INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

DRAFT ICH HARMONISED GUIDELINE

DETECTION OF TOXICITY TO REPRODUCTION FOR HUMAN PHARMACEUTICALS

S5(R3)

Current *Step 2* draft version
dated 5 July 2017

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 Assembly to the regulatory authorities of the ICH regions for internal and external consultation, according to national or regional procedures.

TABLE OF CONTENTS

1 SCOPE OF THE GUIDELINE

This guideline applies to pharmaceuticals, including biotechnology-derived pharmaceuticals, vaccines (and their novel constitutive ingredients) for infectious diseases, and novel excipients that are part of the final pharmaceutical product. It does not apply to cellular therapies, gene therapies and tissue-engineered products. The methodological principles (e.g., study design, dose selection and species selection) outlined in this guideline can also apply to pharmaceuticals intended for the treatment of serious and life threatening diseases, such as advanced malignancies (i.e., see ICH S9 (3)). This guideline should be read in conjunction with ICH M3(R2) (1), ICH S6(R1) (2) and ICH S9 (3) regarding whether and when non-clinical reproductive toxicity studies are warranted.

2 INTRODUCTION & GENERAL PRINCIPLES

The purpose of this guideline is to provide key considerations for developing a testing strategy to identify hazard and characterize reproductive risk for human pharmaceuticals. The guidance informs on the use of existing data and identifies potential study designs to supplement available data to identify, assess, and convey risk. General concepts and recommendations are provided that should be considered when interpreting study data and making an assessment of reproductive risk in support of clinical development and marketing approval.

To assess a human pharmaceutical's effects on reproduction and development, the information should generally include exposure of adult animals and the impact on all stages of development from conception to sexual maturity. No guideline can provide sufficient information to cover all possible cases, and flexibility in testing strategy is warranted. Regardless of the pharmaceutical modality (see Glossary), key factors to consider when developing an overall integrated testing strategy include:

- The anticipated pharmaceutical use in the target population (especially in relation to reproductive potential and severity of disease);
- The formulation of the pharmaceutical and route(s) of administration intended for humans;
- The use of any existing data on toxicity, pharmacodynamics, pharmacokinetics, and similarity to other compounds in structure or activity;
- Selection of specific studies, test species/test system and dose levels.

These concepts are discussed in more detail throughout the guideline, which defines a thoughtful approach for developing a testing strategy. This guideline recommends the use of information about the pharmaceutical and the patient population in order to perform only those studies essential to evaluate the stages (see below) for which there is insufficient knowledge to inform about the risk to reproduction and development.

As appropriate, observations through one complete life cycle (i.e., from conception in one generation through conception in the following generation) permit detection of immediate and latent adverse effects. For the purposes of this guidance, gestation day 0 (GD 0; see Glossary) is when positive evidence of mating is detected. The following stages of reproduction are generally assessed:

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

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, pre-weaning development and growth).

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

The stages covered in individual studies are left to the discretion of the Sponsor, although the timing of studies within the pharmaceutical development process is dependent on study populations and phase of pharmaceutical development (see ICH M3(R2) (1), ICH S6(R1) (2) and ICH S9 (3)).

This guideline also provides considerations for interpreting all available nonclinical information as part of the risk characterization.

3 STRATEGIES FOR REPRODUCTIVE TOXICITY ASSESSMENT

3.1 Considerations/Principles

The initial step is to determine if reproductive toxicity testing for each of the various reproductive stages is warranted and, if so, what are the most appropriate studies to conduct. The considerations should include: a) the target patient population and duration of dosing, b) the known pharmacology of the compound, c) the known toxicity of the compound, d) any existing knowledge of the impact of the target(s) on reproductive risk (e.g., human and/or animal genetics, or class effects), and e) data from *in vitro* and non-mammalian assays (alternative assays, see Glossary) that could be relied upon to identify hazard and/or risk (see Section 3.3.2). Approaches for qualifying and use of alternative assays in assessing reproductive risk are discussed below (Sections 3.3.2 and 9.5). Generally, most alternative assays being developed address endpoints related to Embryo-Fetal Development (EFD) and are thus discussed in section 3.3.2. However, as new assays are developed for other reproductive endpoints, they can be similarly deployed with appropriate qualification.

The experimental strategy to generate the data should consider minimizing the use of animals. Alternative assays and/or *in vivo* studies with fewer animals can be used to identify hazards in a tiered manner. Reductions in animal use can also be achieved by deferring definitive EFD studies (see Section 9.4.3.3) until later in pharmaceutical development (see below). Alternative assays can replace definitive assays in some circumstances where as in others they can be used to defer traditional assays until later in development (see Section 3.3). An important component of the overall strategy is the timing for the additional information to support ongoing clinical development (e.g., developmental toxicity (see Glossary) data to support dosing women of childbearing potential).

Reproductive and developmental studies should in general be conducted according to Good Laboratory Practice (GLP) as they will contribute to risk assessment. However, if a human developmental or reproductive risk is defined during the conduct of a relevant non-GLP study, repetition of the study to confirm the finding(s) under GLP conditions is not warranted. Preliminary EmbryoFetal Development (pEFD; see Glossary) studies should be conducted under high-quality scientific standards with data collection records readily available or under GLP conditions. It is recognized that GLP compliance is not expected for some study types, or aspects of some studies, employing specialized test systems or methods, such as disease models or surrogate molecules (see Glossary), or literature. However, high quality scientific standards should be applied, with data collection records readily available. Areas of non-compliance should be identified and their significance evaluated relative to the overall safety assessment.

3.1.1 ***Target Patient Population/ Therapeutic Indication Considerations***

The patient population or therapeutic indication can influence the extent of reproductive toxicity testing. For example:

- If the female patient population is post-menopausal there is no utility in evaluating any of the reproduction stages;
- A pharmaceutical for use in an elderly male does not warrant conduct of studies to evaluate stages E and F;
- If the disease indicates that reproductive toxicity will have minimal impact on the usage of the pharmaceutical in the target population, studies evaluating only stages C and D can be warranted;
- Short-term therapies under highly controlled settings.

3.1.2 ***Pharmacology Considerations***

Before testing, it should be determined if the pharmacologic effects are incompatible with fertility, normal EFD, or measurement of endpoints of the study being considered (e.g., a general anesthetic and measurement of mating behavior). This assessment could be based on data with other pharmaceuticals with similar pharmacology on the pathways affected, or

on knowledge of effects in humans with related genetic diseases. Based on these considerations, sometimes no testing for a particular reproductive endpoint can be warranted. In contrast, testing for only off-target effects can be warranted if the expected pharmacologic effects on reproductive endpoints are non-adverse. Examples include patients with a condition that mimics the target pharmacology who have normal reproductive capability and healthy offspring; or when other pharmaceuticals have similar pharmacology or pathways affected but have no demonstrated reproductive risk.

3.1.3 *Toxicity Considerations*

Repeat–dose toxicity studies with sexually mature animals can provide important information on toxicity to reproductive organs. The existing toxicology data for the compound should always be considered, taking into account the dose levels, toxicokinetic profile, and dosing duration. For example, the evaluation of fertility effects for a pharmaceutical that damages testicular tissue might warrant modifications to the standard fertility study, if such a study would be appropriate.

Sometimes, toxicity in animals precludes attaining a systemic exposure relevant to the human exposure under conditions of use and this should be addressed.

3.1.4 *Timing Considerations*

General guidance on the timing for conduct of reproductive toxicity studies covering Stages A-F relative to clinical studies is described in the ICH M3(R2) and ICH S9 guidelines (1,3). The timing for when to conduct specific reproductive toxicity assessments should take into consideration the points discussed above. Based on these factors, it can sometimes be appropriate to consider altering timing of the assessment of specific reproductive stages. For example, if there is an equivocal observation from a preliminary study and other compounds in the class are without risk, then consideration should be given to accelerating the definitive studies. In contrast, there can be circumstances for deferring studies. For example, when other studies have revealed a risk and appropriate precautions in clinical trials have been taken, the conduct of definitive studies evaluating the relevant reproductive stages can be deferred to later in development than is recommended in ICH M3(R2) (1). When conducting enhanced Pre- and PostNatal Development (ePPND) studies in NonHuman Primates (NHP) see ICH S6(R1) (2) for timing.

Additional options that include study deferral are discussed in Section 3.3.3.

3.1.5 *Other Considerations for Reproductive Toxicity Studies*

For some species and compounds, it can be more appropriate to test multiple reproductive stages in a single study (e.g., monoclonal antibodies in NHPs; see ICH S6(R1) (2)). Consideration can also be given to evaluation of reproductive toxicity endpoints as a component of another study type (e.g., male fertility as part of a repeat-dose toxicity study, see Section 3.2).

When designing a pre- and post-natal development (PPND) or ePPND study, thought should be given to the value for juvenile animal endpoints for supporting the safety of pediatric use (see Section 9.4.2.1).

Alternative assays are described as part of an integrated testing strategy for assessing embryo-fetal developmental endpoints as described in the examples below (see Section 3.3.2.1).

3.2 Strategy to Address Fertility and Early Embryonic Development

The aim of the fertility study is to test for disturbances resulting from treatment from before mating of males and/or females through mating and implantation. This comprises evaluation of Stages A and B of the reproductive process (see Sections 6 and 9.4).

Fertility studies are generally only performed in rodents or rabbits. Mating evaluations are not generally feasible in non-rodents such as dogs and NHPs. For example if NHPs are the only pharmacologically relevant species (as for many monoclonal antibodies, see ICH S6(R1) (2)), fertility evaluations can be based on the results of the repeat-dose toxicity studies (e.g., histopathological examinations).

Histopathology of the reproductive organs from the repeat-dose toxicity studies is a sensitive method of detecting the majority of effects on male and female fertility, provided animals are sexually mature.

Dogs and minipigs used in long-term repeat-dose studies should have, in general, sexually matured by the end of the study. If NHPs are to be used to assess effects on fertility, there should be a sufficient number of sexually mature animals at study termination.

If repeat-dose toxicity studies are used to assess effects on fertility, a comprehensive histopathological examination of the reproductive organs from both male and female animals should be performed (Note 1).

When there is cause for concern based on mode of action or data from previous studies, additional examinations can be included in repeat-dose toxicity studies, e.g., sperm collection, or monitoring of the estrous or menstrual cycle. Studies of two to four weeks treatment duration can be expected to provide an initial evaluation of effects on the reproductive organs. This information will later be supplemented with similar evaluations in the subchronic and chronic toxicity studies.

A dedicated fertility study includes a mating phase and serves to detect effects that cannot be assessed by histopathology of the reproductive organs. However, if the drug has clinically relevant adverse effects on male or female reproductive organs in the repeat-dose toxicity studies, a routine fertility study in the affected sex will be of limited value and not warranted. Likewise, a fertility study is not warranted for pharmaceuticals that will not be used in subjects of reproductive age. Generally, the repeated-dose toxicity study results can be used to design the fertility study without the need for further dose ranging studies.

If no adverse effects on fertility are anticipated, male and female rodents can be evaluated in the same fertility study. However, if effects on fertility are identified, the affected sex should then be determined. In addition, if it cannot be determined whether effects are reversible based on the pathophysiological evaluation, then reversibility of induced effects should be evaluated. These determinations can have an important impact on risk assessment.

3.3 Strategies to Address Embryo Fetal Development (EFD)

The aim of the EFD studies is to detect adverse effects on the pregnant female and development of the embryo and fetus consequent to exposure of the female during the period of major organogenesis (Stage C). EFD studies include full evaluation of fetal development and survival. For most non-highly targeted pharmaceuticals (e.g., small molecules), effects on EFD are typically evaluated in two species (i.e., rodent and non-rodent). There are cases where testing for effects on EFD in a single species can suffice. General strategies to address EFD studies are shown in Figure 3-1.

3.3.1 ***Routine Approach for Addressing EFD Risk***

In situations where the use of rodent or rabbit species is appropriate, at least one of the test species should exhibit the desired pharmacodynamic (PD) response (Section 4). If the pharmaceutical is not pharmacodynamically active in any routinely used species (Section 9.3), genetically modified (GM) animals or use of a surrogate molecule can be considered. If it is a highly-targeted pharmaceutical these data can be sufficient. If the pharmaceutical is non-highly targeted, it can be appropriate to also administer it to a rodent or a rabbit to test for off-target effects.

However, under some circumstances other approaches can be used to defer (Table 3-1) or replace (Section 9.5.5) definitive studies. Alternatively, there can be adequate information to communicate risk without conducting additional studies. Evidence suggesting an adverse effect of the intended pharmacological mechanism on EFD (e.g., mechanism of action, phenotypic data from genetically modified animals, class effects) can be sufficient to communicate risk.

Non-routine animal models or a surrogate molecule can be considered in place of NHPs for either small molecules or biotechnology-derived products, if appropriate scientific justification indicates that results will inform the assessment of reproductive risk (Section 4.3).

In certain justified cases, testing for effects on embryo-fetal development in a single species can suffice. One example is for highly targeted pharmaceuticals (e.g., for biotechnology-derived products, see ICH S6(R1)) when there is only one relevant species that can be used in reproductive testing (2). Another circumstance is for non-highly targeted pharmaceuticals when it can be shown that a single species is a relevant model for the human, based on pharmacodynamics, pharmacokinetics and metabolite profiles, as well as toxicology data. If the result is clearly positive (teratogenic and/or embryofetal lethal; TEFL; see Glossary) under relevant exposure, testing in a second species is not warranted.

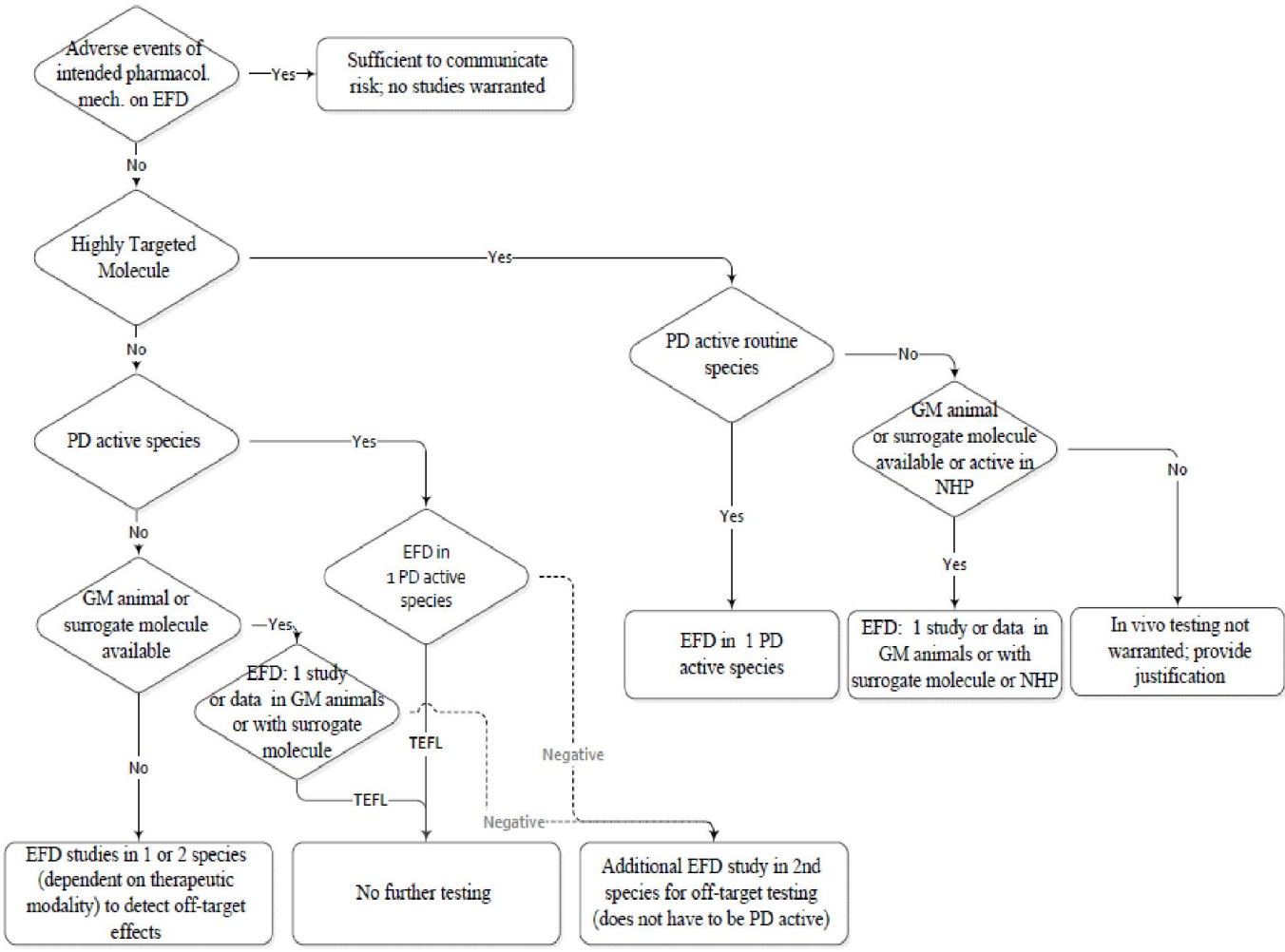
When there are no pharmacologically relevant species (e.g., the pharmacological target only exists in humans), EFD studies in two species can still be warranted to detect off-target effects or secondary pharmacology as appropriate based on the therapeutic modality and the indication.

