper P, Uno Y, Mauthe R, Asano N, Douglas G, Matthews E et al. Follow-up testing of
Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use


Kirkland DJ, Pfuhler S, Tweets D, Aardema M, Corvi R, Darroudi F et al. How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of the ECVAM workshop. Mutat Res 2007;628:31-55.


OECD Guidelines for Genetic Toxicology 1997.


This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
### S3A

#### Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3A</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>27 October 1993</td>
<td>November 2005 S3A</td>
</tr>
</tbody>
</table>

#### Current Step 4 version

| S3A | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 27 October 1994 | S3A |
NOTE FOR GUIDANCE ON TOXICOKinetics:
THE ASSESSMENT OF SYSTEMic EXPOSURE IN TOXICity STUDIES

ICH Harmonised Tripartite Guideline
Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 27 October 1994, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. Introduction ............................................................................................................. 1

2. The objectives of toxicokinetics and the parameters which may be determined ............................................................................................................... 2

3. General principles to be considered .................................................................. 2
   3.1 Introduction .......................................................................................................... 2
   3.2 Quantification of exposure ................................................................................... 2
   3.3 Justification of time points for sampling ............................................................. 3
   3.4 Contribution to the setting of dose levels in order to produce adequate exposure .......................................................................................................................... 3
   3.5 Extent of exposure assessment in toxicity studies ............................................. 4
   3.6 Complicating factors in exposure interpretation ................................................ 4
   3.7 Route of administration ....................................................................................... 4
   3.8 Determination of metabolites ............................................................................ 5
   3.9 Statistical evaluation of data ............................................................................... 5
   3.10 Analytical methods ............................................................................................... 5
   3.11 Reporting ............................................................................................................... 6

4. Toxicokinetics in the various areas of toxicity testing - specific aspects .................. 6
   4.1 Introduction .......................................................................................................... 6
   4.2 Single-dose toxicity studies .................................................................................. 6
   4.3 Repeated-dose toxicity studies ............................................................................. 7
   4.4 Genotoxicity studies ............................................................................................. 7
   4.5 Carcinogenicity (oncogenicity) studies1 .............................................................. 7
   4.6 Reproductive toxicity studies2 ............................................................................. 8

5. Supplementary Notes ............................................................................................. 9

6. References (other ICH Guidance) ....................................................................... 11
NOTE FOR GUIDANCE ON TOXICOkinetics:
THE ASSESSMENT OF SYSTEMIC EXPOSURE IN TOXICITY STUDIES

1. Introduction

This Note for Guidance concerns toxicokinetics only with respect to the development of pharmaceutical products intended for use in human subjects.

In this context, toxicokinetics is defined as the generation of pharmacokinetic data, either as an integral component in the conduct of non-clinical toxicity studies or in specially designed supportive studies, in order to assess systemic exposure. These data may be used in the interpretation of toxicology findings and their relevance to clinical safety issues (see Note 1 for definitions of other terms used in this document).

The Note for Guidance has been developed in order to provide an understanding of the meaning and application of toxicokinetics and to provide guidance on developing test strategies in toxicokinetics. The guidance highlights the need to integrate pharmacokinetics into toxicity testing, which should aid in the interpretation of the toxicology findings and promote rational study design development.

Toxicokinetic measurements are normally integrated within the toxicity studies and as such are described in this document as 'concomitant toxicokinetics' (Note 1). Alternatively, data may be generated in other supportive studies conducted by mimicking the conditions of the toxicity studies.

Toxicokinetic procedures may provide a means of obtaining multiple dose pharmacokinetic data in the test species, if appropriate parameters are monitored, thus avoiding duplication of such studies; optimum design in gathering the data will reduce the number of animals required.

Various components of the total non-clinical pharmacokinetics and metabolism programme may be of value in contributing to the interpretation of toxicology findings. However, the toxicokinetic data focus on the kinetics of a new therapeutic agent under the conditions of the toxicity studies themselves.

Toxicokinetics is thus an integral part of the non-clinical testing programme; it should enhance the value of the toxicological data generated, both in terms of understanding the toxicity tests and in comparison with clinical data as part of the assessment of risk and safety in humans. Due to its integration into toxicity testing and its bridging character between non-clinical and clinical studies, the focus is primarily on the interpretation of toxicity tests and not on characterising the basic pharmacokinetic parameters of the substance studied.

As the development of a pharmaceutical product is a dynamic process which involves continuous feedback between non-clinical and clinical studies, no rigid detailed procedures for the application of toxicokinetics are recommended. It may not be necessary for toxicokinetic data to be collected in all studies and scientific judgement should dictate when such data may be useful. The need for toxicokinetic data and the extent of exposure assessment in individual toxicity studies should be based on a flexible step-by-step approach and a case-by-case decision making process to provide sufficient information for a risk and safety assessment.
2. The objectives of toxicokinetics and the parameters which may be determined

The primary objective of toxicokinetics is:

- to describe the systemic exposure achieved in animals and its relationship to dose level and the time course of the toxicity study.

Secondary objectives are:

- to relate the exposure achieved in toxicity studies to toxicological findings and contribute to the assessment of the relevance of these findings to clinical safety.

- to support (Note 1) the choice of species and treatment regimen in non-clinical toxicity studies.

- to provide information which, in conjunction with the toxicity findings, contributes to the design of subsequent non-clinical toxicity studies.

These objectives may be achieved by the derivation of one or more pharmacokinetic parameters (Note 2) from measurements made at appropriate time points during the course of the individual studies. These measurements usually consist of plasma (or whole blood or serum) concentrations for the parent compound and/or metabolite(s) and should be selected on a case-by-case basis. Plasma (or whole blood or serum) AUC, $C_{\text{max}}$ and $C_{\text{time}}$ (Note 2) are the most commonly used parameters in assessing exposure in toxicokinetics studies. For some compounds it will be more appropriate to calculate exposure based on the (plasma protein) unbound concentration.

These data may be obtained from all animals on a toxicity study, in representative subgroups, in satellite groups (see 3.5 and Note 1) or in separate studies.

Toxicity studies which may be usefully supported by toxicokinetic information include single and repeated-dose toxicity studies, reproductive, genotoxicity and carcinogenicity studies. Toxicokinetic information may also be of value in assessing the implications of a proposed change in the clinical route of administration.

3. General principles to be considered

3.1 Introduction

In the following paragraphs some general principles are set out which should be taken into consideration in the design of individual studies.

It should be noted that for those toxicity studies whose performance is subject to Good Laboratory Practice (GLP) the concomitant toxicokinetics must also conform to GLP. Toxicokinetic studies retrospectively designed to generate specific sets of data under conditions which closely mimic those of the toxicity studies should also conform to GLP when they are necessary for the evaluation of safety.

3.2 Quantification of exposure

The quantification of systemic exposure provides an assessment of the burden on the test species and assists in the interpretation of similarities and differences in toxicity across species, dose groups and sexes. The exposure might be represented by plasma (serum or blood) concentrations or the AUCs of parent compound and/or metabolite(s). In some circumstances, studies may be designed to investigate tissue concentrations. When designing the toxicity studies, the
exposure and dose-dependence in humans at therapeutic dose levels (either expected or established), should be considered in order to achieve relevant exposure at various dose levels in the animal toxicity studies. The possibility that there may be species differences in the pharmacodynamics of the substance (either qualitative or quantitative) should also be taken into consideration.

Pharmacodynamic effects or toxicity might also give supporting evidence of exposure or even replace pharmacokinetic parameters in some circumstances.

Toxicokinetic monitoring or profiling of toxicity studies should establish what level of exposure has been achieved during the course of the study and may also serve to alert the toxicologist to non-linear, dose-related changes in exposure (Note 3) which may have occurred. Toxicokinetic information may allow better interspecies comparisons than simple dose/body weight (or surface area) comparisons.

3.3 Justification of time points for sampling
The time points for collecting body fluids in concomitant toxicokinetic studies should be as frequent as is necessary, but not so frequent as to interfere with the normal conduct of the study or to cause undue physiological stress to the animals (Note 4). In each study, the number of time points should be justified on the basis that they are adequate to estimate exposure (see 3.2). The justification should be based on kinetic data gathered from earlier toxicity studies, from pilot or dose range-finding studies, from separate studies in the same animal model or in other models allowing reliable extrapolation.

3.4 Contribution to the setting of dose levels in order to produce adequate exposure
The setting of dose levels in toxicity studies is largely governed by the toxicology findings and the pharmacodynamic responses of the test species. However, the following toxicokinetic principles may contribute to the setting of the dose levels.

3.4.1 Low dose levels
At the low dose, preferably a no-toxic-effect dose level (Note 5), the exposure in the animals of any toxicity study should ideally equal or just exceed the maximum expected (or known to be attained) in patients. It is recognised that this ideal is not always achievable and that low doses will often need to be determined by considerations of toxicology; nevertheless, systemic exposure should be determined.

3.4.2 Intermediate dose levels
Exposure at intermediate dose levels should normally represent an appropriate multiple (or fraction) of the exposure at lower (or higher) dose levels dependent upon the objectives of the toxicity study.

3.4.3 High dose levels
The high dose levels in toxicity studies will normally be determined by toxicological considerations. However, the exposure achieved at the dose levels used should be assessed.
Where toxicokinetic data indicate that absorption of a compound limits exposure to parent compound and/or metabolite(s) (Note 6), the lowest dose level of the substance producing the maximum exposure should be accepted as the top dose level to be used (when no other dose-limiting constraint applies, Note 7).

Very careful attention should be paid to the interpretation of toxicological findings in toxicity studies (of all kinds) when the dose levels chosen result in non-linear kinetics (Note 3). However, non-linear kinetics should not necessarily result in dose limitations in toxicity studies or invalidate the findings; toxicokinetics can be very helpful in assessing the relationship between dose and exposure in this situation.

### 3.5 Extent of exposure assessment in toxicity studies

In toxicity studies, systemic exposure should be estimated in an appropriate number of animals and dose groups (Note 8) to provide a basis for risk assessment.

Concomitant toxicokinetics may be performed either in all or a representative proportion of the animals used in the main study or in special satellite groups (Notes 1 and 5). Normally, samples for the generation of toxicokinetic data may be collected from main study animals, where large animals are involved, but satellite groups may be required for the smaller (rodent) species.

The number of animals to be used should be the minimum consistent with generating adequate toxicokinetic data. Where both male and female animals are utilised in the main study it is normal to estimate exposure in animals of both sexes unless some justification can be made for not so doing.

Toxicokinetic data are not necessarily required from studies of different duration if the dosing regimen is essentially unchanged (see also 4.3).

### 3.6 Complicating factors in exposure interpretation

Although estimating exposure as described above may aid in the interpretation of toxicity studies and in the comparison with human exposure, a few caveats should be noted.

Species differences in protein binding, tissue uptake, receptor properties and metabolic profile should be considered. For example, it may be more appropriate for highly protein bound compounds to have exposure expressed as the free (unbound) concentrations. In addition, the pharmacological activity of metabolites, the toxicology of metabolites and antigenicity of biotechnology products may be complicating factors. Furthermore, it should be noted that even at relatively low plasma concentrations, high levels of the administered compound and/or metabolite(s) may occur in specific organs or tissues.

### 3.7 Route of administration

The toxicokinetic strategy to be adopted for the use of alternative routes of administration, for example by inhalation, topical or parenteral delivery, should be based on the pharmacokinetic properties of the substance administered by the intended route.

It sometimes happens that a proposal is made to adopt a new clinical route of administration for a pharmaceutical product; for example, a product initially
developed as an oral formulation may subsequently be developed for intravenous administration. In this context, it will be necessary to ascertain whether changing the clinical route will significantly reduce the safety margin.

This process may include a comparison of the systemic exposure to the compound and/or its relevant metabolite(s) (AUC and Cmax) in humans generated by the existing and proposed routes of administration. If the new route results in increased AUC and/or C_{max}, or a change in metabolic route, the continuing assurance of safety from animal toxicology and kinetics should be reconsidered. If exposure is not substantially greater, or different, by the proposed new route compared to that for the existing route(s) then additional non-clinical toxicity studies may focus on local toxicity.

3.8 Determination of metabolites
A primary objective of toxicokinetics is to describe the systemic exposure to the administered compound achieved in the toxicology species. However, there may be circumstances when measurement of metabolite concentrations in plasma or other body fluids is especially important in the conduct of toxicokinetics (Note 9).

- When the administered compound acts as a 'pro-drug' and the delivered metabolite is acknowledged to be the primary active entity.

- When the compound is metabolised to one or more pharmacologically or toxicologically active metabolites which could make a significant contribution to tissue/organ responses.

- When the administered compound is very extensively metabolised and the measurement of plasma or tissue concentrations of a major metabolite is the only practical means of estimating exposure following administration of the compound in toxicity studies (Note 10).

3.9 Statistical evaluation of data
The data should allow a representative assessment of the exposure. However, because large intra- and inter-individual variation of kinetic parameters may occur and small numbers of animals are involved in generating toxicokinetic data, a high level of precision in terms of statistics is not normally needed. Consideration should be given to the calculation of mean or median values and estimates of variability, but in some cases the data of individual animals may be more important than a refined statistical analysis of group data.

If data transformation (e.g. logarithmic) is performed, a rationale should be provided.

3.10 Analytical methods
Integration of pharmacokinetics into toxicity testing implies early development of analytical methods for which the choice of analytes and matrices should be continually reviewed as information is gathered on metabolism and species differences.
The analytical methods to be used in toxicokinetic studies should be specific for the entity to be measured and of an adequate accuracy and precision. The limit of quantification should be adequate for the measurement of the range of concentrations anticipated to occur in the generation of the toxicokinetic data.

The choice of analyte and the matrix to be assayed (biological fluids or tissue) should be stated and possible interference by endogenous components in each type of sample (from each species) should be investigated. Plasma, serum or whole blood are normally the matrices of choice for toxicokinetic studies.

If the drug substance is a racemate or some other mixture of enantiomers, additional justification should be made for the choice of the analyte [racemate or enantiomer(s)].

The analyte and matrix assayed in non-clinical studies should ideally be the same as in clinical studies. If different assay methods are used in non-clinical and clinical studies they should all be suitably validated.

3.11 Reporting
A comprehensive account of the toxicokinetic data generated, together with an evaluation of the results and of the implications for the interpretation of the toxicology findings should be given.

An outline of the analytical method should be reported or referenced. In addition, a rationale for the choice of the matrix analysed and the analyte measured (see 3.8 and 3.10) should be given.

The positioning of the report within the application will depend upon whether the data are specific to any one toxicity study or are supportive of all toxicity testing.

4. Toxicokinetics in the various areas of toxicity testing - specific aspects

4.1 Introduction
Based on the principles of toxicokinetics outlined above the following specific considerations refer to individual areas of toxicity testing. The frequency of exposure monitoring or profiling may be extended or reduced where necessary.

It may be appropriate to take samples from some individual animals only, where this may help in the interpretation of the toxicology findings for these animals.

4.2 Single-dose toxicity studies
These studies are often performed in a very early phase of development before a bioanalytical method has been developed and toxicokinetic monitoring of these studies is therefore not normally possible. Plasma samples may be taken in such studies and stored for later analysis, if necessary; appropriate stability data for the analyte in the matrix sampled would then be required.

Alternatively, additional toxicokinetic studies may be carried out after completion of a single-dose toxicity study in order to respond to specific questions which may arise from the study.

Results from single-dose kinetic studies may help in the choice of formulation and in the prediction of rate and duration of exposure during a dosing interval. This may assist in the selection of appropriate dose levels for use in later studies.
4.3 Repeated-dose toxicity studies

The treatment regimen (Note 11) and species should be selected whenever possible with regard to pharmacodynamic and pharmacokinetic principles. This may not be achievable for the very first studies, at a time when neither animal nor human pharmacokinetic data are normally available.

Toxicokinetics should be incorporated appropriately into the design of the studies. It may consist of exposure profiling or monitoring (Note 1) at appropriate dose levels at the start and towards the end of the treatment period of the first repeat dose study (Note 12). The procedure adopted for later studies will depend on the results from the first study and on any changes in the proposed treatment regimen. Monitoring or profiling may be extended, reduced or modified for specific compounds where problems have arisen in the interpretation of earlier toxicity studies.

4.4 Genotoxicity studies

For negative results of in vivo genotoxicity studies, it may be appropriate to have demonstrated systemic exposure in the species used or to have characterised exposure in the indicator tissue.

4.5 Carcinogenicity (oncogenicity) studies

4.5.1 Sighting or dose-ranging studies

Appropriate monitoring or profiling of these studies should be undertaken in order to generate toxicokinetic data which may assist in the design of the main studies (see 4.5.2.). Particular attention should be paid to species and strains which have not been included in earlier toxicity studies and to the use of routes or methods of administration which are being used for the first time.

Particular attention should be paid to the establishment of appropriate toxicokinetic data when administration is to be in the diet (Note 13).

Toxicokinetic data may assist in the selection of dose levels in the light of information about clinical exposure and in the event that non-linear kinetics (Note 3) may complicate the interpretation of the study.

In principle, the ideal study design would ensure that dose levels in oncogenicity studies generate a range of systemic exposure values that exceed the maximum therapeutic exposure for humans by varying multiples. However, it is recognised that this idealised selection of dose levels may be confounded by unavoidable species-specific problems. Thus, the emphasis of this guidance is on the need to estimate systemic exposure, to parent compound and/or metabolite(s) at appropriate dose levels and at various stages of an oncogenicity study, so that the findings of the study may be considered in the perspective of comparative exposure for the animal model and humans.

A highest dose based on knowledge of probable systemic exposure in the test species and in humans may be an acceptable end-point in testing for carcinogenic potential. Historically, a toxicity end-point\(^1\) has been often used to select the top dose level.
4.5.2 The main studies

The treatment regimen, species and strain selection should, as far as is feasible, be determined with regard to the available pharmacokinetic and toxicokinetic information. In practice, the vast majority of these studies is conducted in the rat and mouse.

As mentioned in the Introduction to this section, it is recommended that reassurance be sought from monitoring that the exposure in the main study is consistent with profiles of kinetics established in free-standing or specific dose-ranging studies. Such monitoring will be appropriate on a few occasions during the study, but it is not considered essential to continue beyond six months.

4.6 Reproductive toxicity studies

4.6.1 Introduction

It is preferable to have some information on pharmacokinetics before initiating reproduction studies since this may suggest the need to adjust the choice of species, study design and dosing schedules. At this time the information need not be sophisticated or derived from pregnant or lactating animals. At the time of study evaluation further information on pharmacokinetics in pregnant or lactating animals may be required depending on the results obtained.

The limitation of exposure in reproductive toxicity is usually governed by maternal toxicity. Thus, while toxicokinetic monitoring in reproductive toxicity studies may be valuable in some instances, especially with compounds with low toxicity, such data are not generally needed for all compounds.

Where adequate systemic exposure might be questioned because of absence of pharmacological response or toxic effects, toxicokinetic principles could usefully be applied to determine the exposures achieved by dosing at different stages of the reproductive process.

A satellite group of female animals may be used to collect the toxicokinetic data.

4.6.2 Fertility studies

The general principles for repeated-dose toxicity studies apply (see 4.3). The need to monitor these studies will depend on the dosing regimen used and the information already available from earlier studies in the selected species.

4.6.3 Studies in pregnant and lactating animals

The treatment regimen during the exposure period should be selected on the basis of the toxicological findings and on pharmacokinetic and toxicokinetic principles.

Consideration should be given to the possibility that the kinetics will differ in pregnant and non-pregnant animals.
Toxicokinetics may involve exposure assessment of dams, embryos, foetuses or newborn at specified days (Note 14). Secretion in milk may be assessed to define its role in the exposure of newborns. In some situations, additional studies may be necessary or appropriate in order to study embryo/foetal transfer and secretion in milk.

Consideration should be given to the interpretation of reproductive toxicity tests in species in which placental transfer of the substance cannot be demonstrated.

5. Supplementary Notes

Note 1 Definitions of expressions appearing in this "Note for Guidance":

**Analyte:** the chemical entity assayed in biological samples.

**Matrix:** blood, plasma, urine, serum or other fluid or tissue selected for assay.

**Concomitant toxicokinetics:** toxicokinetic measurements performed in the toxicity study, either in all animals or in representative subgroups or in satellite groups.

**Exposure:** exposure is represented by pharmacokinetic parameters demonstrating the local and systemic burden on the test species with the test compound and/or its metabolites. The area under the matrix level concentration-time curve (AUC) and/or the measurement of matrix concentrations at the expected peak-concentration time $C_{\text{max}}$, or at some other selected time $C_{\text{time}}$, are the most commonly used parameters. Other parameters might be more appropriate in particular cases.

**Monitor:** to take a small number of matrix samples (e.g. 1-3) during a dosing interval to estimate $C_{\text{time}}$ or $C_{\text{max}}$.

**Profile:** to take (e.g. 4-8) matrix samples during a dosing interval to make an estimate of $C_{\text{max}}$ and/or $C_{\text{time}}$ and area under matrix concentration-time curve (AUC).

**Satellite groups:** groups of animals included in the design and conduct of a toxicity study, treated and housed under conditions identical to those of the main study animals, but used primarily for toxicokinetics.

**Support:** in the context of a toxicity study - to ratify or confirm the design of a toxicity study with respect to pharmacokinetic and metabolic principles. This process may include 2 separate steps:

a) confirmation using toxicokinetic principles that the animals on a study were exposed to appropriate systemic levels of the administered compound (see 3.4) and/or its metabolite(s).

b) confirmation that the metabolic profile in the species used was acceptable; data to support this will normally be derived from metabolism studies in animals and in humans.
**Validate:** in the context of an analytical method - to establish the accuracy, precision, reproducibility, response function and the specificity of the analytical method with reference to the biological matrix to be examined and the analyte to be quantified.

**Note 2** Symbols and definitions according to "Manual of Symbols, Equations and Definitions in Pharmacokinetics", Committee for Pharmacokinetic Nomenclature of the American College of Clinical Pharmacology, Philadelphia, PA, May 1982:

- $C_{\text{max}}$: Maximum (peak) concentration.
- $C_{(\text{time})}$: Maximum concentration at a specified time after administration of a given dose.
- $t_{\text{max}}$: Time to reach peak or maximum concentration following administration.
- $\text{AUC}_{(0-t)}$: Area under concentration-time curve from zero to time $t$. It should be noted that $\text{AUC}_{(0-\infty)}$ is a special case of $\text{AUC}_{(0-t)}$.

Other measurements, for example urinary excretion, may be more appropriate for some compounds. Other derived parameters, for example bioavailability, half-life, fraction of unbound drug and volume of distribution may be of value in interpreting toxicokinetic data. Thus, the selection of parameters and time points has to be made on a case-by-case basis considering the general principles as outlined in Section 3.

**Note 3** Increases in exposure may arise unexpectedly as a result of non-linear kinetics due to saturation of a clearance process. Increasing exposure may also occur during the course of a study for those compounds which have a particularly long plasma half-life. Careful attention should also be paid to compounds which achieve high $C_{\text{max}}$ values over comparatively short time periods within the dosing interval. Conversely, unexpectedly low exposure may occur during a study as a result of auto-induction of metabolising enzymes.

**Note 4** If samples are taken from main study animals it should be considered whether samples should be taken from all the dosed animals and the controls in order to treat all animals on the study in the same way, or whether samples should be taken from representative subgroups of the same size.

**Note 5** In this context, a 'no-toxic-effect dose level' (deemed to be the same as 'no-observed-adverse-effect dose level') is defined as a dose level at which some pharmacological response may be observed, but at which no adverse effect is found.

**Note 6** In these circumstances it should be established that absorption is the rate limiting step and that limitations in exposure to the administered substance are not due to an increased clearance.

**Note 7** The limits placed on acceptable volumes which can be administered orally to animals may constrain the dose levels achievable for comparatively non-toxic compounds administered as solutions or suspensions.
Note 8  It is often considered unnecessary to assay samples from control groups. Samples may be collected and then assayed if it is deemed that this may help in the interpretation of the toxicity findings, or in the validation of the assay method.

Note 9  Measurement of metabolite concentrations may be especially important when documentation of exposure to human metabolite(s) is needed in the non-clinical toxicity studies in order to demonstrate adequate toxicity testing of these metabolites.

Note 10  It is recognised that measurement of metabolite(s) as a part of toxicokinetic evaluation serves only to assess exposure and cannot account for possible reactive intermediate metabolites.

Note 11  Treatment regimen encompasses dosage, formulation, route of administration and dosing frequency.

Note 12  The first repeat dose study incorporating toxicokinetic data for each species is normally of 14 days duration or longer.

Note 13  Additional studies may be required in order to compare exposure to the compound administered in diet and by gavage or by routes different from the intended clinical route.

Note 14  It should be noted that while it is important to consider the transfer of substances entering the embryo-foetal compartment, foetal exposure is the parameter which is most often assessed in practice in separate studies and expressed as 'placental transfer'.

6. References (other ICH Guidance)
   1. Code S 1C Carcinogenicity : Dose Selection for Carcinogenicity Studies of Pharmaceuticals.
   2. Code S 5A Detection of Toxicity to Reproduction for Medicinal Products.
This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
## S3B Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3B</td>
<td>Approval by the Steering Committee under <em>Step 2</em> and release for public consultation.</td>
<td>27 October 1993</td>
<td>November 2005</td>
</tr>
</tbody>
</table>

### Current Step 4 version

| S3B | Approval by the Steering Committee under *Step 4* and recommendation for adoption to the three ICH regulatory bodies. | 27 October 1994 | S3B |
PHARMACOKINETICS:
GUIDANCE FOR REPEATED DOSE TISSUE DISTRIBUTION STUDIES

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 27 October 1994, this guideline is recommended for adoption to the three regulatory parties to ICH

Introduction
A comprehensive knowledge of the absorption, distribution, metabolism and elimination of a compound is important for the interpretation of pharmacology and toxicology studies. Tissue distribution studies are essential in providing information on distribution and accumulation of the compound and/or metabolites, especially in relation to potential sites of action; this information may be useful for designing toxicology and pharmacology studies and for interpreting the results of these experiments.

In the EC, US and Japan, there has been a general agreement on the need to conduct single dose tissue distribution studies as part of the non-clinical programme. These studies often provide sufficient information about tissue distribution.

There has been no consistent requirement for repeated dose tissue distribution studies. However, there may be circumstances when assessments after repeated dosing may yield important information.

This paper provides guidance on circumstances when repeated dose tissue distribution studies should be considered and on the conduct of such studies.

Circumstances Under Which Repeated Dose Tissue Distribution Studies Should be Considered

1. When single dose tissue distribution studies suggest that the apparent half-life of the test compound (and/or metabolites) in organs or tissues significantly exceeds the apparent half life of the elimination phase in plasma and is also more than twice the dosing interval in the toxicity studies, repeated dose tissue distribution studies may be appropriate.

2. When steady-state levels of a compound/metabolite in the circulation, determined in repeated dose pharmacokinetic or toxicokinetic studies, are markedly higher than those predicted from single dose kinetic studies, then repeated dose tissue distribution studies should be considered.

3. When histopathological changes, critical for the safety evaluation of the test substances, are observed that would not be predicted from short term toxicity studies, single dose tissue distribution studies and pharmacological studies, repeated dose tissue distribution studies may aid in the interpretation of these findings. Those organs or tissues which were the site of the lesions should be the focus of such studies.

4. When the pharmaceutical is being developed for site-specific targeted delivery, repeated dose tissue distribution studies may be appropriate.
Design and Conduct of Repeated Dose Tissue Distribution Studies

The objectives of these studies may be achieved using radiolabelled compounds or alternative methods of sufficient sensitivity and specificity.

Dose level(s) and species should be chosen to address the problem that led to the consideration of the repeated dose tissue distribution study.

Information from previous pharmacokinetic and toxicokinetic studies should be used in selecting the duration of dosing in repeated dose tissue distribution studies. One week of dosing is normally considered to be a minimum period. A longer duration should be selected when the blood/plasma concentration of the compound and/or its metabolites does not reach steady state. It is normally considered unnecessary to dose for longer than three weeks.

Consideration should be given to measuring unchanged compound and/or metabolites in organs and tissues in which extensive accumulation occurs or if it is believed that such data may clarify mechanisms of organ toxicity.

Summary

Tissue distribution studies are an important component in the non-clinical kinetics programme. For most compounds, it is expected that single dose tissue distribution studies with sufficient sensitivity and specificity will provide an adequate assessment of tissue distribution and the potential for accumulation. Thus, repeated dose tissue distribution studies should not be required uniformly for all compounds and should only be conducted when appropriate data cannot be derived from other sources. Repeated dose studies may be appropriate under certain circumstances based on the data from single dose tissue distribution studies, toxicity and toxicokinetic studies. The studies may be most appropriate for compounds which have an apparently long half life, incomplete elimination or unanticipated organ toxicity. The design and timing of repeated dose tissue distribution studies should be determined on a case-by-case basis.
ICH Harmonised Tripartite Guideline

Duration of Chronic Toxicity Testing in Animals
(Rodent and Non Rodent Toxicity Testing)

S4

Current Step 4 version
dated 2 September 1998

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
S4
Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4A</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>16 July 1997</td>
<td>November 2005</td>
</tr>
</tbody>
</table>

**Current Step 4 version**

| S4                | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 2 September 1998 | S4               |
1. OBJECTIVE
The objective of this guidance is to set out the considerations that apply to chronic toxicity testing in rodents and non-rodents as part of the safety evaluation of a medicinal product. Since guidance is not legally binding, an applicant may submit justification for an alternative approach.

2. SCOPE
This guidance has been prepared for the development of medicinal products with the exception of those already covered by the ICH Guideline on Safety Studies for Biotechnological Products, e.g., Monoclonal antibodies, recombinant DNA proteins.

3. BACKGROUND
During the first International Conference on Harmonisation in 1991, the practices for the testing of chronic toxicity in the 3 regions (EU, Japan, and US) had been reviewed. Arising from this it emerged that there was a scientific consensus on the approach for chronic testing in rodents, supporting the harmonised duration of testing of 6 months. However, for chronic toxicity testing in non-rodents, there were different approaches to the duration of testing.

The lack of harmonised duration led to the need for pharmaceutical companies to perform partially duplicative studies for both 6 and 12 months duration when developing new medicinal products. As the objective of ICH is to reduce or eliminate the need to duplicate testing during development of medicinal products and to ensure a more economical use of material, animal and human resources, while at the same time maintaining safeguards to protect public health, further scientific evaluation was undertaken.

Each of the regulatory authorities in EU, Japan and US undertook a review to determine whether a single duration for chronic toxicity testing in non-rodents could be identified. From this analysis it emerged that in 16 cases a more detailed evaluation of 6 versus 12 months data should be undertaken.

This evaluation was conducted as a joint exercise by the competent authorities in the 3 regions.

In some of the cases analysed at the tripartite meetings, there were no additional findings at 12 months. For some other cases, there was not complete agreement among the regulators with respect to the comparability in study design and conduct to allow assessment of whether there were differences in the findings at 6 and 12 months due to duration of treatment alone.

