09 April 2018
EMA/181369/2018

# Overview of comments received on ICH S5 (R3) guideline on reproductive toxicology: detection of toxicity to reproduction for human pharmaceuticals (EMA/CHMP/ICH/544278/1998)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

| Stakeholder no. | Name of organisation or individual |
|---|---|
| 1 | Julian French - Morphology Consulting Ltd, Biddulph, Stoke on Trent, ST8 6DQ, UK |
| 2 | International Council on Animal Protection in Pharmaceutical Programs (ICAPPP) |
| 3 | Gilead Sciences International Ltd. |
| 4 | PRA Health Sciences |
| 5 | Society of Toxicologic Pathology |
| 6 | Swissmedic |
| 7 | PMA |

Please note that comments will be sent to the **ICH S5 EWG** for consideration in the context of Step 3 of the ICH process.

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
**Telephone** +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
**Send a question via our website** www.ema.europa.eu/contact

# 1. General comments – overview

| Stakeholder no. | General comment (if any) |
|---|---|
| 1 | Thank you for this opportunity to look at the draft ICH S5(R3) guideline.<br><br>My comments are based on the document located at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/08/WC500233917.pdf (last accessed on 01 November 2017)<br><br>The premise for my comments on the preclinical EFD study (or phase) guidance contained within the draft ICH S5(R3) guideline is that these studies effectively represent the only opportunity to assess the potential of the test compound (pharmaceutical) to affect embryo/fetal development, prior to marketing. The complexity and detail of these studies belies the often repeated mantra that they are merely "screening studies".<br><br>The first line of paragraph 1198 – 1202 states that "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.)." However, the draft guideline goes on to "allow" 50/50 examination as an alternative. If this position is adopted, it will mean that the current status quo of establishments examining either 50% of the fetuses (destroying the remaining tissue without examination/rendering it unexaminable) or examining 100% of fetuses (by either fresh visceral dissection) will continue. Is it justifiable to permit such variation between establishments? If there are examination methods/regimes that allow 100% examination then these should be regarded as best practice and used consistently.<br><br>It has been argued that examination of fixed [rat] specimens is superior to examination following dissection of the fresh specimen because the techniques used are either non - destructive (e.g. head and body (Wilson) sections) or allow tissue to be re-examined (e.g. microdissection following Bouin's/Harrison's fluid fixation). However, use of fresh microdissection techniques for identification of soft tissue alterations in rabbit fetuses were regarded as preferable, in S6 (R1 and 2), to examination of fixed tissue. Examination of fresh tissue was, therefore, universally adopted and accepted as the "norm", with none of the "disadvantages" of fresh dissection being regarded as major issues.<br><br>Mathematically using the litter as the experimental unit to account for the "litter effect" (i.e. the propensity for fetuses of a given litter to exhibit similar responses to toxic injury) requires a determination of the percentage of embryos/fetuses within each litter that are affected. A grand mean should then be calculated from the individual litter means. Use of the litter proportion calculation is considered to be appropriate to analyse fetal malformation data. A common failing is to simply determine a percentage of litters with at least one malformed fetus, failing to apply correct litter based statistics. With this approach the number of malformed fetuses within each litter is not taken into account (e.g. a litter with 1 of 12 fetuses malformed is given the same weighting as a litter with 6 of 12 fetuses malformed). Therefore, variance amongst |

| Stakeholder no. | General comment (if any) |
|---|---|
| | litters cannot be determined. Initially determining the percentage of malformed fetuses within a litter, followed by calculating a litter grand mean, is the most appropriate way to analyse the fetal malformation data (Developmental and Reproductive Toxicology: A Practical Approach, Third Edition, Hood et al; 2012). However, the power of these calculations is reduced by failing to examine half of the viscera/half of the skeletons, potentially producing false negative results.<br><br>I believe that it should be regarded as ethically/morally indefensible not to gain as much information as possible (and practical) from each specimen. A lot of time, effort and money goes into dosing these pregnant animals. Therefore, shouldn't an effort be made to use all the material from them? A (repro) analogy to this situation might be examining the left testis or ovary only and discarding the right one. |
| 2 | ICAPPP welcomes the inclusion in this revised guideline of opportunities to avoid animal use as well as to use alternative assays as part of integrated testing strategies to reduce the number of animals used in reproductive toxicology studies.<br><br>In the interests of avoiding unnecessary use of animals and the principle of the 3Rs (which is not mentioned in the guideline), the structure and language of the guideline could be further improved. The current layout of the guideline does not prioritise methods to replace, reduce and/or refine animal use i.e. alternative assays, and combined study designs using fewer animals are presented towards the end of the document as 'options'. The language could also be improved to make it clear to readers that animal use should be minimised, without changing the outcome of the scientific discussions that have already been concluded in the guideline.<br><br>Our main concerns with the guideline as it stands are;<br><br>1. The implied use of NHPs as a recommended species,<br><br>2. The presentation of 'optional' strategies to reduce animal use,<br><br>3. The continued need for a second species,<br><br>4. Need for examples of commonly used alternative assays,<br><br>5. The qualification criteria for alternative assays,<br><br>6. Avoidance of duplicative animal testing<br><br>**1. The implied use of NHPs as a recommended species**<br><br>We are concerned that the guideline is not clear enough that testing in NHPs is non-routine and may only be applicable for biologicals.<br><br>There is repeated mention of NHP tests within the guideline even though they are acknowledged as a 'non-routine' testing species and come with a long list of limitations. It should be made clear that NHP tests should only be used in very rare circumstances. A description of their use should be restricted to section 4.2 on non-routine species.<br><br>Furthermore, the addition of the ePPND study is a concern as this may lead to increased NHP use. |