For biotechnology-derived products, when no relevant species can be identified because the biopharmaceutical agent does not interact with the orthologous target in any species relevant to reproductive toxicity testing, use of surrogate molecules or transgenic models can be considered, as described in detail in ICH S6(R1) (2). If there are no relevant species, genetically modified animals, or surrogate, *in vivo* reproductive toxicity testing is not meaningful; however, the approach used should be justified.

For other therapeutic modalities that lack orthologous target engagement in useful reproductive toxicology species and also have anticipated off-target effects, use of surrogate molecules or transgenic models can be considered.

Several scenarios of use for integrated testing strategies are described in Annex 9.5.5.

Figure 3-1: General Strategy to Address EFD



3.3.2 *Optional Approaches for Addressing EFD Risk*

3.3.2.1 Use of Alternative Assays

Use of alternative *in vitro*, ex vivo, and non-mammalian *in vivo* assays (alternative assays) can reduce animal use while preserving the ability to detect relevant reproductive risks. The use of qualified (Note 2) alternative assays can be an appropriate approach in lieu of the routine approach discussed above. Use of qualified alternative assays is appropriate for risk assessment under certain circumstances where they are interpreted in conjunction with *in vivo* reproductive testing. Although they are not a replacement for all *in vivo* reproductive testing, they can reduce *in vivo* mammalian animal studies and/or animal usage (Section 3.3.2.1). Several scenarios of use for integrated testing strategies are described in Annex 9.5.5. Furthermore, while a study in a second species could be conducted under the routine approach, the use of an alternative assay could be more informative in some circumstances, taking into consideration route of administration, exposure, and mechanism of action.

The circumstances justifying the incorporation of alternative assays in an integrated testing strategy for assessing EFD risk will be dependent upon a number of factors. These could include the severity of the disease, the characteristics of the patient population, or the limitations of some traditional test systems for specific therapeutic targets. The pharmacological or biological plausibility for developmental toxicity is a key consideration.

This guideline does not recommend specific assays, but basic principles are included to assist in assay qualification for potential regulatory use (Section 9.5.2).

For appropriate use of alternative assays it is important to know the reliability and predictivity for *in vivo* reproductive outcomes. The Annex provides information on various reference compounds that can be used to assess alternative methods for embryo-fetal development/deaths (Note 3). It is possible that a suite of assays/assessments will show improved predictivity.

Where applicable, testing strategies can take into consideration data from qualified alternative assays in combination with one or more *in vivo* mammalian EFD studies. Any alternative assay integrated into a testing strategy should be qualified for its intended context of use (Section 9.5). When alternative assays are used to contribute to the risk assessment they should generally be conducted according to GLP, particularly when the assay results do not identify a hazard. Contexts of use (see Glossary) could include, but are not limited to:

a. Being part of an integrated testing strategy for assessing embryo-fetal developmental endpoints as described in the scenarios in Section 9.5.5;

b. Deferral of definitive studies as discussed in Section 3.3.3;

c. Complete replacement of one species when used in conjunction with an enhanced pEFD study in one species (see Scenarios in Section 9.5.5);

d. There is evidence (e.g., a mechanism of action affecting fundamental pathways in developmental biology, phenotypic data from genetically modified animals, class effects) suggesting an adverse effect on EFD, or contributing to the weight of evidence when animal data are equivocal;

e. Toxicity (on-target related and/or off-target) in a routine animal species precludes attaining a systemic exposure relevant to the human exposure under conditions of use, but higher exposures can be attained in an alternative assay;

f. Low systemic exposure (e.g., no embryo-fetal exposure) in humans such as following ophthalmic administration.

The information from the alternative qualified test systems should be used with all available *in vivo* nonclinical and human data as part of an integrated risk assessment approach (see Principles of Risk assessment; Section 7).

3.3.2.2 *In vitro* and Non-mammalian Exposure Information

As stated in section 7 of this guideline, for the purposes of risk assessment, it is important to consider exposure in the interpretation of non-clinical studies assessing reproductive toxicity. This also applies to assays conducted using *in vitro* or non-mammalian systems. The pharmacokinetic parameter used is dependent upon how the assay was qualified in relation to the *in vivo* concentrations at which the EFD observations were made, considering any normalization factors used in the assay qualification. For example, the maximum concentration tested without an adverse effect in the *in vitro* system can be compared to the C_{max} in humans for the determination of potential human risk, applying the normalization factor used in the assay qualification.

3.3.3 *Potential Approaches to Defer in vivo Testing as Part of an Integrated Testing Strategy*

Table 3-1 illustrates approaches to support inclusion of Women Of Child-Bearing Potential (WOCBP) in clinical studies while deferring conduct of definitive assays. This applies to circumstances where 2 definitive EFD studies are warranted for the pharmaceutical.

One such approach is the use of an enhanced pEFD study for one of the species. In this case, the pEFD study (see ICH M3(R2)) should be conducted in accordance with GLP regulations, the number of pregnant animals should be increased from 6 to ≥ 8 per group, and include fetal skeletal examinations.

Table 3-1. Approaches for Deferral of Definitive EFD Studies in 2 Species

<table>
<tr><th rowspan="2">Approach</th><th colspan="4">Stage of Development</th></tr>
<tr><th>Limited inclusion of WOCBP[a]</th><th>Unlimited inclusion of WOCBP up to start of Phase 3 (supports Phase 2a/b)[b]</th><th>Unlimited inclusion of WOCBP up to marketing (supports Phase 3)</th><th>To support marketing[c]</th></tr>
<tr><td>A</td><td colspan="2">1st species EFD (enhanced pEFD or definitive) + Qualified alternative assay</td><td>2nd species definitive EFD</td><td>1st species definitive EFD if not conducted earlier</td></tr>
<tr><td>B</td><td colspan="2">1st species pEFD + 2nd species EFD (enhanced pEFD or definitive)</td><td>1st species definitive EFD</td><td>2nd species definitive EFD if not conducted earlier</td></tr>
<tr><td>C[d]</td><td>2 species pEFD</td><td colspan="3">2 species definitive EFD</td></tr>
<tr><td colspan="5">[a] Up to 150 WOCBP receiving investigational treatment for a relatively short duration (up to 3 months).
[b] All approaches include “where precautions to prevent pregnancy in clinical trials (see above) are used.”
[c] For monoclonal antibodies, the ePPND is generally conducted before marketing approval (see ICH S6(R1)).
[d] See ICH M3(R2) for regional differences.</td></tr>
</table>

3.4 Strategy to Address Effects on PPND

The aim of the PPND study is to detect adverse effects following exposure of the mother from implantation through weaning on the pregnant or lactating female and development of the offspring. Since manifestations of effects induced during this period can be delayed, development of the offspring is monitored through sexual maturity (i.e., Stages C to F). The usual species used for PPND is the rat; however, other species can be used as appropriate with modifications of the endpoints assessed.

In most cases, a preliminary PPND study is optional because the appropriate information is generally available from prior studies to design the definitive study. However, a preliminary PPND study with termination of the pups before or at weaning can be used to select dose levels or inform study design and to provide pup exposure data.

For pharmaceuticals that can only be tested in the NHP, the ePPND study can provide a limited assessment of post-natal effects, but it is not feasible to follow the offspring through maturity. For the timing of the ePPND study see ICH S6(R1) (2).

3.5 Toxicokinetics (TK)

TK investigations are generally expected and the use of the data is discussed throughout this document. General concepts regarding TK data collection are discussed in ICH S3A.

Determination of the pharmaceutical's concentration in the fetus can be of interest to facilitate interpretation of discordant or equivocal evidence of developmental hazard. However, determination of placental transfer is generally not warranted because of limited ability to translate data to human fetal exposures.

Many pharmaceuticals are excreted in milk, although lactational excretion data in animals are of uncertain value for human risk assessment. Therefore, measurement of drug concentrations in the milk of animals is generally not warranted. However, determination of a pharmaceutical's concentrations in the offspring can support interpretation of findings observed during the pre-weaning period.

4 TEST SYSTEM SELECTION

4.1 Routine Test Species

When a study is warranted, a mammalian species should be used. For the primary species, it is generally desirable to use the same species and strain as in other toxicity studies to avoid additional studies to characterize pharmacokinetics and metabolism, and/or for dose-range finding. The species used should be well-characterized with respect to health, fertility, fecundity, and background rates of malformation and embryo-fetal death. Generally, within and between reproductive studies animals should be of comparable age, weight and parity at the start. The easiest way to fulfil these factors is to use animals that are young, sexually mature adults at the time of the start of dosing with the females being virgin, with the

exception of NHP where proven mothers can be an advantage for ePPND studies.

The species chosen for testing should be relevant and justified based on their advantages and disadvantages (see Table 9-1 in Section 9.3). If the species selected differs considerably from the human in regard to the considerations below, the impact should be considered when interpreting the reproductive toxicity data (see Principles of Risk Assessment, Section 7). Assessing all of the reproductive endpoints or parameters of interest in a single test species, however, is not always possible.

Additional points to consider in selection of a species relate to the interaction of the pharmaceutical with the species including:

a. The pharmacokinetic and metabolite profile (including adequate exposure to major human metabolites, as discussed in ICH M3(R2) (1));
b. Whether the species expresses the pharmacologic target (e.g., is an endogenous or exogenous target) and whether the pharmaceutical has adequate affinity for the target in the species selected;
c. Whether the functional pharmacological activity of the pharmaceutical is exhibited in the test species.

For highly targeted molecules, selection of a pharmacologically relevant species is particularly important as described in more detail in ICH S6(R1) (2).

4.1.1 *Rat as the Primary Species for Reproductive Toxicity Testing*

The rat is the most often used rodent species for reasons of practicality, general knowledge of pharmacology in this species, the extensive toxicology data usually available for interpretation of nonclinical observations from development of the pharmaceutical, and the large amount of historical background data. Thus, in many cases based on how species are selected for general toxicity studies, the rat is generally appropriate for reproductive toxicity testing.

4.1.2 *Rabbit as the Secondary Species for EFD studies*

For assessment of EFD only, a second mammalian non-rodent species is often warranted, although there are exceptions (e.g., vaccines, therapeutic antibodies, etc., see Sections 4.1.3 and 4.2, respectively). The rabbit has proven to be useful in identifying human teratogens that have not been detected in rodents; and the rabbit is routinely used as the non-rodent species based on the extensive historical background data, availability of animals, and practicality.

4.1.3 *Species Selection for Preventative and Therapeutic Vaccines*

The animal species selected for testing of vaccines (with or without adjuvants) should demonstrate an immune response to the vaccine. Typically, rabbits, rats, and mice are used. Nonhuman primates should be used only if no other relevant animal species is available, even though quantitative and qualitative differences can exist in the responses (e.g., in

humoral and cellular endpoints). It is usually sufficient to conduct developmental toxicity studies using only one animal model.

Rabbits are the most common species used for vaccine developmental toxicity studies, but other species are also appropriate. In primates (as in humans), the transfer of maternal antibodies across the placenta is limited, but generally increases over the course of gestation. In other species routinely used in reproductive testing the time course of transfer differs. The type of developmental toxicity study conducted and the choice of the animal model should be justified based on the immune response observed and the ability to administer an appropriate dose.

When there is a lack of an appropriate animal model (including NHP), a developmental toxicity study in rabbits, rats, or mice can still provide important information regarding potential embryo/fetal toxic effects of the vaccine components/formulation and safety of the product during pregnancy.

4.2 Non-routine Test Species

There are cases where it can be appropriate to use strategies other than those involved using the routine species discussed above. A commonly encountered example is where the rabbit is unsuitable for EFD testing. In situations like this, one can consider alternative species or approaches that can inform the risk assessment.

Many other species have been used to evaluate the effects of pharmaceuticals on the various reproductive stages. The suitability of alternative species will depend on the reproductive endpoints to be assessed (see Table 9-1 in Section 9.3).

NHPs can also be used for evaluating reproductive toxicity, especially for biotechnology-derived products, as described in ICH S6(R1) (2). NHPs should be considered if they are the only pharmacologically relevant species, provided that it is not already clear that the pharmacology of the pharmaceutical is incompatible with normal development or maintenance of pregnancy. There are additional factors that further limit the utility of studies in NHPs for reproductive risk assessment (see Annex 9.3 and ICH S6(R1)). An alternative animal model can be considered in place of NHPs for either small molecules or biotechnology-derived products by using a surrogate molecule that elicits the appropriate pharmacologic activity in the animal model, or data from genetically modified animals. The results of the studies can inform the assessment of reproductive risk (see Sections 4.3 and 7).

For biotechnology-derived products, when no relevant species can be identified because the biopharmaceutical agent does not interact with the orthologous target in any species relevant to reproductive toxicity testing, use of surrogate molecules or genetically modified models can be considered, as described in ICH S6(R1) (2) and Section 4.3.2. For some therapeutic modalities that lack orthologous target engagement in useful reproductive toxicology species and also have anticipated off-target effects, the testing strategy should address both of these situations.

In lieu of, or in addition to, the use of an *in vivo* mammalian study for assessment of reproductive toxicity, alternative approaches that can be considered include assessment of pharmacologic or mechanistic information, non-mammalian *in vivo* studies, or *in vitro* assays that predict reproductive toxicity (see Principles of Risk assessment Section 7).

4.3 Other Test Systems

4.3.1 ***Use of Disease Models***

Disease animal models are not routinely used in reproductive toxicity testing; however, there are some cases where they can be informative. Studies in disease models can be of value in cases where the data obtained from healthy animals could be misleading or otherwise not apply to the disease conditions in the clinical setting. Examples of situations where a reproductive toxicity study in a disease model could contribute information to the risk assessment include studies with pharmaceuticals that are replacement therapies, when the target is only present in disease state, or when the pharmacologic activity of the test article could yield confounding results in healthy animals (e.g., causes hypoglycemia or hypotension).

Recognizing that no animal model perfectly replicates human disease, there are several factors to be considered in choosing to study toxicity to reproduction in a disease animal model. The model should be pharmacologically relevant and appropriate for the reproductive endpoints being assessed. The pathophysiology of the disease course in the model should be characterized. Some differences from the human pathophysiology would not preclude its use provided that these are unlikely to confound data interpretation. Animal to animal variability should be characterized and appropriate within the context of the study. Reference data for the study endpoints should be available or should be generated during the study to aid data interpretation.

Although disease animal models can be used in definitive reproductive toxicity studies, they are more likely to be used as supplementary approaches to understand the relevance of adverse reproductive effects of the pharmaceutical in normal animals. The use of disease animal models and the design of the study for reproductive toxicity testing should be justified.