In a number of cases there were findings observed by 12 months, but not by 6 months. It was concluded that these would, or could have been detected in a study of nine months duration. Varying degrees of concern for the differences in findings detected...
between the studies of different durations were expressed. An agreement on the clinical relevance of these findings could not be reached.

Studies of 12 months duration are usually not necessary and studies of shorter than 9 months duration may be sufficient.

In the EU, studies of 6 months duration in non-rodents are acceptable according to Council Directive 75/318/EEC, as amended. To avoid duplication, where studies with a longer duration have been conducted, it would not be necessary to conduct a study of 6 months.

4. **GUIDANCE ON DURATION OF CHRONIC TOXICITY TESTING FOR TRIPARTITE DEVELOPMENT PLAN**

Arising from the extensive analysis and review of the above mentioned data in non-rodents and based upon the achievements of ICH1 for testing in rodents, and so as to avoid duplication and follow a single development plan for chronic toxicity testing of new medicinal products, the following studies are considered acceptable for submission in the 3 Regions:

1) **Rodents:**
   - a study of 6 months duration;

2) **Non-rodents:**
   - a study of nine months duration.
ICH Harmonised Tripartite Guideline

Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility
S5(R2)

Current Step 4 version
Parent Guideline dated 24 June 1993
(Addendum dated 9 November 2000 incorporated in November 2005)

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
### S5(R2)
**Document History**

<table>
<thead>
<tr>
<th>First Codification</th>
<th>Date</th>
<th>New Codification As of November 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Parent Guideline: Detection of Toxicity to Reproduction for Medicinal Products**

<table>
<thead>
<tr>
<th>S5A</th>
<th>Approval of the Guideline by the Steering Committee under <em>Step 2</em> and release for public consultation.</th>
<th>19 September 1992</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5A</td>
<td>Approval of the Guideline by the Steering Committee under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>24 June 1993</td>
<td>S5</td>
</tr>
</tbody>
</table>

**Addendum to the Parent Guideline: Toxicity to Male Fertility**

<table>
<thead>
<tr>
<th>S5B</th>
<th>Approval of the Addendum by the Steering Committee under <em>Step 2</em> and release for public consultation.</th>
<th>30 March 1995</th>
<th>in S5(R1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5B</td>
<td>Approval of the Addendum by the Steering Committee under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>29 November 1995</td>
<td>in S5(R1)</td>
</tr>
<tr>
<td>S5B(M)</td>
<td>Approval of minor revisions to the Addendum by the Steering Committee without further public consultation under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>9 November 2000</td>
<td>in S5(R2)</td>
</tr>
</tbody>
</table>

**Current *Step 4* version**

| S5A and S5B(M) | The parent guideline is now renamed S5(R2) as the Addendum and its first revision have been added to the parent guideline. The new title is: “Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility.” | November 2005 | S5(R2) |
Contents

1. Introduction
   1.1 Purpose of the Guideline
   1.2 Aim of studies
   1.3 Choice of studies

2. Animal Criteria
   2.1 Selection and number of species
   2.2 Other test systems

3. General Recommendations Concerning Treatment
   3.1 Dosages.
   3.2 Route and frequency of administration.
   3.3 Kinetics
   3.4 Control groups

4. PROPOSED STUDY DESIGNS - COMBINATION OF STUDIES
   4.1 The most probable option
      4.1.1 Study of Fertility and Early Embryonic Development to Implantation
      4.1.2 Study for effects on pre- and postnatal development, including maternal function
      4.1.3 Study for effects on embryo-fetal development
   4.2 Single study design (rodents)
   4.3 Two study design (rodents)

5. STATISTICS

6. DATA PRESENTATION

7. TERMINOLOGY
1. INTRODUCTION

1.1 Purpose of the Guideline

There is a considerable overlap in the methodology that could be used to test chemicals and medicinal products for potential reproductive toxicity. As a first step to using this wider methodology for efficient testing, this guideline attempts to consolidate a strategy based on study designs currently in use for testing of medicinal products; it should encourage the full assessment on the safety of chemicals on the development of the offspring. It is perceived that tests in which animals are treated during defined stages of reproduction better reflect human exposure to medicinal products and allow more specific identification of stages at risk. While this approach may be useful for most medicines, long term exposure to low doses does occur and may be represented better by a one or two generation study approach.

The actual testing strategy should be determined by:
- anticipated drug use especially in relation to reproduction,
- the form of the substance and route(s) of administration intended for humans and
- making use of any existing data on toxicity, pharmaco-dynamics, kinetics, and similarity to other compounds in structure/activity.

To employ this concept successfully, flexibility is needed (Note 1). No guideline can provide sufficient information to cover all possible cases, all persons involved should be willing to discuss and consider variations in test strategy according to the state of the art and ethical standards in human and animal experimentation. Areas where more basic research would be useful for optimization of test designs are male fertility assessment, and kinetic and metabolism in pregnant/lactating animals.

Note 1 (1.1) Scientific flexibility

These guidelines are not mandatory rules, they are a starting point rather than an end point. They provide a basis from which an investigator can devise a strategy for testing according to available knowledge of the test material and the state of the art. For encouragement some alternative test designs have been mentioned in this document but there are others that can be sought out or devised. In devising a strategy, the primary objective should be to detect and bring to light any indication of toxicity to reproduction.

Fine details of study design and technical procedures have been omitted from the text. Such decisions rightly belong in the field of the investigator since a technique that may be suitable for one laboratory may not be suitable in another. The investigator needs to utilize staff and resources to do the best he or she can achieve and should know how to do this better than any outsider; human attributes of attitude, ability, and consistency are more important than material facilities. For necessary compliance to GLP, reference is made to such regulations.
1.2 Aim of studies

The aim of reproduction toxicity studies is to reveal any effect of one or more active substance(s) on mammalian reproduction. For this purpose both the investigations and the interpretation of the results should be related to all other pharmacological and toxicological data available to determine whether potential reproductive risks to humans are greater, lesser or equal to those posed by other toxicological manifestations. Further, repeated dose toxicity studies can provide important information regarding potential effects on reproduction, particularly male fertility. To extrapolate the results to humans (assess the relevance), data on likely human exposures, comparative kinetics, and mechanisms of reproductive toxicity may be helpful.

The combination of studies selected should allow exposure of mature adults and all stages of development from conception to sexual maturity. To allow detection of immediate and latent effects of exposure, observations should be continued through one complete life cycle, i.e. from conception in one generation through conception in the following generation. For convenience of testing this integrated sequence can be subdivided into the following stages.

A. Premating to conception (adult male and female reproductive functions, development and maturation of gametes, mating behavior, fertilisation).

B. Conception to implantation (adult female reproductive functions, preimplantation development, implantation).

C. Implantation to closure of the hard palate (adult female reproductive functions, embryonic development, major organ formation).

D. Closure of the hard palate to the end of pregnancy (adult female reproductive functions, fetal development and growth, organ development and growth).

E. Birth to weaning (adult female reproductive functions, neonate adaptation to extrauterine life, preweaning development and growth).

F. Weaning to sexual maturity (postweaning development and growth, adaptation to independent life, attainment of full sexual function).

For timing conventions see Note 2.

Note 2 (1.2) Timing conventions

In this guideline the convention for timing of pregnancy is to refer to the day that a sperm positive vaginal smear and/or plug is observed as day 0 of pregnancy even if mating occurs overnight. Unless shown otherwise it is assumed that, for rats, mice and rabbits implantation occurs on day 6-7 of pregnancy, and closure of the hard palate on day 15-18 of pregnancy.

Other conventions are equally acceptable but MUST be defined in reports. Also, the investigator must be consistent in different studies to assure that no gaps in treatment occur. It is an advisable precaution to provide an overlap of at least one day in the exposure period of related studies.

The accuracy of the time of mating should be specified since this will affect the variability of fetal and neonatal parameters.

Similarly, for reared litters, the day offspring are born will be considered as postnatal or lactation day 0 unless otherwise specified. However, particularly with regard to delays in, or prolongation of parturition, reference to a postcoital time frame may be useful.
1.3. Choice of studies

The guideline addresses the design of studies primarily for detection of effects on reproduction. When an effect is detected, further studies to characterise fully the nature of the response have to be designed on a case by case basis (Note 3). The rationale for the set of studies chosen should be given and should include an explanation for the choice of dosages.

Studies should be planned according to the "state of the art", and take into account preexisting knowledge of class-related effects on reproduction. They should avoid suffering and should use the minimum number of animals necessary to achieve the overall objectives. If a preliminary study is performed the results should be considered and discussed in the overall evaluation (Note 4).

Note 3 (1.3) First pass and secondary testing

To a greater or lesser degree all first pass (guideline) tests are apical in nature, i.e. an effect on one endpoint may have several different origins. A reduced litter size at birth may be due to a reduced ovulation rate (corpora lutea count), higher rate of preimplantation deaths, higher rate of postimplantation deaths or immediate postnatal deaths. In turn these deaths may be the consequence of an earlier physical malformation that can no longer be observed due to subsequent secondary changes and so on. Particularly for effects with a natural low frequency among controls, discrimination between treatment induced and coincidental occurrence is dependent upon association with other types of effects.

A toxicant usually induces more than one type of effect in a dose dependent manner. For example, induction of malformation is almost invariably associated with increased embryonic death and an increased incidence of less severe structural changes. Given an effect on one endpoint secondary investigations for possible associations should be considered i.e. the nature, scope and origins of the substance's toxicity should be characterized. Characterization should also include identification of dose-response relationships to facilitate risk assessment; this is different from the situation in first pass tests where the presence or absence of a dose-response assists discrimination between treatment related and coincidental differences.

Note 4 (1.3) Preliminary studies

At the time most reproduction studies are planned or initiated there is usually information available from acute and repeated dose toxicity studies of at least one month's duration. This information can be expected to be sufficient in identifying doses for reproductive studies. If adequate preliminary studies are performed, they are part of the justification of the choice of dose for the main study. Such studies should be submitted regardless of their GLP-status in principle. This may avoid unnecessary use of animals.
2. ANIMAL CRITERIA

The animals used must be well defined with respect to their health, fertility, fecundity, prevalence of abnormalities, embryofetal deaths and the consistency they display from study to study. Within and between studies animals should be of comparable age, weight and parity at the start; the easiest way to fulfill these criteria is to use animals that are young, mature adults at the time of mating with the females being virgin.

2.1. Selection and number of species

Studies should be conducted in mammalian species. It is generally desirable to use the same species and strain as in other toxicological studies. Reasons for using rats as the predominant rodent species are practicality, comparability with other results obtained in this species and the large amount of background knowledge accumulated.

In embryotoxicity studies only, a second mammalian species traditionally has been required, the rabbit being the preferred choice as a "non-rodent". Reasons for using rabbits in embryotoxicity studies include the extensive background knowledge that has accumulated, as well as availability and practicality. Where the rabbit is unsuitable, an alternative non-rodent or a second rodent species may be acceptable and should be considered on a case by case basis (Note 5).

Note 5 (2.1) Selection of species and strains

In choosing an animal species and strain for reproductive toxicity testing care should be given to select a relevant model. Selection of the species and strain used in other toxicology studies may avoid the need for additional preliminary studies. If it can be shown - by means of kinetic, pharmacological and toxicological data - that the species selected is a relevant model for the human, a single species can be sufficient. There is little value in using a second species if it does not show the same similarities to humans. Advantages and disadvantages of species (strains) should be considered in relation to the substance to be tested, the selected study design and in the subsequent interpretation of the results.

All species have their advantages. Rats, and to a lesser extent mice, are good, general purpose models; the rabbit has been somewhat neglected as a "non-rodent" species for repeated dose toxicity and other reproduction studies than embryotoxicity testing. It has attributes that would make it a useful model for fertility studies, especially male fertility. For both rabbits and dogs (which are often used as a second species for chronic toxicity studies) it is feasible to obtain semen samples without resorting to painful techniques (electro ejaculation) for longitudinal semen analysis. Most of the other species are not good, general purpose models and probably are best used for very specific investigations only.

All species have their disadvantages, for example:
Detection of Toxicity to Reproduction

**Rats:** sensitivity to sexual hormones, unsuitable for dopamine agonists due to dependence on prolactin as the primary hormone for establishment and maintenance of early pregnancy, highly susceptible to non-steroidal anti-inflammatory drugs in late pregnancy.

**Mice:** fast metabolic rate, stress sensitivity, malformation clusters (which occur in all species) particularly evident, small fetus.

**Rabbits:** often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, clinical signs can be difficult to interpret.

**Guinea pigs:** often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, long fetal period, insufficient historical background data.

**Domestic and/or mini pigs:** malformation clusters with variable background rate, large amounts of compound required, large housing necessary, insufficient historical background data.

**Ferrets:** seasonal breeder unless special management systems used (success highly dependent on human/animal interaction), insufficient historical background data.

**Hamsters:** intravenous route difficult if not impossible, can hide doses in the cheek pouches and can be very aggressive, sensitive to intestinal disturbance, overly sensitive teratogenic response to many chemicals, small foetus.

**Dogs:** seasonal breeders, inbreeding factors, insufficient historical background data.

**Non-human primates:** kinetically they can differ from humans as much as other species, insufficient historical background data, often numbers too low for detection of risk. They are best used when the objective of the study is to characterize a relatively certain reproductive toxicant, rather than detect a hazard.
2.2. Other test systems

Other test systems are considered to be any developing mammalian and non-mammalian cell systems, tissues, organs, or organism cultures developing independently in vitro or in vivo. Integrated with whole animal studies either for priority selection within homologous series or as secondary investigations to elucidate mechanisms of action, these systems can provide invaluable information and, indirectly, reduce the numbers of animals used in experimentation. However, they lack the complexity of the developmental processes and the dynamic interchange between the maternal and the developing organisms. These systems cannot provide assurance of the absence of effect nor provide perspective in respect of risk/exposure. In short, there are no alternative test systems to whole animals currently available for reproduction toxicity testing with the aims set out in the introduction (Note 6).

3. GENERAL RECOMMENDATIONS CONCERNING TREATMENT

3.1 Dosages

Selection of dosages is one of the most critical issues in design of the reproductive toxicity study. The choice of the high dose should be based on data from all available studies (pharmacology, acute and chronic toxicity and kinetic studies, Note 7). A repeated dose toxicity study of about 2 - 4 weeks duration provides a close approximation to the duration of treatment in segmental designs of reproductive studies. When sufficient information is not available preliminary studies are advisable (see Note 4).

Note 6 (2.2) Uses of other test systems than whole animals

Other test systems have been developed and used in preliminary investigations ("pre-screening" or priority selection) and secondary testing.

For preliminary investigation of a range of analogue series of substances it is essential that the potential outcome in whole animals is known for at least one member of the series to be studied (by inference, effects are expected). With this strategy substances can be selected for higher level testing.

For secondary testing or further substance characterisation other test systems offer the possibility to study some of the observable developmental processes in detail, e.g. to reveal specific mechanisms of toxicity, to establish concentration-response relationships, to select ‘sensitive periods’, or to detect effects of defined metabolites.

Note 7 (3.1) Selection of dosages

Using similar doses in the reproductive toxicity studies as in the repeated dose toxicity studies will allow interpretation of any potential effects on fertility in context with general systemic toxicity.

Some minimal toxicity is expected to be induced in the high dose dams.

According to the specific compound, factors limiting the high dosage determined from repeat dose toxicity studies or from preliminary reproduction studies could include:

- reduction in bodyweight gain
- increased bodyweight gain, particularly when related to perturbation of homeostatic mechanisms
- specific target organ toxicity
- haematology, clinical chemistry
- exaggerated pharmacological response, which may or may not be reflected as marked clinical reactions (e.g. sedation, convulsions)
- the physico-chemical properties of the test substance or dosage formulation which, allied to the route of administration, may impose practical limitations in the amount that can be administered. Under most circumstances 1 g/kg/day should be an adequate limit dose.
- kinetics, they can be useful in determining high dose exposure for low toxicity compounds. There is, however, little point in increasing administered dosage if it does not result in increased plasma or tissue concentration
- marked increase in embryo-fetal lethality in preliminary studies
Having determined the high dosage, lower dosages should be selected in a descending sequence, the intervals depending on kinetic and other toxicity factors. Whilst it is desirable to be able to determine a "no observed adverse effect level", priority should be given to setting dosage intervals close enough to reveal any dosage related trends that may be present (Note 8).

3.2 Route and frequency of administration.

In general the route or routes of administration should be similar to those intended for human usage. One route of substance administration may be acceptable if it can be shown that a similar distribution (kinetic profile) results from different routes (Note 9).

The usual frequency of administration is once daily but consideration should be given to use either more frequent or less frequent administration taking kinetic variables into account (see also Note 10).

3.3 Kinetics

It is preferable to have some information on kinetics before initiating reproduction studies since this may suggest the need to adjust choice of species, study design and dosing schedules. At this time the information need not be sophisticated nor derived from pregnant or lactating animals.

At the time of study evaluation further information on kinetics in pregnant or lactating animals may be required according to the results obtained (Note 10).

Note 8 (3.1) Determination of dose-response relationships

For many of the variables in reproduction studies the power to discriminate between random variation and treatment effect is poor and the presence or absence of a dosage related trend can be a critical means of determining the probability of a treatment effect. It has to be kept in mind that in these studies dose-responses may be steep, and wide intervals between doses would be inadvisable. If an analysis of dose-response relationships for the effects observed is attempted in a single study, it is recommended to use at least three dose levels and appropriate control groups. If in doubt, a fourth dose group should be added to avoid excessive dosage intervals. Such a strategy should provide a "no observed adverse effect level" for reproductive aspects. If not, the implication is that the test substance merits a greater depth of investigation and further studies.

Note 9 (3.2) Exposure by different routes of administration

If it can be shown that one route provides a greater body burden, e.g. area under the curve (AUC), there seems little reason to investigate routes that would provide a lesser body burden or which present severe practical difficulties (e.g. inhalation). Before designing new studies for a new route of administration existing data on kinetics should be used to determine the necessity of another study.

Note 10 (3.3) Kinetics in pregnant animals

Kinetic investigations in pregnant and lactating animals may pose some problems due to the rapid changes in physiology. It is best to consider this as a two or three phase approach. In planning studies kinetic data (often from non-pregnant animals) provide information on the general suitability of the species, and can assist in deciding study designs and choice of dosage. During a study kinetic investigations can provide assurance of accurate dosing or indicate marked deviations from expected patterns.
3.4. Control groups

It is recommended that control animals be dosed with the vehicle at the same rate as test group animals. When the vehicle may cause effects or affect the action of the test substance, a second (sham- or untreated) control group should be considered.

4. PROPOSED STUDY DESIGNS - COMBINATION OF STUDIES

All available pharmacological, kinetic, and toxicological data for the test compound and similar substances should be considered in deciding the most appropriate strategy and choice of study design. It is anticipated that, initially, preference will be given to designs that do not differ too radically from those of established guidelines for medicinal products (The most probable option). For most medicinal products the 3-study design will usually be adequate. Other strategies, combinations of studies and study designs could be as valid or more valid as the "most probable option" according to circumstances. The key factor is that, in total, they leave no gaps between stages and allow direct or indirect evaluation of all stages of the reproductive process (Note 11).

Designs should be justified.

4.1 The most probable option

The most probable option can be equated to a combination of studies for effects on

- Fertility and early embryonic development
- Pre- and postnatal development, including maternal function
- Embryo-fetal development

Note 11 (4) Examples for choosing other options

For compounds causing no lethality at 2 g/kg and no evidence of repeated dose toxicity at 1 g/kg, conduct of a single 2 generation study with one control and 2 test groups (0.5 and 1.0 g/kg) would seem sufficient. However, it might pose the question as to whether the correct species had been chosen or whether the compound was an effective medicine.

For compounds that may be given as a single dose, once in a life time, (e.g. diagnostics, medicines used in operations) it may be impossible to administer repeated dosages more than twice the human therapeutic dosage for any length of time. A reduced period of treatment allowing a higher dose would seem more appropriate. For females considerations of human exposure suggest little or no need for exposures beyond the embryonic period.

For dopamine agonists or compounds reducing circulating prolactin levels female rats are poor models; the rabbit would probably make a better choice for all the reproductive toxicity studies, but it does not appear to have been attempted. This also applies to other types of compound when the rabbit shows a pattern of metabolism considerably closer to humans than the rat.

For drugs where alterations in plasma kinetics are seen following repeated administration, the potential for adverse effects on embryo-fetal development may not be fully evaluated in studies according to 4.1.3. In such cases it may be desirable to extend the period of drug administration to females in a 4.1.1 study to day 17. With sacrifice at term, both fertility and embryo-fetal development can be assessed.
4.1.1 Study of Fertility and Early Embryonic Development to Implantation

**AIM**
To test for toxic effects/disturbances resulting from treatment from before mating (males/females) through mating and implantation. This comprises evaluation of stages A and B of the reproductive process (see 1.2). For females this should detect effects on the oestrous cycle, tubal transport, implantation, and development of preimplantation stages of the embryo. For males it will permit detection of functional effects (e.g. on libido, epididymal sperm maturation) that may not be detected by histological examinations of the male reproductive organs (Note 12).

**ASSESSMENT OF**
- maturation of gametes,
- mating behavior,
- fertility,
- preimplantation stages of the embryo,
- implantation.

**ANIMALS**
At least one species, preferably rats.

---

**Note 12 (4.1.1) Premating treatment**

The design of the fertility study, especially the reduction in the premating period for males, is based on evidence accumulated and reappraisal of the basic research on the process of spermatogenesis that originally prompted the demand for a prolonged premating treatment period. Compounds inducing selective effects on male reproduction are rare; mating with females is an insensitive means of detecting effects on spermatogenesis; good pathological and histopathological examination (e.g. by employing Bouin’s fixation, paraffine embedding, transverse sections of 2-4 microns for testes, longitudinal sections for epididymides, PAS and haematoxylin staining) of the male reproductive organs provides a more sensitive and quicker means of detecting effects on spermatogenesis; compounds affecting spermatogenesis almost invariably affect post meiotic stages; there is no conclusive example of a male reproductive toxicant the effects of which could be detected only by dosing males for 9-10 weeks and mating them with females.

Information on potential effects on spermatogenesis can be derived from repeated dose toxicity studies. This allows the investigations in the fertility study to be concentrated on other, more immediate, causes of effect. It is noted that the full sequence of spermatogenesis (including sperm maturation) in rats lasts 63 days. When the available evidence, or lack of it, suggests that the scope of investigations in the fertility study should be increased, or extended from detection to characterization, appropriate studies should be designed to further characterise the effects.
NUMBER OF ANIMALS
The number of animals per sex per group should be sufficient to allow meaningful interpretation of the data (Note 13).

ADMINISTRATION PERIOD
The design assumes that, especially for effects on spermatogenesis, use will be made of data from repeated dose toxicity studies of at least one month duration. Provided no effects have been found that preclude this, a premating treatment interval of 2 weeks for females and 4 weeks for males can be used (Note 12). Selection of the length of the premating administration period should be stated and justified (see also chapter 1.1, pointing out the need for research). Treatment should continue throughout mating to termination of males and at least through implantation for females. This will permit evaluation of functional effects on male fertility that cannot be detected by histologic examination in repeated dose toxicity studies and effects on mating behavior in both sexes. If data from other studies show there are effects on weight or histologic appearance of reproductive organs in males or females, or if the quality of examinations is dubious or if there are no data from other studies, then a more comprehensive study should be designed (Note 12).

Note 13 (4.1.1, 4.1.2, 4.1.3) Number of animals
There is very little scientific basis underlying specified group sizes in past and existing guidelines nor in this one. The numbers specified are educated guesses governed by the maximum study size that can be managed without undue loss of overall study control. This is indicated by the fact that the more expensive the animal is to obtain or keep, the smaller the group size proposed. Ideally, at least the same group size should be required for all species and there is a case for using larger group sizes for less frequently used species such as primates.

It should also be made clear that the numbers required depend on whether or not the group is expected to demonstrate an effect. For a high frequency effect few animals are required, to presume the absence of an effect the number required varies according to the variable (endpoint) being considered, its prevalence in control populations (rare or categorical events) or dispersion around the central tendency (continuous or semi-continuous variables). See also Note 23.

For all but the rarest events (such as malformations, abortions, total litter loss), evaluation of between 16 to 20 litters for rodents and rabbits tends to provide a degree of consistency between studies. Below 16 litters per evaluation, between study results become inconsistent, above 20 -24 litters per group consistency and precision is not greatly enhanced. These numbers relate to evaluation. If groups are subdivided for different evaluations the number of animals starting the study should be doubled. Similarly, in studies with 2 breeding generations, 16-20 litters would be required for the final evaluation of the litters of the F1 generation. To allow for natural wastage, the starting group size of the F0 generation must be larger.
**MATING**
A mating ratio of 1:1 is advisable and procedures should allow identification of both parents of a litter (Note 14).

**TERMINAL SACRIFICE**
Females may be sacrificed at any point after mid pregnancy. Males may be sacrificed at any time after mating but it is advisable to ensure successful induction of pregnancy before taking such an irrevocable step (Note 15).

**Note 14 (4.1.1) Mating**

*Mating ratios:* When both sexes are being dosed or are of equal consideration in separate male and female studies, the preferred mating ratio is 1:1 since this is the safest option in respect of obtaining good pregnancy rates and avoiding incorrect analysis and interpretation of results.

*Mating period and practices:* most laboratories would use a mating period of between 2-3 weeks, some remove females as soon as a positive vaginal smear or plug is observed whilst others leave the pairs together. Most rats will mate within the first 5 days of cohabitation (i.e. at the first available estrus), but in some cases females may become pseudopregnant. Leaving the female with the male for about 20 days allows these females to restart estrus cycles and become pregnant.

**Note 15 (4.1.1) Terminal sacrifice**

**FEMALES**
When exposure of the females ceases at implantation, termination of females between days 13-15 of pregnancy in general is adequate to assess effects on fertility or reproductive function, e.g. to differentiate between implantation and resorption sites.

In general, for detection of adverse effects, it is not thought necessary, in a fertility study, to sacrifice females at day 20/21 of pregnancy in order to gain information on late embryo loss, fetal death, and structural abnormalities.

**MALES**
It would be advisable to delay sacrifice of the males until the outcome of mating is known. In the event of an equivocal result, males could be mated with untreated females to ascertain their fertility or infertility. The males treated as part of study 4.1.1 may also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and sacrifice delayed.
OBSERVATIONS
During study:
- signs and mortalities at least once daily,
- body weight and body weight change at least twice weekly (Note 16),
- food intake at least once weekly (except during mating),
- record vaginal smears daily, at least during the mating period, to determine whether there are effects on mating or precoital time
- observations that have proved of value in other toxicity studies.

At terminal examination:
- necropsy (macroscopic examination) of all adults,
- preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison,
- preserve testes, epididymides, ovaries and uteri from all animals for possible histological examination and evaluation on a case by case basis. Tissues can be discarded after completion and reporting of the study.
- sperm count in epididymides or testes, as well as sperm viability
- count corpora lutea, implantation sites (Note 16),
- live and dead conceptuses.

4.1.2. Study for effects on pre- and postnatal development, including maternal function

**AIM**
To detect adverse effects on the pregnant/lactating female and on development of the conceptus and the offspring following exposure of the female from implantation through weaning. Since manifestations of effect induced during this period may be delayed, observations should be continued through sexual maturity (i.e. stages C to F listed in 1.2) (Notes 17, 18).

**Note 16 (4.1.1, 4.1.2, 4.1.3) Observations**

Daily weighing of pregnant females during treatment can provide useful information. Weighing an animal more frequently than twice weekly during periods other than pregnancy (premating, mating, lactation) may also be advisable for some compounds. For apparently non-pregnant rats or mice (but not rabbits), ammonium sulphide staining of the uterus might be useful to identify peri-implantation death of embryos.

**Note 17 (4.1.2) Treatment of offspring**

Consequent to derivation from existing guidelines for medicines this guideline does not fully cover exposures from weaning through puberty, nor does it deal with the possibility of reduced reproductive life span.
To detect adverse effects for medicinal products that may be used in infants and juveniles, special studies (case by case designs) involving direct treatment of offspring, at ages to be specified, should be considered.

**Note 18 (4.1.2) Separate embryotoxicity and peri-postnatal studies**

If a pre- and postnatal study is separated into two studies, one covering the embryonic period the other the fetal period, parturition, and lactation, postnatal evaluation of offspring is required in both studies.
ADVERSE EFFECTS TO BE ASSESSED
- enhanced toxicity relative to that in non-pregnant females,
- pre- and postnatal death of offspring,
- altered growth and development,
- functional deficits in offspring, including
  - behavior, maturation (puberty) and reproduction (F1).

ANIMALS
At least one species, preferably rats;

NUMBER OF ANIMALS
The number of animals per sex per group should be sufficient to allow meaningful interpretation of the data (Note 13).

ADMINISTRATION PERIOD
Females are exposed to the test substance from implantation to the end of lactation (i.e. stages C to E listed in 1.2)

EXPERIMENTAL PROCEDURE
The females are allowed to deliver and rear their offspring to weaning at which time one male and one female offspring per litter should be selected (document method used) for rearing to adulthood and mating to assess reproductive competence (Note 19).

OBSERVATIONS
During study (for maternal animals):
- signs and mortalities at least once daily,
- body weight and body weight change at least twice weekly (Note 16),
- food intake at least once weekly at least until delivery,
- observations that have proved of value in other toxicity studies,
- duration of pregnancy,
- parturition.

Note 19 (4.1.2) F1-animals
The guideline suggests selection of 1 male and 1 female per litter on the evidence that it is feasible to conduct behavioral and other functional tests on the same F1 individuals that will be used for assessment of reproductive function. This has the advantage of allowing cross referencing of performance in different tests at the individual level. It is recognised, however, that some laboratories prefer to select separate sets of animals for behavior testing and for assessment of reproductive function. Which is the most suitable for an individual laboratory will depend upon the combination of tests used and the resources available.
At terminal examination (for maternal animals and where applicable for offspring):
- necropsy (macroscopic examination) of all adults,
- preservation and possibly histological evaluation of organs with macroscopic findings; keep corresponding organs of sufficient controls for comparison
- implantations (Note 16)
- abnormalities
- live offspring at birth,
- dead offspring at birth,
- body weight at birth,
- pre- and postweaning survival and growth/body weight (Note 20), maturation and fertility,
- physical development (Note 21),
- sensory functions and reflexes (Note 21),
- behavior (Note 21).

4.1.3 Study for effects on embryo-fetal development

**AIM**
To detect adverse effects on the pregnant female and development of the embryo and fetus consequent to exposure of the female from implantation to closure of the hard palate (i.e. stages C to D listed in 1.2).

**ADVERSE EFFECTS TO BE ASSESSED**
- enhanced toxicity relative to that in non-pregnant females,
- embryofetal death,
- altered growth
- structural changes.