4.3.2 ***Use of Genetically Modified Models and Use of Surrogate Molecules***

For both genetically modified models and for surrogate molecules the effect of the intended pharmacology on reproduction is being investigated and thus informs the assessment of risk. For example, if the pharmacology is linked to adverse effects on reproduction, it can reasonably be concluded that the adverse effects would be experienced in some proportion of pregnant women receiving the pharmaceutical. However, the actual proportion of individuals affected (incidence) cannot be determined from animal studies, even if the actual pharmaceutical and a pharmacologically relevant species are used.

Genetically modified models can be used to create disease models or to characterize the on-target and off-target effects of a pharmaceutical on reproductive toxicity parameters. Such models can inform on whether the pharmacology of the target is closely linked to adverse effects on reproduction and development. When these models are used and off-target effects are anticipated based on therapeutic modality, the clinical candidate should be evaluated with this model to assess both on- and off-target effects.

When the clinical candidate does not have adequate activity against the target receptor in the routine test species, surrogate molecules can be used for any modality to assess potential adverse effects on reproductive toxicity. Using surrogate molecules is analogous to identifying class-effects from structurally diverse molecules with similar pharmacology. The overall approach is comparable to using a surrogate antibody that is pharmacologically active in the species being tested rather than using the humanized antibody that is pharmacologically active only in the NHP.

If there are no adverse effects on reproduction associated with the target pharmacology, evaluation of off-target reproductive toxicity using the clinical candidate is warranted.

5 DOSE LEVEL SELECTION, ROUTE OF ADMINISTRATION AND SCHEDULE

As part of the dose selection process, route of administration and schedule are important components in the design of reproductive toxicity studies. The dose selection should optimize exposure relative to humans considering route, schedule, and pharmacokinetics profile, to the extent that is practical.

The choice of dose levels, schedule and route of administration should be based on all available information (e.g., pharmacology, repeated-dose toxicity, pharmaco-/toxicokinetics, and Dose Range Finding studies) and a rationale should be provided. Guidance on the principles of dose selection is given in ICH M3(R2) Q&A (1) and ICH S6(R1) (2), and all available data should be used. Dose levels should be selected to investigate dose-response relationships for the primary endpoints of the study. Using doses similar to those used in the repeat dose toxicity studies of comparable duration permits interpretation of potential effects on reproductive and/or developmental endpoints within the context of general systemic toxicity and enables integration of data. When sufficient information on tolerability and pharmaco-/toxicokinetics in the test system is not available, appropriately designed exploratory studies are advisable.

Dosing schedules used in the toxicity studies influence the exposure profile which can be important in the risk assessment. Usually mimicking the clinical schedule is sufficient, but is not always warranted. A more frequent (e.g., twice a day) or a less frequent schedule can be appropriate to provide an exposure profile more relevant to the clinical exposure. When a more frequent schedule is contemplated, pragmatic factors (e.g., study logistics, stress on animals) should be considered.

In general the route of administration should be similar to the clinical route, provided the relevant human reproductive risk can be assessed. In circumstances where systemic exposure cannot be achieved or only small multiples of the clinical systemic exposure are achieved in the absence of maternal toxicity, a different route of administration should be considered. Use of a route of administration other than the clinical route should be justified in the context of the general toxicology program. When multiple routes of administration are being evaluated in humans, a single route in the test species can be adequate provided sufficient systemic exposure is achieved compared to that of the clinical routes.

It is not always warranted to use pregnant animals for dose selection, even if the reproductive study assesses pregnant animals. However, when exposure-based endpoints are used as the basis for selection of the dose levels (Section 5.1.3), it can be important to have TK from pregnant animals. If the TK is derived from non-pregnant animals for dose selection, then the achievement of the TK endpoint should be confirmed in pregnant animals.

5.1 Dose Selection Common to all Pharmaceuticals, Including Biotechnology-derived Pharmaceuticals

There are a number of dose selection endpoints that can be used for reproductive toxicity studies. All the endpoints discussed in this section are considered equally appropriate in terms of study design. The high dose in the definitive study should be one that is predicted to produce the anticipated change in the endpoint as described below in Sections 5.1.1 to 5.1.6. The selected high dose should be based on the observations made in appropriately designed studies, including the effects observed at higher dose levels in other studies (e.g., repeat-dose, TK, pEFD).

Justification for high dose selection using other endpoints than specified below, can be made on a case-by-case basis.

5.1.1 ***Toxicity–based Endpoints***

This endpoint is based on the prediction of minimal toxicity in the parental animals at the high dose. Minimal toxicity is defined as having an adverse effect on the parental animals without having an anticipated direct effect on the reproductive outcome. Factors limiting the high dose determined from previously conducted studies could include:

- Alterations in body weight (gain or absolute; either reductions or increases). Minor, transient changes in body weight gain or in body weight are not considered dose limiting. When assessing weight change effects, the entire dosing duration of the study should be considered and the absolute change that is appropriate is dependent on the parameter being measured, the species, strain, and the window of development being evaluated.
- Specific target organ toxicity (e.g., ovarian, uterine) or clinical pathology perturbations (e.g., changes in glucose) that would interfere with the study endpoints within the duration of the planned reproductive or developmental toxicity study.
- Exaggerated pharmacological responses (e.g., excessive sedation or hypoglycemia)

- Toxicological responses (e.g., convulsions, increased TEFL).

5.1.2 *Absorption, Distribution, Metabolism and Excretion (ADME)-based Saturation of Systemic Exposure Endpoint*

High dose selection based on saturation of systemic exposure measured by systemic availability of pharmaceutical-related substances can be appropriate (see ICH M3(R2) (1)). There is, however, little value in increasing the administered dose if it does not result in increased plasma concentration. For the purposes of this guideline, saturation of exposure is defined as substantial increases in dose that result in minimal increases in total exposure (e.g., a doubling of the dose resulting in only an approximate 20% increase in exposure).

5.1.3 *Exposure-based Endpoint*

It can be appropriate to select doses based on exposure margins above the exposure at the maximum recommended human dose (MRHD). For pharmaceuticals having primary and secondary pharmacology (or off-target effects) in the test species (e.g., small molecules), a systemic exposure representing a large multiple of the human AUC (area under the exposure curve) or C_{max} can be an appropriate endpoint for high-dose selection. This dose selection approach can be applied when there are qualitatively similar metabolite profiles between humans and the test species. The rationale for the metric used should be provided. Doses anticipated to provide an exposure > 25-fold of the clinical systemic exposure at the MRHD are generally considered appropriate as the maximum dose for reproductive toxicity studies (Note 4). Usually this is based on the parent moiety if it is the pharmacologically active agent. There are other cases (e.g., prodrugs, pharmacologically active metabolites) for which the Sponsor should provide a justification for the moieties included in the exposure multiple calculations.

When evaluating a pharmaceutical against a human endogenous target using an exposure-based endpoint, it is recommended to choose at least one species with pharmacodynamic activity. For studies using a surrogate molecule a dose should be used that has adequate pharmacodynamic activity in the test species. In addition to testing the surrogate, if the clinical candidate is anticipated to have secondary pharmacology or off-target effects, the clinical candidate should also be tested at doses anticipated to provide an exposure > 25-fold at the MRHD in the routine species.

Alternatively, instead of using a surrogate, for clinical candidates that have some demonstrated pharmacodynamic activity in the test species only at exposures > 25-fold, doses that achieve pharmacodynamic activity in the routine test species can be used. However, it should be noted that irrelevant off-target effects are likely to be observed.

If none of the routine test species are pharmacodynamically relevant, but the target is endogenous and the clinical candidate is anticipated to have off-target effects, an alternative endpoint rather than the exposure-based endpoints should be considered (e.g., limit dose, maximum feasible dose, toxicity-based endpoints).

When there is no human endogenous target (e.g., viral target), a > 25-fold exposure multiple of the MRHD is sufficient for high dose selection.

5.1.3.1 Considerations for Total vs. Fraction Unbound Pharmaceutical Exposure

The choice for the use of total vs. fraction unbound pharmaceutical exposures should be justified. The total exposure can be used as the default, unless the fraction unbound results in a lower exposure margin than that of the total; in this case the lower exposure multiple should be used for the comparison of animal vs. human exposures. Alternatively, the fraction unbound pharmaceutical exposure can be used regardless of whether it generates a lower or greater exposure multiple than that of the total exposure provided the following applies:

- The fractions unbound can be calculated accurately from the total pharmaceutical exposure, is reproducible at the effective concentrations in humans and at the toxicological concentrations in animals, and the fractions unbound are statistically significantly different.

Two examples of how this calculation might impact the exposure multiples are provided below.

- 25 fold exposure multiple not met: If the total exposure is 25 µM-hr in animals and 1 µM-hr in humans and unbound protein fraction is 5% and the unbound fraction in animals is 1%, then the margin would be 5.

- 25 fold exposure multiple exceeded: If the exposure is 10 µM-hr in animals and 5 µM-hr in humans and unbound protein fraction is 1% in human and 20% in animals, then the unbound ratio would be 40 rather than the apparent ratio of 2 based on total.

5.1.3.2 Exposure-based Approach for Highly Targeted Therapeutics

Highly targeted therapies (e.g., monoclonal antibodies, therapeutic proteins) are those that exhibit no or minimal off-target effect. For these therapeutics that exhibit pharmacodynamic effects in the test species, high dose selection can be accomplished by either identifying a dose which provides the maximum intended pharmacological effect in the preclinical species or a dose which provides an approximately 10-fold exposure multiple over the maximum exposure to be achieved in the clinic, whichever one is higher (ICH S6(R1)) (2). Corrections for large differences in target binding affinity and *in vitro* pharmacological activity between the nonclinical species and humans should be considered in dose selection such that a higher dose can be appropriate to elicit pharmacodynamic effects, if not limited by toxicity or feasibility. If the routine species do not exhibit pharmacological activity and a surrogate molecule is used, a dose of the surrogate that is 10-fold that which elicits the intended pharmacological activity in the test species can be appropriate.

5.1.4 *Maximum Feasible Dose (MFD) Endpoint*

Use of the MFD should maximize exposure in the test species, rather than maximize the administered dose (see also ICH M3(R2) (1)).

The MFD can be used for high dose selection when the physico-chemical properties of the test substance (or formulation) associated with the route/frequency of administration and the anatomical/physiological attributes of the test species limit the amount of test substance that can be administered.

5.1.5 *Limit Dose Endpoint*

A limit dose of 1 g/kg/day can be applied when other dose selection factors have not been achieved with lower dose levels (see also ICH M3(R2) (1) for other considerations).

5.1.6 *Selection of Lower Dose Levels*

It is generally desirable to establish a "no observed adverse effect level" for developmental and reproductive toxicity. Having selected the high dose, lower doses should be selected taking into account exposure, pharmacology, and toxicity, such that there is separation in anticipated outcomes between groups. Any dose level that yields a sub-therapeutic exposure is not generally informative to risk assessment, unless it is the highest dose that can be achieved without toxicity in the parental animals. For some of the variables in reproductive toxicity studies the ability to discriminate between background and treatment effects can be difficult and the presence or absence of a dose-related trend can be informative. The low dose should generally provide a low multiple (e.g., 1 to 5-fold) of the human exposure MRHD. The exposure at the mid dose should be intermediate between the exposures at the low and the high doses; however, dose spacing that results in less than 3-fold increase in exposure is not generally recommended.

5.2 Dose Selection and Study Designs for Vaccines

This guideline covers vaccines (adjuvanted or not) used in both preventative and therapeutic indications against infectious diseases. The principles outlined can be applicable to the nonclinical testing of vaccines for other indications as well (e.g., cancer). The types of studies depend on the target population for the vaccine and the relevant reproductive risk. Generally, reproductive studies are not warranted for vaccines being developed for neonates, pre-pubertal children, or geriatric populations.

For reproductive toxicity studies of vaccines it is typically sufficient to assess a single dose level capable of inducing an immune response in the animal model (Section 4.1.3). This single dose level should be the maximum human dose without correcting for bodyweight (i.e., 1 human dose = 1 animal dose). If it is not feasible to administer the maximum human dose to the animal because of a limitation in total volume that can be administered or because of dose-limiting toxicity (e.g., local, systemic), a dose that exceeds the human dose on a mg/kg basis can be used. To use a reduced dose, justification as to why a full human dose cannot be used in an animal model should be provided.

The vaccination regimen should maximize maternal antibody titers and /or immune response throughout the embryonic, fetal, and early postnatal periods. Timing and number of doses will depend on the onset and duration of the immune response of the particular vaccine. When developing vaccines to be given during pregnancy, the sponsor should justify the specific study design based upon its intended use (e.g., protecting the mother during pregnancy or protecting the child early postnatally).

Daily dosing regimens can lead to overexposure to the vaccine constituents. Episodic dosing of pregnant animals rather than daily dosing is recommended. Also, episodic dosing better approximates the proposed clinical immunization schedule for most preventive and therapeutic vaccines for infectious disease indications. Considering the short gestational period of routine animal species, it is generally recommended to administer a priming dose(s) to the animals several days or weeks prior to mating in order to elicit peak immune response during the critical phases of pregnancy (i.e., the period of organogenesis). The dosing regimen can be modified according to the intended vaccination schedule in humans.

At least one dose should be administered during early organogenesis to evaluate potential direct embryotoxic effects of the components of the vaccine formulation and to maintain a high antibody response throughout the remainder of gestation. If EFD toxicity is observed, this can be further assessed using subgroups of animals that are dosed at certain time points.

In cases where a vaccine includes a novel, active constitutive ingredient (including novel adjuvants) consideration of additional testing strategies similar to those for non-vaccine products can be appropriate.

It is recommended that the route of administration be similar to the clinical route of administration.

6 *DESIGN AND EVALUATION OF IN VIVO* MAMMALIAN STUDIES

The testing strategy to evaluate the potential reproductive risk of a pharmaceutical can include one or more *in vivo* studies. Although three separate study designs have been employed for the development of the majority of pharmaceuticals, various combinations of these study designs can be conducted to reduce animal use. All available pharmacological, kinetic, and toxicological data for the pharmaceutical should be considered in determining which study design(s) should be used. Study details for fertility, EFD, and PPND studies, and combinations thereof, can be found in Annex 9.4. Different approaches are listed below.

6.1 Three separate studies to assess all stages (A→F)

- Fertility and Early Embryo Development (FEED)
  - If effects on fertility are suspected, based on mode of action or on the results of repeat dose studies, it can be advisable to dose males and females in separate arms or separate studies comprising mating with untreated animals of the opposite sex.

- Embryo-Fetal Development (EFD)

- Pre- and PostNatal Development, including maternal function (PPND)

6.2 Single study design

A combination of fertility, gestation, and postnatal development (Stages A→F).

A single study design in rodents might be appropriate when reproductive toxicity is not expected. If such a study provides clearly negative results at appropriately selected doses, no further reproduction studies in that species are warranted. In this study, all newborns and pups, including stillbirths and culled pups, should be examined for morphological abnormalities. If reproductive and developmental toxicity is observed, these toxicity risks should be assessed in detail.

6.3 Two study design

- Combination of FEED and EFD (Stages A→D) + PPND (Stages C→F) studies.
  This combination of the FEED and EFD, in addition to the PPND study provides all the information obtained from conducting separate FEED and EFD and PPND studies, but uses fewer animals.
- Combination of EFD (Stages C→D) + FEED and PPND (Stages A→C + D→F) studies.

  This combination study design does not include an assessment of external, soft tissues, or skeletal morphology. It is most useful when no treatment-related TEFL effects were observed in the EFD study. The fertility and PPND combined study together with an EFD study, provide all the desired information for all stages of development, but uses fewer animals than the three study design.

6.4 Combination design of repeat-dose and fertility studies

In cases where no effects on male or female fertility are expected, or where extending the dosing period is appropriate due to observation of reproductive organ toxicity in long term repeated dose toxicity study, a combination design of repeat-dose and fertility studies can be considered. If effects on fertility are suspected, based on mode of action or on the results of repeat dose studies, it can be advisable to dose males and females in separate studies comprising mating with untreated animals of the opposite sex.

After a defined dosing period within the longer term repeat-dose toxicity study (e.g., 13- or 26-week repeat-dose study), males from the repeat dose study can be cohabited with sexually mature females from a separate study arm (untreated sexually mature females or where the female are treated for at least two weeks prior to mating). This combination study can reduce the number of animals used; however, the number of male animals in the repeat-dose study should be approximately 16 per group. Female animals and their fetuses will be examined for endpoints described in the procedures of the fertility study (Annex Section 9.4.1).

The male dose duration period which precedes the period of cohabitation should be determined based on the design principles of the fertility study described in Sections 3.2 and 9.4.1. The dosed males used for this assessment can come from any repeat-dose study (e.g., 4-, 13-, or 26-week study) provided the dose duration is sufficient for the project aims, the males are sexually mature, and the number of males available for cohabitation is sufficient to assess effects on male fertility and implant survival. The group size selected to assess male fertility should be justified based on species / strain characteristics. This combination study can reduce the number of dosed males which can be particularly useful with technically challenging exposure routes. It is also particularly useful where evaluation of the long term effects on male reproductive performance is desired.

It is possible to assess both male and female fertility simultaneously using males from the repeat-dose toxicity study by cohabiting the males with sexually mature females from a separate study arm that have been treated with drug for at least two weeks. The females and fetuses are assessed as described for the fertility study (Section 9.4.1). However, to detect drug effects on the oestrus cycle, group size should be at least 16 unless justification for smaller group sizes can be provided.

6.5 Evaluation of Data

6.5.1 *Data Handling/Data Presentation/Statistics for in vivo Studies*

The key to good reporting is the tabulation of individual values in a clear concise manner to account for all animals that are being assessed. Because the data are derived from offspring that are often not directly treated, clear and concise tabulation that permits any individual animal from initiation to termination to be followed should be presented. This will enable assessment of the contribution that the individual has made to any group summary values. Group summary values should be presented with significant figures that avoid false precision and that reflect the distribution of the variable.

For the presentation of data on structural changes (e.g., 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, as appropriate.

Graphical presentations that depict mean values for data collected on multiple days (e.g., mean body weights) are useful in visualizing a large amount of data. Annex or tabulations of individual values such as bodyweight, food consumption, and litter values, should be concise. While the presentation of absolute values should be the default, calculated values such as bodyweight gain or litter survival indices can provide further support. Where data from non-pregnant animals have been excluded from summary tables, this should be clearly indicated.

Presentation of fetal abnormality findings should utilize terminology that is consistent and easily understood.

Interpretation of study data should rely primarily on comparison with the concurrent control group. Historical control/reference data are most useful when an interpretation of the data relies on the knowledge of variability within the larger control population and specifically among control groups in previous studies. For example, when trying to understand relevance of malformations, historical control data are useful in interpreting the significance of rare events. The individual laboratory´s recent historical control database, if available, is preferred over data compilations from other laboratories. Ideally, the historical data should reflect data from contemporary studies (e.g., from years immediately preceding or following the study conduct, if available) as genetic drift can be an issue.

Comparison of study data to the historical mean and standard deviation or range is often performed. It can be important to take into consideration the frequency of the occurrence of an event. If so, then the frequency should be presented.

6.5.2 ***Statistics***

Developmental and reproductive toxicity studies usually show a distribution of response that does not follow a normal distribution, but can vary from any continuous to any discrete distribution. As a result, this should inform the statistical method used. When employing inferential statistics (determination of statistical significance) the basic unit of comparison should be used. The experimental unit is a concept that is oftentimes misinterpreted but refers to the units that have been randomized and treated. Therefore, cesarean and fetal data should be calculated for the litter as the unit of measure; study result inferences are made back to the mother, not to fetuses. This is because the pregnant females have been allocated to different dose groups (not the fetuses or neonates) and the development of individual offspring in a given litter is not independent. The responses of individual offspring in a given litter are expected to be more alike than responses of offspring from different litters. Similarly, for fertility studies the mating pair should be used as the basic unit of comparison.

In most cases, inferential statistics ("significance tests") will evaluate the relationship between a response and treatment factor. The key outputs from a statistical model are then the p-values and confidence intervals for assessing treatment effects – typically pairwise comparisons back to vehicle and/or a trend test across all the groups. The output of such significance tests should only be used as a support for the interpretation of results. Any biologically meaningful difference in treated animals compared with concurrent controls should be discussed. Statistical significance alone does not always constitute a positive signal nor does lack of statistical significance constitute a lack of effect; historical controls, biological plausibility, and reproducibility should be considered in this context. Use of statistical significance alone for drawing inferences when dealing with studies with small group sizes (e.g., NHP) should be approached with caution.

7 PRINCIPLES OF RISK ASSESSMENT

All available data on the pharmaceutical and any related compounds (e.g., surrogates or class alerts), as well as information on human genetics, transgenic animals and the role of the target in reproduction should be considered in this assessment. The amount of information available can depend on the stage of pharmaceutical development, the nature of the pharmaceutical and its intended use. The (projected) human exposure, comparative kinetics between species and plausible mechanism of reproductive toxicity, if available, should be considered.

Therapeutic benefit considerations can influence the appropriate level of human risk. For instance, a higher degree of risk could be appropriate for a pharmaceutical intended to treat a life-threatening disease for which all existing therapies have known adverse effects on reproduction than for a life-style pharmaceutical. Human data (e.g., known effects of human genetic variations, clinical trial experience) can greatly influence the overall assessment of human risk of reproductive or developmental toxicity. Definitive human data will supersede nonclinical data.

Any limitations (*e.g.,* test system relevance, achieved exposure), uncertainties and data gaps in the available nonclinical reproductive toxicity data package should be addressed and their impact assessed.

Risk assessment should generate conclusions relevant for risk communication and management for the intended patient population.

7.1 Risk Assessment for Reproductive and Developmental Toxicities

For human pharmaceuticals, an assessment should be conducted to identify potential risks on human reproduction throughout pharmaceutical development.

Endpoints reflecting the full range of potential reproductive and developmental effects as described in Section 2 should be addressed, if not otherwise justified.

Not all observations from nonclinical studies are considered to be adverse. An identified effect of the pharmaceutical can also be considered as non-adverse if it is an adaptive change (e.g., enzyme induction) which does not impact on reproductive or developmental function.

Adverse nonclinical effects should be evaluated to estimate the likelihood of increased reproductive or developmental risk for humans under the proposed conditions of use of the pharmaceutical. An analysis considering various factors that can increase or decrease the level of concern is recommended. Such factors include animal-human exposure ratio, level of maternal toxicity, dose-response relationship, type of observed effect(s), cross-species concordance, or similarity between pharmacologic and toxicological mechanisms. For example, concern for a reproductive or developmental risk would be increased in the event of a finding observed under any of the following conditions: low relative exposure in animals, cross-species concordance, absence of maternal toxicity, or similarity between pharmacologic and reproductive/developmental toxicological mechanisms. Conversely,

concern can be decreased by high relative exposure in animals, absence of cross-species concordance, excessive maternal toxicity or species-specific mechanisms.

When assessing effects on embryo-fetal development, one particular difficulty arises when fetal toxicity is observed at dose levels that were also toxic for the mother. It cannot be assumed that developmental toxicity was secondary to maternal toxicity unless such a relationship can be demonstrated either de novo or from published precedence. One way of doing this is to assess the degree of concordance between the severity of toxicity seen in the individual dams and the effects on their litters.

Also, the consistency between studies can provide further evidence of an adverse effect of the pharmaceutical (e.g., increased fetal lethality seen in a rodent EFD study consistent with decreased live litter sizes in the PPND study). It is important to consider the exposure at which specific effects were seen across studies and species. Knowledge of the mechanism of reproductive or developmental effects identified in animal studies can help to explain differences in response between species and provide information on the human relevance of the effect (e.g., rodent-specific effects of prostaglandin synthetase inhibitors on cardiovascular fetal development).

In general, TEFL are considered to be the critical endpoints in assessing prenatal developmental toxicity. In contrast, reversible or minor manifestations of developmental toxicity (e.g., changes in fetal weight, skeletal variations) by themselves are of minimal concern from a risk assessment perspective. However, an increased incidence of variations can influence the interpretation of an equivocal increase in related malformations. The extent of concern will be influenced by other factors (e.g., exposure multiple at which the findings occurred, cross-species concordance).

As in the case of developmental toxicity, reversible or minor manifestations of reproductive toxicity (e.g., a transient inhibition of spermatogenesis) by themselves are of minimal concern from a risk assessment perspective.

Comparison of pharmaceutical exposure at the No Observable Adverse Effect Level (NOAEL) in the test species to that at the MRHD is a critical determination. This comparison should be based on the most relevant metric (e.g., AUC, C_{max}, C_{min}, body surface area-adjusted dose). In general, there is increased concern for reproductive or developmental toxicity in humans when effects are seen in a relevant animal species and exposure at the NOAEL is < 10-fold the human exposure at the MRHD. When exposure at the NOAEL is > 10-fold the human exposure at the MRHD, the concern is reduced. When the exposure in animals at the NOAEL is > 25-fold the exposure at the MRHD, there is minimal concern for the clinical use of the pharmaceutical (Note 4). If a significant difference in relative exposures is observed between multiple test species, the appropriateness of the metric (e.g., AUC, C_{max}) being used for the interspecies exposure comparisons should be reassessed. When an alternative metric fails to reduce the disparity between species, the assessment of risk should be based on the most sensitive species. When applicable, the relative exposure ratio should consider both the parent compound and its metabolites.

Generally, the results from definitive *in vivo* studies with adequate exposures compared to the exposure at the MRHD carry more weight than those from alternative assays or preliminary studies. Also, the exposure data obtained from *in vivo* studies can be used to determine whether a positive signal identified in an alternative assay presents a risk at the MRHD under the clinical conditions of use of the pharmaceutical.

7.2 Risk Assessment for Lactation

Generally, evaluations of a pharmaceutical's effects on lactation and its presence in milk in animal studies have little relevance for human risk assessment. Pharmaceuticals can alter the process of lactation in the nursing mother. While the outcome of the PPND (or ePPND) study can inform the risk assessment and can inform as to whether there was extensive systemic exposure in the suckling infant, information on the quantity of the pharmaceutical in milk and production of milk is best derived from human experience, given that the composition of milk varies significantly between rodents and humans. The risk for direct adverse effects on the nursing infant depends on the concentrations of the pharmaceutical and its metabolites in the milk, their absorption, and the age of the infant. Premature infants and neonates have a different capacity to absorb, metabolize and excrete pharmaceuticals compared to older infants.

8 ENDNOTES

Note 1: In particular, the testes and epididymides should be sampled and processed using methods which preserve the tissue architecture and permits visualization of the spermatic cycles. A detailed qualitative microscopic evaluation with awareness of the spermatogenic cycle is sufficient to detect effects on spermatogenesis. A quantitative analysis of spermatic stages (i.e., staging) is not generally recommended but can be useful to further characterize any identified effects. In females, a detailed qualitative microscopic examination of the ovary (including follicles, corpora lutea, stroma, interstitium, and vasculature), uterus and vagina (rodents) should be conducted with special attention given to the qualitative assessment of primordial and primary follicles.

Note 2: Qualified alternative assays within the context of this guideline can only be applied under certain specific circumstances and have not been subject to formal validation. The EU requires the use of non-animal approaches as soon as they are validated and accepted for regulatory purposes (Directive 2010/63/EU, sector legislation and related guidance). However, this EU directive does not apply to alternative assays qualified according to this guideline.

Note 3: The ICH Reference Compound List in Annex 9.5.4 is not complete and as such we are soliciting data for additional reference compounds (positive and negative) for potential inclusion into the list, including relevant information as discussed below. These compounds can be either pharmaceuticals or non-pharmaceuticals and should be commercially available. Data to be submitted should include:

- Name, structure of the compound, suggested compound category, and CAS identifier (if available);

- The specific TEFL observed in nonclinical test species;

- Exposures (C_{max} and AUC) at the Lowest Observed Adverse Effect Level (LOAEL) if applicable and the NOAEL;

- References/sources for the specific data provided (will be made publicly available, if it is not already):

See examples in Table 9-7 in Annex 9.5.4 for the type of data being requested, as exemplified by four positive compounds (carbamazepine, fluconazole, 5-fluorouracil, and topiramate) and one negative compound (saxagliptin). Data should be summarized using a similar format as that shown in those examples.

This is not a request for data for the compounds listed in the Table 9-6 in Annex 9.5.4, nor is this a request for examples of assays that could be used.

Note 4: An analysis of 20 known human teratogens showed that if malformations were observed, exposure at the LOAEL in at least one species was < 25-fold the exposure at the MRHD. This indicates that using a > 25-fold exposure ratio for high dose selection in the development toxicity studies would have been sufficient to detect the teratogenic hazard for all these therapeutics. The analysis also showed that for all human teratogens that were detected in animal species the exposure at the NOAEL in at least one species was < 10-fold the exposure at the MRHD.

In addition, a survey was conducted on EFD toxicity studies by the IQ DruSafe Leadership Group. This survey identified 163 and 152 definitive rat and rabbit EFD studies, respectively, that achieved ≥ 15-fold animal to human parent drug exposure ratios (using human exposure at the intended therapeutic dose) in the absence of confounding (i.e., dose-limiting) maternal toxicity. An analysis showed that:

- Of the 163 rat studies, 51 (31%) achieved exposures ≥ 25-fold human and only 6 (3.7% of total cases) of these had TEFL findings. For all 6 rat cases, the LOAEL was ≥ 50-fold human exposure, one of which was predicted to be positive based on its mechanism of action.

- Of 152 rabbit EFD studies, 35 (23%) achieved exposures ≥ 25-fold human exposure and only 2 (1.3%) of these had TEFL findings. For the 2 rabbit cases, the LOAEL was ≥ 50-fold human exposure.

These data show that dosing animals to achieve exposures ≥ 25-fold human exposures when there is no maternal toxicity (that would otherwise limit the high dose), only infrequently detects a TEFL. In all these cases, TEFL findings were not observed until exposures exceeded 50-fold and findings at such high exposures are not believed to be relevant to human risk assessment. In the absence of confounding (i.e., dose-limiting maternal toxicity),

963 the selection of a high dose for EFD and PPND studies that represents a > 25-fold exposure
964 ratio to human plasma exposure of total parent compound at the intended maximal
965 therapeutic dose is therefore considered pragmatic and sufficient for detecting outcomes
966 relevant for human risk assessment.

967 9 GLOSSARY

968 **Alternative assay(s):** *In-vitro, ex-vivo* or non-mammalian *in-vivo* assay(s) intended to
969 evaluate a developmental endpoint (i.e., teratogenicity or embryo/fetal lethality; see TEFL).

970 **Applicability domain**: This describes the types of substances in terms of their physical
971 properties or specific types of substances for which the assay is appropriate. This applies to
972 what types of chemicals can meaningfully be tested in an assay, the applicable chemical
973 space. Examples of applicability could include physicochemical properties of the
974 pharmaceutical such as solubility, volatility, or assay interference by the molecule. The
975 applicability domain also refers to reasons why and conditions under which an assay can be
976 informative or cannot provide useful results. It could include the Training Set of the model
977 for which it is applicable to make predictions for new compounds.

978 **Assay qualification (for regulatory use):** Confirmation of the predictivity of an alternative
979 assay(s) to identify a defined adverse developmental outcome (i.e., TEFL), as outlined in this
980 guideline.

981 **Constitutive ingredients:** Chemicals or biologic substances used as excipients, diluents, or
982 adjuvants in a vaccine, including any diluent provided as an aid in the administration of the
983 product and supplied separately.