**ANIMALS**
Usually, two species: one rodent, preferably rats; one non-rodent, preferably rabbits (Note 5). Justification should be provided when using one species.

**NUMBER OF ANIMALS**
The number of animals should be sufficient to allow meaningful interpretation of the data (Note 13).

---

**Note 20 (4.1.2) Reduction of litter size**

The value of culling or not culling for detection of effects on reproduction is still under discussion. Whether or not culling is performed, it should be explained by the investigator.

**Note 21 (4.1.2) Physical development, sensory functions, reflexes, and behavior**

The best indicator of physical development is bodyweight. Achievement of preweaning landmarks of development such as pinna unfolding, coat growth, incisor eruption etc, is highly correlated with pup bodyweight. This weight is better related to postcoital time than postnatal time, at least when significant differences in gestation length occur. Reflexes, surface righting, auditory startle, air righting, and response to light are also dependent on physical development.

Two postweaning landmarks of development that are advised are vaginal opening of females and cleavage of the balanopreputial gland of males. The latter is associated with increasing testosterone levels whereas testis descent is not. These landmarks indicate the onset of sexual maturity and it is advised that bodyweight be recorded at the time of attainment to determine whether any differences from control are specific or related to general growth.

**Functional tests:** To date, functional tests have been directed almost exclusively to behavior. Even though a great deal of effort has been expended in this direction it is not possible to recommend specific test methods. Investigators are encouraged to find methods that will assess sensory functions, motor activity, learning and memory.
ADMINISTRATION PERIOD
The treatment period extends from implantation to the closure of the hard palate (i.e. end of C) - see 1.2.

EXPERIMENTAL PROCEDURE

Females should be killed and examined about one day prior to parturition. All fetuses should be examined for viability and abnormalities. To allow subsequent assessment of the relationship between observations made by different techniques fetuses should be individually identified (Note 22).

When using techniques requiring allocation to separate examination for soft tissue or skeletal changes, it is preferable that 50% of fetuses from each litter be allocated for skeletal examination. A minimum of 50% rat fetuses should be examined for visceral alterations, regardless of the technique used. When using fresh microdissection techniques for soft tissue alterations - which is the strongly preferred method for rabbits - 100% of rabbit fetuses should be examined for soft tissue and skeletal abnormalities.

OBSERVATIONS
During study (for maternal animals):
- signs and mortalities at least once daily,
- body weight and body weight change at least twice weekly (Note 16),
- food intake at least once weekly, and
- observations that have proved of value in other toxicity studies.

At terminal examination:
- necropsy (macroscopic examination) of all adults,
- preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison,
- count corpora lutea, numbers of live and dead implantations (Note 16),
- individual fetal body weight,
- fetal abnormalities (Note 22),
- gross evaluation of placenta

Note 22 (4.1.3) Individual identification and evaluation of fetuses

It must be possible to relate all findings by different techniques (i.e. body weight, external inspection, visceral and/or skeletal examinations) to single specimen in order to detect patterns of abnormalities. The examination of mid and low dose fetuses for visceral and/or skeletal abnormalities may not be necessary where the evaluation of the high dose and the control groups did not reveal any relevant differences. It is advisable, however, to store the fixed specimen for possible later examination. If fresh dissection techniques are normally used, difficulties with later comparisons involving fixed fetuses should be anticipated.
4.2 Single study design (rodents)

If the dosing period of the fertility study and pre- and postnatal study are combined into a single investigation, this comprises evaluation of stages A to F of the reproductive process (see 1.2). If such a study, if it includes fetal examinations, provided clearly negative results at sufficiently high exposure no further reproduction studies in rodents should be required. Fetal examinations for structural abnormalities can also be supplemented with an embryo-fetal development study (or studies) to make a 2-study approach (Note 3, 11).

Results from a study for effects on embryo-fetal development in a second species are expected (see also 4.1.3).

4.3 Two study design (rodents)

The simplest two segment design would consist of the fertility study and the pre- and postnatal development study, if it includes fetal examinations. It can be assumed, however, that if the pre- and postnatal development study provided no indication of prenatal effects at adequate margins above human exposure, the additional fetal examinations (4.1.3.) are most unlikely to provide a major change in the assessment of risk.

Alternatively, female treatment in the fertility study (4.1.1.) could be continued until closure of the hard palate and fetuses examined according to the procedures of the embryo-fetal development study (4.1.3.). This, combined with the pre- and postnatal study (4.1.2.), would provide all the examinations required in “the most probable option” but use considerably less animals (Note 3, 11).

Results from a study for effects on embryo-fetal development in a second species are expected (see also 4.1.3).
5. STATISTICS

Analysis of the statistics of a study is the means by which results are interpreted. The most important part of this analysis is to establish the relationship between the different variables and their distribution (descriptive statistics), since these determine how groups should be compared. The distributions of the endpoints observed in reproductive tests are usually non-normal and extend from almost continuous to the extreme categorical.

When employing inferential statistics (determination of statistical significance) the mating pair or litter, not the foetus or neonate, should be used as the basic unit of comparison. The tests used should be justified (Note 23).

6. DATA PRESENTATION

The key to good reporting is the tabulation of individual values in a clear concise manner to account for every animal that was entered into the study. A reader should be able to follow the history of any individual animal from initiation to termination and should be able to deduce with ease the contribution that the individual has made to any group summary values. Group summary values should be presented in a form that is biologically plausible (i.e. avoid false precision) and that reflects the distribution of the variable. Appendices or tabulations of individual values such as bodyweight, food consumption, litter values should be concise and, as far as possible, consist of absolute rather than calculated values; unnecessary duplication should be avoided.

For tabulation of low frequency observations such as clinical signs, autopsy findings, abnormalities etc. it is advisable to group together the (few) individuals with a positive recording. Especially in the presentation of data on structural changes (fetal abnormalities) the primary listing (tabulation) should clearly identify the litters containing abnormal fetuses, identify the affected fetuses in the litter and report all the changes observed in the affected fetus. Secondary listings by type of change can be derived from this, if necessary.

Note 23 (5.) Inferential statistics

"Significance" tests (inferential statistics) can be used only as a support for the interpretation of results. The interpretation itself must be based on biological plausibility. It is unwise to assume that a difference from control values is not biologically relevant simply because it is not "statistically significant". To a lesser extent it can be unwise to assume that a "statistically significant" difference must be biologically relevant. Particularly for low frequency events (e.g. embryonic death, malformations) with one sided distributions, the statistical power of studies is low. Confidence intervals for relevant quantities can indicate the likely size of the effect. When using statistical procedures, experimental units of comparison should be considered: the litter, not the individual conceptus, the mating pair, when both sexes are treated, the mating pair of the parent generation in a two generation study.
7. TERMINOLOGY

Besides effects on the reproductive competence of adult animals toxicity to reproduction includes:

**Developmental toxicity**: Any adverse effect induced prior to attainment of adult life. It includes effects induced or manifested in the embryonic or fetal period and those induced or manifested postnatally.

**Embryotoxicity, fetotoxicity, embryo-fetal toxicity**: Any adverse effect on the conceptus resulting from prenatal exposure, including structural or functional abnormalities or postnatal manifestations of such effects. Terms like "embryotoxicity", "fetotoxicity" relate to the timepoint/-period of induction of adverse effects, irrespective of the time of detection.

**One, two or three generation studies**: are defined according to the number of adult breeding generations directly exposed to the test material. For example, in a one generation study there is direct exposure of the F0 generation and indirect exposure (via the mother) of the F1 generation and the study is usually terminated at weaning of the F1 generation. In a two generation study as used for agro-chemicals and industrial chemicals there is direct exposure of the F0 generation, indirect and direct exposure of the F1 generation and indirect exposure of the F2 generation. A three generation study is defined accordingly.

**Body burden**: The total internal dosage of an individual arising from the administration of a substance, comprising parent compound and metabolites, taking distribution and accumulation into account.

**Kinetics**: The term "kinetics" is used consistently throughout this guideline, irrespective of intending to mean pharmaco- and/or toxicokinetics. No better single term was available.
PART II:
TOXICITY TO MALE FERTILITY

An Addendum to the ICH Tripartite Guideline on
DETECTION OF TOXICITY TO REPRODUCTION FOR
MEDICINAL PRODUCTS

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 29 November 1995, amended on 9 November 2000 and incorporated into the core guideline in November 2005, this guideline is recommended for adoption to the three regulatory parties to ICH

Background Note (not part of the guideline)

In June 1993 an ICH Harmonised Tripartite Guideline on Detection of Toxicity to Reproduction for Medicinal Products was adopted by the ICH Steering Committee and has since been implemented in the three ICH regions. Additional data on the duration of the premating treatment and on the parameters suited for the detection of effects on the male reproductive system and on fertility have been collected and are presented in two papers that were published in J. Amer. Coll. Toxicol. in August 1995 and also in J. Toxicol. Sci. as Special Issue in September 1995. The Japanese collaborative study which was conducted came to the conclusion that histopathology of reproductive organs is the most sensitive method for detecting effects on spermatogenesis. Sperm analysis gives similar information and could be a useful method where histopathology is not feasible, as well as for confirmative purposes and further characterisation. Mating trials were found to be sensitive for the detection of effects of sperm maturation, sperm motility and of behavioural changes (e.g., libido). The Japanese parties to ICH compared premating treatments of 4 and 9 weeks duration and found no differences in detection rate, i.e., at least 4 weeks is necessary. Data from the European literature survey show equal detection efficiency with premating treatments of 2 and 4 weeks duration. Validation studies conducted in Japan recently indicated that the 2 week study with an appropriate dose setting has equal potential in detecting the toxicity on male reproductive organs (Sakai et al 2000).

References

TOXICITY TO MALE FERTILITY

An Addendum to the ICH Tripartite Guideline on DETECTION OF TOXICITY TO REPRODUCTION FOR MEDICINAL PRODUCTS

Introduction
This text is an addendum to the ICH Tripartite Guideline on Detection of Toxicity to Reproduction for Medicinal Products\(^1\) and provides amendments to the published text.

At the time of adoption, it was accepted that the male fertility investigation, as included in the currently harmonised guideline, would require scientific and regulatory improvement and optimisation of test designs.

The amendments are intended to provide a better description of the testing concept and recommendations, especially those addressing:

- flexibility
- premating treatment duration
- observations.

The general principles and background in two papers were published in the *J. Amer. Coll. Toxicol*. These papers contain the necessary experimental data (prospective and retrospective) for reaching consensus, and have been commented. The individual data from the Japanese collaborative study were also published in the *J. Toxicol. Sci*.

AMENDMENTS

INTRODUCTION (Last Paragraph)
To employ this concept successfully, flexibility is needed (Note 1). No guideline can provide sufficient information to cover all possible cases. All persons involved should be willing to discuss and consider variations in test strategy according to the state of the art and ethical standards in human and animal experimentation.

4.1.1. Study of Fertility and Early Embryonic Development to Implantation

ADMINISTRATION PERIOD
It is assumed that, especially for effects on spermatogenesis, use will be made of available data from toxicity studies (e.g., histopathology, weight of reproductive organs, in some cases hormone assays and genotoxicity data). Provided no effects have been found in repeated dose toxicity studies of at least 2 weeks duration that preclude this, a premating treatment interval of 2 weeks for females and 2 weeks for males can be used (Note 12). Selection of the length of the premating administration period should be stated and justified. Treatment should continue throughout mating to termination of males and at least through implantation for females. This will permit evaluation of functional effects on male fertility that cannot be detected by histopathological examination in repeated dose toxicity studies and effects on mating behaviour in both sexes. If data from other studies show there are effects on weight or histology of

\(^1\) ICH Harmonised Tripartite Guideline recommended for adoption, at Step 4 of the ICH process, 24 June 1993.
reproductive organs in males or females, or if the quality of examinations is dubious or if there are no data from other studies, then the need for a more comprehensive study should be considered (Note 12).

**OBSERVATIONS**

At terminal examination:

- perform necropsy (macroscopic examination) of all adults,
- preserve organs with macroscopic findings for possible histopathological evaluation; keep corresponding organs of sufficient controls for comparison,
- preserve testes, epididymides, ovaries and uteri from all animals for possible histopathological examination and evaluation on a case-by-case basis,
- count corpora lutea, implantation sites (Note 16),
- count live and dead conceptuses.

Sperm analysis can be used as an optional procedure for confirmation or better characterisation of the effects observed (Note 12).

**Note 12 (4.1.1) Premating treatment**

The design of the fertility study, especially the reduction in the pre-mating period for males, is based on evidence accumulated and on re-appraisal of the basic research on the process of spermatogenesis. Compounds inducing selective effects on male reproduction are rare; compounds affecting spermatogenesis almost invariably affect post meiotic stages and weight of testis; mating with females is an insensitive means of detecting effects on spermatogenesis. Histopathology of the testis has been shown to be the most sensitive method for the detection of effects on spermatogenesis. Good pathological and histopathological examination (e.g., by employing Bouin's fixation, paraffin embedding, transverse section of 2-4 microns for testes, longitudinal section for epididymides, PAS and haematoxylin staining) of the male reproductive organs provides a direct means of detection. Sperm analysis (sperm counts, sperm motility, sperm morphology) can be used as an optional method to confirm findings by other methods and to characterise effects further. Sperm analysis data are considered more relevant for fertility assessment when samples from vas deferens or from cauda epididymis are used. Information on potential effects on spermatogenesis (and female reproductive organs) can be derived from repeated dose toxicity studies or reproductive toxicity studies.

For detection of effects not detectable by histopathology of male reproductive organs and sperm analysis, mating with females after a premating treatment of 4 weeks has been shown to be at least as efficient as mating after a longer duration of treatment. Since 2 week study was validated to be as effective as a 4 week study, 2 weeks treatment before mating is also acceptable. When the available evidence suggests that the scope of investigations in the fertility study should be increased, appropriate studies should be designed to characterise the effects further.
ICH Harmonised Tripartite Guideline

Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals
S6(R1)

Parent Guideline dated 16 July 1997
Current Step 4 version
Addendum dated 12 June 2011 incorporated at the end of June 2011

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
## S6(R1) Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>November 2005</td>
</tr>
</tbody>
</table>

### Parent Guideline: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

<table>
<thead>
<tr>
<th>S6</th>
<th>Approval by the Steering Committee under Step 2 and release for public consultation.</th>
<th>6 November 1996</th>
<th>S6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6</td>
<td>Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies.</td>
<td>16 July 1997</td>
<td>S6</td>
</tr>
</tbody>
</table>

### Addendum to the Parent Guideline

<table>
<thead>
<tr>
<th>S6(R1)</th>
<th>Approval by the Steering Committee under Step 2 and release for public consultation.</th>
<th>29 October 2009</th>
<th>S6(R1)</th>
</tr>
</thead>
</table>

**Current Step 4 version**

<table>
<thead>
<tr>
<th>S6(R1)</th>
<th>Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Addendum has been incorporated into the parent Guideline which is now renamed S6(R1).</td>
</tr>
<tr>
<td></td>
<td>12 June 2011</td>
</tr>
<tr>
<td></td>
<td>S6(R1)</td>
</tr>
</tbody>
</table>
# Table of Contents

## Part I: Introduction

1. **Introduction** ................................................................. 1  
1.1 Background .............................................................................. 1  
1.2 Objectives ............................................................................... 1  
1.3 Scope ..................................................................................... 1  

## Part II: Specific Considerations

4. **Specific Considerations** ...................................................... 6  
4.1 Safety Pharmacology ............................................................ 6  
4.2 Exposure Assessment .......................................................... 6  
4.2.1 Pharmacokinetics and Toxicokinetics ................................. 6  
4.2.2 Assays ............................................................................... 6  
4.2.3 Metabolism ........................................................................ 7  
4.3 Single Dose Toxicity Studies .................................................. 7  
4.4 Repeated Dose Toxicity Studies .............................................. 7  
4.5 Immunotoxicity Studies ........................................................ 7  
4.6 Reproductive Performance and Developmental Toxicity Studies .............................................................................. 8  
4.7 Genotoxicity Studies .............................................................. 8  
4.8 Carcinogenicity Studies ........................................................ 8  
4.9 Local Tolerance Studies ......................................................... 9  

## Notes

NOTES ..................................................................................... 9  

## Part II: Introduction

1. **Introduction** ................................................................. 10  
1.1 Purpose of the Addendum .................................................... 10  
1.2 Background .......................................................................... 10  
1.3 Scope of the Guideline ........................................................ 10
2. **SPECIES SELECTION** ........................................................................................................ 11
   2.1 General Principles ....................................................................................................... 11
   2.2 One or Two Species .................................................................................................... 11
   2.3 Use of Homologous Proteins .................................................................................... 12
3. **STUDY DESIGN** ........................................................................................................... 12
   3.1 Dose Selection and Application of PK/PD Principles ............................................... 12
   3.2 Duration of Studies .................................................................................................... 12
   3.3 Recovery .................................................................................................................... 13
   3.4 Exploratory Clinical Trials ....................................................................................... 13
4. **IMMUNOGENICITY** ................................................................................................ 13
5. **REPRODUCTIVE AND DEVELOPMENTAL TOXICITY** ........................................ 13
6. **CARCINOGENICITY** ............................................................................................... 16
   **NOTES** ...................................................................................................................... 16
   **REFERENCES** ......................................................................................................... 19
PART I:

PRECLINICAL SAFETY EVALUATION OF
BIOTECHNOLOGY-DERIVED PHARMACEUTICALS
ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 16 July 1997, this Guideline is recommended for adoption to the three regulatory parties to ICH

1. INTRODUCTION

1.1 Background
Biotechnology-derived pharmaceuticals (biopharmaceuticals) were initially developed in the early 1980s. The first marketing authorisations were granted later in the decade. Several guidelines and points-to-consider documents have been issued by various regulatory agencies regarding safety assessment of these products. Review of such documents, which are available from regulatory authorities, may provide useful background in developing new biopharmaceuticals.

Considerable experience has now been gathered with submission of applications for biopharmaceuticals. Critical review of this experience has been the basis for development of this guidance that is intended to provide general principles for designing scientifically acceptable preclinical safety evaluation programs.

1.2 Objectives
Regulatory standards for biotechnology-derived pharmaceuticals have generally been comparable among the European Union, Japan and United States. All regions have adopted a flexible, case-by-case, science-based approach to preclinical safety evaluation needed to support clinical development and marketing authorisation. In this rapidly evolving scientific area, there is a need for common understanding and continuing dialogue among the regions.

The primary goals of preclinical safety evaluation are: 1) to identify an initial safe dose and subsequent dose escalation schemes in humans; 2) to identify potential target organs for toxicity and for the study of whether such toxicity is reversible; and 3) to identify safety parameters for clinical monitoring. Adherence to the principles presented in this document is intended to improve the quality and consistency of the preclinical safety data supporting the development of biopharmaceuticals.

1.3 Scope
This guidance is intended primarily to recommend a basic framework for the preclinical safety evaluation of biotechnology-derived pharmaceuticals. It applies to products derived from characterised cells through the use of a variety of expression systems including bacteria, yeast, insect, plant, and mammalian cells. The intended indications may include in vivo diagnostic, therapeutic, or prophylactic uses. The active substances include proteins and peptides, their derivatives and products of which they are components; they could be derived from cell cultures or produced using recombinant DNA technology including production by transgenic plants and animals. Examples include but are not limited to: cytokines, plasminogen activators,
recombinant plasma factors, growth factors, fusion proteins, enzymes, receptors, hormones, and monoclonal antibodies.

The principles outlined in this guidance may also be applicable to recombinant DNA protein vaccines, chemically synthesised peptides, plasma derived products, endogenous proteins extracted from human tissue, and oligonucleotide drugs.

This document does not cover antibiotics, allergenic extracts, heparin, vitamins, cellular blood components, conventional bacterial or viral vaccines, DNA vaccines, or cellular and gene therapies.

2. Specification of the Test Material

Safety concerns may arise from the presence of impurities or contaminants. It is preferable to rely on purification processes to remove impurities and contaminants rather than to establish a preclinical testing program for their qualification. In all cases, the product should be sufficiently characterised to allow an appropriate design of preclinical safety studies.

There are potential risks associated with host cell contaminants derived from bacteria, yeast, insect, plants, and mammalian cells. The presence of cellular host contaminants can result in allergic reactions and other immunopathological effects. The adverse effects associated with nucleic acid contaminants are theoretical but include potential integration into the host genome. For products derived from insect, plant and mammalian cells, or transgenic plants and animals there may be an additional risk of viral infections.

In general, the product that is used in the definitive pharmacology and toxicology studies should be comparable to the product proposed for the initial clinical studies. However, it is appreciated that during the course of development programs, changes normally occur in the manufacturing process in order to improve product quality and yields. The potential impact of such changes for extrapolation of the animal findings to humans should be considered.

The comparability of the test material during a development program should be demonstrated when a new or modified manufacturing process or other significant changes in the product or formulation are made in an ongoing development program. Comparability can be evaluated on the basis of biochemical and biological characterisation (i.e., identity, purity, stability, and potency). In some cases additional studies may be needed (i.e., pharmacokinetics, pharmacodynamics and/or safety). The scientific rationale for the approach taken should be provided.

3. Preclinical Safety Testing

3.1 General Principles

The objectives of the preclinical safety studies are to define pharmacological and toxicological effects not only prior to initiation of human studies but throughout clinical development. Both in vitro and in vivo studies can contribute to this characterisation. Biopharmaceuticals that are structurally and pharmacologically comparable to a product for which there is wide experience in clinical practice may need less extensive toxicity testing.

Preclinical safety testing should consider:

1) selection of the relevant animal species;
2) age;
3) physiological state;
4) the manner of delivery, including dose, route of administration, and treatment regimen; and
5) stability of the test material under the conditions of use.

Toxicity studies are expected to be performed in compliance with Good Laboratory Practice (GLP); however, it is recognised that some studies employing specialised test systems which are often needed for biopharmaceuticals, may not be able to comply fully with GLP. Areas of non-compliance should be identified and their significance evaluated relative to the overall safety assessment. In some cases, lack of full GLP compliance does not necessarily mean that the data from these studies cannot be used to support clinical trials and marketing authorisations.

Conventional approaches to toxicity testing of pharmaceuticals may not be appropriate for biopharmaceuticals due to the unique and diverse structural and biological properties of the latter that may include species specificity, immunogenicity, and unpredicted pleiotropic activities.

### 3.2 Biological Activity/Pharmacodynamics

Biological activity may be evaluated using *in vitro* assays to determine which effects of the product may be related to clinical activity. The use of cell lines and/or primary cell cultures can be useful to examine the direct effects on cellular phenotype and proliferation. Due to the species specificity of many biotechnology-derived pharmaceuticals, it is important to select relevant animal species for toxicity testing. *In vitro* cell lines derived from mammalian cells can be used to predict specific aspects of *in vivo* activity and to assess quantitatively the relative sensitivity of various species (including human) to the biopharmaceutical. Such studies may be designed to determine, for example, receptor occupancy, receptor affinity, and/or pharmacological effects, and to assist in the selection of an appropriate animal species for further *in vivo* pharmacology and toxicology studies. The combined results from *in vitro* and *in vivo* studies assist in the extrapolation of the findings to humans. *In vivo* studies to assess pharmacological activity, including defining mechanism(s) of action, are often used to support the rationale of the proposed use of the product in clinical studies.

For monoclonal antibodies, the immunological properties of the antibody should be described in detail, including its antigenic specificity, complement binding, and any unintentional reactivity and/or cytotoxicity towards human tissues distinct from the intended target. Such cross-reactivity studies should be carried out by appropriate immunohistochemical procedures using a range of human tissues.

### 3.3 Animal Species/Model Selection

The biological activity together with species and/or tissue specificity of many biotechnology-derived pharmaceuticals often preclude standard toxicity testing designs in commonly used species (e.g., rats and dogs). Safety evaluation programs should include the use of relevant species. A relevant species is one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies). A variety of techniques (e.g., immunochemical or functional tests) can be used to identify a relevant species. Knowledge of receptor/epitope distribution can provide greater understanding of potential *in vivo* toxicity.
Relevant animal species for testing of monoclonal antibodies are those that express the desired epitope and demonstrate a similar tissue cross-reactivity profile as for human tissues. This would optimise the ability to evaluate toxicity arising from the binding to the epitope and any unintentional tissue cross-reactivity. An animal species which does not express the desired epitope may still be of some relevance for assessing toxicity if comparable unintentional tissue cross-reactivity to humans is demonstrated.

Safety evaluation programs should normally include two relevant species. However, in certain justified cases one relevant species may suffice (e.g., when only one relevant species can be identified or where the biological activity of the biopharmaceutical is well understood). In addition even where two species may be necessary to characterise toxicity in short term studies, it may be possible to justify the use of only one species for subsequent long term toxicity studies (e.g., if the toxicity profile in the two species is comparable in the short term).

Toxicity studies in non-relevant species may be misleading and are discouraged. When no relevant species exists, the use of relevant transgenic animals expressing the human receptor or the use of homologous proteins should be considered. The information gained from use of a transgenic animal model expressing the human receptor is optimised when the interaction of the product and the humanised receptor has similar physiological consequences to those expected in humans. While useful information may also be gained from the use of homologous proteins, it should be noted that the production process, range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use. Where it is not possible to use transgenic animal models or homologous proteins, it may still be prudent to assess some aspects of potential toxicity in a limited toxicity evaluation in a single species, e.g., a repeated dose toxicity study of ≤ 14 days duration that includes an evaluation of important functional endpoints (e.g., cardiovascular and respiratory).

In recent years, there has been much progress in the development of animal models that are thought to be similar to the human disease. These animal models include induced and spontaneous models of disease, gene knockout(s), and transgenic animals. These models may provide further insight, not only in determining the pharmacological action of the product, pharmacokinetics, and dosimetry, but may also be useful in the determination of safety (e.g., evaluation of undesirable promotion of disease progression). In certain cases, studies performed in animal models of disease may be used as an acceptable alternative to toxicity studies in normal animals (Note 1). The scientific justification for the use of these animal models of disease to support safety should be provided.

3.4 Number/Gender of Animals

The number of animals used per dose has a direct bearing on the ability to detect toxicity. A small sample size may lead to failure to observe toxic events due to observed frequency alone regardless of severity. The limitations that are imposed by sample size, as often is the case for non-human primate studies, may be in part compensated by increasing the frequency and duration of monitoring. Both genders should generally be used or justification given for specific omissions.

3.5 Administration/Dose Selection

The route and frequency of administration should be as close as possible to that proposed for clinical use. Consideration should be given to pharmacokinetics and bioavailability of the product in the species being used, and the volume which can be
safely and humanely administered to the test animals. For example, the frequency of administration in laboratory animals may be increased compared to the proposed schedule for the human clinical studies in order to compensate for faster clearance rates or low solubility of the active ingredient. In these cases, the level of exposure of the test animal relative to the clinical exposure should be defined. Consideration should also be given to the effects of volume, concentration, formulation, and site of administration. The use of routes of administration other than those used clinically may be acceptable if the route must be modified due to limited bioavailability, limitations due to the route of administration, or to size/physiology of the animal species.

Dosage levels should be selected to provide information on a dose-response relationship, including a toxic dose and a no observed adverse effect level (NOAEL). For some classes of products with little to no toxicity it may not be possible to define a specific maximum dose. In these cases, a scientific justification of the rationale for the dose selection and projected multiples of human exposure should be provided. To justify high dose selection, consideration should be given to the expected pharmacological/physiological effects, availability of suitable test material, and the intended clinical use. Where a product has a lower affinity to or potency in the cells of the selected species than in human cells, testing of higher doses may be important. The multiples of the human dose that are needed to determine adequate safety margins may vary with each class of biotechnology-derived pharmaceutical and its clinical indication(s).

3.6 Immunogenicity

Many biotechnology-derived pharmaceuticals intended for human are immunogenic in animals. Therefore, measurement of antibodies associated with administration of these types of products should be performed when conducting repeated dose toxicity studies in order to aid in the interpretation of these studies. Antibody responses should be characterised (e.g., titer, number of responding animals, neutralising or non-neutralising), and their appearance should be correlated with any pharmacological and/or toxicological changes. Specifically, the effects of antibody formation on pharmacokinetic/pharmacodynamic parameters, incidence and/or severity of adverse effects, complement activation, or the emergence of new toxic effects should be considered when interpreting the data. Attention should also be paid to the evaluation of possible pathological changes related to immune complex formation and deposition.

The detection of antibodies should not be the sole criterion for the early termination of a preclinical safety study or modification in the duration of the study design unless the immune response neutralises the pharmacological and/or toxicological effects of the biopharmaceutical in a large proportion of the animals. In most cases, the immune response to biopharmaceuticals is variable, like that observed in humans. If the interpretation of the data from the safety study is not compromised by these issues, then no special significance should be ascribed to the antibody response.

The induction of antibody formation in animals is not predictive of a potential for antibody formation in humans. Humans may develop serum antibodies against humanised proteins, and frequently the therapeutic response persists in their presence. The occurrence of severe anaphylactic responses to recombinant proteins is rare in humans. In this regard, the results of guinea pig anaphylaxis tests, which are generally positive for protein products, are not predictive for reactions in humans; therefore, such studies are considered of little value for the routine evaluation of these types of products.
4. **Specific Considerations**

4.1 **Safety Pharmacology**

It is important to investigate the potential for undesirable pharmacological activity in appropriate animal models and, where necessary, to incorporate particular monitoring for these activities in the toxicity studies and/or clinical studies. Safety pharmacology studies measure functional indices of potential toxicity. These functional indices may be investigated in separate studies or incorporated in the design of toxicity studies. The aim of the safety pharmacology studies should be to reveal any functional effects on the major physiological systems (e.g., cardiovascular, respiratory, renal, and central nervous systems). Investigations may also include the use of isolated organs or other test systems not involving intact animals. All of these studies may allow for a mechanistically-based explanation of specific organ toxicities, which should be considered carefully with respect to human use and indication(s).

4.2 **Exposure Assessment**

4.2.1 **Pharmacokinetics and Toxicokinetics**

It is difficult to establish uniform guidelines for pharmacokinetic studies for biotechnology-derived pharmaceuticals. Single and multiple dose pharmacokinetics, toxicokinetics, and tissue distribution studies in relevant species are useful; however, routine studies that attempt to assess mass balance are not useful. Differences in pharmacokinetics among animal species may have a significant impact on the predictiveness of animal studies or on the assessment of dose response relationships in toxicity studies. Alterations in the pharmacokinetic profile due to immune-mediated clearance mechanisms may affect the kinetic profiles and the interpretation of the toxicity data. For some products there may also be inherent, significant delays in the expression of pharmacodynamic effects relative to the pharmacokinetic profile (e.g., cytokines) or there may be prolonged expression of pharmacodynamic effects relative to plasma levels.