984 **Context of use:** For this guideline, context of use applies to regulatory conditions under
985 which the results of an assay can be relied upon. Examples could be: a stand-alone
986 replacement for an *in vivo* study under specified conditions, inclusion in a suite of
987 assays/assessments to replace *in vivo* studies, or to defer definitive studies to later in clinical
988 development.

989 **Developmental toxicity:** Any adverse effect induced prior to attainment of adult life. It
990 includes effects induced or manifested from conception to postnatal life.

991 **GD:** Gestation Day.

992 **GD 0:** The day on which positive evidence of mating is detected (e.g., sperm is found in the
993 vaginal smear / vaginal plug in rodents, or observed mating in rabbits).

994 **Highly targeted or highly selective pharmaceutical/therapeutic:** Therapeutics that exhibit
995 no or minimal off-target effects due to the nature of target binding (e.g., monoclonal
996 antibodies, therapeutic proteins).

997 **ICH Reference Compound List Categories Based on Intended Mechanism of Action:**

- **Channel modulator:** Compounds with a primary mode of action of targeting cellular channels or transporters.

- **DNA modifiers:** Compounds with a primary mode of action of either DNA intercalation or DNA modification (direct [e.g., alkylation, methylation] or indirect [e.g., based on enzyme modulation]).

- **Enzyme Modulator:** Inhibitor, activator, or inducer of enzymes not covered by other categories (e.g., Kinase Modulator).

- **Hormone/Steroids:** Compounds with a primary mode of action of mimicking, modulating, or antagonizing paracrine, endocrine, or exocrine function.

- **Kinase Modulator:** A specific subset of Enzyme Modulators specifically affecting kinases.

- **Nucleoside Modulator/Nutrient Blocker/Central Metabolite Inhibitor:** Anti-metabolites of nucleosides, nutrients, or metabolic pathway intermediates.

- **Oligonucleotide-based Modulators:** DNA or RNA-based oligonucleotides affecting transcription or translation.

- **Receptor Modulator:** Compound that binds to a receptor, either nuclear- or membrane-based (non-kinase receptor modulators), to elicit a response.

- **Secondary Messenger Modulator:** Binding to a target that directly alters cellular communications between intra- and extra-cellular compartments.

- **Others:** Any other compounds that are not part of any of the above categories or for which there is no intended biological activity (e.g., industrial chemicals).

Malformation: Permanent structural deviation that generally is incompatible with or severely detrimental to normal postnatal development or survival.

Modality: Type of pharmaceutical such as small chemical entity, monoclonal antibody, oligonucleotide, nanobody, peptide, protein, vaccine.

Normalization Factor: For the purposes of this guideline; a mathematical algorithm used to relate the alternative assay result and the *in vivo* observations to the exposures at which they occur.

Off-target or Secondary Pharmacological Activity: Action or effect of a pharmaceutical not related to its intended therapeutic effect.

Pharmacologically Active or Primary Pharmacological Activity: Eliciting the desired effects by either directly impacting the target (e.g., inhibition, activation, up regulation, or

down regulation) or resulting in the intended physiological outcome (e.g., lower blood pressure).

PND: Postnatal day.

PND 0: Day last offspring of a litter is confirmed as delivered.

Preliminary EFD (pEFD): A developmental toxicity study that includes exposure over the period of organogenesis, has adequate dose levels, uses a minimum of 6 pregnant animals per group, and includes assessments of fetal survival, fetal weight, and external and soft tissue alterations (see ICH M3(R2) (1)).

Enhanced pEFD: A pEFD study that is GLP compliant, increases the number of pregnant animals to ≥ 8 per group, and includes fetal skeletal examinations.

Surrogate molecule: A molecule showing similar pharmacologic activity in the test species as that shown by the human pharmaceutical in the human; for a biologic, is can also be referred to as a homologous protein.

TEFL: Teratogenic and/or embryofetal lethal.

Teratogen: For the purpose of this guideline; a pharmaceutical that causes malformations.

Training Set: A set of data used to discover potentially predictive relationships.

Test Set: A set of data used to assess the strength and utility of a predictive relationship.

Vaccine: For the purpose of this guideline, this term refers to preventative or therapeutic vaccines for infectious diseases. Vaccine (inclusive of the term vaccine product) is defined as the complete formulation and includes antigen(s) (or immunogen(s)) and any additives such as adjuvants, excipients or preservatives. The vaccine is intended to stimulate the immune system and result in an immune response to the vaccine antigen(s). The primary pharmacological effect of the vaccine is the prevention and/or treatment of an infection or infectious disease.

Variation: Structural change that does not impact viability, development, or function (e.g., delays in ossification) which can be reversible, and are found in the normal population under investigation.

10 REFERENCES

1. International Conference on Harmonisation M3(R2): Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals (2009) together with ICH M3(R2) Questions & Answers (2012)

2. International Conference on Harmonisation S6(R1): Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (2011)

3. International Conference on Harmonisation (2009). S9: Nonclinical Evaluation for Anticancer Pharmaceuticals.

11 ANNEX

11.1 Table of species advantages/disadvantages

Table 9-1. Species for Developmental and Reproductive Toxicity Testing

Species	Advantages	Disadvantages
Routine Species		

Rat	• Well-understood biology
	• Widely used for pharmacodynamics and drug discovery
	• Robust reproductive capacity with short gestation
	• Large group sizes and litter size
	• Suitable for all stages of testing
	• Widespread laboratory experience and high capacity
	• Extensive historical data
		• Different placentation (e.g., timing, inverted yolk sac)
		• Dependence on prolactin as the primary hormone for establishment and maintenance of early pregnancy, which makes them sensitive to some pharmaceuticals (e.g., dopamine agonists)
		  • Highly sensitive to pharmaceuticals that disrupt parturition (e.g., Nonsteroidal anti-inflammatory drugs in late pregnancy)
		• Less sensitive than humans to fertility perturbations
		• Limited application for humanized monoclonal antibodies
		  o Limited or no pharmacologic activity
		  o Limited or no binding
		  o Significant anti-drug immune response

Rabbit	• Similar advantages to rats plus • Non-rodent model • Readily amenable to semen collection • Placental transfer of antibodies more closely approximates primates than does rodents	• Limitations similar to rat for biologics • Limited historical data for fertility and pre-/postnatal studies • Sensitive to gastrointestinal disturbances; (e.g., some antibiotics) • Prone to spontaneous abortion • Clinical signs difficult to interpret • Not generally used for general toxicology (except for vaccines), lack of kinetic or toxicity data • Limited use for pharmacodynamics

Mouse	• Similar advantages to rats • Genetically modified models available or readily generated • Amenable to surrogate approaches • Uses small amounts of test material	• Similar limitations to rats • Small fetus size and tissue volumes • Stress sensitivity • Malformation clusters particularly evident • Less historical data with certain strains • Different placentation (e.g., timing, inverted yolk sac) • Less sensitive than humans to fertility perturbations

1071

Species	Advantages	Disadvantages
Non-routine Species		
NHP (Details are for Cyno)	• Phylogenetically and physiologically more similar to humans • More likely than rodents to show pharmacology and tissue reactivity to human proteins • Placentation similar to human • Larger size and tissue samples • Used in repeat-dose toxicity • Transfer of mAb across the placenta similar to humans	• Low fecundity o High background pregnancy loss o Single offspring • Long menstrual cycle (30 days) and gestation (165 days) • Impractical for fertility (mating) studies • Sexual maturity occurs around 3 to 6 years of age • Separation of mother and neonate during postpartum bonding period can be detrimental to neonate • F1 reproduction function difficult to evaluate • Small group size (ethical considerations), hence low statistical power • Animal welfare considerations • Kinetics can differ from humans as much as other species • Limited historical control and laboratory experience/capability • Limited availability of breeding animals • Highly variable age, weight and parity at the start • Uses a large amount of test material

1072

Species	Advantages	Disadvantages
Mini-pigs	• Alternate non-rodent for general and reproductive toxicity testing • Susceptibility to some human teratogens • Short period of organogenesis (GD 11-35) • Defined genetic background and specific-pathogen-free animals • Short dose range-finding studies possible (mid-term) • Bred in and adapted to laboratory conditions • Sexual maturity at 3 to 5 months • Good litter size compared to NHP • Suitable for serial semen sampling and mating studies • Monitor pregnancy by ultrasound • Sufficient historical background data on reproductive endpoints	• Limited number of experienced laboratories • Long gestation • Uses a large amount of test material • Large housing requirement • Minimal to no prenatal transfer of antibodies
Limited Use Species (primarily used for investigative purposes)		
Guinea pig	• Alternate rodent model that can demonstrate efficacy and cross-reactivity • Placental transfer of antibodies in the last part of gestation is at a similar level in humans	• Historical control and laboratory experience limited to few laboratories • Sensitive to GI disturbances; susceptibility to some antibiotics • Validation of postnatal behavioral and functional tests is limited • Long fetal period • Lack of kinetic or toxicity data • Blood sampling more difficult

1073

Species	Advantages	Disadvantages
Hamster	• Alternate rodent model that can demonstrate efficacy and cross-reactivity	• Higher postnatal loss due to cannibalization • Limited historical control and laboratory experience • Validation of postnatal behavioral and functional tests is limited • IV route difficult, can hide orally administered doses in cheek pouches • Aggressive • Sensitive to GI disturbances • Overly sensitive teratogenic response to many chemicals • Lack of kinetic or toxicity data • Blood sampling more difficult
Dog	• Usually have repeat-dose toxicity data • Large tissue volume • Readily amendable to semen collection	• Twice yearly ovulators and long gestation (63 days) • Limited historical control and laboratory experience • Validation of postnatal behavioral and function tests is limited • Uses a large amount of test material • Immunogenicity/anaphylaxis concerns
Ferrets	• Alternate model that can demonstrate efficacy and cross-reactivity	• Seasonal breeder unless special management system used (success highly dependent on human/animal interactions) • Minimal historical control data and laboratory experience

11.2 *In vivo* Study Designs

The number of animals per group specified in individual studies is a balance based on scientific judgment from many years of experience with these study designs, and ethical considerations on the appropriate use of animals. Numbers group sizes can be adjusted when there is evidence either from the pharmacological action of the compound or from existing studies that the dosages used are expected to elicit an effect at a high frequency and therefore fewer animals are warranted to confirm the presence of an effect. The number of animals can differ 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).

For all but the rarest events (such as malformations, abortions, total litter loss), evaluation of 16 to 20 litters for rodents and rabbits tends to provide a degree of consistency among

studies. Below 16 litters per evaluation, between study results become inconsistent, and above 20 to 24 litters per group, consistency and precision is not greatly enhanced. These numbers relate litters available for evaluation. If groups are subdivided for different evaluations the number of animals starting the study should be adjusted accordingly. Similarly, in studies with 2 breeding generations, 16 to 20 litters should be available for the final evaluation of the litters of the F1 generation. To permit for natural attrition, starting group size of the F0 generation of at least 20 is recommended.

Provided below are representative study designs that could be utilized. However, parameters, timings, and assessments can be readily modified and still meet the study goals. Expert judgment should be used for adapting these framework designs for individual laboratories and purposes.

11.2.1 *Fertility and Early Embryonic Development (FEED) Study*

A fertility assessment in rodents is generally recommended (see Sections 3.2 and 4.1). The aim of the FEED study is 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 Section 2). For females, this should detect effects on the estrous cycle, tubal transport, implantation, and development of preimplantation stages of the embryo. For males, it will permit detection of functional effects (e.g., epididymal sperm maturation) that cannot be detected by histological examinations of the male reproductive organs. The fertility study is designed to assess the maturation of gametes, mating behavior, fertility, preimplantation stages of the embryo, and implantation.

A combined male/female FEED study is commonly used (See Table 9-2), but separate male only or female only options are possible by substituting the appropriate number of untreated males or females in the study designs and should be considered case-by-case.

Table 9-2: FEED Study Design: Rats, combined male and female study

Parameter	Male and Female
Typical Group size	20 + 20
Number of dose groups	4
Administration period[a]	M: ≥ 2 weeks prior to cohabitation through at least confirmation of mating F: ≥ 2 weeks prior to cohabitation through implantation (GD6)
Mating ratio	1 male:1 female
Mating period[b]	≥ 2 weeks
Estrous cycle evaluation	Daily, commencing 2 weeks before cohabitation and until confirmation of mating
Clinical observations/mortality	At least once daily
Body weight	At least twice weekly
Food consumption	At least once weekly (except during mating)
Male euthanasia[c]	Perform macroscopic examination and preserve macroscopic findings, testes and epididymides for possible microscopic examination
Sperm analysis[d]	Optional
Mated female euthanasia[e]	Perform macroscopic examination and cesarean section; preserve macroscopic findings, ovaries and uteri for possible microscopic examination
Scheduled cesarean section: uterine implantation data	Corpora lutea counts, number of implantation sites, live and dead embryos

a: Available data (e.g., histopathology, weight of reproductive organs, in some cases hormone assays and genotoxicity data) from toxicity studies should be used to justify dosing duration, especially for detecting effects on spermatogenesis. 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. Treatment of males should continue throughout confirmation of mating, although termination following confirmation of female fertility can be valuable. Treatment of females should continue through at least implantation. This will permit evaluation of functional effects on fertility that cannot be detected by histopathological examination in repeated dose toxicity studies and effects on mating behaviour. If data from other studies show there are effects on weight or histology of reproductive organs in males or females, then a more comprehensive study should be considered.

b: Most rats will mate within the first 5 days of cohabitation (i.e., at the first available estrus), but in some cases females can become pseudopregnant. Leaving the female with the male for up to 3 weeks permits these females to restart estrous cycles and become pregnant.

c: It can be of value to delay sacrifice of the males until the outcome of mating is known. In the event of an effect on fertility, males could be mated with untreated females to ascertain any potential male mediation of the effect. The males can also be used for evaluation of toxicity to the male reproductive system if dosing is continued beyond mating and euthanasia delayed (e.g., histopathology, sperm analysis (see footnote d).

d: Sperm analysis (e.g., sperm counts, motility, and/or morphology) can be used as an optional method to confirm findings by other methods and to characterize effects further.

e: 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).

11.2.2 *Pre- and Postnatal Developmental (PPND) toxicity study*

A PPND study in rodents is generally warranted (see Sections 3.4 and 4.1). The aim of the PPND is 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 effects induced during this period can be delayed, observations should be continued through sexual maturity (i.e., stages C through F of the reproductive process, see Section 2). The PPND toxicity study is designed to assess enhanced toxicity relative to that in non-pregnant females, pre- and postnatal death of offspring, altered growth and development, and functional deficits in offspring, including maturation (puberty), reproductive capacity at maturity, sensory functions, motor activity, and learning and memory.

The females are permitted to deliver and rear their offspring to weaning at which time at least one male and one female offspring per litter should be selected for rearing to adulthood and mating to assess reproductive competence (see Table 9-3).