Pharmacokinetic studies should, whenever possible, utilise preparations that are representative of that intended for toxicity testing and clinical use, and employ a route of administration that is relevant to the anticipated clinical studies. Patterns of absorption may be influenced by formulation, concentration, site, and/or volume. Whenever possible, systemic exposure should be monitored during the toxicity studies.

When using radiolabeled proteins, it is important to show that the radiolabeled test material maintains activity and biological properties equivalent to that of the unlabeled material. Tissue concentrations of radioactivity and/or autoradiography data using radiolabeled proteins may be difficult to interpret due to rapid *in vivo* metabolism or unstable radiolabeled linkage. Care should be taken in the interpretation of studies using radioactive tracers incorporated into specific amino acids because of recycling of amino acids into non-drug related proteins/peptides.

Some information on absorption, disposition and clearance in relevant animal models should be available prior to clinical studies in order to predict margins of safety based upon exposure and dose.

4.2.2 **Assays**

The use of one or more assay methods should be addressed on a case-by-case basis and the scientific rationale should be provided. One validated method is usually considered sufficient. For example, quantitation of TCA-precipitable radioactivity
following administration of a radiolabeled protein may provide adequate information, but a specific assay for the analyte is preferred. Ideally the assay methods should be the same for animals and humans. The possible influence of plasma binding proteins and/or antibodies in plasma/serum on the assay performance should be determined.

4.2.3 Metabolism
The expected consequence of metabolism of biotechnology-derived pharmaceuticals is the degradation to small peptides and individual amino acids. Therefore, the metabolic pathways are generally understood. Classical biotransformation studies as performed for pharmaceuticals are not needed.

Understanding the behaviour of the biopharmaceutical in the biologic matrix, (e.g., plasma, serum, cerebral spinal fluid) and the possible influence of binding proteins is important for understanding the pharmacodynamic effect.

4.3 Single Dose Toxicity Studies
Single dose studies may generate useful data to describe the relationship of dose to systemic and/or local toxicity. These data can be used to select doses for repeated dose toxicity studies. Information on dose-response relationships may be gathered through the conduct of a single dose toxicity study, as a component of pharmacology or animal model efficacy studies. The incorporation of safety pharmacology parameters in the design of these studies should be considered.

4.4 Repeated Dose Toxicity Studies
For consideration of the selection of animal species for repeated dose studies see Section 3.3. The route and dosing regimen (e.g., daily versus intermittent dosing) should reflect the intended clinical use or exposure. When feasible, these studies should include toxicokinetics.

A recovery period should generally be included in study designs to determine the reversal or potential worsening of pharmacological/toxicological effects, and/or potential delayed toxic effects. For biopharmaceuticals that induce prolonged pharmacological/toxicological effects, recovery group animals should be monitored until reversibility is demonstrated. The duration of repeated dose studies should be based on the intended duration of clinical exposure and disease indication. This duration of animal dosing has generally been 1-3 months for most biotechnology-derived pharmaceuticals. For biopharmaceuticals intended for short-term use (e.g., ≤ to 7 days) and for acute life-threatening diseases, repeated dose studies up to two weeks duration have been considered adequate to support clinical studies as well as marketing authorisation. For those biopharmaceuticals intended for chronic indications, studies of 6 months duration have generally been appropriate although in some cases shorter or longer durations have supported marketing authorisations. For biopharmaceuticals intended for chronic use, the duration of long term toxicity studies should be scientifically justified.

4.5 Immunotoxicity Studies
One aspect of immunotoxicological evaluation includes assessment of potential immunogenicity (see Section 3.6). Many biotechnology-derived pharmaceuticals are intended to stimulate or suppress the immune system and therefore may affect not only humoral but also cell-mediated immunity. Inflammatory reactions at the injection site may be indicative of a stimulatory response. It is important, however, to recognise that simple injection trauma and/or specific toxic effects caused by the formulation vehicle may also result in toxic changes at the injection site. In addition,
the expression of surface antigens on target cells may be altered, which has implications for autoimmune potential. Immunotoxicological testing strategies may require screening studies followed by mechanistic studies to clarify such issues. Routine tiered testing approaches or standard testing batteries, however, are not recommended for biotechnology-derived pharmaceuticals.

4.6 Reproductive Performance and Developmental Toxicity Studies
The need for reproductive/developmental toxicity studies is dependent upon the product, clinical indication and intended patient population (Note 2). The specific study design and dosing schedule may be modified based on issues related to species specificity, immunogenicity, biological activity and/or a long elimination half-life. For example, concerns regarding potential developmental immunotoxicity, which may apply particularly to certain monoclonal antibodies with prolonged immunological effects, could be addressed in a study design modified to assess immune function of the neonate.

4.7 Genotoxicity Studies
The range and type of genotoxicity studies routinely conducted for pharmaceuticals are not applicable to biotechnology-derived pharmaceuticals and therefore are not needed. Moreover, the administration of large quantities of peptides/proteins may yield uninterpretable results. It is not expected that these substances would interact directly with DNA or other chromosomal material (Note 3).

Studies in available and relevant systems, including newly developed systems, should be performed in those cases where there is cause for concern about the product (e.g., because of the presence of an organic linker molecule in a conjugated protein product). The use of standard genotoxicity studies for assessing the genotoxic potential of process contaminants is not considered appropriate. If performed for this purpose, however, the rationale should be provided.

4.8 Carcinogenicity Studies
Standard carcinogenicity bioassays are generally inappropriate for biotechnology-derived pharmaceuticals. However, product-specific assessment of carcinogenic potential may still be needed depending upon duration of clinical dosing, patient population and/or biological activity of the product (e.g., growth factors, immunosuppressive agents, etc.) When there is a concern about carcinogenic potential a variety of approaches may be considered to evaluate risk.

Products that may have the potential to support or induce proliferation of transformed cells and clonal expansion possibly leading to neoplasia should be evaluated with respect to receptor expression in various malignant and normal human cells that are potentially relevant to the patient population under study. The ability of the product to stimulate growth of normal or malignant cells expressing the receptor should be determined. When in vitro data give cause for concern about carcinogenic potential, further studies in relevant animal models may be needed. Incorporation of sensitive indices of cellular proliferation in long term repeated dose toxicity studies may provide useful information.

In those cases where the product is biologically active and non-immunogenic in rodents and other studies have not provided sufficient information to allow an assessment of carcinogenic potential then the utility of a single rodent species should be considered. Careful consideration should be given to the selection of doses. The use of a combination of pharmacokinetic and pharmacodynamic endpoints with consideration of comparative receptor characteristics and intended human exposures
represents the most scientifically based approach for defining the appropriate doses. The rationale for the selection of doses should be provided.

4.9 Local Tolerance Studies
Local tolerance should be evaluated. The formulation intended for marketing should be tested; however, in certain justified cases, the testing of representative formulations may be acceptable. In some cases, the potential adverse effects of the product can be evaluated in single or repeated dose toxicity studies thus obviating the need for separate local tolerance studies.

NOTES

Note 1 Animal models of disease may be useful in defining toxicity endpoints, selection of clinical indications, and determination of appropriate formulations, route of administration, and treatment regimen. It should be noted that with these models of disease there is often a paucity of historical data for use as a reference when evaluating study results. Therefore, the collection of concurrent control and baseline data is critical to optimise study design.

Note 2 There may be extensive public information available regarding potential reproductive and/or developmental effects of a particular class of compounds (e.g., interferons) where the only relevant species is the non-human primate. In such cases, mechanistic studies indicating that similar effects are likely to be caused by a new but related molecule, may obviate the need for formal reproductive/developmental toxicity studies. In each case, the scientific basis for assessing the potential for possible effects on reproduction/development should be provided.

Note 3 With some biopharmaceuticals there is a potential concern about accumulation of spontaneously mutated cells (e.g., via facilitating a selective advantage of proliferation) leading to carcinogenicity. The standard battery of genotoxicity tests is not designed to detect these conditions. Alternative in vitro or in vivo models to address such concerns may have to be developed and evaluated.
PART II:
ADDENDUM TO S6
PRECLINICAL SAFETY EVALUATION OF
BIOTECHNOLOGY-DERIVED PHARMACEUTICALS
ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process on 12 June 2011 and been incorporated in the parent Guideline at the end of June 2011, this Guideline is recommended for adoption to the three regulatory parties to ICH

Preamble:
This addendum should be read in close conjunction with the original ICH S6 Guideline. In general the addendum is complementary to the guideline, and where the addendum differs from the original guideline, the guidance in the addendum prevails.

1. INTRODUCTION

1.1 Purpose of the Addendum
The purpose of the addendum is to complement and provide clarification on and update the following topics discussed in the original ICH S6 Guideline: species selection, study design, immunogenicity, reproductive and developmental toxicity and assessment of carcinogenic potential. Scientific advances and experience gained since publication of the original ICH S6 Guideline call for this addendum. This harmonised addendum will help to define the current recommendations and reduce the likelihood that substantial differences will exist among regions.

This guidance should facilitate the timely conduct of clinical trials, reduce the use of animals in accordance with the 3Rs (reduce/refine/replace) principles and reduce the use of other drug development resources. Although not discussed in this guidance, consideration should be given to the use of appropriate in vitro alternative methods for safety evaluation. These methods, if accepted by all ICH regulatory authorities, can be used to replace current standard methods.

This guidance promotes safe and ethical development and availability of new pharmaceuticals.

1.2 Background
The recommendations of this addendum further harmonise the nonclinical safety studies to support the various stages of clinical development among the regions of European Union (EU), Japan, and the United States. The present addendum represents the consensus that exists regarding the safety evaluation of biotechnology-derived pharmaceuticals.

1.3 Scope of the Guideline
This addendum does not alter the scope of the original ICH S6 Guideline. For biotechnology-derived products intended to be used in oncology the Guidance on Nonclinical Evaluation for Anticancer Pharmaceuticals (ICH S9 Guideline) should be consulted.
2. **Species Selection**

2.1 **General Principles**

A number of factors should be taken into account when determining species relevancy. Comparisons of target sequence homology between species can be an appropriate starting point, followed by *in vitro* assays to make qualitative and quantitative cross-species comparisons of relative target binding affinities and receptor/ligand occupancy and kinetics.

Assessments of functional activity are also recommended. Functional activity can be demonstrated in species-specific cell-based systems and/or *in vivo* pharmacology or toxicology studies. Modulation of a known biologic response or of a pharmacodynamic (PD) marker can provide evidence for functional activity to support species relevance.

Consideration of species differences in target binding and functional activity in the context of the intended dosing regime should provide confidence that a model is capable of demonstrating potentially adverse consequences of target modulation. When the target is expressed at very low levels in typical healthy preclinical species (e.g., inflammatory cytokines or tumour antigens), binding affinity and activity in cell-based systems can be sufficient to guide species selection.

Assessment of tissue cross-reactivity in animal tissues is of limited value for species selection (see *Note 1*). However, in specific cases (i.e., where the approaches described above cannot be used to demonstrate a pharmacologically relevant species) tissue cross-reactivity (TCR) studies can be used to guide selection of toxicology species by comparison of tissue binding profiles in human and those animal tissues where target binding is expected.

As described in ICH S6 Guideline, when no relevant species can be identified because the biopharmaceutical does not interact with the orthologous target in any species, use of homologous molecules or transgenic models can be considered.

For monoclonal antibodies and other related antibody products directed at foreign targets (i.e., bacterial, viral targets etc.), a short-term safety study (see ICH S6 Guideline) in one species (choice of species to be justified by the sponsor) can be considered; no additional toxicity studies, including reproductive toxicity studies, are appropriate. Alternatively, when animal models of disease are used to evaluate proof of principle, a safety assessment can be included to provide information on potential target-associated safety aspects. Where this is not feasible, appropriate risk mitigation strategies should be adopted for clinical trials.

Species selection for an antibody-drug/toxin conjugate (ADC) incorporating a novel toxin/toxicant should follow the same general principles as an unconjugated antibody (see above and see *Note 2*).

2.2 **One or Two Species**

If there are two pharmacologically relevant species for the clinical candidate (one rodent and one non-rodent), then both species should be used for short-term (up to 1 month duration) general toxicity studies. If the toxicological findings from these studies are similar or the findings are understood from the mechanism of action of the product, then longer-term general toxicity studies in one species are usually considered sufficient. The rodent species should be considered unless there is a scientific rationale for using non-rodents. Studies in two non-rodent species are not appropriate.
The use of one species for all general toxicity studies is justified when the clinical candidate is pharmacologically active in only one species. Studies in a second species with a homologous product are not considered to add further value for risk assessment and are not recommended.

2.3 Use of Homologous Proteins
Use of homologous proteins is one of the alternative approaches described under ICH S6 Guideline Section 3.3. Studies with homologous proteins can be used for hazard detection and understanding the potential for adverse effects due to exaggerated pharmacology, but are generally not useful for quantitative risk assessment. Therefore, for the purposes of hazard identification it can be possible to conduct safety evaluation studies using a control group and one treatment group provided there is a scientific justification for the study design and dose selected (e.g., maximum pharmacological dose).

3. STUDY DESIGN

3.1 Dose Selection and Application of PK/PD Principles
The toxicity of most biopharmaceuticals is related to their targeted mechanism of action; therefore, relatively high doses can elicit adverse effects which are apparent as exaggerated pharmacology.

A rationale should be provided for dose selection taking into account the characteristics of the dose-response relationship. Pharmacokinetic-pharmacodynamic (PK-PD) approaches (e.g., simple exposure-response relationships or more complex modeling and simulation approaches) can assist in high dose selection by identifying 1) a dose which provides the maximum intended pharmacological effect in the preclinical species; and 2) a dose which provides an approximately 10-fold exposure multiple over the maximum exposure to be achieved in the clinic. The higher of these two doses should be chosen for the high dose group in preclinical toxicity studies unless there is a justification for using a lower dose (e.g., maximum feasible dose).

Where in vivo/ex vivo PD endpoints are not available, the high dose selection can be based on PK data and available in vitro binding and/or pharmacology data. Corrections for differences in target binding and in vitro pharmacological activity between the nonclinical species and humans should be taken into account to adjust the exposure margin over the highest anticipated clinical exposure. For example, a large relative difference in binding affinity and/or in vitro potency might suggest that testing higher doses in the nonclinical studies is appropriate. In the event that toxicity cannot be demonstrated at the doses selected using this approach, then additional toxicity studies at higher multiples of human dosing are unlikely to provide additional useful information.

3.2 Duration of Studies
For chronic use products, repeat dose toxicity studies of 6 months duration in rodents or non-rodents are considered sufficient, providing the high dose is selected in accordance with the principles above in Section 3.1. Studies of longer duration have not generally provided useful information that changed the clinical course of development.

For chronic use of biopharmaceutical products developed for patients with advanced cancer, the principles for duration of toxicology studies are outlined in ICH S9 Guideline.
3.3 Recovery
Recovery from pharmacological and toxicological effects with potential adverse clinical impact should be understood when they occur at clinically relevant exposures. This information can be obtained by an understanding that the particular effect observed is generally reversible/non-reversible or by including a non-dosing period in at least one study, at at least one dose level, to be justified by the sponsor. The purpose of the non-dosing period is to examine reversibility of these effects, not to assess delayed toxicity. The demonstration of complete recovery is not considered essential. The addition of a recovery period just to assess potential for immunogenicity is not required.

3.4 Exploratory Clinical Trials
The flexible approaches to support exploratory clinical trials as outlined in ICH M3(R2) Guideline can be applicable to biopharmaceuticals. It is recommended that these approaches be discussed and agreed upon with the appropriate regulatory authority.

4. IMMUNOGENICITY
Immunogenicity assessments are conducted to assist in the interpretation of the study results and design of subsequent studies. Such analyses in nonclinical animal studies are not relevant in terms of predicting potential immunogenicity of human or humanized proteins in humans.

Measurement of anti-drug antibodies (ADA) in nonclinical studies should be evaluated when there is 1) evidence of altered PD activity; 2) unexpected changes in exposure in the absence of a PD marker; or 3) evidence of immune-mediated reactions (immune complex disease, vasculitis, anaphylaxis, etc.). Since, it is difficult to predict whether such analysis will be called for prior to completion of the in-life phase of the study, it is often useful to obtain appropriate samples during the course of the study, which can subsequently be analyzed when warranted to aid in interpretation of the study results. When ADAs are detected, their impact on the interpretation of the study results should be assessed (see also Part I, Section 3.6, Paragraph 2 for further guidance on the impact of immunogenicity).

Characterization of neutralizing potential is warranted when ADAs are detected and there is no PD marker to demonstrate sustained activity in the in vivo toxicology studies. Neutralizing antibody activity can be assessed indirectly with ex vivo bioactivity assay or an appropriate combination of assay formats for PK-PD, or directly in a specific neutralizing antibody assay.

5. REPRODUCTIVE AND DEVELOPMENTAL TOXICITY
5.1 General Comments
Reproductive toxicity studies should be conducted in accordance with the principles outlined in ICH S5(R2) Guideline; however, the specific study design and dosing schedule can be modified based on an understanding of species specificity, the nature of the product and mechanism of action, immunogenicity and/or pharmacokinetic behaviour and embryo-fetal exposure.

An assessment of reproductive toxicity with the clinical candidate in a relevant species is generally preferred. The evaluation of toxicity to reproduction should be conducted only in pharmacologically relevant species. When the clinical candidate is pharmacologically active in rodents and rabbits, both species should be used for
embryo-fetal development (EFD) studies, unless embryo-fetal lethality or teratogenicity has been identified in one species.

Developmental toxicity studies should only be conducted in non-human primates (NHPs) when they are the only relevant species.

When the clinical candidate is pharmacologically active only in NHPs, there is still a preference to test the clinical candidate. However an alternative model can be used in place of NHPs if appropriate scientific justification is provided.

When no relevant animal species exists for testing the clinical candidate, the use of transgenic mice expressing the human target or homologous protein in a species expressing an ortholog of the human target can be considered, assuming that sufficient background knowledge exists for the model (e.g., historical background data) (see Part I, Note 1). For products that are directed at a foreign target such as bacteria and viruses, in general no reproductive toxicity studies would be expected (see Section 2.1).

When the weight of evidence (e.g., mechanism of action, phenotypic data from genetically modified animals, class effects) suggests that there will be an adverse effect on fertility or pregnancy outcome, these data can provide adequate information to communicate risk to reproduction, and under appropriate circumstances additional nonclinical studies might not be warranted.

### 5.2 Fertility

For products where mice and rats are pharmacologically relevant species, an assessment of fertility can be made in one of these rodent species (see ICH S5 Guideline). ICH S5 Guideline study designs can be adapted for other species provided they are pharmacologically relevant; in addition, the design of the study should be amended as appropriate, for example to address the nature of the product and potential for immunogenicity.

It is recognized that mating studies are not practical for NHPs. However, when the NHP is the only relevant species, the potential for effects on male and female fertility can be assessed by evaluation of the reproductive tract (organ weights and histopathological evaluation) in repeat dose toxicity studies of at least 3 months duration using sexually mature NHPs. If there is a specific cause for concern based on pharmacological activity or previous findings, specialized assessments such as menstrual cyclicity, sperm count, sperm morphology/motility, and male or female reproductive hormone levels can be evaluated in a repeat dose toxicity study.

If there is a specific concern from the pharmacological activity about potential effects on conception/implantation and the NHP is the only relevant species, the concern should be addressed experimentally. A homologous product or transgenic model could be the only practical means to assess potential effects on conception or implantation when those are of specific concern. However, it is not recommended to produce a homologous product or transgenic model solely to conduct mating studies in rodents. In absence of nonclinical information, the risk to patients should be mitigated through clinical trial management procedures, informed consent and appropriate product labeling.

### 5.3 Embryo-Fetal Development (EFD) and Pre/Post-Natal Development (PPND)

Potential differences in placental transfer of biopharmaceuticals should be considered in the design and interpretation of developmental toxicity studies (see Note 3).
For products pharmacologically active only in NHPs, several study designs can be considered based on intended clinical use and expected pharmacology. Separate EFD and/or PPND studies, or other study designs (justified by the sponsor) can be appropriate, particularly when there is some concern that the mechanism of action might lead to an adverse effect on embryo-fetal development or pregnancy loss. However, one well-designed study in NHPs which includes dosing from day 20 of gestation to birth (enhanced PPND, ePPND) can be considered, rather than separate EFD and/or PPND studies.

For the single ePPND study design described above, no Caesarian section group is warranted, but assessment of pregnancy outcome at natural delivery should be performed. This study should also evaluate offspring viability, external malformations, skeletal effects (e.g., by X-ray) and, ultimately, visceral morphology at necropsy. Ultrasound is useful to track maintenance of pregnancy but is not appropriate for detecting malformations. These latter data are derived from post-partum observations. Because of confounding effects on maternal care of offspring, dosing of the mother post-partum is generally not recommended. Other endpoints in the offspring can also be evaluated if relevant for the pharmacological activity. The duration of the post-natal phase will be dependent on which additional endpoints are considered relevant based on mechanism of action (see Note 4).

Developmental toxicity studies in NHPs can only provide hazard identification. The number of animals per group should be sufficient to allow meaningful interpretation of the data (see Note 5).

The sponsor should justify the study design if other NHP species are used. The developmental toxicity studies in NHPs as outlined above are just hazard identification studies; therefore it might be possible to conduct these studies using a control group and one dose group, provided there is a scientific justification for the dose level selected. An example of an appropriate scientific justification would be a monoclonal antibody which binds a soluble target with a clinical dosing regimen intended to saturate target binding. If such a saturation of target binding can be demonstrated in the animal species selected and there is an up to 10-fold exposure multiple over therapeutic drug levels, a single dose level and control group would provide adequate evidence of hazard to embryo-fetal development.

### 5.4 Timing of Studies

If women of child-bearing potential are included in clinical trials prior to acquiring information on effects on embryo-fetal development, appropriate clinical risk management is appropriate, such as use of highly effective methods of contraception (see ICH M3(R2) Guideline).

For biopharmaceuticals pharmacologically active only in NHPs, where there are sufficient precautions to prevent pregnancy (see ICH M3(R2) Guideline, Section 11.3, Paragraph 2), an EFD or ePPND study can be conducted during Phase III, and the report submitted at the time of marketing application. When a sponsor cannot take sufficient precaution to prevent pregnancy in clinical trials, either a complete report of an EFD study or an interim report of an ePPND study should be submitted before initiation of Phase III (see Note 6). Where the product is pharmacologically active only in NHPs and its mechanism of action raises serious concern for embryo-fetal development, the label should reflect the concern without warranting a developmental toxicity study in NHPs and therefore administration to women of child-bearing potential should be avoided.
If the rodent or rabbit is a relevant species, see ICH M3(R2) Guideline for timing of reproductive toxicity studies. ICH M3(R2) Guideline should also be followed for the timing of data on fertility for products where rodents are relevant species.

For oncology products which fall within the scope of ICH S9 Guideline, see that guidance for aspects relating to timing of study conduct.

6. CARCINOGENICITY

The need for a product-specific assessment of the carcinogenic potential for biopharmaceutical should be determined with regard to the intended clinical population and treatment duration (see ICH S1A Guideline). When an assessment is warranted, the sponsor should design a strategy to address the potential hazard.

This strategy could be based on a weight of evidence approach, including a review of relevant data from a variety of sources. The data sources can include published data (e.g., information from transgenic, knock-out or animal disease models, human genetic diseases), information on class effects, detailed information on target biology and mechanism of action, in vitro data, data from chronic toxicity studies and clinical data. In some cases, the available information can be sufficient to address carcinogenic potential and inform clinical risk without additional nonclinical studies.

The mechanism of action of some biopharmaceuticals might raise concern regarding potential for carcinogenicity (e.g., immunosuppressives and growth factors). If the weight of evidence (see above) supports the concern regarding carcinogenic potential, rodent bioassays are not warranted. In this case potential hazard can be best addressed by product labeling and risk management practices. However, when the weight of evidence is unclear, the sponsor can propose additional studies that could mitigate the mechanism-based concern (see Part I, Section 4.8).

For products where there is insufficient knowledge about specific product characteristics and mode of action in relation to carcinogenic potential, a more extensive assessment might be appropriate (e.g., understanding of target biology related to potential carcinogenic concern, inclusion of additional endpoints in toxicity studies).

If the weight of evidence from this more extensive assessment does not suggest carcinogenic potential, no additional nonclinical testing is recommended. Alternatively, if the weight of evidence suggests a concern about carcinogenic potential, then the sponsor can propose additional nonclinical studies that could mitigate the concern, or the label should reflect the concern.

The product-specific assessment of carcinogenic potential is used to communicate risk and provide input to the risk management plan along with labeling proposals, clinical monitoring, post-marketing surveillance, or a combination of these approaches.

Rodent bioassays (or short-term carcinogenicity studies) with homologous products are generally of limited value to assess carcinogenic potential of the clinical candidate. Alternative approaches can be considered as new strategies/assays are developed.

NOTES

Note 1 Tissue cross-reactivity (TCR) studies are in vitro tissue-binding assays employing immunohistochemical (IHC) techniques conducted to characterize binding of monoclonal antibodies and related antibody-like products to antigenic determinants in tissues. Other technologies can be employed in place of IHC techniques to demonstrate target/binding site distribution.
A TCR study with a panel of human tissues is a recommended component of the safety assessment package supporting initial clinical dosing of these products. However, in some cases the clinical candidate is not a good IHC reagent and a TCR study might not be technically feasible.

TCR studies can provide useful information to supplement knowledge of target distribution and can provide information on potential unexpected binding. Tissue binding \textit{per se} does not indicate biological activity \textit{in vivo}. In addition, binding to areas not typically accessible to the antibody \textit{in vivo} (i.e., cytoplasm) is generally not relevant. Findings should be evaluated and interpreted in the context of the overall pharmacology and safety assessment data package.

When there is unexpected binding in human tissues an evaluation of selected animal tissues can provide supplemental information regarding potential correlations or lack thereof with preclinical toxicity. TCR using a full panel of animal tissues is not recommended.

Since a bi-specific antibody product will be evaluated in a TCR study using a panel of human tissues, there is no need to study the individual binding components.

Evaluating the tissue binding of homologous products does not provide additional value when TCR studies have been conducted with the clinical candidate in a human tissue panel, and is not recommended. TCR studies cannot detect subtle changes in critical quality attributes. Therefore TCR studies are not recommended for assessing comparability of the test article as a result of process changes over the course of a development program.

\textbf{Note 2} If two species have been used to assess the safety of the ADC, an additional short-term study or arm in a short-term study should be conducted in at least one species with the unconjugated toxin. In these cases a rodent is preferred unless the toxin is not active in the rodent. If only one pharmacologically relevant species is available, then the ADC should be tested in this species. A novel toxicant calls for an approach to species selection similar to that used for a new chemical entity on a case-by-case approach (e.g., for anticancer products in accordance with ICH S9 Guideline). For toxins or toxicants which are not novel and for which there is a sufficient body of scientific information available, separate evaluation of the unconjugated toxin is not warranted. Data should be provided to compare the metabolic stability of the ADC in animals with human.

\textbf{Note 3} The species-specific profile of embryo-fetal exposure during gestation should be considered in interpreting studies. High molecular weight proteins (>5,000 D) do not cross the placenta by simple diffusion. For monoclonal antibodies with molecular weight as high as 150,000 D, there exists a specific transport mechanism, the neonatal Fc receptor (FcRn) which determines fetal exposure and varies across species.

In the NHPs and humans, IgG placental transfer is low in the period of organogenesis and begins to increase in early second trimester, reaching highest levels late in the third trimester. (5) Therefore, standard embryo-fetal studies in NHPs, which are dosed from early pregnancy up to Gestation Day 50, might not be of value to assess direct embryo-fetal effects in the period of organogenesis, although effects on embryo-fetal development as an indirect result of maternal effects can be evaluated. Furthermore, maternal
dosing in NHPs after delivery is generally without relevance as IgG is only excreted in the milk initially (i.e., in the colostrum), and not later during the lactation and nursing phase.

Rodents differ from the NHPs and humans, as IgG crosses the yolk sac in rodents by FcRn transport mechanisms and exposure can occur relatively earlier in gestation than with NHPs and humans. In addition, delivery of rodents occurs at a stage of development when the pups are not as mature as the NHP or the human neonate. Therefore, rat/mouse dams should be dosed during lactation in order to expose pups via the milk up to at least day 9 of lactation when the offspring are at an equivalent stage of development as human neonates.

**Note 4**  The minimum duration of post-natal follow-up should be one month to cover early functional testing (e.g., growth and behaviour).

In general, if there is evidence for adverse effects on the immune system (or immune function) in the general toxicology studies, immune function testing in the offspring during the post-partum phase of the enhanced Pre/Post-Natal Development (ePPND) study is warranted. When appropriate, immunophenotyping can be obtained as early as post-natal day 28. The duration of post-natal follow-up for assessment of immune function can be 3-6 months depending on the functional test used.

Neurobehavioural assessment can be limited to clinical behavioural observations. Instrumental learning calls for a training period, which would result in a post-natal duration of at least 9 months and is not recommended.

**Note 5**  A detailed discussion of the approach to determine group sizes in cynomolgus monkey ePPND studies can be found in Jarvis *et al.*, 2010 (6). Group sizes in ePPND studies should yield a sufficient number of infants (6-8 per group at post-natal day 7) in order to assess post-natal development and provide the opportunity for specialist evaluation if necessary (e.g., immune system).

Most ePPND studies accrue pregnant animals over weeks and months. Consideration should be given to terminating further accrual of pregnant animals into the study, and adapting the study design (e.g., by Caesarian section) when pre-natal losses in a test item group indicate a treatment-related effect.

Reuse of vehicle-control treated maternal animals is encouraged.

If there is some cause for concern that the mechanism of action might lead to an effect on EFD or pregnancy loss, studies can be conducted in a limited number of animals in order to confirm the hazard.

**Note 6**  Endpoints to be included in an interim report of an ePPND study in NHPs:

- Dam data: survival, clinical observations, bodyweight, gestational exposure data (if available), any specific PD endpoints;
- Pregnancy data: number of pregnant animals started on study, pregnancy status at both the end of organogenesis (gestation day (GD) 50) and at GD100, occurrence of abortions and timing of abortions. There is no need for ultrasound determinations of fetal size in the interim report; these are not considered essential since actual birth weight will be available;
- Pregnancy outcome data: number of live births/still births, infant birth weight, infant survival and bodyweight at day 7 post-partum, qualitative external morphological assessment (i.e., confirming appearance is within normal limits), infant exposure data (if available), any specific PD endpoints in the infant if appropriate.
REFERENCES


2. ICH S1A Guideline: Guideline on the Need for Carcinogenicity Studies for Pharmaceuticals; November 1995.


International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use

ICH Harmonised Tripartite Guideline

Safety Pharmacology Studies for Human Pharmaceuticals

S7A

Current Step 4 version

dated 8 November 2000

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
## S7A Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification November 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7A</td>
<td>Approval by the Steering Committee under <em>Step 2</em> and release for public consultation.</td>
<td>2 March 2000</td>
<td>S7A</td>
</tr>
</tbody>
</table>

## Current *Step 4* version

| S7A                | Approval by the Steering Committee under *Step 4* and recommendation for adoption to the three ICH regulatory bodies. | 8 November 2000 | S7A                             |
SAFETY PHARMACOLOGY STUDIES FOR HUMAN PHARMACEUTICALS

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 8 November 2000, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. INTRODUCTION .......................................................................................................................... 1
   1.1 Objectives of the Guideline ................................................................................................. 1
   1.2 Background ......................................................................................................................... 1
   1.3 Scope of the Guideline ........................................................................................................ 1
   1.4 General Principle ............................................................................................................... 1
   1.5 Definition of Safety Pharmacology ................................................................................... 1

2. GUIDELINE .................................................................................................................................... 2
   2.1 Objectives of Studies .......................................................................................................... 2
   2.2 General Considerations in Selection and Design of Safety Pharmacology Studies .......... 2
   2.3 Test Systems ....................................................................................................................... 2
      2.3.1 General Considerations on Test Systems ................................................................... 3
      2.3.2 Use of In Vivo and In Vitro Studies ........................................................................... 3
      2.3.3 Experimental Design ............................................................................................... 3
         2.3.3.1 Sample Size and Use of Controls ..................................................................... 3
         2.3.3.2 Route of Administration .................................................................................... 3
   2.4 Dose Levels or Concentrations of Test Substance ......................................................... 4
      2.4.1 In Vivo Studies .......................................................................................................... 4
      2.4.2 In Vitro Studies ......................................................................................................... 4
   2.5 Duration of Studies ............................................................................................................ 4
   2.6 Studies on Metabolites, Isomers and Finished Products .............................................. 4
   2.7 Safety Pharmacology Core Battery ................................................................................. 5
      2.7.1 Central Nervous System ............................................................................................ 5
      2.7.2 Cardiovascular System ............................................................................................. 5
      2.7.3 Respiratory System .................................................................................................. 5
2.8 Follow-up and Supplemental Safety Pharmacology Studies
2.8.1 Follow-up Studies For Safety Pharmacology Core Battery
2.8.1.1 Central Nervous System
2.8.1.2 Cardiovascular System
2.8.1.3 Respiratory System
2.8.2 Supplemental Safety Pharmacology Studies
2.8.2.1 Renal/Urinary System
2.8.2.2 Autonomic Nervous System
2.8.2.3 Gastrointestinal System
2.8.2.4 Other Organ Systems
2.9 Conditions under which Studies are not Necessary
2.10 Timing of Safety Pharmacology Studies in Relation to Clinical Development
2.10.1 Studies Prior to First Administration in Humans
2.10.2 Studies During Clinical Development
2.10.3 Studies Before Approval
2.11 Application of Good Laboratory Practice (GLP)
3. NOTES
4. REFERENCES
SAFETY PHARMACOLOGY STUDIES FOR HUMAN PHARMACEUTICALS

1. INTRODUCTION

1.1 Objectives of the Guideline
This guideline was developed to help protect clinical trial participants and patients receiving marketed products from potential adverse effects of pharmaceuticals, while avoiding unnecessary use of animals and other resources.

This guideline provides a definition, general principles and recommendations for safety pharmacology studies.

1.2 Background
Pharmacology studies have been performed worldwide for many years as part of the non-clinical evaluation of pharmaceuticals for human use. There have been, however, no internationally accepted definitions, objectives or recommendations on the design and conduct of safety pharmacology studies. (Note 1)

The term “safety pharmacology studies” first appeared in the ICH topics, “Timing of Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (M3)” and “Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (S6)” as studies that should be conducted to support use of therapeutics in humans (1, 2). Details of the safety pharmacology studies, including their definition and objectives, were left for future discussion.

1.3 Scope of the Guideline
This guideline generally applies to new chemical entities and biotechnology-derived products for human use. This guideline can be applied to marketed pharmaceuticals when appropriate (e.g., when adverse clinical events, a new patient population, or a new route of administration raises concerns not previously addressed).

1.4 General Principle
It is important to adopt a rational approach when selecting and conducting safety pharmacology studies. The specific studies that should be conducted and their design will vary based on the individual properties and intended uses of the pharmaceuticals. Scientifically valid methods should be used, and when there are internationally recognized methods that are applicable to pharmaceuticals, these are preferable. Moreover, the use of new technologies and methodologies in accordance with sound scientific principles is encouraged.

Some safety pharmacology endpoints can be incorporated in the design of toxicology, kinetic, clinical studies, etc., while in other cases these endpoints should be evaluated in specific safety pharmacology studies. Although adverse effects of a substance may be detectable at exposures that fall within the therapeutic range in appropriately designed safety pharmacology studies, they may not be evident from observations and measurements used to detect toxicity in conventional animal toxicity studies.

1.5 Definition of Safety Pharmacology
Pharmacology studies can be divided into three categories: primary pharmacodynamic, secondary pharmacodynamic and safety pharmacology studies.

For the purpose of this document, safety pharmacology studies are defined as those studies that investigate the potential undesirable pharmacodynamic effects of a substance on
physiological functions in relation to exposure in the therapeutic range and above. (See Note 2 for definitions of primary pharmacodynamic and secondary pharmacodynamic studies.)

In some cases, information on the primary and secondary pharmacodynamic properties of the substance may contribute to the safety evaluation for potential adverse effect(s) in humans and should be considered along with the findings of safety pharmacology studies.

2. GUIDELINE

2.1 Objectives of Studies

The objectives of safety pharmacology studies are: 1) to identify undesirable pharmacodynamic properties of a substance that may have relevance to its human safety; 2) to evaluate adverse pharmacodynamic and/or pathophysiological effects of a substance observed in toxicology and/or clinical studies; and 3) to investigate the mechanism of the adverse pharmacodynamic effects observed and/or suspected. The investigational plan to meet these objectives should be clearly identified and delineated.

2.2 General Considerations in Selection and Design of Safety Pharmacology Studies

Since pharmacological effects vary depending on the specific properties of each test substance, the studies should be selected and designed accordingly. The following factors should be considered (the list is not comprehensive):

1) Effects related to the therapeutic class of the test substance, since the mechanism of action may suggest specific adverse effects (e.g., proarrhythmia is a common feature of antiarrhythmic agents);

2) Adverse effects associated with members of the chemical or therapeutic class, but independent of the primary pharmacodynamic effects (e.g., anti-psychotics and QT prolongation);

3) Ligand binding or enzyme assay data suggesting a potential for adverse effects;

4) Results from previous safety pharmacology studies, from secondary pharmacodynamic studies, from toxicology studies, or from human use that warrant further investigation to establish and characterize the relevance of these findings to potential adverse effects in humans.

During early development, sufficient information (e.g., comparative metabolism) may not always be available to rationally select or design the studies in accordance with the points stated above; in such circumstances, a more general approach in safety pharmacology investigations can be applied.

A hierarchy of organ systems can be developed according to their importance with respect to life-supporting functions. Vital organs or systems, the functions of which are acutely critical for life, such as the cardiovascular, respiratory and central nervous systems, are considered to be the most important ones to assess in safety pharmacology studies. Other organ systems, such as the renal or gastrointestinal system, the functions of which can be transiently disrupted by adverse pharmacodynamic effects without causing irreversible harm, are of less immediate investigative concern. Safety pharmacology evaluation of effects on these other systems may be of particular importance when considering factors such as the likely clinical trial or patient population (e.g. gastrointestinal tract in Crohn’s disease, renal function in primary renal hypertension, immune system in immunocompromised patients).
2.3.1 General Considerations on Test Systems
Consideration should be given to the selection of relevant animal models or other test systems so that scientifically valid information can be derived. Selection factors can include the pharmacodynamic responsiveness of the model, pharmacokinetic profile, species, strain, gender and age of the experimental animals, the susceptibility, sensitivity, and reproducibility of the test system and available background data on the substance. Data from humans (e.g., in vitro metabolism), when available, should also be considered in the test system selection. The time points for the measurements should be based on pharmacodynamic and pharmacokinetic considerations. Justification should be provided for the selection of the particular animal model or test system.

2.3.2 Use of In Vivo and In Vitro Studies
Animal models as well as ex vivo and in vitro preparations can be used as test systems. Ex vivo and in vitro systems can include, but are not limited to: isolated organs and tissues, cell cultures, cellular fragments, subcellular organelles, receptors, ion channels, transporters and enzymes. In vitro systems can be used in supportive studies (e.g., to obtain a profile of the activity of the substance or to investigate the mechanism of effects observed in vivo).

In conducting in vivo studies, it is preferable to use unanesthetized animals. Data from unrestrained animals that may be chronically instrumented for telemetry, other suitable instrumentation methods for conscious animals, or animals conditioned to the laboratory environment are preferable to data from restrained or unconditioned animals. In the use of unanesthetized animals, the avoidance of discomfort or pain is a foremost consideration.

2.3.3 Experimental Design

2.3.3.1 Sample Size and Use of Controls
The size of the groups should be sufficient to allow meaningful scientific interpretation of the data generated. Thus, the number of animals or isolated preparations should be adequate to demonstrate or rule out the presence of a biologically significant effect of the test substance. This should take into consideration the size of the biological effect that is of concern for humans. Appropriate negative and positive control groups should be included in the experimental design. In well-characterized in vivo test systems, positive controls may not be necessary. The exclusion of controls from studies should be justified.

2.3.3.2 Route of Administration
In general, the expected clinical route of administration should be used when feasible. Regardless of the route of administration, exposure to the parent substance and its major metabolites should be similar to or greater than that achieved in humans when such information is available. Assessment of effects by more than one route may be appropriate if the test substance is intended for clinical use by more than one route of administration (e.g. oral and parenteral), or where there are observed or anticipated significant qualitative and quantitative differences in systemic or local exposure.
2.4 Dose Levels or Concentrations of Test Substance

2.4.1 In Vivo Studies
Safety pharmacology studies should be designed to define the dose-response relationship of the adverse effect observed. The time course (e.g., onset and duration of response) of the adverse effect should be investigated, when feasible. Generally, the doses eliciting the adverse effect should be compared to the doses eliciting the primary pharmacodynamic effect in the test species or the proposed therapeutic effect in humans, if feasible. It is recognized that there are species differences in pharmacodynamic sensitivity. Therefore, doses should include and exceed the primary pharmacodynamic or therapeutic range. In the absence of an adverse effect on the safety pharmacology parameter(s) evaluated in the study, the highest tested dose should be a dose that produces moderate adverse effects in this or in other studies of similar route and duration. These adverse effects can include dose-limiting pharmacodynamic effects or other toxicity. In practice, some effects in the toxic range (e.g., tremors or fasciculation during ECG recording) may confound the interpretation of the results and may also limit dose levels. Testing of a single group at the limiting dose as described above may be sufficient in the absence of an adverse effect on safety pharmacology endpoints in the test species.

2.4.2 In Vitro Studies
In vitro studies should be designed to establish a concentration-effect relationship. The range of concentrations used should be selected to increase the likelihood of detecting an effect on the test system. The upper limit of this range may be influenced by physico-chemical properties of the test substance and other assay specific factors. In the absence of an effect, the range of concentrations selected should be justified.

2.5 Duration of Studies
Safety pharmacology studies are generally performed by single dose administration. When pharmacodynamic effects occur only after a certain duration of treatment, or when results from repeat dose non-clinical studies or results from use in humans give rise to concerns about safety pharmacological effects, the duration of the safety pharmacology studies to address these effects should be rationally based.

2.6 Studies on Metabolites, Isomers and Finished Products
Generally, any parent compound and its major metabolite(s) that achieve, or are expected to achieve, systemic exposure in humans should be evaluated in safety pharmacology studies. Evaluation of major metabolites is often accomplished through studies of the parent compound in animals. If the major human metabolite(s) is (are) found to be absent or present only at relatively low concentrations in animals, assessment of the effects of such metabolite(s) on safety pharmacology endpoints should be considered. Additionally, if metabolites from humans are known to substantially contribute to the pharmacological actions of the therapeutic agent, it may be important to test such active metabolites. When the in vivo studies on the parent compound have not adequately assessed metabolites, as discussed above, the tests of metabolites can use in vitro systems based on practical considerations.

In vitro or in vivo testing of the individual isomers should also be considered when the product contains an isomeric mixture.

Safety pharmacology studies with the finished product formulation(s) should be conducted only for formulations that substantially alter the pharmacokinetics and/or
pharmacodynamics of the active substance in comparison to formulations previously tested (i.e. through active excipients such as penetration enhancers, liposomes, and other changes such as polymorphism).

2.7 Safety Pharmacology Core Battery

The purpose of the safety pharmacology core battery is to investigate the effects of the test substance on vital functions. In this regard, the cardiovascular, respiratory and central nervous systems are usually considered the vital organ systems that should be studied in the core battery. In some instances, based on scientific rationale, the core battery should be supplemented (see section 2.8) or need not be implemented (see also section 2.9).

The exclusion of certain test(s) or exploration(s) of certain organs, systems or functions should be scientifically justified.

2.7.1 Central Nervous System

Effects of the test substance on the central nervous system should be assessed appropriately. Motor activity, behavioral changes, coordination, sensory/motor reflex responses and body temperature should be evaluated. For example, a functional observation battery (FOB) (3), modified Irwin’s (4), or other appropriate test (5) can be used.

2.7.2 Cardiovascular System

Effects of the test substance on the cardiovascular system should be assessed appropriately. Blood pressure, heart rate, and the electrocardiogram should be evaluated. In vivo, in vitro and/or ex vivo evaluations, including methods for repolarization and conductance abnormalities, should also be considered. (Note 3)

2.7.3 Respiratory System

Effects of the test substance on the respiratory system should be assessed appropriately. Respiratory rate and other measures of respiratory function (e.g., tidal volume (6) or hemoglobin oxygen saturation) should be evaluated. Clinical observation of animals is generally not adequate to assess respiratory function, and thus these parameters should be quantified by using appropriate methodologies.

2.8 Follow-up and Supplemental Safety Pharmacology Studies

Adverse effects may be suspected based on the pharmacological properties or chemical class of the test substance. Additionally, concerns may arise from the safety pharmacology core battery, clinical trials, pharmacovigilance, experimental in vitro or in vivo studies, or from literature reports. When such potential adverse effects raise concern for human safety, these should be explored in follow-up or supplemental safety pharmacology studies, as appropriate.

2.8.1 Follow-up Studies For Safety Pharmacology Core Battery

Follow-up studies are meant to provide a greater depth of understanding than, or additional knowledge to, that provided by the core battery on vital functions. The following subsections provide lists of studies to further evaluate these organ systems for potential adverse pharmacodynamic effects. These lists are not meant to be comprehensive or prescriptive, and the studies should be selected on a case-by-case basis after considering factors such as existing non-clinical or human data. In some cases, it may be more appropriate to address these effects during the conduct of other non-clinical and/or clinical studies.
2.8.1.1 Central Nervous System
Behavioral pharmacology, learning and memory, ligand-specific binding, neurochemistry, visual, auditory and/or electrophysiology examinations, etc.

2.8.1.2 Cardiovascular System
Cardiac output, ventricular contractility, vascular resistance, the effects of endogenous and/or exogenous substances on the cardiovascular responses, etc.

2.8.1.3 Respiratory System
Airway resistance, compliance, pulmonary arterial pressure, blood gases, blood pH, etc.

2.8.2 Supplemental Safety Pharmacology Studies
Supplemental studies are meant to evaluate potential adverse pharmacodynamic effects on organ system functions not addressed by the core battery or repeated dose toxicity studies when there is a cause for concern.

2.8.2.1 Renal/Urinary System
Effects of the test substance on renal parameters should be assessed. For example, urinary volume, specific gravity, osmolality, pH, fluid/electrolyte balance, proteins, cytology, and blood chemistry determinations such as blood urea nitrogen, creatinine and plasma proteins can be used.

2.8.2.2 Autonomic Nervous System
Effects of the test substance on the autonomic nervous system should be assessed. For example, binding to receptors relevant for the autonomic nervous system, functional responses to agonists or antagonists in vivo or in vitro, direct stimulation of autonomic nerves and measurement of cardiovascular responses, baroreflex testing, and heart rate variability can be used.

2.8.2.3 Gastrointestinal System
Effects of the test substance on the gastrointestinal system should be assessed. For example, gastric secretion, gastrointestinal injury potential, bile secretion, transit time in vivo, ileal contraction in vitro, gastric pH measurement and pooling can be used.

2.8.2.4 Other Organ Systems
Effects of the test substance on organ systems not investigated elsewhere should be assessed when there is a reason for concern. For example, dependency potential or skeletal muscle, immune and endocrine functions can be investigated.

2.9 Conditions under which Studies are not Necessary
Safety pharmacology studies may not be needed for locally applied agents (e.g., dermal or ocular) where the pharmacology of the test substance is well characterized, and where systemic exposure or distribution to other organs or tissues is demonstrated to be low.

Safety pharmacology studies prior to the first administration in humans may not be needed for cytotoxic agents for treatment of end-stage cancer patients. However, for cytotoxic agents with novel mechanisms of action, there may be value in conducting safety pharmacology studies.

For biotechnology-derived products that achieve highly specific receptor targeting, it is often sufficient to evaluate safety pharmacology endpoints as a part of toxicology and/or
pharmacodynamic studies, and therefore safety pharmacology studies can be reduced or eliminated for these products.

For biotechnology-derived products that represent a novel therapeutic class and/or those products that do not achieve highly specific receptor targeting, a more extensive evaluation by safety pharmacology studies should be considered.

There may be additional exceptions where safety pharmacology testing is not needed, for example, in the case of a new salt having similar pharmacokinetics and pharmacodynamics.

2.10 Timing of Safety Pharmacology Studies in Relation to Clinical Development

When planning a safety pharmacology program, section 2.9 should be reviewed to determine whether or not specific studies are recommended.

2.10.1 Studies Prior to First Administration in Humans

The effects of a test substance on the functions listed in the safety pharmacology core battery should be investigated prior to first administration in humans. Any follow-up or supplemental studies identified as appropriate, based on a cause for concern, should also be conducted. Information from toxicology studies adequately designed and conducted to address safety pharmacology endpoints can result in reduction or elimination of separate safety pharmacology studies.

2.10.2 Studies During Clinical Development

Additional studies may be warranted to clarify observed or suspected adverse effects in animals and humans during clinical development.

2.10.3 Studies Before Approval

Safety pharmacology effects on systems listed in section 2.8 should be assessed prior to product approval, unless not warranted, in which case this should be justified. Available information from toxicology studies adequately designed and conducted to address safety pharmacology endpoints, or information from clinical studies, can support this assessment and replace safety pharmacology studies.

2.11 Application of Good Laboratory Practice (GLP)

It is important to ensure the quality and reliability of non-clinical safety studies. This is normally accomplished through the conduct of the studies in compliance with GLP. Due to the unique design of, and practical considerations for, some safety pharmacology studies, it may not be feasible to conduct these in compliance with GLP. It has to be emphasized that data quality and integrity in safety pharmacology studies should be ensured even in the absence of formal adherence to the principles of GLP. When studies are not conducted in compliance with GLP, study reconstruction should be ensured through adequate documentation of study conduct and archiving of data. Any study or study component not conducted in compliance with GLP should be adequately justified, and the potential impact on evaluation of the safety pharmacology endpoints should be explained.

The safety pharmacology core battery should ordinarily be conducted in compliance with GLP. Follow-up and supplemental studies should be conducted in compliance with GLP to the greatest extent feasible. Safety pharmacology investigations can be part of toxicology studies; in such cases, these studies would be conducted in compliance with GLP.

Primary pharmacodynamic studies do not need to be conducted in compliance with GLP.
Generally, secondary pharmacodynamic studies do not need to be conducted in compliance with GLP. Results from secondary pharmacodynamic studies conducted during the compound selection process may contribute to the safety pharmacology evaluation; when there is no cause for concern (e.g., there are no findings for the safety pharmacological endpoint or the chemical or therapeutic class), these studies need not be repeated in compliance with GLP. In some circumstances, results of secondary pharmacodynamic studies may make a pivotal contribution to the safety evaluation for potential adverse effects in humans, and these are normally conducted in compliance with GLP.

3. **NOTES**

1. General pharmacology studies have been considered an important component in drug safety assessment. General pharmacology studies were originally referred to as those designed to examine effects other than the primary therapeutic effect of a drug candidate. Safety pharmacology studies were focused on identifying adverse effects on physiological functions. All three regions have accepted data from general pharmacology studies (Japan and EC) or safety pharmacology studies (USA) in the assessment of a marketing application. The Japanese Ministry of Health and Welfare (MHW) issued the “Guideline for General Pharmacology” in 1991. In this MHW guideline, general pharmacology studies include those designed to identify unexpected effects on organ system function, and to broaden pharmacological characterization (pharmacological profiling). However, there has been no internationally accepted definition of the terms “primary pharmacodynamics”, “secondary pharmacodynamics” and “safety pharmacology.” The need for international harmonization of the nomenclature and the development of an international guideline for safety pharmacology has been recognized.

2. Studies on the mode of action and/or effects of a substance in relation to its desired therapeutic target are primary pharmacodynamic studies. Studies on the mode of action and/or effects of a substance not related to its desired therapeutic target are secondary pharmacodynamic studies (these have sometimes been referred to as part of general pharmacology studies).

3. There is no scientific consensus on the preferred approach to, or internationally recognized guidance on, addressing risks for repolarization-associated ventricular tachyarrhythmia (e.g., Torsade de Pointes). A guideline (S7B) will be prepared to present some currently available methods and discuss their advantages and disadvantages. Submission of data to regulatory authorities to support the use of these methods is encouraged.
4. REFERENCES


2) ICH Harmonized Tripartite Guideline (S6) “Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals” (1997)


This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
S7B
Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7B</td>
<td>Approval by the Steering Committee under <em>Step 2</em> and release for public consultation.</td>
<td>7 February 2002</td>
<td>November 2005</td>
</tr>
<tr>
<td>S7B</td>
<td>Due to scientific and procedural reasons, the guideline was revised, re-approved by the Steering Committee and re-released for public consultation as a <em>Step 2</em> document.</td>
<td>10 June 2004</td>
<td>S7B</td>
</tr>
</tbody>
</table>

**Current Step 4 version**

| S7B                | Approval by the Steering Committee under *Step 4* and recommendation for adoption to the three ICH regulatory bodies. | 12 May 2005           | S7B               |
THE NON-CLINICAL EVALUATION OF THE POTENTIAL FOR DELAYED VENTRICULAR REPOLARIZATION (QT INTERVAL PROLONGATION) BY HUMAN PHARMACEUTICALS

ICH Harmonised Tripartite Guideline

Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on 12 May 2005, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. INTRODUCTION ......................................................................................................... 1
   1.1 Objective of the Guideline ...................................................................................... 1
   1.2 Background ............................................................................................................. 1
   1.3 Scope of the Guideline ............................................................................................ 2
   1.4 General Principles .................................................................................................. 2

2. GUIDELINE .................................................................................................................. 2
   2.1 Objectives of S7B Studies ....................................................................................... 2
   2.2 Considerations for Selection and Design of Studies ............................................. 3
   2.3 Non-clinical Testing Strategy ................................................................................. 4
       2.3.1 In Vitro Ikr Assay ........................................................................................... 4
       2.3.2 In Vivo QT Assay .......................................................................................... 4
       2.3.3 Chemical/Pharmacological Class .................................................................... 4
       2.3.4 Relevant Non-Clinical and Clinical Information ............................................ 5
       2.3.5 Follow-up Studies .......................................................................................... 5
       2.3.6 Integrated Risk Assessment ........................................................................... 6
       2.3.7 Evidence of Risk ............................................................................................ 6
   2.4 Timing of S7B Non-clinical Studies and Integrated Risk Assessment in Relation to Clinical Development ..................................................................................................... 6

3. TEST SYSTEMS ........................................................................................................... 6
   3.1 Considerations for Test Systems .......................................................................... 6
       3.1.1 Use of Positive Control Substances and Reference Compounds ................ 7
       3.1.2 In Vitro Electrophysiology Studies ................................................................ 7
       3.1.3 In Vivo Electrophysiology Studies ................................................................ 8
       3.1.4 Simulated Pathological Conditions and Arrhythmias ............................... 10
THE NON-CLINICAL EVALUATION OF THE POTENTIAL FOR DELAYED VENTRICULAR REPOLARIZATION (QT INTERVAL PROLONGATION) BY HUMAN PHARMACEUTICALS

1. INTRODUCTION
The assessment of the effects of pharmaceuticals on ventricular repolarization and proarrhythmic risk is the subject of active investigation. When additional data (non-clinical and clinical) are accumulated in the future, they will be evaluated and this guideline might be revised.

1.1 Objective of the Guideline
This guideline describes a non-clinical testing strategy for assessing the potential of a test substance to delay ventricular repolarization. This guideline includes information concerning non-clinical assays and integrated risk assessments.

1.2 Background
The QT interval (time from the beginning of the QRS complex to the end of the T wave) of the electrocardiogram (ECG) is a measure of the duration of ventricular depolarization and repolarization. QT interval prolongation can be congenital or acquired (e.g., pharmaceutical-induced). When ventricular repolarization is delayed and the QT interval is prolonged, there is an increased risk of ventricular tachyarrhythmia, including torsade de pointes, particularly when combined with other risk factors (e.g., hypokalemia, structural heart disease, bradycardia). Thus, much emphasis has been placed on the potential proarrhythmic effects of pharmaceuticals that are associated with QT interval prolongation.

Ventricular repolarization, determined by the duration of the cardiac action potential, is a complex physiological process. It is the net result of the activities of many membrane ion channels and transporters. Under physiological conditions, the functions of these ion channels and transporters are highly interdependent. The activity of each ion channel or transporter is affected by multiple factors including, but not limited to, intracellular and extracellular ion concentrations, membrane potential, cell-to-cell electrical coupling, heart rate, and autonomic nervous system activity. The metabolic state (e.g., acid-base balance) and location and type of cardiac cell are also important. The human ventricular action potential consists of five sequential phases:

- phase 0: The upstroke of the action potential is primarily a consequence of a rapid, transient influx of Na+ (INa) through Na+ channels;
- phase 1: The termination of the upstroke of the action potential and early repolarization phase result from the inactivation of Na+ channels and the transient efflux of K+ (Ito) through K+ channels;
- phase 2: The plateau of the action potential is a reflection of a balance between the influx of Ca2+ (ICa) through L-type Ca2+ channels and outward repolarizing K+ currents;
- phase 3: The sustained downward stroke of the action potential and the late repolarization phase result from the efflux of K+ (IKr and IKs) through delayed rectifier K+ channels;
• phase 4: The resting potential is maintained by the inward rectifier K+ current (I_K1).

Prolongation of the action potential can result from decreased inactivation of the inward Na+ or Ca2+ currents, increased activation of the Ca2+ current, or inhibition of one or more of the outward K+ currents. The rapidly and slowly activating components of the delayed rectifier potassium current, I_Kr and I_Ks, seem to have the most influential role in determining the duration of the action potential and thus the QT interval. The human ether-a-go-go-related gene (hERG) and KvLQT1 gene encode pore-forming proteins KCNH2 and KCNQ1 that are thought to represent the α-subunits of the human potassium channels responsible for I_Kr and I_Ks, respectively. These α-subunit proteins can form hetero-oligomeric complexes with auxiliary β-subunits (i.e. MiRP and MinK gene products), which have been speculated to modulate the gating properties of the channel proteins. The most common mechanism of QT interval prolongation by pharmaceuticals is inhibition of the delayed rectifier potassium channel that is responsible for I_Kr.

1.3 Scope of the Guideline
This guideline extends and complements the “ICH Guideline on Safety Pharmacology Studies for Human Pharmaceuticals” (ICH S7A). This guideline applies to new chemical entities for human use and marketed pharmaceuticals when appropriate (e.g., when adverse clinical events, a new patient population, or a new route of administration raises concerns not previously addressed). Conditions under which studies are not called for are described in ICH S7A.

1.4 General Principles
Principles and recommendations described in ICH S7A also apply to the studies conducted in accordance with the present guideline. In vitro I_Kr and in vivo QT assays described in Sections 2.3.1 and 2.3.2 when performed for regulatory submission should be conducted in compliance with good laboratory practice (GLP). Follow-up studies described in Section 2.3.5 should be conducted in compliance with GLP to the greatest extent feasible.

In vitro and in vivo assays are complementary approaches; therefore, according to current understanding, both assay types should be conducted.

The investigational approach and evidence of risk should be individualized for the test substance, depending on its pharmacodynamic, pharmacokinetic and safety profiles.

2. GUIDELINE

2.1 Objectives of S7B Studies
The objectives of studies are to: 1) identify the potential of a test substance and its metabolites to delay ventricular repolarization, and 2) relate the extent of delayed ventricular repolarization to the concentrations of a test substance and its metabolites. The study results can be used to elucidate the mechanism of action and, when considered with other information, estimate risk for delayed ventricular repolarization and QT interval prolongation in humans.
2.2 Considerations for Selection and Design of Studies

Non-clinical methodologies can address the following:

- Ionic currents measured in isolated animal or human cardiac myocytes, cultured cardiac cell lines, or heterologous expression systems for cloned human ion channels;
- Action potential parameters in isolated cardiac preparations or specific electrophysiology parameters indicative of action potential duration in anesthetized animals;
- ECG parameters measured in conscious or anesthetized animals;
- Proarrhythmic effects measured in isolated cardiac preparations or animals.

As indicated above, these four functional levels can be investigated by in vitro and/or in vivo methods. Findings from the functional levels listed above are considered useful and complementary.

In vitro electrophysiology studies can explore potential cellular mechanisms that might not be evident from in vivo data. Changes in other cardiovascular parameters or effects on multiple ion channels can complicate interpretation of data. Complementary assessments in other systems can address this issue. Although delay of repolarization can occur through modulation of several types of ion channels, inhibition of I_Kr is the most common mechanism responsible for pharmaceutical-induced prolongation of QT interval in humans.

In vivo models that possess the full complement of molecular, biochemical, and physiological systems can also be informative with regard to the response in humans to the test substance. Carefully designed and conducted in vivo studies allow evaluation of parent substance and metabolites, and can enable estimation of safety margins. In vivo ECG evaluations provide information on conduction properties and non-cardiac influences (e.g., autonomic nervous system tone). Studies of action potential parameters provide information on the integrated activity of multiple ion channels in the heart.
2.3 Non-clinical Testing Strategy

The following sections describe a general non-clinical testing strategy for assessing risk for delayed ventricular repolarization and QT interval prolongation that is pragmatic and based on currently available information. The figure illustrates the component elements of the testing strategy, but not specific test systems or their designs.

2.3.1 In Vitro $I_{Kr}$ Assay

An in vitro $I_{Kr}$ assay evaluates the effects on the ionic current through a native or expressed $I_{Kr}$ channel protein, such as that encoded by hERG (see section 3.1.2).