Table 9-3: PPND Toxicity Study Design: Rats

Parameter	
Typical Group size[a]	Approximately 20 females
Number of dose groups	4
Administration period	From implantation (GD 6/7) through weaning (PND 20/21)
F0 Females	
Clinical observations/mortality	At least once daily
Body weight	At least twice weekly
Food consumption	At least once weekly at least until delivery
Parturition observations	GD 21 until complete
Necropsy	PND 21 At necropsy, preserve and retain tissues with macroscopic findings and corresponding control tissues for possible histological evaluation
F1 Pre-weaning	
Clinical observations/mortality	Daily from PND 0
Litter size, live and dead	Daily from PND 0
Body weights and sex	PND 1, 4, 7, 14, and 21
Optional Standardization of litter size	≥ PND 4, to 4 or 5 pups per sex
Physical development and reflex ontogeny[b]	Depending on landmark
F1 Post-weaning	
Selection for post-weaning evaluation and group size[c]	PND 21, at least 1 male and 1 female/litter where possible to achieve 20 animals per group/sex
Clinical observations/mortality	Daily
Body weight	Weekly
Optional Food consumption	Weekly
Maturation (puberty)[d]	Females: vaginal opening, from PND 30 until complete Males: preputial separation, from Day 40 until complete

Other functional tests[e]	According to standard procedures
Reproductive performance	At least 10 weeks old, paired for mating (1M:1F) within the same group (not siblings)
Terminal procedures of males and females	Preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison Cesarean section: uterine implantation data, corpora lutea counts, number of implantation sites, live and dead embryos

a: In studies with 2 breeding generations, 16-20 litters should be available for the final evaluation of the litters of the F1 generation. To permit for natural wastage, the starting group size of the F0 generation should be approximately 20.
b: The best indicator of physical development is bodyweight. Achievement of preweaning landmarks of development such as eye opening and pinna unfolding as well as others is highly correlated with pup bodyweight. Reflexes, surface righting, auditory startle, air righting, and response to light are also dependent on physical development. Therefore, attention should be paid to differences in these parameters when observed in the absence of effects on bodyweight.
c: One animal per sex per litter are retained to conduct behavioral and other functional tests, and to assess reproductive function. There can be circumstances where more animals per litter can be retained for independent functional assessments.
d: Bodyweight should be recorded at the time of attainment to determine whether any differences from control are specific or related to general growth.
e: Investigators are encouraged to adopt methods that would assess sensory functions, motor activity, and learning and memory. Learning and memory should be evaluated in a complex learning task. Assessments of locomotor activity and startle reflex with prepulse inhibition (if conducted) should be evaluated over a sufficient period of time to demonstrate habituation.

11.2.2.1 Optional Modification of Rodent PPND Study to Assess Juvenile Toxicity Endpoints

In certain cases when a juvenile animal study is warranted, a PPND study can be modified to add juvenile toxicity endpoints to potentially reduce animal use and address a specific issue of concern (1). The following should be considered to support this approach:

- Determine the period of exposure appropriate to support the pediatric use.
- Demonstrate adequate exposure in the pups *via* the milk and/or consider direct dosing of pups for the period of developmental interest (TK sampling of the F1 generation using culled animals during the early post-partum period or study animals shortly before weaning can provide exposure data and can avoid pre-weaning dosing).

Endpoints included in this modified PPND study should be based on the principles appropriate for juvenile animal study designs supporting pediatric uses and are not discussed in this (S5) guidance.

11.2.2.2 Enhanced Pre- and Postnatal Developmental toxicity study (ePPND) in NHP

The ePPND toxicity study (Table 9-4) is a study in NHP that combines the endpoints from both the EFD and PPND studies in which dosing is extended throughout the gestation period to parturition (i.e., GD20 to parturition). See ICH S6(R1) for information on timing and additional parameters to be evaluated.

Table 9-4: ePPND Toxicity Study Design: for cynomolgus monkey[a]

Parameter	
Group size[b]	Generally ≥ 16 presumed pregnant
Number of dose groups	At least one treatment group plus a control group
Administration period	Initiates upon detection of pregnancy (approximately GD 20) to parturition
F0 Females	
Clinical observations/mortality	At least once daily
Body weight	At least weekly
Parturition observations	Document day of completion
Ultrasound evaluations	Only to track pregnancy status
Necropsy and tissue evaluation	Only as warranted
F1	
Clinical observations/mortality	Daily from PND 0
Body weights	Weekly
Morphometry/Physical development	After PND 0 and at regular intervals
Mother-infant interaction	Minimally in early postnatal period to confirm nursing; as appropriate thereafter
External evaluation	After PND 0 and at regular intervals
Skeletal evaluation	Month 1 and/or later
Visceral evaluation	At necropsy
Necropsy	Variable timing, depends on aim of the evaluations Preserve and retain tissues for possible histological evaluation

a: If an NHP other than the cynomolgus monkey is used, the study design should be adapted accordingly and a rationale provided.

b: Group sizes in ePPND studies should yield a sufficient number of infants (6-8 per group at postnatal day 7) in order to assess postnatal development and provide the opportunity for specialist evaluation if warranted (e.g., immune system). Most ePPND studies accrue pregnant animals over several months. See ICH S6(R1) regarding accrual of animals.

11.2.3 *Embryo-Fetal Developmental (EFD) Toxicity Study*

The aim of the EFD toxicity study is 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 (Table 9-5). This comprises evaluation of stages C through D of the reproductive process (see Section 2). The embryo-fetal developmental toxicity study is designed to assess enhanced maternal toxicity relative to that in non-pregnant females, embryo-fetal death, altered growth, and structural changes.

11.2.3.1 Dose Range Finding (DRF) Study

DRF studies in mated females are most often used to select appropriate dose levels, or dose schedules, for the definitive EFD studies but tolerability and TK data from existing repeat-dose toxicity can be sufficient for this purpose.

11.2.3.2 pEFD Study

The preliminary embryo-fetal developmental toxicity study (Table 9-5) is similar in design to the definitive embryo-fetal developmental toxicity study. A typical pEFD study design includes dosing over the period of organogenesis, has adequate dose levels, evaluates a minimum of 6 pregnant females per group, and includes assessments of fetal survival and weight, as well as external and soft tissue examinations (see ICH M3(R2)).

11.2.3.3 Definitive Embryo-fetal Developmental Toxicity Study

The females are cesarean sectioned near term and includes assessments of fetal survival and weight, as well as external, soft tissue and skeletal examinations (Table 9-5). The timing given in Table 9-5 is for rat and rabbit. For other species appropriate timing should be used.

Table 9-5: Embryo-Fetal Developmental Toxicity Study Designs for Rat and Rabbit

	EFD		
Parameter	**Rat**	**Rabbit**	**pEFD[a]**
GLP Status	Yes	Yes	**No**
Minimum number of litters	16	16	6 (pregnant animal)[g]
Number of dose groups	4	4	4
Administration period[b]	GD6-17	GD7-19	Species appropriate
Antemortem endpoints			
Clinical observations/mortality	At least once daily	At least once daily	At least once daily
Body weight[c]	At least twice weekly	At least twice weekly	At least twice weekly
Food consumption	At least once weekly	At least once weekly	At least once weekly
Toxicokinetics	Yes	Yes	Optional
Postmortem endpoints			
Cesarean section[d]	GD20/21	GD28/29	Species appropriate
Macroscopic examination	✓	✓	✓
Uterine weight	Optional	Optional	Optional
Corpora lutea	Optional	Optional	Optional
Implant sites	✓	✓	✓
Live and dead conceptuses	✓	✓	✓
Early and Late resorptions	✓	✓	✓
Gross evaluation of placenta	✓	✓	✓
Fetal body weight	✓	✓	✓
Fetal sex	✓	✓	✓
Fetal external evaluations[e,f]	Yes	Yes	Yes
Fetal soft tissue evaluations[e,f]	Yes	Yes	Yes
Fetal skeletal evaluations[e,f]	Yes	Yes	**No**

a: In an enhanced pEFD study the number of pregnant animals should be increased from 6 to ≥ 8 per group, include fetal skeletal examinations, and it should be conducted in accordance with GLP regulations.

b: Females are dosed with the test substance from implantation to closure of the hard palate (i.e., stage C of the reproductive process, see Section 2).

c: Daily weighing of pregnant females during treatment can provide useful information.

d: Cesarean sections should be conducted approximately one day prior to parturition. Preserve organs with macroscopic findings for possible histological evaluation; keep corresponding organs of sufficient controls for comparison.

e: All fetuses should be examined for viability and abnormalities. To permit subsequent assessment of the relationship between observations made by different techniques fetuses should be individually identified. It is critical to be able to relate all findings by different examination techniques (i.e., body weight, external inspection, soft tissue and/or skeletal examinations) to a single specimen in order to detect patterns of abnormalities.

f: It is preferable to examine all fetuses for both soft tissue and skeletal alterations, if permitted by the methods employed (e.g. fresh dissection or μCT, MRI, etc.). When using techniques precluding evaluation of both soft tissue and skeletal changes in the same fetus, 50% of fetuses from each litter should be allocated to each examination. The internal soft tissues of the head should be examined in at least 50% of the fetuses.

g: Minimum number of litters equals the number of pregnant animals per group, not the number of litters for pEFD studies.

11.2.4 *Combination Studies*

11.2.4.1 Fertility and Embryonic Development (FEFD)

The aim of the combined FEFD study is to test for toxic effects/disturbances resulting from treatment from before mating (males/females) through mating, implantation and until the end of organogenesis. This comprises evaluation of stages A to C of the reproductive process (see Section 2).

A combined male/female FEFD is commonly used, but a separate female only option is possible where male fertility is assessed in a separate study such as a repeat dose study of suitable duration. The study would then use untreated males for mating purposes only. For specific study design and observational parameters see Sections 9.4.1 and 9.4.3 (FEED and EFD).

11.2.4.2 Fertility and PPND (FPPND)

The aim of the combined Fertility and Pre-and Postnatal Development study (FPPND) study is to test for toxic effects/disturbances resulting from treatment from before mating (males/females) and 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 effects induced during this period can be delayed, observations should be continued through sexual maturity. This comprises evaluation of stages A to F of the reproductive process (see Section 2). The pre- and postnatal developmental toxicity study is designed to assess enhanced toxicity relative to that in non-pregnant females, pre- and postnatal death of offspring, altered growth and development, and functional deficits in offspring, including behavior, maturation (puberty) and reproductive capacity at maturity.

The study design features should encompass those of the individual studies in terms of the number of animals used and the parameters assessed. For specific study design and observational parameters see Sections 9.4.1 and 9.4.2 (FEED and PPND, respectively).

A combined male/female FPPND can be used, but a separate female only option is possible where male fertility is assessed in a separate study such as a repeat dose study of suitable duration. The study would then use untreated males for mating purposes only.

11.3 Qualification of Alternative Test Systems for Regulatory Acceptance

A framework and testing scheme to facilitate the qualification of alternative assays, including a list of test compounds (ICH Reference Compound List), is provided in this section. The ICH Reference Compound List provides information on embryo-fetal toxicity for various reference compounds, organized by overarching categories. This list is generated recognizing that the context of use will inform on acceptability of particular alternative assessments. Performance factors for assay acceptance are also outlined. The ICH Reference Compound List is intended to be periodically updated.

The applicability domain (see Glossary) together with the intended regulatory context of use influences the factors for assay qualification and the rigor for achieving regulatory acceptance.

11.3.1 ***Selection Factors for the ICH Reference Compound List***

The ICH Reference Compound List aims to cover reference compounds known for their TEFL effects in animals or humans, even if the mode of action is uncertain.

Availability of data showing clear TEFL effects in rats and/or rabbits in the absence of maternal toxicity represents an essential inclusion criterion for the selected positive compounds. This includes, when available, the multiples comparing human exposure to animal exposures where effects were seen.

Availability of pharmacokinetic and toxicokinetic data in the test species is an important criterion for the selection of reference compounds. Thus, all compounds used should have non-clinical exposure data (C_{max} and/or AUC) under the approximate conditions tested yielding either negative or positive results in the *in vivo* studies for the species being predicted. While pharmaceuticals are preferred, other chemicals can be considered. The ICH Reference Compound List does not currently include biotechnology-derived pharmaceuticals. The list favors compounds with direct effects on the fetus; however, a few are known to depend on cytochrome P450 metabolic activation to cause TEFL. Cytotoxic and/or genotoxic compounds are included to a limited extent because they are expected to induce TEFL through their intrinsic property of preferentially damaging rapidly dividing cells.

The performance of alternative assay(s) to detect species-specific differences can be evaluated by testing reference compounds known to cause TEFL in a single species; however, the number of such compounds available in the public domain is limited.

Compounds not causing TEFL (negative compounds) are also included in the ICH Reference Compound List to permit assessment of assay specificity. These compounds can be negative at all *in vivo* doses tested, or can be positive (TEFL observed) at higher doses/exposures, provided the alternative assay predicts the transition from negative to positive. The alternative assay should predict a negative result at some extrapolated multiple under the conditions for which the *in vivo* study yielded a negative result (no TEFL).

Further, the ICH Reference Compound List includes compounds from different chemical/pharmacologic classes with overlap with both negative and positive compounds to enable adequate coverage of the alternative assay for pharmaceuticals and diverse chemical structures and mode of action.

It is not critical for assay qualification purposes that the exposures achieved in animals that resulted in negative or positive TEFL outcome exceed the human exposures. This is in contrast to application of assay results for risk extrapolation where preferably the highest doses/exposures tested are at or above MRHD.

Finally, the commercial availability of the selected compounds of appropriate quality was considered in the generation of the list.

11.3.2 ***Performance Factors***

To be appropriate for regulatory use, the alternative assay(s) should be characterized using the ICH Reference Compound List. The list is not exhaustive and the recommendations provided are based on available information and pragmatic considerations. At least 45 compounds in total should be tested. Other compounds can substitute for the non-core compounds, but their use should be justified according to the inclusion factors mentioned above.

The compounds are distributed into multiple classes, covering a wide range of biological and chemical activities. All classes should be tested (at least 2 or 3 compounds from each class). An approximate 2:1 ratio of positive to negative compounds should be tested because it is important to identify positive compounds, but this ratio also ensures selectivity with the limited number of compounds available. For safety assessment purposes, and for some contexts of use, the false negative rate can be more important than the false positive rate.

The sensitivity to detect a positive signal in an assay(s), should be at least 80%, with evidence of selectivity (i.e., differentiating between true positives and true negatives).

The evaluation should identify the applicability domain and any limitations of the assay(s), and include assessments of accuracy, and reproducibility over time. Inter-laboratory reproducibility and transferability should be established if a particular assay is to be used in more than one laboratory.

Individual assays or combinations of assays can be used to predict TEFL. The performance characteristics of each individual assay as well as the performance of the combined battery, if used, should be specified. Various statistical methods are available for determining which combination of assessments will give the best predictivity.

11.3.3 ***Assay Qualification Information to be Provided to Health Authorities***

To enable evaluation of an alternative assay(s) for use in risk assessment for regulatory purposes, the following information should be provided.

A detailed description should be presented concerning what the predictive model is, what species (e.g., rat, rabbit, and/or human outcomes) it is trying to predict, and what reproductive endpoint it assesses. The predictive model can consist of a single assay or a battery of assays used together to predict the endpoint of interest (e.g., TEFL) in the respective species such as rat. If a battery of assays is used, each should be fully described. The specific endpoint(s) used (e.g., gene signature, morphology) should be described and how the assessment is made, including how the endpoints were selected and the specific factors for positive and negative determinations, should be discussed.

The details of the algorithm employed for determining positive and negative outcomes from assay observations should also be presented. The predictive model should correlate concentrations tested in the alternative assay(s) to the *in vivo* exposure that results in an adverse outcome in the species being predicted. For example, concentrations associated with positive effects on the endpoint should take into consideration *in vivo* exposure such as C_{max} or AUC. This permits the model to be used for exposure-based risk assessment. The pharmacokinetic parameter used including any normalization factors employed to correlate with *in vivo* results should be presented (Section 3.5.3).

The compound list used to qualify the assay performance should be presented. Documentation should include a clear identification of the compound list used as the Training Set (see Glossary) to develop the assay, and the compound list used as the Test Set (see Glossary) to evaluate the assay's performance. The assay Training Set can include compounds of the sponsor's choice not on the ICH Reference Compound List. Additional compounds not in the ICH Reference Compound list can be used as part of the Training Set or the Test set, but not both. No more than 15% compounds from the ICH Reference Compound List can be used for the Training Set. This permits an adequate number of compounds from the ICH Reference Compound List to be used as part of the Test Set for qualification purposes. Reserving ≥ 85% of compounds from the ICH Reference Compound List for the Test Set permits a sufficiently robust evaluation of the assay's predictivity.

The performance of the Training and Test sets should be evaluated separately and together and the results of each analysis presented. The performance summary should list the sensitivity, specificity, positive predictive value, and negative predictive value. If more than one assay is used, the performance of each assay should be provided separately in addition to the integrated assessment used for the predictive model. In the case of integration of more than one assay in the model, a clear description should be presented of how the integration of the individual assays is conducted to arrive at the integrated predictive model.