2.3.2 In Vivo QT Assay

An in vivo QT assay measures indices of ventricular repolarization such as QT interval (see section 3.1.3). This assay can be designed to meet the objective of both ICH S7A (cardiovascular core battery study) and S7B. This will reduce the use of animals and other resources.

2.3.3 Chemical/Pharmacological Class

Consideration should be given to whether the test substance belongs to a chemical/pharmacological class in which some members have been shown to induce QT interval prolongation in humans (e.g., antipsychotics, histamine H-1 receptor antagonists, fluoroquinolones). This factor should, where appropriate, influence the choice of reference compound(s) and be included in the integrated risk assessment.
2.3.4 Relevant Non-Clinical and Clinical Information

Additional information for the integrated risk assessment can include results from:

- Pharmacodynamic studies;
- Toxicology/safety studies;
- Pharmacokinetic studies, including plasma levels of parent substance and metabolites (including human data if available);
- Drug interaction studies;
- Tissue distribution and accumulation studies;
- Post-marketing surveillance.

2.3.5 Follow-up Studies

Follow-up studies are intended to provide greater depth of understanding or additional knowledge regarding the potential of test substance for delayed ventricular repolarization and QT interval prolongation in humans. Such studies can provide additional information concerning potency, mechanism of action, slope of the dose-response curve, or magnitude of the response. Follow-up studies are designed to address specific issues, and, as a result, various in vivo or in vitro study designs can be applicable.

In circumstances where results among non-clinical studies are inconsistent and/or results of clinical studies differ from those for non-clinical studies, retrospective evaluation and follow-up non-clinical studies can be used to understand the basis for the discrepancies. Results from follow-up studies can be a significant component of an integrated risk assessment.

Relevant non-clinical and clinical information along with the following should be considered in the selection and design of follow-up studies:

- Use of ventricular repolarization assays that measure action potential parameters in isolated cardiac preparations (see section 3.1.2);
- Use of specific electrophysiological parameters indicative of action potential duration in anesthetized animals (see section 3.1.3);
- Repeated administration of test substance;
- Selection of animal species and gender(s);
- Use of metabolic inducers or inhibitors;
- Use of concurrent positive control substances and reference compounds (see section 3.1.1);
- Inhibition of other channels not previously evaluated;
- Measurement of electrophysiological parameters at multiple time points;
- Confounding effects in conscious animals that limit the interpretation of data such as test substance-induced effects on heart rate or autonomic tone, or toxicities such as tremor, convulsion, or emesis.
2.3.6 Integrated Risk Assessment

The integrated risk assessment is the evaluation of non-clinical study results including the results from follow-up studies and other relevant information. The integrated risk assessment should be scientifically based and individualized for the test substance. Such an assessment can contribute to the design of clinical investigations and interpretation of their results. When available, the integrated risk assessment should be included in the Investigator’s Brochure and the Non-clinical Overview (ICH M4). Depending on the stage of drug development, the integrated risk assessment should also consider:

- Assay sensitivity and specificity;
- Potencies of test substance in S7B assays relative to reference compound(s);
- Relationship between the exposures associated with an effect on repolarization and those eliciting the primary pharmacodynamic effect in the non-clinical test species or the proposed therapeutic effect in humans;
- Contribution of metabolites to QT interval prolongation as well as metabolic differences between humans and animals.

2.3.7 Evidence of Risk

Evidence of risk is the overall conclusion from the integrated risk assessment for a test substance to delay ventricular repolarization and prolong QT interval in humans.

2.4 Timing of S7B Non-clinical Studies and Integrated Risk Assessment in Relation to Clinical Development

Conduct of S7B non-clinical studies assessing the risk for delayed ventricular repolarization and QT interval prolongation prior to first administration in humans should be considered. These results, as part of an integrated risk assessment, can support the planning and interpretation of subsequent clinical studies.

3. TEST SYSTEMS

3.1 Considerations for Test Systems

This section provides an overview of methodologies currently used to assess the potential for a test substance to delay ventricular repolarization and to prolong QT interval. The following should be considered in selecting the most appropriate test systems:

- Assay methodology and experimental endpoints are scientifically valid and robust;
- Assays and preparations are standardized;
- Results are reproducible;
- Endpoints/parameters of the assays are relevant for assessing human risk.
3.1.1 Use of Positive Control Substances and Reference Compounds

A sub-maximally effective concentration of a positive control substance should be used to demonstrate the responsiveness of in vitro preparations for ion channel and action potential duration assays and should be included in every study. In the case of in vivo studies, positive control substances should be used to validate and define the sensitivity of the test system, but need not be included in every study.

For test substances belonging to a chemical/pharmacological class that is associated with QT interval prolongation in humans, the use of concurrent reference compound(s) (member(s) of the same class) in in vitro and in vivo studies should be considered to facilitate ranking the potency of the test substance in relation to its comparators.

3.1.2 In Vitro Electrophysiology Studies

In vitro electrophysiology studies can provide valuable information concerning the effect of a test substance on action potential duration and/or cardiac ionic currents. These assays have an important role in assessing the potential for QT interval prolongation and elucidating cellular mechanisms affecting repolarization. In vitro electrophysiology studies employ either single cell (e.g., heterologous expression systems, disaggregated cardiomyocytes) or multicellular (e.g., Purkinje fiber; papillary muscle; trabeculae; perfused myocardium; intact heart) preparations. Heterologous expression systems, where human ion channel protein(s) are expressed in noncardiac cell lines, are used to assess the effects of a test substance on a specific ion channel. Disaggregated myocytes are technically more challenging than the expression systems but have the advantage of being suitable for assessing effects on both action potential duration and ionic currents. Although single cell preparations are more fragile, they minimize diffusional barriers to the site of action. Multicellular preparations are stable test systems to study action potential duration. The analysis of parameters for each phase of the action potential such as $V_{max}$ for phase 0 (I_{Na}), APD_{30} or APD_{40} for phase 2 (I_{Ca}) and “triangulation” for phase 3 (I_{K}) can be useful to investigate the effects on specific channels responsible for these phases. In addition, some parameters derived from the Langendorff preparation have been reported to provide information regarding proarrhythmic risk.

Tissue and cell preparations for in vitro assays are obtained from different laboratory animal species including rabbit, ferret, guinea pig, dog, swine, and occasionally from humans. The ionic mechanisms of repolarization in adult rats and mice differ from larger species, including humans (the primary ion currents controlling repolarization in adult rats and mice is I_{so}); therefore, use of tissues from these species is not considered appropriate. Species differences in terms of which cardiac ion channels contribute to cardiac repolarization and to the duration of the action potential should be considered in selecting a test system. When native cardiac tissues or cells are used, the characteristics and source of the preparation should be considered because the distribution of ion channel types varies according to the region and type of cell.

Test substance concentrations for in vitro studies should span a broad range, covering and exceeding the anticipated maximal therapeutic plasma concentration. Ascending concentrations should be tested until a concentration-response curve has been characterized or physicochemical effects become concentration-limiting. Ideally, the duration of exposure should be sufficient to obtain steady-state electrophysiological
effects, unless precluded by the viability of the cell or tissue preparation. The duration of exposure should be indicated. Appropriate positive control substances should be used to establish the sensitivity of the \textit{in vitro} assay system.

Factors that can confound or limit the interpretation of \textit{in vitro} electrophysiology studies include the following:

- The testing of high concentrations of the test substance can be precluded by limited solubility in aqueous physiological salt solutions;
- Adsorption to glass or plastic or non-specific binding to the test matrix can reduce the concentration of the test substance in the incubation or perfusion medium;
- Test substance concentrations can be limited by cytotoxic or physicochemical attributes of the test substance that disrupt cell membrane integrity so that electrophysiological endpoints cannot be obtained;
- Cardiac cells and tissues have limited capacity for drug metabolism; therefore, \textit{in vitro} studies using the parent substance do not provide information on the effects of metabolites. When \textit{in vivo} non-clinical or clinical studies reveal QT interval prolongation that is not consistent with data from \textit{in vitro} studies using the parent substance, testing metabolites in the \textit{in vitro} test systems should be considered.

New technologies for potassium channel assays are being developed. Novel ion channel activity assays can be useful in preliminary screening of test substances to identify lead candidates. It is important to demonstrate concordance between conventional and new technologies before adopting new technologies for regulatory purposes.

Competition binding protocols, in which test substances are studied for their ability to displace a radiolabeled hERG channel blocker from a cell line expressing hERG, are used. However, competition for radioligand-binding sites provides no information on agonistic or antagonistic effects of the test substance on I\textsubscript{Kr}. Moreover, this assay will not identify test substances that bind to hERG at sites other than the radioligand binding sites. Based upon these potential limitations, this assay is not considered a substitute for voltage clamp assays described above.

### 3.1.3 \textbf{In Vivo Electrophysiology Studies}

Intact animal models allow investigation of ventricular repolarization or associated arrhythmias where integrated effects on the full complement of ion channel and cell types are assessed. Also, potential neuronal and hormonal influences on the pharmacodynamic effect of the pharmaceuticals are present in animals.

The QT interval of the ECG is the most commonly used endpoint to gauge effects of a test substance on ventricular repolarization. In specialized electrophysiology studies, information regarding the ventricular repolarization (e.g., monophasic action potential duration and effective refractory period) can also be obtained from \textit{in vivo} models. Additional safety parameters of interest, including blood pressure, heart rate, PR interval, QRS duration, and arrhythmias, can be assessed simultaneously.

The QT interval and heart rate have an inverse, non-linear relationship, which varies among species and between animals within a species. Thus, a change in heart rate
exerts an effect on QT interval, which can confound the assessment of the effect of the
test substance on ventricular repolarization and the QT interval. There are two
important situations where there is variability in heart rate among animals: one is
due to difference in autonomic tone, and the other is due to effects of test substances
on heart rate. Therefore, the interpretation of data from in vivo test systems should
take into account the effect of coincident changes in heart rate. Ideally, QT interval
data obtained after administration of a test substance should be compared with
control and baseline data at similar heart rates. When the heart rate variability is
not due to the test substance, it can be reduced by acclimatization, or the use of
anesthetized animal models. When the effects are due to a test substance, the most
common approach is to correct the QT interval for heart rate (QTc) using formulae
such as Bazett or Fridericia. The choice of heart rate correction formula should be
justified with data from the test system. When differences in heart rate between
treatment and control are large, the correction formulae may not be effective for
assessing risk of QT interval prolongation. An alternative approach is to maintain a
constant heart rate using cardiac pacing. An analysis of QT/RR relationship,
including correction of the QT interval using formulae for individual animals, may be
more appropriate.

Laboratory animal species used for in vivo electrophysiology studies include dog,
monkey, swine, rabbit, ferret, and guinea pig. The ionic mechanisms of repolarization
in adult rats and mice differ from larger species, including humans (the primary ion
currents controlling repolarization in adult rats and mice is I_{to}); therefore, use of these
species is not considered appropriate. The most appropriate in vivo test systems and
species should be selected and justified.

The dose range should be in accord with that discussed in ICH S7A and, whenever
feasible, should include and exceed the anticipated human exposure. The dose range
can be limited by animal intolerance to the test substance, e.g., emesis, tremor, or
hyperactivity. For studies designed to relate the extent of delayed ventricular
repolarization to concentrations of the parent test substance and its metabolites,
controlled exposure via constant intravenous infusion can be used. Monitoring
exposure to the test substance and metabolites (see ICH S3A) provides opportunities
to interpret dose- and concentration-response data and to design follow-up studies, if
appropriate.

Factors that should be considered in conducting studies and interpreting the results
include the following:

- Data acquisition and analysis methods;
- Sensitivity and reproducibility of the test systems;
- Dosing period and measurement points;
- Heart rate and other effects that confound interpretation of QT interval data;
- Inter-species and gender differences, e.g., cardiac electrophysiology,
hemodynamics, or metabolism of pharmaceuticals;
- Pharmaceuticals that have effects on several ion channels can yield complex
dose-response relationships that could be difficult to interpret.
3.1.4 Simulated Pathological Conditions and Arrhythmias

The precise relationship between test substance-induced delay of ventricular repolarization and risk of proarrhythmia is not known. Directly assessing the proarrhythmic risk of pharmaceuticals that prolong the QT interval would be a logical undertaking. Indices of proarrhythmic activity (e.g., electrical instability, temporal and/or spatial dispersion of refractoriness, reverse use-dependency, changes in action potential configuration) and animal models might have utility in assessing proarrhythmia. Interested parties are encouraged to develop these models and test their usefulness in predicting risk in humans.
This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
### S8 Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
<th>New Codification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S8</td>
<td>Approval by the Steering Committee under <em>Step 2</em> and release for public consultation.</td>
<td>18 November 2004</td>
<td>S8 November 2005</td>
</tr>
</tbody>
</table>

#### Current *Step 4* version

<table>
<thead>
<tr>
<th>S8</th>
<th>Approval by the Steering Committee under <em>Step 4</em> and recommendation for adoption to the three ICH regulatory bodies.</th>
<th>15 September 2005</th>
<th>S8</th>
</tr>
</thead>
</table>
IMMUNOTOXICITY STUDIES FOR HUMAN PHARMACEUTICALS

ICH Harmonised Tripartite Guideline

Having reached *Step 4* of the ICH Process at the ICH Steering Committee meeting on 15 September 2005, this guideline is recommended for adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

1. INTRODUCTION ..................................................................................................... 1
   1.1 Objectives of the Guideline ........................................................................... 1
   1.2 Background .................................................................................................. 1
   1.3 Scope of the Guideline ................................................................................ 2
   1.4 Overview ..................................................................................................... 2

2. GUIDELINE ............................................................................................................. 2
   2.1 Factors to Consider in the Evaluation of Potential Immunotoxicity .......... 2
      2.1.1 Standard Toxicity Studies ................................................................. 3
      2.1.2 Pharmacological Properties .............................................................. 3
      2.1.3 Intended Patient Population .............................................................. 4
      2.1.4 Structural Similarity ......................................................................... 4
      2.1.5 Disposition of the Drug .................................................................. 4
      2.1.6 Signs Observed in Clinical Trials or Clinical Use ......................... 4
   2.2 Weight of Evidence Review ............................................................................ 4

3. SELECTION AND DESIGN OF ADDITIONAL IMMUNOTOXICITY STUDIES ................................................................................................................... 4
   3.1 Objectives .................................................................................................... 4
   3.2 Selection of assays ........................................................................................ 4
   3.3 Study Design ............................................................................................... 5
   3.4 Evaluation of Additional Immunotoxicity Studies and Need for Further Studies ..................................................................................................................... 5

4. TIMING OF IMMUNOTOXICITY TESTING IN RELATION TO CLINICAL STUDIES ............................................................................................................. 5

5. REFERENCES ...................................................................................................... 5

Figure 1: Flow Diagram for Recommended Immunotoxicity Evaluation .......... 6

APPENDIX: Methods to Evaluate Immunotoxicity ............................................ 7
1. INTRODUCTION

1.1 Objectives of the Guideline

The objectives of this guideline are to provide (1) recommendations on nonclinical testing approaches to identify compounds which have the potential to be immunotoxic, and (2) guidance on a weight-of-evidence decision making approach for immunotoxicity testing. Immunotoxicity is, for the purpose of this guideline, defined as unintended immunosuppression or enhancement. Drug-induced hypersensitivity and autoimmunity are excluded.

1.2 Background

Evaluation of potential adverse effects of human pharmaceuticals on the immune system should be incorporated into standard drug development. Toxicity to the immune system encompasses a variety of adverse effects. These include suppression or enhancement of the immune response. Suppression of the immune response can lead to decreased host resistance to infectious agents or tumor cells. Enhancing the immune response can exaggerate autoimmune diseases or hypersensitivity. Drug or drug-protein adducts might also be recognized as foreign and stimulate an anti-drug response. Subsequent exposures to the drug can lead to hypersensitivity (allergic) reactions. Much of the science and method development and validation efforts in the past have been focused on evaluating drug development candidates for their potential for either immunosuppression or contact sensitization. No standard approaches for human pharmaceuticals are currently available for testing for respiratory or systemic allergenicity (antigenicity) or drug-specific autoimmunity; testing for these endpoints is not currently required in any region. There are no regional differences in testing approaches of skin sensitization.

Immunosuppression or enhancement can be associated with two distinct groups:

1) Drugs intended to modulate immune function for therapeutic purposes (e.g., to prevent organ transplant rejection) where adverse immunosuppression can be considered exaggerated pharmacodynamics;

2) Drugs not intended to affect immune function but cause immunotoxicity due, for instance, to necrosis or apoptosis of immune cells or interaction with cellular receptors shared by both target tissues and non-target immune system cells.

Anti-proliferative agents used to treat cancer are an example of drugs that produce unintended immunosuppression. In such instances, adverse findings in nonclinical studies are predictive of human immunotoxicity in a rather straightforward manner. That is, specific assays to determine immunotoxicity are probably not valuable in drug risk assessment since the target tissues are usually rapidly dividing cell types, such as bone marrow-derived immune system progenitor cells. Hence, the adverse effects on immune function can be predicted based on pharmacologic activity and can usually be reliably evaluated in non-clinical studies. For other types of compounds not intended to suppress the immune response, distinction between exaggerated pharmacodynamics and non-target effects can be less obvious. As an example, some anti-inflammatory compounds have an effect on certain innate immune functions but do not necessarily affect the adaptive immune response.
1.3  **Scope of the Guideline**
This guideline is focused on providing recommendations on nonclinical testing for immunotoxicity induced by human pharmaceuticals. It is restricted to unintended immunosuppression and immunoenhancement, excluding allergenicity or drug-specific autoimmunity.

This guideline applies to new pharmaceuticals intended for use in humans, as well as to marketed pharmaceuticals proposed for different indications or other variations on the current product label in which the change could result in unaddressed and relevant immunotoxicity issues. In addition, the guideline might also apply to drugs for which clinical signs of immunotoxicity are observed during clinical trials and following approval to market. The guideline does not apply to biotechnology-derived pharmaceutical products covered by ICH S6 Guideline¹ and other biologicals.

Existing guidance documents on sensitization or hypersensitivity remain in force and are not affected by this document. It is beyond the scope of this guideline to provide specific guidance on how each immunotoxicity study should be performed. General methodology guidance is provided in the Appendix.

1.4  **Overview**
The general principles that apply to this guideline are:

1) All new human pharmaceuticals should be evaluated for the potential to produce immunotoxicity;

2) Methods include standard toxicity studies (STS) and additional immunotoxicity studies conducted as appropriate. Whether additional immunotoxicity studies are appropriate should be determined by a weight of evidence review of factor(s) in section 2.1.

The description of the guideline below will follow the recommended decision process in immunotoxicity evaluation as shown in the flow diagram (Figure 1). More detailed descriptions of the testing methods are in the Appendix.

2.  **GUIDELINE**

2.1  **Factors to Consider in the Evaluation of Potential Immunotoxicity**
Factors to consider that might prompt additional immunotoxicity studies can be identified in the following areas: (1) findings from STS; (2) the pharmacological properties of the drug; (3) the intended patient population; (4) structural similarities to known immunomodulators; (5) the disposition of the drug; and (6) clinical information.

The initial screen for potential immunotoxicity involves standard toxicity studies. Data from rodent and non-rodent studies from early short term to more chronic repeat-dose studies should be taken into consideration. Additional details on the parameters that should be evaluated and the reporting of histopathology findings are provided in the Appendix.
2.1.1 Standard Toxicity Studies

Data from STS should be evaluated for signs of immunotoxic potential. Signs that should be taken into consideration are the following:

1) Hematological changes such as leukocytopenia/leukocytosis, granulocytopenia/granulocytosis, or lymphopenia/lymphocytosis;
2) Alterations in immune system organ weights and/or histology (e.g., changes in thymus, spleen, lymph nodes, and/or bone marrow);
3) Changes in serum globulins that occur without a plausible explanation, such as effects on the liver or kidney, can be an indication that there are changes in serum immunoglobulins;
4) Increased incidence of infections;
5) Increased occurrence of tumors can be viewed as a sign of immunosuppression in the absence of other plausible causes such as genotoxicity, hormonal effects, or liver enzyme induction.

Changes in these parameters could reflect immunosuppression or enhanced activation of the immune system. Immunosuppression is usually reflected by reduced values of immune parameters, whereas immunoenhancement is usually reflected by increased values. However, these relationships are not absolute and can be inverted in some cases.

Similar to the assessment of risk with toxicities in other organ systems, the assessment of immunotoxicity should include the following:

- Statistical and biological significance of the changes;
- Severity of the effects;
- Dose/exposure relationship;
- Safety factor above the expected clinical dose;
- Treatment duration;
- Number of species and endpoints affected;
- Changes that may occur secondarily to other factors (e.g., stress, see the Appendix, section 1.4);
- Possible cellular targets and/or mechanism of action;
- Doses which produce these changes in relation to doses which produce other toxicities; and
- Reversibility of effect(s).

2.1.2 Pharmacological Properties

If the pharmacological properties of a test compound indicate it has the potential to affect immune function (e.g., anti-inflammatory drugs), additional immunotoxicity testing should be considered. Information obtained from the nonclinical pharmacology studies on the ability of the compound to affect the immune system could be used in a weight of evidence approach to decide if additional immunotoxicity studies are needed.
2.1.3 Intended Patient Population
Additional immunotoxicity studies might be warranted if the majority of the patient population for whom the drug is intended is immunocompromised by a disease state or concurrent therapy.

2.1.4 Structural Similarity
Compounds structurally similar to compounds with known immunosuppressive properties should also be considered for additional immunotoxicity testing.

2.1.5 Disposition of the Drug
If the compound and/or its metabolites are retained at high concentrations in cells of the immune system, additional immunotoxicity testing should be considered.

2.1.6 Signs Observed in Clinical Trials or Clinical Use
Clinical findings suggestive of immunotoxicity in patients exposed to the drug could call for additional nonclinical immunotoxicity testing.

2.2 Weight of Evidence Review
A weight of evidence review should be performed on information from all the factors outlined above to determine whether a cause for concern exists. A finding of sufficient magnitude in a single area should trigger additional immunotoxicity studies. Findings from two or more factors, each one of which would not be sufficient on its own, could trigger additional studies. If additional immunotoxicity studies are not performed, the sponsor should provide justification.

3. SELECTION AND DESIGN OF ADDITIONAL IMMUNOTOXICITY STUDIES
3.1 Objectives
If a cause for concern is identified, additional immunotoxicity studies should be performed to verify the immunotoxic potential of the compound. These studies can also help determine the cell type affected reversibility, and the mechanism of action. This type of information can also provide more insight into potential risk and possibly lead to biomarker selection for clinical studies.

3.2 Selection of assays
If the weight-of-evidence review indicates that additional immunotoxicity studies are called for, there are a number of assays which can be used. If there are changes in standard toxicity testing data suggesting immunotoxicity, the type of additional immunotoxicity testing that is considered appropriate will depend on the nature of the immunological changes observed and the concerns raised by the class of compound. It is recommended that an immune function study be conducted, such as a T-cell dependent antibody response (TDAR). If specific cell types that are affected in STS are not known to participate in a TDAR, assays that measure function of that specific affected cell type might be conducted (see the Appendix). Where a specific target is not identified, an immune function study such as the TDAR is recommended. In addition, immunophenotyping of leukocyte populations, a non-functional assay, can be conducted to identify the specific cell populations affected and might provide useful clinical biomarkers.
3.3 Study Design
To assess drug-induced immunotoxicity, a generally accepted study design in rodents is a 28 day study with consecutive daily dosing. Adaptations of immunotoxicity assays have been described using non-rodent species. The species, strain, dose, duration, and route of administration used in additional immunotoxicity studies should be consistent, where possible, with the standard toxicity study in which an adverse immune effect was observed. Usually both sexes should be used in these studies, excluding nonhuman primates. Rationale should be given when one sex is used in other species. The high dose should be above the no observed adverse effect level (NOAEL) but below a level inducing changes secondary to stress (see Appendix, section 1.4). Multiple dose levels are recommended in order to determine dose-response relationships and the dose at which no immunotoxicity is observed.

3.4 Evaluation of Additional Immunotoxicity Studies and Need for Further Studies
Results from additional immunotoxicity studies should be evaluated as to whether sufficient data are available to reasonably determine the risk of immunotoxicity:

1. Additional studies might show that no risk of immunotoxicity can be detected and no further testing is called for;
2. Additional studies might demonstrate a risk of immunotoxicity but fail to provide sufficient data to make a reasonable risk-benefit decision. In this case further testing might help provide sufficient information for the risk-benefit decision;
3. If the overall risk-benefit analysis suggests that the risk of immunotoxicity is considered acceptable and/or can be addressed in a risk management plan (see ICH E2E Guideline²), then no further testing in animals might be called for.

4. TIMING OF IMMUNOTOXICITY TESTING IN RELATION TO CLINICAL STUDIES
If the weight-of-evidence review indicates that additional immunotoxicity studies are appropriate, these should be completed before exposure of a large population of patients, usually Phase III. This will allow for the incorporation of monitoring immune system parameters in the clinical studies if appropriate. The timing of the additional immunotoxicity testing might be determined by the nature of the effect by the test compound and the type of clinical testing that would be called for if a positive finding is observed with the additional immunotoxicity testing. If the target patient population is immunocompromised, immunotoxicity testing can be initiated at an earlier time point in the development of the drug.

5. REFERENCES
1. ICH Harmonised Tripartite Guideline (S6) “Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals”
2. ICH Harmonised Tripartite Guideline (E2E) “Pharmacovigilance Planning”
Figure 1: Flow Diagram for Recommended Immunotoxicity Evaluation

All human pharmaceuticals (non-biologicals)

(2.1) Identify factors to consider

(2.2) Weight of evidence (WoE) review

WoE review warrants additional immunotoxicity Testing ?

NO → Additional nonclinical immunotoxicity testing not needed

YES → (3.0) Conduct additional immunotoxicity studies

(3.4) Significant changes observed?

NO → (3.4 Pt 1) Further nonclinical immunotoxicity testing not needed

YES

(3.4) Sufficient data for risk assessment / risk management?

YES → (3.4 Pt 3) Further nonclinical immunotoxicity testing not needed

NO → (3.4 Pt 2) Consider further immunotoxicity testing
APPENDIX: Methods to Evaluate Immunotoxicity

1. **Standard Toxicity Studies**

The following table lists the parameters that should be evaluated in standard toxicity studies for signs of immunotoxicity. These parameters (excluding hematology and clinical chemistry) and methods for obtaining samples and evaluating tissue sections are described in more detail in documents from professional toxicological pathology societies.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specific Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematology</td>
<td>Total and absolute differential leukocyte counts</td>
</tr>
<tr>
<td>Clinical Chemistry</td>
<td>Globulin levels(^1) and A/G ratios</td>
</tr>
<tr>
<td>Gross pathology</td>
<td>Lymphoid organs / tissues</td>
</tr>
<tr>
<td>Organ weights</td>
<td>Thymus, spleen (optional: lymph nodes)</td>
</tr>
<tr>
<td>Histology</td>
<td>Thymus, spleen, draining lymph node and at least one additional lymph node, bone marrow(^2), Peyer’s patch(^3), BALT(^4), NALT(^4)</td>
</tr>
</tbody>
</table>

\(^1\) Unexplained alterations in globulin levels could call for measurement of immunoglobulins.

\(^2\) Unexplained alterations in peripheral blood cell lines or histopathologic findings might suggest that cytologic evaluation of the bone marrow would be appropriate.

\(^3\) Oral administration only.

\(^4\) For inhalation or nasal route only. BALT: bronchus-associated lymphoid tissues. NALT: nasal-associated lymphoid tissues

1.1 **Hematology and Clinical Chemistry**

Total leukocyte counts and absolute differential leukocyte counts are recommended to assess immunotoxicity. When evaluating changes in globulin levels, other factors should be taken into account (e.g., liver toxicity, nephrotoxicity). Changes in serum globulins can be an indication that there are changes in serum immunoglobulins. Although serum immunoglobulins are an insensitive indicator of immunosuppression, changes in immunoglobulins levels can be useful in certain situations in order to better understand target cell populations or mechanism of action.

1.2 **Gross Pathology and Organ Weights**

All lymphoid tissues should be evaluated for gross changes at necropsy. However, this can be more difficult for the Peyer’s patches of rodents due to the small size. Spleen and thymus weights should be recorded. To minimize variability of spleen weights in dogs and monkeys, bleeding the animals thoroughly at necropsy is recommended. Atrophy of the thymus with aging can preclude obtaining accurate thymus weight.

1.3 **Histopathological Examination**

Histopathological changes of the spleen and thymus should be evaluated as an indicator of systemic immunotoxicity. The lymphoid tissue that drains or contacts the
site of drug administration (and therefore is exposed to the highest concentration of
the drug) should be examined. These sites include the Peyer’s patches and mesenteric
lymph nodes for orally administered drugs, bronchus-associated lymphoid tissues
(BALT) for drugs administered by the inhalation route, nasal-associated lymphoid
tissues (NALT) for drugs administered by the inhalation or nasal route (if possible),
and the most proximal regional draining lymph nodes for drugs administered by the
dermal, intramuscular, intradermal, intrathecal, or subcutaneous routes. The specific
node selected and the additional lymph node should be at the discretion of the sponsor
based on the sponsor’s experience. For intravenously administered drugs, the spleen
can be considered the draining lymphoid tissue.

It is recommended that a “semi-quantitative” description of changes in compartments
of lymphoid tissues be used in recording changes and reporting treatment-related
changes in lymphoid tissues.

1.4 Interpretation of Stress Related Changes
With standard toxicity studies, doses near or at the maximum tolerated dose can
result in changes to the immune system related to stress (e.g., by exaggerated
pharmacodynamic action). These effects on the immune system might be mediated by
increased corticosterone or cortisol release or other mediators. Commonly observed
stress-related immune changes include increases in circulating neutrophils, decreases
in circulating lymphocytes, decreases in thymus weight, decreases in thymic cortical
cellularity and associated histopathologic changes, and changes in spleen and lymph
node cellularity. Increases in adrenal gland weight and/or histologic evidence of
adrenal cortical hyperplasia can also be observed. Thymic weight decreases in the
presence of clinical signs, such as decreased body weight and physical activity, are too
often attributed to stress. These findings on their own should not be considered
sufficient evidence of stress-related immunotoxicity. The evidence of stress should be
compelling in order to justify not conducting additional immunotoxicity studies.

2. Additional Immunotoxicity Studies
2.1 Assay Characterization and Validation
In general, the immunotoxicity test selected should be widely used and have been
demonstrated to be adequately sensitive and specific for known immunosuppressive
agents. However, in certain situations, extensive validation might have not been
completed and/or the assay might not be widely used. In these situations, a
scientific/mechanistic basis for use of the assay is called for and, if feasible,
appropriate positive controls should be incorporated.

There can be variations of response for each type of immunotoxicity test used by
different labs. In most situations, these changes do not affect the ability of the assay
to assess immunotoxicity. However, to ensure proper assay performance and lab
proficiency, several standard technical validation parameters should be observed.
These parameters can include determining intra- and inter-assay precision,
technician-to-technician precision, limit of quantitation, linear region of quantitation
and test sample stability. In addition, assay sensitivity to known immunosuppressive
agents should be established. It is recommended that each laboratory test a positive
control concomitantly with an investigational compound or periodically in order to
demonstrate proficiency of performance, except for studies with non-human primates.
For immunophenotyping, if properly validated technically, the addition of positive controls for each study might not be needed.

Immunotoxicity studies are expected to be performed in compliance with Good Laboratory Practice (GLP). It is recognized that some specialized assays, such as those described below, might not comply fully with GLP.

2.2 T-cell Dependent Antibody Response (TDAR)

The TDAR should be performed using a recognized T-cell dependent antigen (e.g., sheep red blood cells (SRBC) or keyhole limpet hemocyanin (KLH)) that results in a robust antibody response. The endpoint selected should be justified as the most appropriate for the chosen assay and the selected species.

Antigens for immunization should not be used with adjuvants without justification. Alum might be considered acceptable for use only in non-human primate studies. The relative TDAR response can be strain-dependent, especially in mice. With outbred rats, there can be significant variability among rats within the same group. Inbred rat strains could be used with provision of sufficient exposure data to bridge to the strain used in the STS.

Antibody can be measured by using an ELISA or other immunoassay methods. One advantage of this method over the antibody forming cell response is that samples can be collected serially during the study. In monkeys, serial blood collection can be important due to the high inter-animal variability in the kinetics of the response. For these studies, data can be expressed as the sum of the antibody response over several collection dates (e.g., area under the curve).

When SRBC antigens are used for an ELISA, the preparation of the capture antigen that is coated on the plates is considered critical. Whole fixed erythrocytes or membrane preparations can be used as the SRBC capture antigen. ELISA results should be expressed either as concentration or as titer, but expression as optical densities is not recommended.

2.3 Immunophenotyping

Immunophenotyping is the identification and/or enumeration of leukocyte subsets using antibodies. Immunophenotyping is usually conducted by flow cytometric analysis or by immunohistochemistry.

Flow cytometry, when employed to enumerate specific cell populations, is not a functional assay. However, flow cytometry can be used to measure antigen-specific immune responses of lymphocytes. Data obtained from peripheral blood can be useful as a bridge for clinical studies in which peripheral blood leukocytes are also evaluated. It is recommended that absolute numbers of lymphocyte subsets as well as percentages be used in evaluating treatment-related changes.

One of the advantages of immunohistochemistry over flow cytometry is that tissues from standard toxicity studies can be analyzed retrospectively if signs of immunotoxicity are observed. In addition, changes in cell types within a specific compartment within the lymphoid tissue can be observed. Some of the lymphocyte markers for certain species are sensitive to formalin fixation and can only be localized in tissue that are either fixed with certain fixatives or flash frozen. Quantitation of leukocytes and intensity of staining is much more difficult with immunohistochemistry.
When immunophenotyping studies are used to characterize or identify alterations in specific leukocyte populations, the choice of the lymphoid organs and/or peripheral blood to be evaluated should be based on changes observed. Immunophenotyping can be easily added to standard repeat dose toxicity studies and changes can be followed during the dosing phase and periods without drug exposure (reversal period).

2.4 Natural Killer Cell Activity Assays

Natural killer (NK) cell activity assays can be conducted if immunophenotyping studies demonstrate a change in number, or if STS studies demonstrate increased viral infection rates, or in response to other factors. In general, all NK cell assays are ex vivo assays in which tissues (e.g., spleen) or blood are obtained from animals that have been treated with the test compound. Cell preparations are co-incubated with target cells that have been labeled with $^{51}$Cr. New methods that involve non-radioactive labels can be used if adequately validated. Different effector to target cell ratios should be evaluated for each assay to obtain a sufficient level of cytotoxicity and generate a curve.

2.5 Host Resistance Studies

Host resistance studies involve challenging groups of mice or rats treated with the different doses of test compound with varying concentrations of a pathogen (bacteria, fungal, viral, parasitic) or tumor cells. Infectivity of the pathogens or tumor burden observed in vehicle versus test compound treated animals is used to determine if the test compound is able to alter host resistance. Models have been developed to evaluate a wide range of pathogens such as Listeria monocytogenes, Streptococcus pneumoniae, Candida albicans, influenza virus, cytomegalovirus, Plasmodium yoelii and Trichinella spiralis. Tumor host resistance models in mice have used the B16F10 melanoma and PYB6 sarcoma tumor cell lines.

Host resistance assays can provide information on the susceptibility to particular classes of infectious agents or tumor cells and can have an impact on the risk management plan. In addition, they can have an important role in identifying or confirming the cell type affected by a test compound. Moreover, host resistance assays involve innate immune mechanisms for which specific immune function assays have not been developed. In conducting host resistance studies, the investigator should carefully consider the direct or indirect (non-immune mediated) effects of the test compound on the growth and pathogenicity of the organism or tumor cell. For instance, compounds that inhibit the proliferation of certain tumor cells can seem to increase host resistance. An in vitro assay to test direct effects on the organism is recommended.

2.6 Macrophage/Neutrophil Function

In vitro macrophage and neutrophil function assays (phagocytosis, oxidative burst, chemotaxis, and cytolytic activity) have been published for several species. These assays assess macrophage/neutrophil function of cells exposed to the test compound in vitro or obtained from animals treated with the test compound (ex vivo assay). In vitro exposure to test compound can also be investigated. An in vivo assay can also be used to assess the effects on the reticuloendothelial cell to phagocytize radioactively or fluorescently labeled targets.
2.7 Assays to Measure Cell-Mediated Immunity

Assays to measure cell-mediated immunity have not been as well established as those used for the antibody response. These are in vivo assays where antigens are used for sensitization. The endpoint is the ability of drugs to modulate the response to challenge. Delayed-type hypersensitivity (DTH) reactions with protein immunization and challenge have been reported for mice and rats. Models in which contact sensitizers are used have been explored in mice but have not been well validated or extensively used. Cytotoxic T cell response can be generated in mice using a virus, tumor cell line, or allograft as the antigenic challenge. Monkey DTH reactions have also been reported. However, these reactions in monkeys are very difficult to consistently reproduce. In addition, one should make sure that the DTH response is not mistaken for an antibody and complement mediated Arthus reaction.
ICH Harmonised Tripartite Guideline

Nonclinical Evaluation for Anticancer Pharmaceuticals

S9

Current Step 4 version
dated 29 October 2009

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.
S9
Document History

<table>
<thead>
<tr>
<th>First Codification</th>
<th>History</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S9</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>13 November 2009</td>
</tr>
</tbody>
</table>

Current Step 4 version

| S9 | Approval by the Steering Committee under Step 4 and recommendation for adoption to the three ICH regulatory bodies. | 29 October 2009   |
Having reached Step 4 of the ICH Process at the ICH Steering Committee meeting on October 29, 2009, this guideline is recommended for adoption to the three regulatory parties to ICH.

TABLE OF CONTENTS

1. INTRODUCTION ..................................................................................................... 1
1.1 Objectives of the Guideline ................................................................................... 1
1.2 Background ........................................................................................................... 1
1.3 Scope ...................................................................................................................... 1
1.4 General Principles ................................................................................................ 2
2. STUDIES TO SUPPORT NONCLINICAL EVALUATION ........................................... 2
2.1 Pharmacology ........................................................................................................ 2
2.2 Safety Pharmacology ............................................................................................ 3
2.3 Pharmacokinetics .................................................................................................. 3
2.4 General Toxicology ................................................................................................. 3
2.5 Reproduction Toxicology ....................................................................................... 4
2.6 Genotoxicity ........................................................................................................... 4
2.7 Carcinogenicity ..................................................................................................... 5
2.8 Immunotoxicity ..................................................................................................... 5
2.9 Photosafety testing ............................................................................................... 5
3. NONCLINICAL DATA TO SUPPORT CLINICAL TRIAL DESIGN AND MARKETING ... 5
3.1 Start Dose for First Administration in Humans ................................................. 5
3.2 Dose Escalation and the Highest Dose in a Clinical Trial .................................. 5
3.3 Duration and Schedule of Toxicology Studies to Support Initial Clinical Trials ................................................................................................................................ 6
3.4 Duration of Toxicology Studies to Support Continued Clinical Development and Marketing ................................................................................................................. 6
3.5 Combination of Pharmaceuticals ........................................................................... 6
3.6 Nonclinical Studies to Support Trials in Pediatric Populations ........................ 7
4. OTHER CONSIDERATIONS .................................................................................... 7
4.1 Conjugated Products ............................................................................................. 7
4.2 Liposomal Products ............................................................................................... 7
4.3 Evaluation of Drug Metabolites ............................................................................ 7
4.4 Evaluation of Impurities ....................................................................................... 7
5. NOTES .................................................................................................................... 9
1. INTRODUCTION

1.1 Objectives of the Guideline

The purpose of this guidance is to provide information to assist in the design of an appropriate program of nonclinical studies for the development of anticancer pharmaceuticals. The guidance provides recommendations for nonclinical evaluations to support the development of anticancer pharmaceuticals in clinical trials for the treatment of patients with advanced disease and limited therapeutic options.

This guideline aims to facilitate and accelerate the development of anticancer pharmaceuticals and to protect patients from unnecessary adverse effects, while avoiding unnecessary use of animals, in accordance with the 3R principles (reduce/refine/replace), and other resources.

As appropriate, the principles described in other ICH guidelines should be considered in the development of anticancer pharmaceuticals. Specific situations where recommendations for nonclinical testing deviate from other guidance are described in this document.

1.2 Background

Because malignant tumors are life-threatening, the death rate from these diseases is high, and existing therapies have limited effectiveness, it is desirable to provide new, effective anticancer drugs to patients more expeditiously.

There have been no internationally accepted objectives or recommendations on the design and conduct of nonclinical studies to support the development of anticancer pharmaceuticals in clinical trials for the treatment of patients with advanced disease and limited therapeutic options. Nonclinical evaluations are conducted to:

1) identify the pharmacologic properties of a pharmaceutical;
2) establish a safe initial dose level for the first human exposure; and
3) understand the toxicological profile of a pharmaceutical (e.g., identification of target organs, exposure-response relationships, and reversibility).

In the development of anticancer drugs, clinical studies often involve cancer patients whose disease condition is progressive and fatal. In addition, the dose levels in these clinical studies often are close to or at the adverse effect dose levels. For these reasons, the type, timing and flexibility called for in the design of nonclinical studies of anticancer pharmaceuticals can differ from those elements in nonclinical studies for other pharmaceuticals.

1.3 Scope

This guideline provides information for pharmaceuticals that are intended to treat cancer in patients with serious and life threatening malignancies. For the purpose of this guideline, this patient population is referred to as patients with advanced cancer.
The guideline applies to both small molecule and biotechnology-derived pharmaceuticals (biopharmaceuticals), regardless of the route of administration. This guideline describes the type and timing of nonclinical studies in relation to the development of anticancer pharmaceuticals in patients with advanced cancer and references other guidance as appropriate. It describes the minimal considerations for initial clinical trials in patients with advanced cancer whose disease is refractory or resistant to available therapy, or where current therapy is not considered to be providing benefit. The nonclinical data to support Phase I and the clinical Phase I data would normally be sufficient for moving to Phase II and into second or first line therapy in patients with advanced cancer. The guideline also describes further non-clinical data to be collected during continued clinical development in patients with advanced cancer. When an anticancer pharmaceutical is further investigated in cancer patient populations with long expected survival (e.g., those administered pharmaceuticals on a chronic basis to reduce the risk of recurrence of cancer), the recommendations for and timing of additional nonclinical studies depend upon the available nonclinical and clinical data and the nature of the toxicities observed.

This guideline does not apply to pharmaceuticals intended for cancer prevention, treatment of symptoms or side effects of chemotherapeutics, studies in healthy volunteers, vaccines, or cellular or gene therapy. If healthy volunteers are included in clinical trials, the ICH M3 guideline should be followed. Radiopharmaceuticals are not covered in this guideline, but some of the principles could be adapted.

1.4 General Principles

The development of each new pharmaceutical calls for studies designed to characterize its pharmacological and toxicological properties according to its intended use in humans. Modification of "standard" nonclinical testing protocols generally is warranted to address novel characteristics associated with the pharmaceutical or with the manner in which it is to be used in humans.

The manufacturing process can change during the course of development. However, the active pharmaceutical substance used in nonclinical studies should be well characterized and should adequately represent the active substance to be used in the clinical trials.

In general, nonclinical safety studies that are used to support the development of a pharmaceutical should be conducted in accordance with Good Laboratory Practices.

2. STUDIES TO SUPPORT NONCLINICAL EVALUATION

2.1 Pharmacology

Prior to Phase I studies, preliminary characterization of the mechanism(s) of action and schedule dependencies as well as anti-tumor activity of the pharmaceutical should have been made. Appropriate models should be selected based on the target and mechanism of action, but the pharmaceutical need not be studied using the same tumor types intended for clinical evaluation.
These studies can:

- provide nonclinical proof of principle;
- guide schedules and dose-escalation schemes;
- provide information for selection of test species;
- aid in start dose selection and selection of investigational biomarkers, where appropriate; and
- if relevant, justify pharmaceutical combinations.

Understanding the secondary pharmacodynamic properties of a pharmaceutical could contribute to the assessment of safety for humans, and those properties might be investigated as appropriate.

### 2.2 Safety Pharmacology

An assessment of the pharmaceutical's effect on vital organ functions (including cardiovascular, respiratory and central nervous systems) should be available before the initiation of clinical studies; such parameters could be included in general toxicology studies. Detailed clinical observations following dosing and appropriate electrocardiographic measurements in non-rodents are generally considered sufficient. Conducting stand-alone safety pharmacology studies to support studies in patients with advanced cancer is not called for. In cases where specific concerns have been identified that could put patients at significant additional risks in clinical trials, appropriate safety pharmacology studies described in ICH S7A and/or S7B should be considered. In the absence of a specific risk, such studies will not be called for to support clinical trials or for marketing.

### 2.3 Pharmacokinetics

The evaluation of limited pharmacokinetic parameters (e.g., peak plasma/serum levels, area under the curve (AUC), and half-life) in the animal species used for nonclinical studies can facilitate dose selection, schedule and escalation during Phase I studies. Further information on absorption, distribution, metabolism and excretion of the pharmaceutical in animals should normally be generated in parallel with clinical development.

### 2.4 General Toxicology

The primary objective of Phase I clinical trials in patients with advanced cancer is to assess the safety of the pharmaceutical. Phase I assessments can include dosing to a maximum tolerated dose (MTD) and dose limiting toxicity (DLT). Toxicology studies to determine a no observed adverse effect level (NOAEL) or no effect level (NOEL) are not considered essential to support clinical use of an anticancer pharmaceutical. As the toxicity of the pharmaceutical can be greatly influenced by its schedule of administration, an approximation of its clinical schedule should be evaluated in toxicology studies. This is further discussed in Section 3.3 and 3.4.

Assessment of the potential to recover from toxicity should be provided to understand whether serious adverse effects are reversible or irreversible. A study that includes a terminal non-dosing period is called for if there is severe toxicity.
Nonclinical Evaluation for Anticancer Pharmaceuticals

at approximate clinical exposure and recovery cannot be predicted by scientific assessment. This scientific assessment can include the extent and severity of the pathologic lesion and the regenerative capacity of the organ system showing the effect. If a study of recovery is called for, it should be available to support clinical development. The demonstration of complete recovery is not considered essential.

For small molecules, the general toxicology testing usually includes rodents and nonrodents. In certain circumstances, determined case-by-case, alternative approaches can be appropriate (e.g., for genotoxic drugs targeting rapidly dividing cells, a repeat-dose toxicity study in one rodent species might be considered sufficient, provided the rodent is a relevant species). For biopharmaceuticals, see ICH S6 for the number of species to be studied.

Toxicokinetic evaluation should be conducted as appropriate.

2.5 Reproduction Toxicology

An embryofetal toxicology assessment is conducted to communicate potential risk for the developing embryo or fetus to patients who are or might become pregnant. Embryofetal toxicity studies of anticancer pharmaceuticals should be available when the marketing application is submitted, but these studies are not considered essential to support clinical trials intended for the treatment of patients with advanced cancer. These studies are also not considered essential for the purpose of marketing applications for pharmaceuticals that are genotoxic and target rapidly dividing cells (e.g., crypt cells, bone marrow) in general toxicity studies or belong to a class that has been well characterized as causing developmental toxicity.

For small molecules, embryofetal toxicology studies are typically conducted in two species as described by ICH S5(R2). In cases where an embryofetal developmental toxicity study is positive for embryofetal lethality or teratogenicity, a confirmatory study in a second species is usually not warranted.

For biopharmaceuticals, an assessment in one pharmacologically relevant species should usually be sufficient. This assessment might be done by evaluating the toxicity during the period of organogenesis or study designs as described by ICH S6. Alternative approaches might be considered appropriate if scientifically justified. The alternative approaches might include a literature assessment, assessment of placental transfer, the direct or indirect effects of the biopharmaceutical, or other factors.

A study of fertility and early embryonic development is not warranted to support clinical trials or for marketing of pharmaceuticals intended for the treatment of patients with advanced cancer. Information available from general toxicology studies on the pharmaceutical’s effect on reproductive organs should be used as the basis of the assessment of impairment of fertility.

A pre- and postnatal toxicology study is generally not warranted to support clinical trials or for marketing of pharmaceuticals for the treatment of patients with advanced cancer.

2.6 Genotoxicity

Genotoxicity studies are not considered essential to support clinical trials for therapeutics intended to treat patients with advanced cancer. Genotoxicity studies
should be performed to support marketing (see ICH S2). The principles outlined in ICH S6 should be followed for biopharmaceuticals. If the \textit{in vitro} assays are positive, an \textit{in vivo} assay might not be warranted.

2.7 Carcinogenicity

The appropriateness of a carcinogenicity assessment for anticancer pharmaceuticals is described in ICH S1A. Carcinogenicity studies are not warranted to support marketing for therapeutics intended to treat patients with advanced cancer.

2.8 Immunotoxicity

For most anticancer pharmaceuticals, the design components of the general toxicology studies are considered sufficient to evaluate immunotoxic potential and support marketing. For immunomodulatory pharmaceuticals, additional endpoints (such as immunophenotyping by flow cytometry) might be included in the study design.

2.9 Photosafety testing

An initial assessment of phototoxic potential should be conducted prior to Phase I, based on photochemical properties of the drug and information on other members in the class. If assessment of these data indicates a potential risk, appropriate protective measures should be taken during outpatient trials. If the photosafety risk cannot be adequately evaluated based on nonclinical data or clinical experience, a photosafety assessment consistent with the principles described in ICH M3 should be provided prior to marketing.

3. Nonclinical Data to Support Clinical Trial Design and Marketing

3.1 Start Dose for First Administration in Humans

The goal of selecting the start dose is to identify a dose that is expected to have pharmacologic effects and is reasonably safe to use. The start dose should be scientifically justified using all available nonclinical data (e.g., pharmacokinetics, pharmacodynamics, toxicity), and its selection based on various approaches (see Note 2). For most systemically administered small molecules, interspecies scaling of the animal doses to an equivalent human dose is usually based on normalization to body surface area. For both small molecules and biopharmaceuticals, interspecies scaling based on body weight, AUC, or other exposure parameters might be appropriate.

For biopharmaceuticals with immune agonistic properties, selection of the start dose using a minimally anticipated biologic effect level (MABEL) should be considered.

3.2 Dose Escalation and the Highest Dose in a Clinical Trial

In general, the highest dose or exposure tested in the nonclinical studies does not limit the dose-escalation or highest dose investigated in a clinical trial in patients with cancer. When a steep dose- or exposure-response curve for severe toxicity is observed in nonclinical toxicology studies, or when no preceding marker of severe
toxicity is available, smaller than usual dose increments (fractional increments rather than dose doubling) should be considered.

3.3 Duration and Schedule of Toxicology Studies to Support Initial Clinical Trials

In Phase I clinical trials, treatment can continue according to the patient’s response, and in this case, a new toxicology study is not called for to support continued treatment beyond the duration of the completed toxicology studies.

The design of nonclinical studies should be appropriately chosen to accommodate different dosing schedules that might be utilized in initial clinical trials. It is not expected that the exact clinical schedule always will be followed in the toxicological study, but the information provided from the toxicity studies should be sufficient to support the clinical dose and schedule and to identify potential toxicity. For example, one factor that can be considered is the half-life in the test species and the projected (or known) half-life in humans. Other factors could include exposure assessment, toxicity profile, saturation of receptors, etc. Table 1 provides examples of nonclinical treatment schedules that are commonly used in anticancer pharmaceutical development and can be used for small molecules or biopharmaceuticals. In cases where the available toxicity information does not support a change in clinical schedules, an additional toxicity study in a single species is usually sufficient.

3.4 Duration of Toxicology Studies to Support Continued Clinical Development and Marketing

The nonclinical data to support Phase I and the clinical Phase I data would normally be sufficient for moving to Phase II and into second or first line therapy in patients with advanced cancer. In support of continued development of an anticancer pharmaceutical for patients with advanced cancer, results from repeat dose studies of 3 months’ duration following the intended clinical schedule should be provided prior to initiating Phase III studies. For most pharmaceuticals intended for the treatment of patients with advanced cancer, nonclinical studies of 3 months duration are considered sufficient to support marketing.

When considering a change in the clinical schedule, an evaluation of the existing clinical data should be conducted to justify such change. If the clinical data alone are inadequate to support the change in schedule, the factors discussed in Section 3.3 above should be considered.

3.5 Combination of Pharmaceuticals

Pharmaceuticals planned for use in combination should be well studied individually in toxicology evaluations. Data to support a rationale for the combination should be provided prior to starting the clinical study. In general, toxicology studies investigating the safety of combinations of pharmaceuticals intended to treat patients with advanced cancer are not warranted. If the human toxicity profile of the pharmaceuticals has been characterized, a nonclinical study evaluating the combination is not usually warranted. For studies in which at least one of these compounds is in early stage development (i.e., the human toxicity profile has not been characterized), a pharmacology study to support the rationale for the combination should be provided. This study should provide evidence of increased activity in the
absence of a substantial increase in toxicity on the basis of limited safety endpoints, such as mortality, clinical signs, and body weight. Based on available information, a determination should be made whether or not a dedicated toxicology study of the combination is warranted.

3.6 Nonclinical Studies to Support Trials in Pediatric Populations

The general paradigm for investigating most anticancer pharmaceuticals in pediatric patients is first to define a relatively safe dose in adult populations and then to assess some fraction of that dose in initial pediatric clinical studies. The recommendations for nonclinical testing outlined elsewhere in this document also apply for this population. Studies in juvenile animals are not usually conducted in order to support inclusion of pediatric populations for the treatment of cancer. Conduct of studies in juvenile animals should be considered only when human safety data and previous animal studies are considered insufficient for a safety evaluation in the intended pediatric age group.

4. OTHER CONSIDERATIONS

4.1 Conjugated Products

Conjugated products are pharmaceuticals covalently bound to carrier molecules, such as proteins, lipids, or sugars. The safety of the conjugated material is the primary concern. The safety of the unconjugated material, including the linker used, can have a more limited evaluation. Stability of the conjugate in the test species and human plasma should be provided. A toxicokinetic evaluation should assess both the conjugated and the unconjugated compound after administration of the conjugated material.

4.2 Liposomal Products

A complete evaluation of the liposomal product is not warranted if the unencapsulated material has been well characterized. As appropriate, the safety assessment should include a toxicological evaluation of the liposomal product and a limited evaluation of the unencapsulated pharmaceutical and carrier (e.g., a single arm in a toxicology study). The principle described here might also apply to other similar carriers. A toxicokinetic evaluation should be conducted as appropriate. If possible, such an evaluation should assess both the liposomal product and the free compound after administration of the liposomal product.

4.3 Evaluation of Drug Metabolites

In some cases, metabolites that have been identified in humans have not been qualified in nonclinical studies. For these metabolites, a separate evaluation is generally not warranted for patients with advanced cancer.

4.4 Evaluation of Impurities

It is recognized that impurity standards have been based on a negligible risk, as discussed in ICH Q3A and Q3B. Exceeding the established limits for impurities identified in these ICH guidelines could be appropriate for anticancer
pharmaceuticals, and a justification should be provided in the marketing application. The justification could include the disease being treated and the patient population, the nature of the parent pharmaceutical (pharmacologic properties, genotoxicity and carcinogenic potential, etc.), duration of treatment, and the impact of impurity reduction on manufacturing. Further, the qualification assessment could include consideration of either the dose or concentration tested in nonclinical study relative to clinical levels. For genotoxic impurities, several approaches have been used to set limits based on increase in lifetime risk of cancer. Such limits are not appropriate for pharmaceuticals intended to treat patients with advanced cancer, and justifications described above should be considered to set higher limits. Impurities that are also metabolites present in animal and/or human studies are generally considered qualified.

### Table 1: Examples of Treatment Schedules for Anticancer Pharmaceuticals to Support Initial Clinical Trials

<table>
<thead>
<tr>
<th>Clinical Schedule</th>
<th>Examples of Nonclinical Treatment Schedule¹,²,³,⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once every 3-4 weeks</td>
<td>Single dose</td>
</tr>
<tr>
<td>Daily for 5 days every 3 weeks</td>
<td>Daily for 5 days</td>
</tr>
<tr>
<td>Daily for 5-7 days, alternating weeks</td>
<td>Daily for 5-7 days, alternating weeks (2-dose cycles)</td>
</tr>
<tr>
<td>Once a week for 3 weeks, 1 week off</td>
<td>Once a week for 3 weeks</td>
</tr>
<tr>
<td>Two or three times a week</td>
<td>Two or three times a week for 4 weeks</td>
</tr>
<tr>
<td>Daily</td>
<td>Daily for 4 weeks</td>
</tr>
<tr>
<td>Weekly</td>
<td>Once a week for 4-5 doses</td>
</tr>
</tbody>
</table>

¹ Table 1 describes the dosing Phase. The timing of the toxicity assessment(s) in the non-clinical studies should be scientifically justified based on the anticipated toxicity profile and the clinical schedule. For example, both a sacrifice shortly after the dosing Phase to examine early toxicity and a later sacrifice to examine late onset of toxicity should be considered.

² For further discussion regarding flexibility in the relationship of the clinical schedule and the non-clinical toxicity studies, see Section 3.3.

³ The treatment schedules described in the table do not specify recovery periods (see Section 2.4 and Note 1 regarding recovery).

⁴ The treatment schedules described in this table should be modified as appropriate for molecules with extended pharmacodynamic effects, long half-lives or potential for anaphylactic reactions. In addition, the potential effects of immunogenicity should be considered (see ICH S6).
5. Notes

1. For nonrodent studies, dose groups usually consist of at least 3 animals/sex/group, with an additional 2/sex/group for recovery, if appropriate (see Section 2.4). Both sexes should generally be used, or justification should be given for specific omissions.

2. A common approach for many small molecules is to set a start dose at 1/10 the Severely Toxic Dose in 10% of the animals (STD 10) in rodents. If the non-rodent is the most appropriate species, then 1/6 the Highest Non-Severely Toxic Dose (HNSTD) is considered an appropriate starting dose. The HNSTD is defined as the highest dose level that does not produce evidence of lethality, life-threatening toxicities or irreversible findings.
At Step 2 of the ICH Process, a consensus draft text or guideline, agreed by the appropriate ICH Expert Working Group, is transmitted by the ICH Steering Committee to the regulatory authorities of the three ICH regions (the European Union, Japan and the USA) for internal and external consultation, according to national or regional procedures.
S10
Document History
Current Step 2 version

<table>
<thead>
<tr>
<th>Code</th>
<th>History</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>S10</td>
<td>Approval by the Steering Committee under Step 2 and release for public consultation.</td>
<td>15 November 2012</td>
</tr>
</tbody>
</table>

**Legal Notice:** This document is protected by copyright and may be used, reproduced, incorporated into other works, adapted, modified, translated or distributed under a public license provided that ICH’s copyright in the document is acknowledged at all times. In case of any adaption, modification or translation of the document, reasonable steps must be taken to clearly label, demarcate or otherwise identify that changes were made to or based on the original document. Any impression that the adaption, modification or translation of the original document is endorsed or sponsored by the ICH must be avoided.

The document is provided "as is" without warranty of any kind. In no event shall the ICH or the authors of the original document be liable for any claim, damages or other liability arising from the use of the document.

The above-mentioned permissions do not apply to content supplied by third parties. Therefore, for documents where the copyright vests in a third party, permission for reproduction must be obtained from this copyright holder.
PHOTOSAFETY EVALUATION OF PHARMACEUTICALS

Draft ICH Consensus Guideline

Released for Consultation on 15 November 2012, at Step 2 of the ICH Process

TABLE OF CONTENTS

1. INTRODUCTION ............................................................................................................ 1
  1.1. Objectives of the Guideline .................................................................................. 1
  1.2. Background .......................................................................................................... 1
  1.3. Scope of the Guideline ....................................................................................... 1
  1.4. General Principles ............................................................................................. 1

2. FACTORS TO CONSIDER IN THE PHOTOSAFETY EVALUATION ......................... 2
  2.1. Photochemical Properties ................................................................................... 2
  2.2. Tissue Distribution/Pharmacokinetics .................................................................. 2
  2.3. Metabolite Considerations .................................................................................. 3
  2.4. Pharmacological Properties ............................................................................... 3

3. NONCLINICAL PHOTOSAFETY TESTING ..................................................................... 3
  3.1. General Considerations ...................................................................................... 3
  3.2. Photoreactivity Testing Using Chemical Assays .................................................. 4
  3.3. Phototoxicity Testing Using in vitro Assays ......................................................... 4
  3.4. Photosafety Testing Using in vivo Assays and Systemic Administration ............... 5
  3.5. Photosafety Testing Using in vivo Assays and Dermal Administration ................. 6
  3.6. Photosafety Testing Using in vivo Assays and Ocular Administration ............... 7

4. CLINICAL PHOTOSAFETY ASSESSMENT ..................................................................... 7

5. ASSESSMENT STRATEGIES ....................................................................................... 7
  5.1. Recommendations for Testing of Pharmaceuticals Given via Systemic Routes ..... 7
    5.1.1 Assessment of Phototoxic Potential ............................................................... 7
    5.1.2 Experimental Evaluation of Phototoxicity ....................................................... 8
  5.2. Recommendations for Testing of Pharmaceuticals Given via Dermal Routes....... 8
    5.2.1 Assessment of Phototoxic Potential ............................................................... 8
    5.2.2 Experimental evaluation of Phototoxicity and Photoallergy ......................... 8
  5.3. Recommendations for Testing of Pharmaceuticals Given via Ocular Routes....... 9

6. ENDNOTES ............................................................................................................... 10

7. GLOSSARY ............................................................................................................... 12

8. REFERENCES ............................................................................................................ 13
PHOTOSAFETY EVALUATION OF PHARMACEUTICALS

1. INTRODUCTION

1.1. Objectives of the Guideline

The purpose of this document is to recommend international standards for photosafety assessment, and to harmonise such assessments supporting human clinical trials and marketing authorization for pharmaceuticals. It includes criteria for initiation of and triggers for additional photosafety testing and should be read in conjunction with ICH M3(R2), Section 14 on Photosafety Testing. (1) This guideline for photosafety assessment should reduce the likelihood that substantial differences in testing requirements and data interpretation will exist among regions.