As part of the assay qualification and predictive model use, the category of compounds the assay can and cannot predict (e.g., a component of the applicability domain) should be defined from the following list of categories included in the ICH Compound Reference List (see Glossary): Channel modulator, DNA modifiers, Enzyme modulator, Hormone/steroids, Kinase modulator, Nucleoside modulator/nutrient blocker/central metabolite inhibitor, Receptor modulator, Oligonucleotide-based modulators, secondary messenger modulator, and Others. Additionally, human teratogens not detected *in vivo* by rat and/or rabbit should

also be evaluated to understand if the assay can detect them, even if the assay(s) intended use is to predict rat or rabbit outcomes. These results should be presented separately and the sponsor should justify whether or not and if so, how, to include these results in their predictivity assessment.

Demonstration of assay reproducibility should be assessed and can be accomplished by inclusion of at least one positive control and one negative control in either each assay run or interspersed over time between test compound runs. The sponsor should justify their approach to inclusion of positive and negative controls. The approach used to demonstrate assay reproducibility should be described in the information provided. Additionally, several of the compounds from the ICH Reference Compound List should be periodically reassessed and the data provided along with compounds being evaluated for therapeutic development.
The source of reagents, biologic materials, and compounds tested should be provided. Likewise, the source/reference of all *in vivo* exposure data used for compounds in the qualification data set should also be presented, except for those compounds in the ICH Reference Compound List since that would be the source (reference) information. Assays should be developed with the understanding there is an expectation that regulatory studies should generally be conducted in compliance with GLP.

The sponsor of the alternative assay should state whether the assay qualification has been previously submitted to any health authority in support of reproductive toxicity assessments and, if so, to which one(s).

11.3.4 *ICH Reference Compound List*

The ICH Reference Compound List (Table 9-6) is not intended to cover tailored approaches studying specific pharmaceutical targets or chemistry of structurally related analogs. For particular pharmaceuticals and contexts of use, justification for use of particular assays/assessments should be given (e.g., the Sponsor has *in vivo* information on other pharmaceuticals in the class). Table 9-7 provides examples of data records for including compounds in the ICH Reference Compound List for qualifying alternative assays.

Table 9-6. ICH Reference Compounds for Qualifying Alternative Assays

Category	Positive Controls	Negative Controls
Channel Modulator	Sotalol	Hydrochlorothiazide
	Almokalant	Chlorthalidone
	Diltiazem	
	Topiramate	
	Trimethadione	
	Phenytoin (Diphenylhydantoin)	
	Carbamazepine	
DNA Modifiers	Cyclophosphamide	

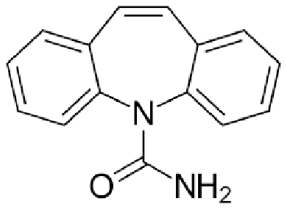
Category	Positive Controls	Negative Controls
	Busulfan	
	Cisplatin	
	Thiotepa	
Enzyme Modulator	Aspirin	
	Captopril	Saxagliptin
	Enalapril	Vildagliptin
	Methimazole (Thiamazole)	
Hormone/Steroid	Dexamethasone	Progesterone
	Fluticasone	
Kinase Modulator	Afatinib	
	Ceritinib	
	Dabrafenib	
	Dasatinib	
	Ibrutinib	
	Pazopanib	
	Tacrolimus	
	Imatinib	
Nucleoside Modulator/ Central metabolite inhibitor	Cytarabine	
	5-Fluorouracil	
	Hydroxyurea	
	Methotrexate	
	Ribavirin	
	Teriflunomide	
	Warfarin	
Other	Artesunate / amodiaquine	Amoxicillin
	Clarithromycin	Clindamycin
	Doxycycline	Cyclobenzaprine
	Fluconazole	Erythromycin
	Pomalidomide	Sulfasalazine
	Tafamidis	
	Telavancin	
	Thalidomide	
	Valproic acid	
Receptor Modulator		Cetirizine
	Bosentan	Cyproheptadine
	Clobazam	Doxylamine
	Fingolimod	Maraviroc
	Plerixafor	Metoclopramide

Category	Positive Controls	Negative Controls
	Sumatriptan	Nizatidine
Second Messenger Modulator	Theophylline	
Transcription Modulator	Acitretin	
	Isotretinoin (13-*cis*-retinoic acid)	
	Vismodegib	

Table 9-7. Examples of Data Records for Including Compounds in Reference List for Qualifying Alternative Assays

Carbamazepine

Proposed Class: Other
CAS No.: 298-46-4
Structure:



Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
250 mg/kg/day Fasted 200 mg/kg single PO dose: C_{max} = 32.7 µg/mL [3] (extrapolates to 41 µg/mL at 250 mg/kg) $AUC_{(0\text{-}24\ h)}$ = 32.8 mg•min/mL = 547 µg•h/mL (extrapolates to 684 µg•h/mL at 250 mg/kg)	400 mg/kg Fasted 200 mg/kg single PO dose: C_{max} = 32.7 µg/mL [3] (extrapolates to 65 µg/mL at 400 mg/kg) $AUC_{(0\text{-}24h)}$ = 32.8 mg•min/mL = 547 µg•h/mL (extrapolates to 1094 µg•h/mL at 400 mg/kg)	650 mg/kg [2] Maternal toxicity increased resorptions, increased skeletal and visceral abnormalities (4/119 offspring showed cleft palate, talipes, or anophthalmos) 600 mg/kg [4] increased resorptions, increased skeletal and visceral abnormalities (edema and kinked tails)	NOAEL was not identified	225 mg/kg/day Exposure data available for 80 mg/kg [5]: C_{max} = 10.4 µg/mL (extrapolates to 29 µg/mL at 225 mg/kg) $AUC_{(0\text{-}24h)}$ = 94.8 µg•h/mL (extrapolates to 267 µg•h/mL at 225 mg/kg)	Dosed 225 – 450 mg/kg [1] No malformations Decreased numbers of fetuses, increased resorptions in all groups Maternal toxicity at 450 mg/kg	Carbamazepine 10,11-epoxide metabolite present

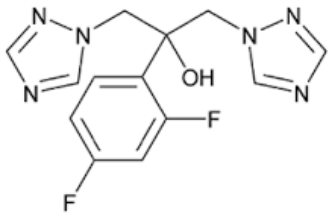
Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
		400 mg/kg [1, 2, 4] Reduced maternal weight gain; increased visceral abnormalities; abortions 250 mg/kg [1, 2] kinked ribs in 2/119 fetuses (not considered a TEFL finding)				

1. Published Pharm/tox review of NDA 16-608 (December 19, 1967), 16608/S-000 Part 02.
2. Equetro (carbamazepine) extended-release capsules Label, Carbamazepine FDA approval package, Label 021710/S-011, S-012.
3. Shi L, Dang XL, Liu XY, Wei HM, Yang MM, Zhang Y. Effect of *Sophora flavescens* on the pharmacokinetics of carbamazepine in rats. Arch Pharm Res. 2014;37:1617-23.
4. Vorhees CV, Acuff KD, Weisenburger WP, Minck DR. Teratogenicity of carbamazepine in rats. Teratology. 1990;41:311-17.
5. Koumaravelou K, Adithan C, Shashindran CH, Asad M, Abraham BK. Effect of honey on carbamazepine kinetics in rabbits. Indian J Exp Biol. 2002;40(5):560-3

1424

FLUCONAZOLE

Proposed Class: Other
CAS No.: 86386-73-4
Structure:



Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
50 mg/kg Following 20 mg/kg single oral dose: C_{max} [2] = 13.5 µg/mL (extrapolates to 34 µg/mL at 50 mg/kg) AUC [1] = 152 µg•hr/mL (extrapolates to 380 µg•h/mL at 50 mg/kg)	80 mg/kg 20 mg/kg single oral dose: C_{max} = 13.5 µg/mL [3] (extrapolates to 54 µg/mL at 80 mg/kg) AUC = 152 µg•h/mL [1] (extrapolates to 608 µg•h/mL at 80 mg/kg)	<u>80 –320 mg/kg [2, 3]</u> Increased embryolethality and fetal abnormalities (wavy ribs, cleft palate, and abnormal cranio-facial ossification) <u>≥25 mg/kg</u> Increases in fetal anatomical variants (supernumerary ribs, renal pelvis dilation) and delays in ossification were observed at 25 and 50 mg/kg and higher doses <u><10 mg/kg</u> No fetal effects	≤ 25 mg/kg 10 mg/kg single oral dose: C_{max} = 10.8 µg/mL (extrapolates to 27 µg/mL at 25 mg/kg)	75 mg/kg [2, 3] 10 mg/kg single oral dose: C_{max} = 10.8 µg/mL (extrapolates to 81 µg/mL at 75 mg/kg)	<u>75 mg/kg</u> Abortions	

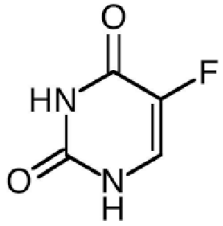
1. Humphrey MJ, Jevons S, Tarbit MH. Pharmacokinetic evaluation of UK-49,858, a metabolically stable triazole antifungal drug, in animals and humans. Antimicrob Agents Chemother. 1985 Nov;28(5):648-53.
2. Published Pharm/tox review of NDA 20322 (June 30, 1994), Part 01
3. Diflucan (Fluconazole) FDA Prescribing Information

5-FLUOROURACIL

Proposed Class: Nucleoside modulator

CAS No.: 51-21-8

Structure:



Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
15 mg/kg single dose IP (Kuwagata) 30 mg/kg , IP (Zhang) C_{max} = 7.74 µg/mL (extrapolates to 3.87 at 15 mg/kg) AUC = 11.66 µg•h/mL (extrapolates to 5.83 at 15 mg/kg)	12 – 37 mg/kg single IP dose on GD11 or 12 (Chaube) 17 mg/kg single dose IP on GD 9 (Kuwagata) 30 mg/kg , IP (Zhang) C_{max} = 7.74 µg/mL (extrapolates to 4.4 at 17 mg/kg) AUC = 11.66 µg•h/mL (extrapolates to 6.6 at 17 mg/kg)	12 – 37 mg/kg (Chaube) Cleft palate and deformed appendages ≥17 mg/kg (Kuwagata) micro-anophthalmos, craniofacial defects, hydrocephaly, brain hernia, edema; embryolethality at 30 mg/kg ≥15 mg/kg decreased fetal weight	Not determined, <40 mg/kg	40 mg/kg SC GD12 (480 mg/m^2) PK: 20 mg/kg IV (Kar) C_{max} = 427 nmol/mL =55 µg/mL (extrapolates to 110 at 40 mg/kg) AUC = 2535 nmol•min/mL = 5.5 µg·h/mL (extrapolates to 11 at 40 mg/kg)	40 mg/kg (DeSesso) 2/5 females died, with fetuses of surviving females exhibiting anomalies of the limb in 85% of cases	5FU is a pro-drug: thymidylate synthetase inhibitor is 5FdUMP MW = 130.077 g/mol

Chaube S, Murphy ML. The teratogenic effects of the recent drugs active in cancer chemotherapy. In: Advances in Teratology. ed. DHM Woolham. Academic Press, New York. 1968

DeSesso, JM, Scialli AR, Goeringer GC. Teratology. 1995;51:172 (abstract)

Kar R, Cohen RA, Terem TM, Nahabedian MY, Wile AG. Pharmacokinetics of 5-fluorouracil in rabbits in experimental regional chemotherapy. Cancer Res. 1986;46(9):4491-5.

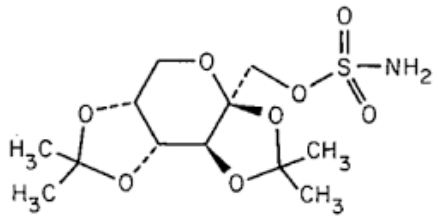
Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
Kuwagata M, Takashima H, Nagao T. A comparison of the *in vivo* and *in vitro* response of rat embryos to 5-fluorouracil. J Vet Med Sci. 1998;60(1):93-9. Zhang C, Li G, Wang Y, Cui F, Zhang J, Huang Q. Preparation and characterization of 5-fluorouracil-loaded PLLA-PEG/PEG nanoparticles by a novel supercritical CO2 technique. Int J Pharm. 2012;436(1-2):272-81.						

TOPIRAMATE

Proposed Class: Channel Modulator

CAS No.: 97240-79-4

Structure:



Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
100 mg/kg Exposure (FDA pharmtox review) 30 mg/kg, female SD, 8 doses C_{max} = 22.2 µg/mL (extrapolates to 74 at 100 mg/kg)	400 mg/kg Exposure (FDA pharmtox review) 30 mg/kg, female SD, 8 doses C_{max} = 22.2 µg/mL (extrapolates to 296 µg/mL at 400 mg/kg)	≥400 mg/kg (FDA pharmtox review and/or topamax label) limb defects (ectrodactyly, micromelia, and amelia) ≥20 mg/kg	10 mg/kg Exposure (FDA pharmtox review) 60 mg/kg, females, 14 doses C_{max} = 39 µg/mL (extrapolates to 6.5 at 10 mg/kg) AUC = 201	35 mg/kg Exposure (FDA pharmtox review) 60 mg/kg, females, 14 doses C_{max} = 39 µg/mL (extrapolates to 23 at 35 mg/kg) AUC = 201 µg•h/mL	≥35 mg/kg (FDA pharmtox review and/or topamax label) Embryofetal mortality increased at ≥35 mg/kg; Teratogenic effects (primarily rib/vertebral malformations) were observed at 120 mg/kg	In rats: maternal toxicity were seen at ≥400 mg/kg and maternal body weight gain was reduced at ≥100 mg/kg In rabbits: maternal toxicity (decreased body weight gain, clinical signs, and/or mortality)

Rat NOAEL Dose AUC C_{max}	Rat LOAEL Dose AUC C_{max}	Rat Findings	Rabbit NOAEL Dose AUC C_{max}	Rabbit LOAEL Dose AUC C_{max}	Rabbit Findings	Notes
AUC = 268 µg•h/mL (extrapolates to 893 at 100 mg/kg) In pregnant rats dosed w/ 200 mg/kg, at GD12-15, $C_{1.5h}$ = 97 µg/mL (extrapolates to 49 at 100)	AUC = 268 µg•h/mL (extrapolates to 3573 at 400 mg/kg) In pregnant rats dosed w/ 400 mg/kg, at GD12-15, $C_{1.5h}$ = 169 µg/mL	reduced fetal body weights and increased incidence of structural variations	µg•h/mL (extrapolates to 33.5 at 10 mg/kg)	(extrapolates to 117 at 35 mg/kg)		was seen at ≥35 mg/kg Rabbit LOAEL margins all <10
Topamax label (US): rat: oral doses of 20, 100, and 500 mg/kg or 0.2, 2.5, 30, and 400 mg/kg; rabbit: oral doses of 20, 60, and 180 mg/kg or 10, 35, and 120 mg/kg Published Pharm/tox review of NDA 20505/S000 (August 1, 1995)						

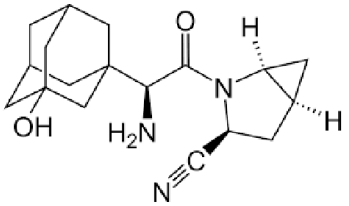
1441

SAXAGLIPTIN

Proposed Class: Enzyme modulator

CAS No.: 361442-04-8

Structure:



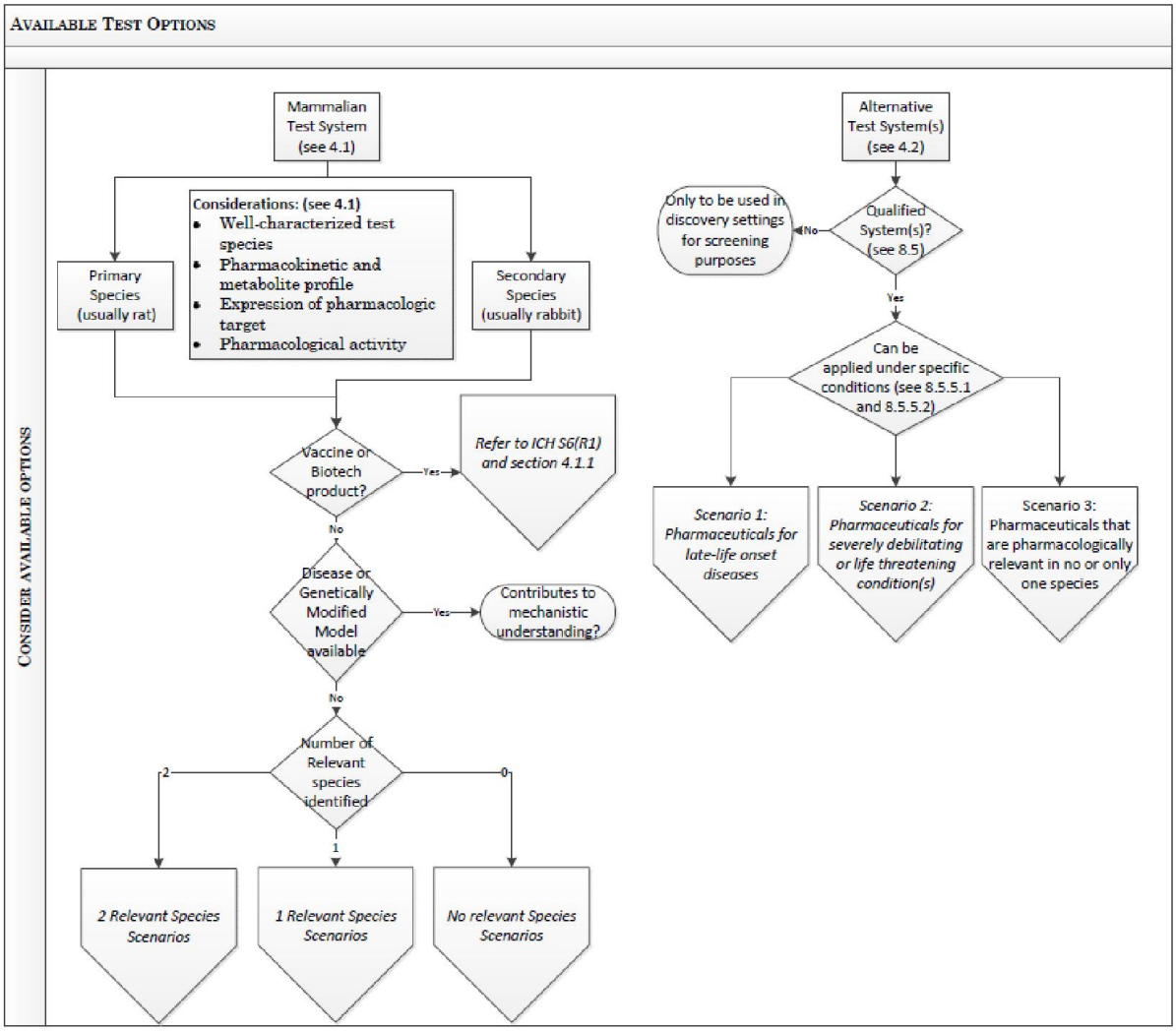
Rat NOAEL (Highest Dose Tested) Dose, AUC, C_{max}	Rat LOAEL	Rat Findings	Rabbit NOAEL (Highest Dose Tested) Dose, AUC, C_{max}	Rabbit LOAEL	Rabbit Findings	Notes
900 mg/kg C_{max} = 62 µg/mL AUC = 647 µg•h/mL	Not relevant	No malformations or embryofetal lethality noted. ≥240 mg/kg delayed ossification	200 mg/kg C_{max} = 34 µg/mL AUC = 111 µg•h/mL	Not relevant	No malformations or embryofetal lethality 200 mg/kg increased ossification	
Published FDA Pharm/tox review of NDA 022350/S000, Parts 2, 3, and 5 (March 3, 2009). Rat: oral dosages of 64, 240 and 900 mg/kg; rabbit: oral dosages of 8, 40 and 200 mg/kg						

11.3.5 ***Examples of EFD Testing Strategies***

This section describes optional integrated testing strategies that can be used to detect adverse effects on EFD. The use of a particular scenario needs to be justified.

In circumstances other than those described in 9.5.5.1 and 9.5.5.2 below and elsewhere in this guideline where use of alternative assays is proposed, positive results in alternative assays can also reduce mammalian *in vivo* testing. In contrast, negative results in alternative assays in most of these other circumstances would not be anticipated to reduce *in vivo* testing. See Figure 9-1.

Figure 9-1: Summary of Available Test Options



11.3.5.1 Scenarios applicable when there are at least 2 relevant mammalian species (crf. Species selection)

This section describes optional integrated testing strategies that can be used to detect adverse effects on embryo-fetal development. The use of a particular testing strategy should be justified.

a) Scenario 1: Pharmaceuticals for late-life onset diseases (Figure 9-2)

1. When a qualified alternative assay predicts TEFL in one species (e.g., rat) or is equivocal, an EFD assessment (e.g., pEFD, enhanced pEFD) in another species (e.g., rabbit) should be conducted to evaluate the multi-species risk and assess the finding *in vivo*.

    a. If TEFL is observed in the *in vivo* study (e.g., rabbit), the pharmaceutical will be considered to induce TEFL in multiple species based on the alternative assay and *in vivo* results.

    b. If no TEFL is detected in the *in vivo* study, a definitive EFD should be conducted in the species corresponding to the alternative assay to further assess the TEFL potential *in vivo*. If TEFL is observed in this definitive *in vivo* EFD study, the pharmaceutical will be considered positive in animal studies based on the positive alternative assay and *in vivo* for the same species. No further EFD studies are warranted, as a hazard has been identified and the risk assessment can be made based on the totality of the information. If no TEFL is observed in both *in vivo* EFD studies, the results from the alternative assay represent a false positive and the pharmaceutical will be considered not likely to induce TEFL, provided adequate exposure was achieved in the *in vivo* testing (e.g., exposures *in vivo* exceed the human exposure).

2. When an alternative assay predicts a negative outcome (i.e., no TEFL) in one species (e.g., rat), an EFD study in another species (e.g., rabbit) should be conducted to determine if the pharmaceutical is positive for TEFL *in vivo*.

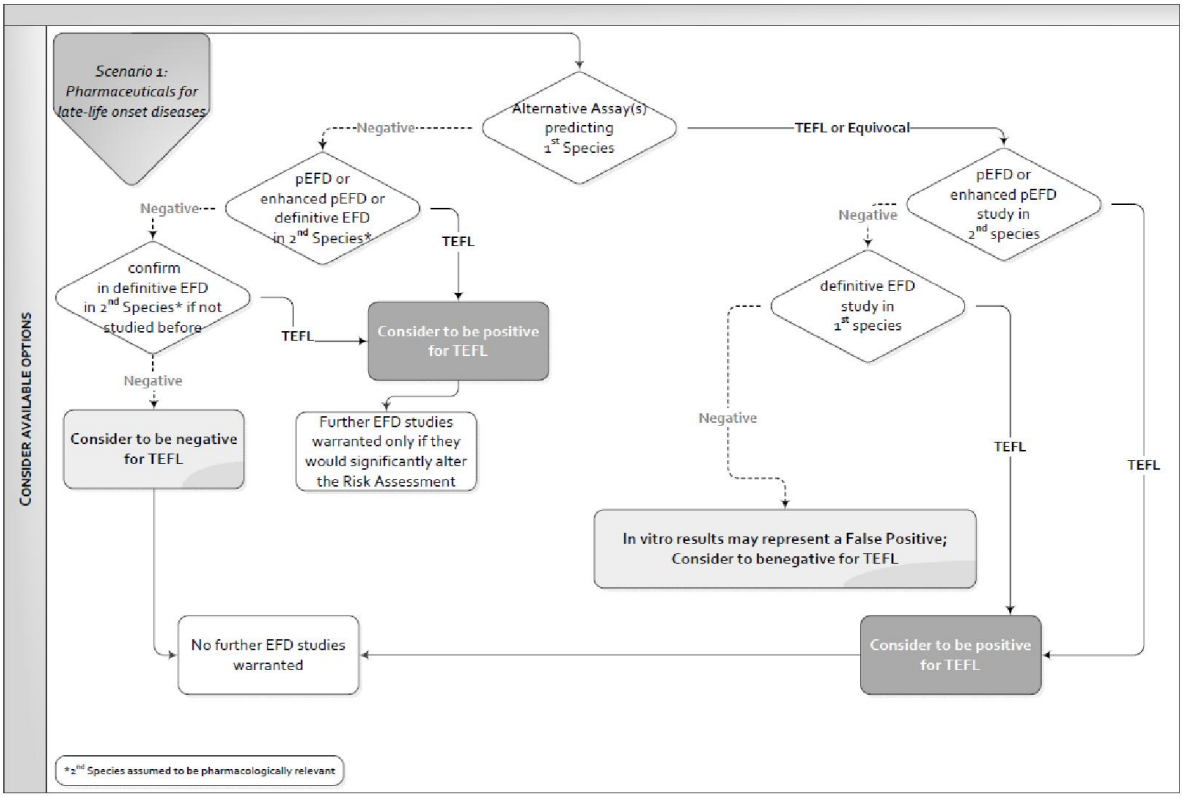
    a. If a TEFL outcome is observed in the second species EFD study, the pharmaceutical will be considered positive in animals. Further EFD studies would be warranted only if they would significantly alter the risk assessment (e.g., positive only at high multiples of the clinical exposure and thus another species could indicate a relevant risk at low exposures).

    b. If no TEFL is detected in the second species definitive EFD study, the pharmaceutical will be considered not likely to induce TEFL in animal studies (*in vitro* and *in vivo*) and no further EFD studies would be warranted.

For the scenarios above where a rat EFD study is not conducted, an additional opportunity to confirm *in vitro* positive outcomes is presented in either rat fertility or pre-and postnatal

development studies where exposure *in vivo* can further inform on developmental reproductive risk.

Figure 9-2: Scenario 1 Showing the Integrated Testing Strategies for EFD for Pharmaceuticals for Late-life Onset Diseases



b) Scenario 2: Pharmaceuticals for severely debilitating or life-threatening disease(s) (Figure 9-3)

Considering the risk/benefit for pharmaceuticals for severely debilitating or life threatening conditions compared to less severe chronic disease, the use of qualified alternative assay(s) contributes to and can be sufficient to assess relevant risk.

1. When a qualified alternative assay predicts TEFL in a species (e.g., rat) or is equivocal (or if a class effect has been identified) additional testing is not warranted (Flow Chart 2) unless the result is suspected to represent a false positive.

    a. If the Sponsor wants to demonstrate that results represent a false positive, definitive EFD studies should be conducted in two species to confirm absence of TEFL *in vivo*.

i. If no TEFL is observed in both species *in vivo*, results from the alternative *in vitro* assay represent a false positive and the pharmaceutical will be considered negative *in vivo* and this information will be used in the risk assessment.

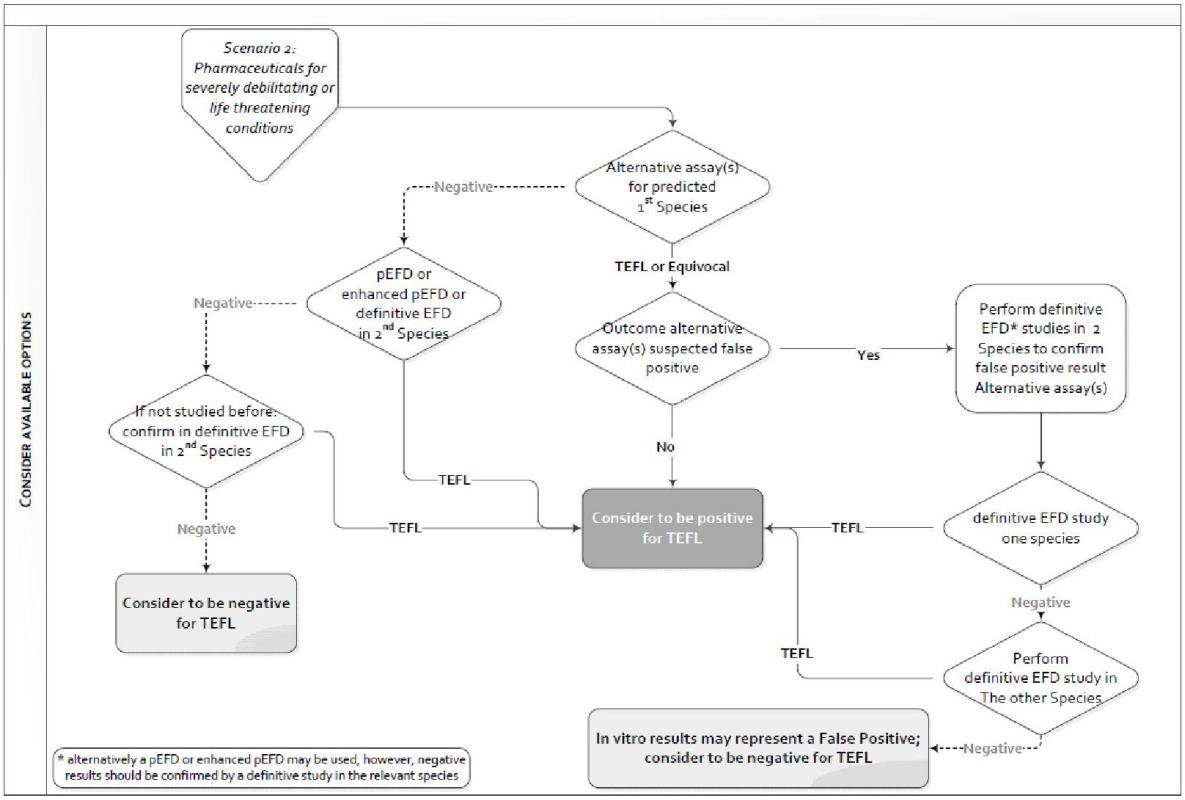
ii. If one or more of these *in vivo* studies has positive TEFL outcome, the pharmaceutical will be considered positive *in vivo* and this will be factored into the risk assessment.

2. If the alternative assay predicts a negative outcome (i.e., no TEFL), an EFD study in the other species (e.g., rabbit) should be conducted to determine if the pharmaceutical is positive *in vivo*.

   a. If a TEFL outcome is observed in the second species EFD study, the pharmaceutical will be considered positive in animals. Further EFD studies would be warranted only if they would significantly alter the risk assessment (e.g., positive only at high multiples of the clinical exposure and thus another species could indicate a relevant risk at low exposures).

   b. If no TEFL is observed in the second species definitive EFD study, the pharmaceutical will be considered negative in animals and no further EFD studies would be warranted.

Figure 9-3: Scenario 2 Showing the Integrated Testing Strategies for EFD for Pharmaceuticals for Severely Debilitating or Life Threatening Diseases



11.3.5.2 Scenarios applicable in case there is no or only 1 relevant mammalian species (crf. Species selection)

a) Scenario 3: Non-highly Targeted pharmaceuticals that are pharmacologically active in only one or no species

If there is evidence (e.g., mechanism of action, phenotypic data from genetically modified animals, class effects) that there will be an adverse effect on pregnancy outcome, these data can provide adequate information to communicate risk to reproduction and nonclinical *in vivo* studies are not warranted. Similar approaches are discussed in other guidelines (ICH S6(R1)(2) and ICH S9 (3)).

If the evidence is lacking, inconclusive or negative for TEFL effects, an EFD study in a single species should be conducted. If that study is positive for TEFL, an EFD study in a second species is not warranted provided the observations occurred at relevant margins of exposure and interpretation is not confounded by maternal toxicity.