Consideration should be given to the use of in vitro alternative methods or clinical data for photosafety assessment which could reduce the use of animals in accordance with the 3R (Replacement/Reduction/Refinement) principles.

1.2. Background

The ICH M3(R2) Guideline provides certain information regarding timing of photosafety assessment relative to clinical development. It recommends that an initial assessment of phototoxic potential be conducted, and if appropriate, an experimental evaluation be undertaken before exposure of large numbers of subjects (Phase III). Similarly, ICH S9 describes the timing of photosafety testing for oncology products. However, neither ICH M3(R2) nor ICH S9 provide specific information regarding testing strategies. This ICH S10 Guideline outlines further details on when photosafety testing is warranted, and on possible assessment strategies.

1.3. Scope of the Guideline

This guideline generally applies to new active pharmaceutical ingredients (APIs) and new excipients for systemic administration, clinical formulations for topical application, dermal patches, ocular products, and photodynamic therapy products.

Photodynamic therapy drugs are developed with photochemical reactivity as an inherent aspect of their intended pharmacology and additional assessment of their phototoxicity is not usually warranted. However, an evaluation of the toxicokinetics and tissue distribution of photodynamic therapy drugs is warranted to enable appropriate risk management in patients.

This guideline does not generally apply to peptides, proteins, antibody drug conjugates, or oligonucleotides. Further, this guideline does not apply to marketed products unless there is a new cause for concern.

1.4. General Principles

The photosafety assessment of a pharmaceutical is an integrated process that can involve an evaluation of photochemical characteristics, data from nonclinical studies and human safety information. This information is used to determine adequate risk minimization measures to prevent adverse events in humans.

Four different effects have been discussed in connection with photosafety testing: phototoxicity, photoallergy, photogenotoxicity and photocarcinogenicity. Testing for photogenotoxicity (Note 1) and photocarcinogenicity (Note 6 of ICH M3 (R2)) is not
currently considered useful for human pharmaceuticals. This guideline addresses only phototoxicity and photoallergy effects as defined below:

- **Phototoxicity** (photoirritation): An acute light-induced tissue response to a photoreactive chemical.

- **Photoallergy**: An immunologically mediated reaction to a chemical, initiated by the formation of photoproducts (e.g., protein adducts) following a photochemical reaction.

Photosensitization is a general term occasionally used to describe all light-induced tissue reactions. However, in order to clearly distinguish between photoallergy and phototoxicity, this term is not used in this guideline.

For a chemical to demonstrate phototoxicity and/or photoallergy, the following characteristics are critical:

- Absorbs light within the range of natural sunlight (290-700 nm);
- Generates a reactive species following absorption of UV/visible light; and
- Distributes sufficiently to light-exposed tissues (e.g., skin, eye).

If one or more of these conditions is not met, a compound will not present a photosafety concern.

2. **Factors to Consider in the Photosafety Evaluation**

2.1. **Photochemical Properties**

The initial consideration for assessment of photoreactive potential is whether a compound absorbs wavelengths between 290 and 700 nm. Absorption with a Molar Extinction Coefficient (MEC) less than 1000 L mol\(^{-1}\) cm\(^{-1}\) (2) is not considered to result in a photosafety concern (see Note 2 for further details).

Excitation of molecules by light can lead to generation of Reactive Oxygen Species (ROS), including superoxide and singlet oxygen \textit{via} energy transfer mechanisms.

Although other mechanisms for phototoxicity are known (e.g., formation of photoadducts or cytotoxic photoproducts), even in these cases, it appears that ROS are typically generated as well. Thus, ROS generation following irradiation with UV or visible light can be an indicator of phototoxic potential.

Photostability testing (see ICH Q1B [3]) can also suggest the potential for photoreactivity. However, not all photoreactive compounds are detected under these conditions, and photodegradation \textit{per se} does not imply that a drug will be phototoxic. Therefore, photostability testing alone should not be used to determine whether further photosafety evaluation is warranted.

Assessments of photochemical properties should be conducted under high-quality scientific standards with data collection records readily available, or in compliance with Good Laboratory Practices/Good Manufacturing Practices (GLP/GMP) regulations.

2.2. **Tissue Distribution/Pharmacokinetics**

The concentration of a photoreactive chemical in tissue at the time of light exposure is a very important pharmacokinetic parameter in determining whether a phototoxic reaction will occur. This concentration depends on a variety of factors, such as plasma concentration, perfusion of the tissue, partitioning from vascular to interstitial and cellular compartments, and binding, retention, and accumulation, of the chemical in the tissue.
Binding, retention, or accumulation of a compound in a tissue is not critical for a phototoxic reaction. If a molecule is sufficiently photoreactive, it might produce a phototoxic reaction at the concentration achieved in plasma or interstitial fluid. However, compounds having longer residence times in sun-exposed tissues or with higher tissue to plasma concentration ratios are more likely to produce a phototoxic tissue reaction than compounds with shorter residence times or lower tissue to plasma ratios. Further, the longer the concentration of a compound is maintained at a level above that critical for a photochemical reaction, the longer a person is at risk for phototoxicity.

Compound binding to melanin is one mechanism by which tissue retention and/or accumulation can occur. Although melanin binding can increase tissue levels, experience with melanin binding drugs suggests such binding alone does not present a photosafety concern.

A single-dose tissue distribution study, with animals assessed at multiple timepoints after dosing, will generally provide an adequate assessment of tissue drug levels and the potential for accumulation.

Although a tissue concentration threshold below which the risk for phototoxic reactions would be negligible is scientifically plausible, there are currently no data to delineate such a generic threshold for all compounds. Nevertheless, on a case-by-case basis it may be possible to justify that further photosafety assessment is not warranted based upon actual or anticipated tissue drug levels, and taking into consideration the factors discussed above. One example could be a low-dose inhaled drug for which overall systemic exposure levels are very low.

For those compounds with potent in vivo phototoxicity (or known to be phototoxic based on their mechanism of action such as photodynamic therapy drugs), distribution to internal as well as external tissues and estimates of tissue-specific half-lives should be assessed. Compounds activated by visible light and exhibiting long elimination half-lives in internal tissues have been demonstrated to cause injury to tissues exposed to intense light during medical procedures. Drugs that only absorb UV light or have short tissue elimination half-lives are not likely to present a risk to internal tissues even if they are known to be photoreactive.

2.3. Metabolite Considerations

Metabolites generally do not warrant separate photosafety evaluations as metabolism does not typically create new chromophores.

2.4. Pharmacological Properties

In most cases, drug-induced phototoxicity is due to the chemical structure and not to the pharmacology. However, certain pharmacologic properties can enhance susceptibility to light-induced effects, including reactions ranging from skin irritation to carcinogenesis (e.g., immunosuppression, perturbation of heme synthesis). The testing strategies outlined in this document are not designed to detect these types of indirect phototoxicity. Many of these mechanisms can be identified and evaluated in nonclinical pharmacology/toxicity testing (see ICH M3(R2)).

3. Nonclinical Photosafety Testing

3.1. General Considerations

Carefully selected conditions that consider both the model system and exposure to a relevant radiation spectrum are critical for nonclinical photosafety testing. Ideally, a nonclinical assay should exhibit both high sensitivity and specificity (i.e., low false
negative and low false positive rates). However, to support the integrated assessment strategy described in this document, it is most important that nonclinical photosafety assays show high sensitivity (i.e., produce a low frequency of false negatives). This is because negative assay results usually do not warrant further photosafety evaluation. It is not essential that positive assay results always predict a clinically relevant phototoxic response. The available nonclinical assays, both in vitro and in vivo, are focused primarily on detecting potential phototoxicity, which might or might not translate into clinically relevant phototoxicity. Therefore, the false positive rate for an assay should still be considered when deciding whether or not to use an assay.

Selection of irradiation conditions is critical for both in vitro and in vivo assays. Natural sunlight represents the broadest range of light exposure that humans might be exposed to regularly. However, sunlight per se is not well defined and depends on many factors (such as latitude, altitude, season, time of day, weather). In addition, sensitivity of human skin to natural sunlight depends on a number of individual factors (e.g., skin type, anatomical site and tanning status). Standardized sunlight exposure conditions have been defined by various organizations. Such standards (e.g., CIE-85-1989 [4]) should be considered in order to assess suitability of a sunlight simulator light source, and irradiance and irradiation dose should be normalized based on the UVA part (320 to 400 nm) of the applied spectrum. UVA doses ranging from 5 to 20 J/cm² have successfully been used to establish in vitro and in vivo phototoxicity assays. These UVA doses are comparable to those obtained during longer outdoor activities on summer days at noon time, in temperate zones, and at sea level. In humans, total sunlight exposure is normally limited by sunburn reactions caused by the UVB part of sunlight. In nonclinical phototoxicity assays, however, the amount of UVB should not limit the overall irradiation and might be attenuated (partially filtered) so that relevant UVA doses can be tested without reducing assay sensitivity. Penetration of UVB light into human skin is mainly limited to the epidermis, while UVA can reach capillary blood. Therefore, clinical relevance of photochemical activation by UVB is considered less important than UVA for systemic drugs. However, UVB irradiation is relevant for topical formulations.

3.2. Photoreactivity Testing Using Chemical Assays

If a drug developer chooses to assess photoreactivity, the assay should be qualified using pharmaceutical agents under appropriate conditions to demonstrate assay sensitivity. One such assay that is subject of a validation exercise is a ROS assay (e.g., [5]). Preliminary data suggest that this assay has high sensitivity for predicting in vivo phototoxicants. However, it has a low specificity, generating a high percentage of false positive results. A negative result in this assay, conducted under the appropriate conditions for the particular assay, would indicate a very low probability of phototoxicity, whereas a positive result would only be a flag for follow-up assessment.

3.3. Phototoxicity Testing Using in vitro Assays

A number of in vitro models have been developed for assessing the phototoxic potential of chemicals. Some of these models have not been qualified for use with pharmaceuticals. Some models involve testing compounds that are dissolved in the culture medium, and such methods are often appropriate for the active ingredient or excipients in systemic drug products, depending on the solubility. Other models involve direct application to the surface of a tissue preparation and can be appropriate for entire topical formulations.

The most widely used in vitro assay for phototoxicity is the “in vitro 3T3 Neutral Red Uptake Phototoxicity Test” (3T3 NRU-PT) for which a guideline (6) is available. This is currently considered the most appropriate in vitro screen for soluble compounds that are not exclusively UVB absorbers.
Although the formal European Centre for the Validation of Alternative Methods (ECVAM) validation exercise conducted on this assay indicated a sensitivity of 93% and a specificity of 84%, experience within the pharmaceutical industry suggests a much lower specificity (Note 3). The original Organisation for Economic Co-operation and Development (OECD) protocol was not validated for pharmaceuticals specifically. Thus, some modifications to the original OECD protocol have been proposed to address the low specificity observed with drug substances (see 3T3 Workshop Report [7], and Note 4). The sensitivity of the 3T3 NRU-PT remains unquestioned, and if a compound is negative in this assay it would have a very low probability of being phototoxic in humans. However, a positive result in the 3T3 NRU-PT should not be regarded as indicative of a likely clinical phototoxic risk, but rather a flag for follow-up assessment.

The BALB/c 3T3 cell line is sensitive to UVB and the recommended irradiation conditions involve the use of filters to attenuate wavelengths below 320 nm. UVB attenuation should not present a problem for systemic pharmaceuticals since these wavelengths minimally penetrate beyond the epidermis and hence UVB absorbers in systemic circulation are unlikely to be photoactivated. However, this is not true for topical products that absorb in the UVB range or for systemically administered compounds that distribute to the epidermis. For topical products that absorb predominately in the UVB range, and where in vitro assessment is desired, alternative models (e.g., reconstructed human skin models) which better tolerate UVB might be used.

Reconstructed human skin models, with the presence of a stratum corneum, permit testing of various types of topically applied materials ranging from neat chemicals to final clinical formulations. The models developed to date measure cell viability in the tissue preparation with and without irradiation. While such models appear to be capable of detecting known human dermal phototoxicants, the sensitivity of some models with respect to the dose eliciting a positive response can be lower than in the in vivo human situation. Consequently, it is important to understand the sensitivity of any model selected and, if appropriate, to adjust the assay conditions accordingly (e.g., testing higher strength formulations, increasing exposure time).

There are no in vitro models that specifically assess ocular phototoxicity. While negative results in the 3T3 NRU-PT or a reconstructed skin model might suggest a low risk, in the absence of data, the predictive value of these assays for ocular phototoxicity is unknown.

### 3.4. Photosafety Testing Using in vivo Assays and Systemic Administration

To date, no nonclinical in vivo phototoxicity or photoallergy assay has been formally validated. Phototoxicity testing for systemically administered compounds has been conducted in a variety of species, including guinea pig, mouse, and rat. No standardized study design has been established and thus the following criteria might be considered as best practices, if a decision is made by the drug developer to conduct in vivo studies in animals.

For species selection, irradiation sensitivity (i.e., minimal erythema dose), heat tolerance, and performance of reference substances should be considered. Models with both pigmented and non-pigmented animals are available. Although non-pigmented skin tends to be more sensitive than pigmented skin for detecting phototoxicity, the influence of melanin-binding (see Section 2.2) should be considered when selecting a species/strain to ensure appropriate exposures in target tissues.

Although phototoxicity is typically an acute reaction, the duration of an in vivo assay should be carefully considered. Accumulation of compound in relevant light-exposed tissues might lead to an increased sensitivity after repeated administration. Similarly, repeated irradiation after each dose might also lead to an increased sensitivity due to the
accumulation of damage. Generally, studies of a few days’ duration of dosing are appropriate, but pharmacokinetic properties as well as the intended clinical treatment regimen should be taken into consideration. Whenever feasible, the clinical route of administration should be used. Single or repeated daily irradiations after dosing (around T_{max}) can be used.

Dose selection for in vivo nonclinical phototoxicity testing of systemic drugs, if conducted, should support a meaningful human risk assessment. For such studies a maximum dose level that complies with the recommendations for general toxicity studies in ICH M3(R2) Section 1.5 is considered appropriate. If a negative result is obtained at the maximum dose, testing of lower doses is usually not warranted. However, if a positive result is anticipated, additional dose groups can support a NOAEL-based risk assessment. A vehicle group as well as non-irradiated controls can support adequate analyses and can distinguish between irradiation-induced and non-irradiation-induced adverse reactions. If the maximum systemic exposure achieved in animals is lower than clinical exposure, the reliability of a negative result in predicting human risk is questionable.

If an in vivo phototoxicity study is conducted, it is desirable to know the pharmacokinetic profile of the compound before designing the study, to ensure that irradiation of the animals is conducted at the approximate T_{max}. Relevant systemic exposure data (e.g., C_{max}), if not already available, should be collected as part of the in vivo phototoxicity study.

The most sensitive early signs of compound-induced phototoxicity are usually erythema followed by edema at a normally sub-erythemogenic irradiation dose. The type of response might vary with the compound. Any identified phototoxicity reaction should be evaluated regarding dose and time dependency and, if possible, the NOAEL should be established. The hazard assessment might be further supported by additional endpoints (e.g., early inflammatory markers in skin or lymph node reactions indicative of acute irritation).

In some cases, phototoxicity in the retina should be assessed (usually only warranted for substances absorbing light above 400 nm considering the optical properties of the human eye [8]). However, wavelength-dependent penetration of light through the eye of typical animal species might vary significantly (related to species, age, and gender) and occurs in some cases even in the UVA range. In such cases it is possible that findings observed in the animal model might not be relevant to humans. If warranted, phototoxicity of the retina should be assessed in established animal models using a careful histopathological analysis. No preference is made whether to restrain the animals during irradiation or whether to enforce open eyelids.

Adequate performance of in vivo phototoxicity models, which are not formally validated, should be demonstrated using suitable reference compounds. Compounds that are phototoxic in humans and that represent different chemical classes and mechanisms of phototoxicity should be evaluated to establish adequacy. For retinal toxicity, a reference compound with a light absorption profile within the visible light range (i.e., above 400 nm) is recommended. The concurrent use of a positive control compound might not be warranted if an in vivo model has been formally validated or has reached general acceptance and is established in the testing facility.

Testing for photoallergy is not recommended for compounds that are administered systemically.

3.5. Photosafety Testing Using in vivo Assays and Dermal Administration

The main recommendations provided for investigating the systemic route of administration also apply to dermal administration, including those for species selection,
study duration, and irradiation conditions. For dermal drug products in general, the clinical formulation should be tested. The intended clinical conditions of administration (e.g., occluded, non-occluded, intradermal) should be used to the extent possible. Irradiation of the exposed area should take place at a specified time after application, and the interval between application and irradiation should be justified based on the specific properties of the formulation to be tested. Signs of phototoxicity should be assessed based on relevant endpoints. The sensitivity of the assay should be demonstrated using appropriate reference compounds. Assessment of systemic drug levels is generally not warranted in dermal phototoxicity studies.

For dermal drug products, acute phototoxicity (photoirritation) and contact photoallergy have often been investigated in conjunction with nonclinical skin sensitization testing. However, no formal validation of such models has been performed and their predictivity for human photoallergy is unknown. For regulatory purposes, such nonclinical photoallergy testing is generally not recommended.

3.6. Photosafety Testing Using in vivo Assays and Ocular Administration

Currently, there are no standardised nonclinical in vivo approaches for assessing phototoxicity following ocular administration.

4. CLINICAL PHOTOSAFETY ASSESSMENT

There are various options for collecting human data, if warranted, ranging from standard reporting of adverse events in clinical studies to a dedicated clinical photosafety study. The precise strategy is determined on a case-by-case basis.

5. ASSESSMENT STRATEGIES

The choice of the photosafety assessment strategy is up to the drug developer. For a compound that has characteristics consistent with photoreactivity, nonclinical in vitro and in vivo tests and clinical alternatives are available for photosafety testing. If any one of the tests, having been conducted in an appropriate way, is negative, a compound is unlikely to elicit phototoxicity and further phototoxicity testing is generally not recommended.

ICH M3(R2) suggests a stepwise approach to photosafety assessment. An initial assessment of phototoxic potential based on photochemical properties and pharmacological/chemical class should be undertaken before outpatient studies. In addition, the distribution to skin and eye can be evaluated to inform further on the human risk and the need for further testing. Then, if appropriate, an experimental evaluation of phototoxic potential (nonclinical, in vitro or in vivo, or clinical) should be undertaken before exposure of large numbers of subjects (Phase III).

5.1. Recommendations for Testing of Pharmaceuticals Given via Systemic Routes

5.1.1 Assessment of Phototoxic Potential

If the substance has a MEC less than 1000 L mol\(^{-1}\) cm\(^{-1}\) (between 290 and 700 nm), no further photosafety testing is recommended and no phototoxicity is anticipated in humans. Any available data on the phototoxicity of class-related compounds should also be assessed, as this could inform on the decision taken for further assessment. If the drug developer chooses to conduct a test for photoreactivity (see Section 3.2) the resulting data can support a decision that no further photosafety assessment is warranted. Similarly, if a drug developer chooses to assess drug distribution to light-exposed tissues
(see Section 2.2), the resulting data can support a decision that no further photosafety assessment is warranted (Note 5). Otherwise, nonclinical and/or clinical photosafety assessment of the substance should be conducted.

5.1.2 Experimental Evaluation of Phototoxicity

If the drug developer chooses an \textit{in vitro} approach, the 3T3 NRU-PT is currently the most widely used assay and in most cases could be considered as an initial test for phototoxicity. In the EU, a validated \textit{in vitro} alternative method should generally be used before considering animal testing. The high sensitivity of the 3T3 NRU-PT results in good negative predictivity, and negative results are generally accepted as sufficient evidence that a substance is not phototoxic. In such cases no further testing is recommended and no phototoxicity is anticipated in humans.

In some situations (e.g., poorly soluble compounds) an initial assessment of phototoxicity in an \textit{in vitro} assay might not be appropriate. In this case, an assessment in animals or in humans could be considered.

If an \textit{in vitro} phototoxicity assay gives a positive result, a phototoxicity study in animals could be conducted to assess whether the potential phototoxicity identified \textit{in vitro} correlates with an \textit{in vivo} response. Alternatively, the photosafety risk could be addressed/managed in the clinical setting. This could include a recommendation for protective measures in clinical trials in lieu of photosafety testing, or until the risk has been assessed (see ICH M3(R2)). A negative result in an appropriately conducted \textit{in vivo} phototoxicity study (either in animals or humans) supersedes a positive \textit{in vitro} result. In such cases no further testing is recommended and no phototoxicity is anticipated in humans. In addition, a robust clinical phototoxicity assessment indicating no concern supersedes any positive nonclinical results.

In cases where an \textit{in vivo} animal phototoxicity study or clinical phototoxicity study had already been conducted, there is no reason to subsequently conduct an \textit{in vitro} phototoxicity assay.

5.2. Recommendations for Testing of Pharmaceuticals Given via Dermal Routes

5.2.1 Assessment of Phototoxic Potential

If the active substance and excipients have MEC values less than 1000 L mol\(^{-1}\) cm\(^{-1}\) (between 290 and 700 nm), no further photosafety testing is recommended and no phototoxicity is anticipated in humans. Any available data on the phototoxicity of chemical class-related compounds should also be assessed as this could inform on the approach taken for further assessment. For compounds with MEC values of 1000 L mol\(^{-1}\) cm\(^{-1}\) or higher, in the EU and Japan, negative photoreactivity test results (e.g., a ROS assay) can support a decision that no further photosafety assessment is warranted. In the US, negative test results in photoreactivity assays do not generally preclude further clinical photosafety assessment using the to-be-marketed formulation.

Tissue distribution is not a consideration for dermal products. Dermal products are administered directly to the skin and hence, unless they are applied to areas not exposed to light, are assumed to be present in light-exposed tissues.

5.2.2 Experimental evaluation of Phototoxicity and Photoallergy

The \textit{in vitro} 3T3 NRU-PT can be used to assess individually the phototoxicity potential of the API and any new excipient(s), provided that appropriate testing conditions can be achieved (e.g., test concentrations not limited by poor solubility, relevant UVB dose can
be applied). In cases where no phototoxic component has been identified in vitro, the overall phototoxicity potential of the clinical formulation can be regarded as low.

Some properties of the clinical formulation which could influence the potential phototoxic response (e.g., penetration into skin, intracellular uptake) cannot be evaluated using the 3T3 NRU-PT alone. Therefore, confirmation of the overall negative result in an evaluation using the clinical formulation and/or monitoring during clinical trials can still be warranted.

Reconstituted 3D skin models can be used to assess the phototoxicity potential of clinical formulations. It is important to understand the sensitivity of the particular 3D skin model selected and, if appropriate, adjust the assay conditions accordingly (e.g., testing higher strength formulations, increasing exposure time). However, under adequate test conditions, a negative result in a 3D skin model indicates that the phototoxicity potential of the formulation can be regarded as low. In this case, in the EU and Japan generally no further phototoxicity testing is recommended. In the US, negative test results do not generally preclude further clinical photosafety assessment using the to-be-marketed formulation.

If an appropriate in vitro model is not available, the initial test could be an in vivo animal phototoxicity test on the clinical formulation. Alternatively, the phototoxic potential in humans can be assessed prior to exposure of large numbers of subjects (ICH M3(R2)). In the EU and Japan, a negative result in an appropriately conducted in vivo animal phototoxicity study would be sufficient evidence that the formulation is not phototoxic and no further phototoxicity testing is recommended. In the US, negative test results do not generally preclude further clinical photosafety assessment using the to-be-marketed formulation.

For dermal products where the API or any new excipient has a MEC value of 1000 L mol⁻¹ cm⁻¹ or higher, a photoallergy assessment is generally warranted in addition to phototoxicity testing. A clinical photoallergy assessment is generally recommended using the to-be-marketed formulation, and a study can be conducted during Phase III, if warranted.

5.3. Recommendations for Testing of Pharmaceuticals Given via Ocular Routes

For compounds that have a MEC value less than 1000 L mol⁻¹ cm⁻¹ (between 290 and 700 nm) no phototoxicity is anticipated in humans. Compounds that only absorb light at wavelengths below 400 nm and are to be administered as intraocular injections behind the lens (e.g., in the vitreous) are of low concern, as only light of wavelengths greater than 400 nm reaches the back of the adult eye. However, the lens in children is not completely protective against wavelengths below 400 nm.

For compounds that absorb at relevant wavelengths and are given via ocular routes (e.g., ocular eye drops, intraocular injections), an assessment of photosafety is generally recommended. The reliability of in vitro approaches in predicting phototoxicity following ocular administration is unknown and there are no standardised in vivo approaches for assessing phototoxicity for products administered via the ocular route. Nevertheless, the basic principles of phototoxicity assessment still apply and any available data on the phototoxicity of the compound in question or of chemical class-related compounds should be considered in the overall assessment. In the US and Japan there are no specific recommendations to experimentally assess the phototoxic potential of ocular products. In the EU, an experimental assessment would be recommended using in vitro approaches or in vivo studies using other routes of administration when the available data are considered insufficient for hazard identification.
6. ENDNOTES

Note 1 Testing of photogenotoxicity is not recommended as a part of the standard photosafety testing programme. In the past, some regional guidances (e.g., CPMP/SWP/398/01) have recommended that photogenotoxicity testing should be conducted, preferentially using a photoclastogenicity assay (chromosomal aberration or micronucleus test) in mammalian cells in vitro. However, experience with these models since the CPMP/SWP guidance was issued has indicated that these tests are substantially oversensitive and even incidences of pseudo-photoclastogenicity have been reported. (9) Furthermore, the interpretation of photogenotoxicity data regarding its meaning for clinically relevant enhancement of UV-mediated skin cancer is unclear in most cases. In most cases, the mechanism by which compounds induce photogenotoxic effects is identical to those that produce phototoxicity, and thus separate testing of both endpoints is not warranted.

Note 2 Standardized conditions for determination of MEC are critical. Selection of an adequate solvent is driven by both analytical requirements (e.g., dissolving power, UV/visible transparency) and physiological relevance (e.g., pH 7.4-buffered aqueous conditions). Methanol has been selected as a preferred solvent and was used to support the MEC threshold of 1000 L mol⁻¹ cm⁻¹ (data to be published). For most compounds, useful UV/visible spectra can be obtained, at concentrations around 100 µM. Nevertheless, potential limitations (e.g., artifacts due to high concentrations or slow precipitation) should be considered. If the chromophore of the molecule appears to be pH-sensitive (e.g., phenolic structure, aromatic amines, carboxylic acids, etc.) an additional spectrum obtained under aqueous, pH 7.4-buffered conditions, could add valuable information regarding differences in the shape of the absorption spectrum and in the MEC. If significant differences are seen between measurements obtained in methanol versus pH-adjusted conditions, the MEC threshold of 1000 L mol⁻¹ cm⁻¹ cannot be used to support a definitive assessment.

Note 3 A survey of EFPIA member companies indicated that the 3T3 NRU-PT, as described in the OECD guideline, generates a high percentage of positive results (approximately 50%), the majority of which do not correlate with phototoxicity responses in animals or humans. (10)

Note 4 Following a retrospective review of data for pharmaceuticals, a reduction of the maximum test concentration from 1000 to 100 µg/mL appears justified. Compounds without any significant cytotoxicity (under irradiation) up to this limit can be considered as being devoid of relevant phototoxicity. In addition, the category named “probable phototoxicity” per OECD (i.e., Photo Irritation Factor (PIF) values between 2 and 5 or Mean Photo Effect (MPE) values between 0.10 and 0.15) is of questionable toxicological relevance for systemic drugs. Compounds falling into this category generally do not warrant further photosafety evaluations. For compounds that give a PIF value between 2 and 5, and for which it is not possible to determine an IC₅₀ in the absence of irradiation, it is important to check that the compound is not classified as positive using the MPE calculation, i.e., that the MPE is less than 0.15.

Systemic drugs that are positive in the 3T3 NRU-PT only at in vitro concentrations that are many times higher than drug concentrations likely to be achieved in light-exposed tissues in humans, can, on a case-by-case basis, and in
consultation with regulatory authorities, be considered to be ‘low risk’ for human phototoxicity, without follow-up in vivo testing.

**Note 5** If a systemically administered drug does not have higher tissue to plasma concentration ratios or does not accumulate in skin, in the US further assessment of the phototoxicity potential is generally not warranted. In the EU and Japan higher tissue to plasma concentration ratios and/or tissue accumulation are also considered to be important. However, the presence of compound in skin is considered to be the critical factor in determining whether further testing is warranted. If a drug developer believes there is a rationale for not testing based on very low tissue levels, this can be discussed with the regulatory authority on a case-by-case basis.
7. GLOSSARY

3T3 NRU-PT:
*In vitro* 3T3 Neutral Red Uptake Phototoxicity Test.

Assessment:
In the context of this document, an assessment is an evaluation of all available information and does not always mean an additional test is conducted.

Chromophore:
The substructure of a molecule that absorbs visible or UV light.

Irradiance:
The intensity of UV or visible light incident on a surface, measured in W/m² or mW/cm².

Irradiation:
The process by which an object/subject is exposed to UV or visible radiation.

MEC:
Molar Extinction Coefficient (also called molar absorptivity) is a constant for any given molecule under a specific set of conditions (e.g., solvent, temperature, wavelength) and reflects the efficiency with which a molecule can absorb a photon (typically expressed as L mol⁻¹ cm⁻¹).

MPE:
The Mean Photo Effect is calculated for results of the 3T3 NRU-PT when two equally effective concentrations (IC₅₀), both with and without irradiation, cannot be determined. The MPE is based on comparison of the complete concentration response curves (see OECD TG 432).

NOAEL:
No observed adverse effect level.

OECD TG:
Organisation for Economic Co-operation and Development Test Guideline.

Photoproducts:
New compounds/structures formed as a result of a photochemical reaction.

Photoreactivity:
The property of chemicals that react with another molecule as a consequence of absorption of photons.

PIF:
Photo Irritation Factor is calculated for results of the 3T3 NRU-PT by comparing the IC₅₀ with and without irradiation.

ROS:
Reactive Oxygen Species, including superoxide anion radicals and singlet oxygen.

UVA:
Ultraviolet light A (wavelengths between 320 and 400 nm).

UVB:
Ultraviolet light B (wavelengths between 290 and 320 nm).
8. References