| Stakeholder no. | General comment (if any) |
|---|---|
| | The previous version of the guideline stated a major disadvantage with the use of NHPs; '[…] numbers too low for detection of risk. They are best used when the objective of the study is to characterise a relatively certain reproductive toxicant, rather than detect a hazard'. This point, which is also evident from the recent scientific literature (Chellman et al., 2009, Faqi, 2012), challenges the value of the NHP test and yet has been inexplicably omitted from the new draft guideline. There are also several other limitations that impact the suitability of NHPs for reproductive toxicity testing e.g. low pregnancy rate, impracticality of conducting fertility of F1 generation studies, impossibility of determining age of onset of puberty, high abortion rate, stillbirths and low statistical power due to limited numbers (Faqi. 2012).<br><br>Furthermore, a 2013 analysis conducted to determine the sensitivity of male reproductive toxicity endpoints in NHPs, found that the addition of triggered non-routine endpoints in NHP studies showed 'poor power (less than 80%) to detect a change from control with a group size of 3' and that 'testosterone and sperm count, 2 of the more commonly suggested endpoints for male reproductive effects, were among the lowest powered endpoints' i.e. 'less than 20% even when the group size was doubled to 6 NHPs' (Cappon et al., 2013). Another more general study found that the use of NHPs in preclinical drug tests are just as unpredictive of human effects as those using any of the other four commonly used species (Bailey et al., 2015).<br><br>Although the use of NHPs appears to be recommended mainly for testing biopharmaceuticals (e.g. mAbs) where they are considered to be the only relevant species, the guideline should still discourage their use (due to the many known limitations as well as animal welfare and ethical issues) and emphasise that they should only be considered in very rare circumstances and as a last resort.<br><br>Rather than developing new and modified testing approaches (e.g. enhanced PPND in NHPs) there is an urgent need to evaluate the need for primate reproductive toxicity studies.<br><br>**2. The presentation of 'optional' strategies to reduce animal use**<br><br>While there are some testing strategies that incorporate the use of alternative assays to potentially reduce animal use given in the guideline, their presentation is at the end of the document and as 'options'. Starting with the introduction section, the guideline should clearly state that the use of alternatives is encouraged to avoid and minimise animal use. This language should continue throughout the guideline and be included wherever a specific call for animal data is made. This addition would allow for flexibility as the science evolves, without changing the outcome of the scientific discussions already included in the guideline.<br><br>Further, the use of alternative testing strategies is entirely absent from Figure 3.1 and there are several points throughout the general scheme where alternative testing strategies are relevant.<br><br>The scenarios listed in which alternative assays can be used are very limited, restricted to severely debilitating or life-threatening diseases and late-life onset diseases. The guideline also provides a list of combined study designs that use fewer animals than the routine approach, but these are also just tacked on as 'options' when they should |

| Stakeholder no. | General comment (if any) |
|---|---|
| | be prioritised in accordance with the 3Rs principles.<br><br>With the goals of human-relevance and minimising animal testing in mind, the guideline should be restructured to present a tiered testing strategy where all existing data and data from non-animal alternatives are considered before new animal tests are discussed. The testing strategies that can waive the need for the second species should be presented as a 'standard consideration' rather than an 'option'.<br><br>For example, a recent Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis stated; 'many alternative systems show high sensitivity (true positive rate), so a positive result in an alternative system may be highly predictive of teratogenicity in in-vivo animal studies. So, while a negative result for teratogenicity in an alternative test may not be completely reassuring of lack of teratogenicity in the human, a positive teratogenicity finding in an alternative test may make further animal studies redundant' (Barrow, 2016). While this is presented as a scenario for pharmaceuticals for older patients or very severe diseases, if the alternative is validated we see no scientific reason why its use should be so restricted.<br><br>The concept of 'weight of evidence' and Integrated Testing and Assessment is now common place within the chemical sector and we are surprised to not see more specific reference to the integration of evidence – not just from in vitro assays – but also from in silico techniques, such as read across, physico/chemical information, etc., to determine whether to proceed with traditional in vivo animal testing. The value of computational toxicology is well recognised (Ford, 2017) and these options should be incorporated into the general testing strategy.<br><br>Similarly, when animal tests are deemed necessary, all of the different study designs (e.g. combination of tests) and waiving options should be considered before the traditional tests are used. These recommendations should appear before discussions of the routine animal tests in order to encourage their implementation.<br><br>**3. Need for a second species**<br><br>While the guideline does provide some examples where testing in a second species can be waived e.g. when 'enhanced' animal studies, existing data or alternatives are used, more scenarios where testing in two species is not considered scientifically necessary should be provided.<br><br>According to a 2012 FDA workshop, participants agreed that the current requirement of a second species test is a concern and suggested a weight-of-evidence approach along with several situations where the second species test could be avoided e.g. for antimicrobials, 'a combination of in vitro assays and testing in one in vivo species might be more informative' (FDA, 2012).<br><br>We strongly encourage ICH members to conduct a proper analysis of the added value of tests on a second species (as has been done in recent years for chemicals). To have not done this prior to the revision of this guideline is a real missed opportunity to maximise efficient use of animals.<br><br>**4. Need for examples of commonly used alternative assays**<br><br>It is not clear why the guideline does not contain a single example of an alternative |

| Stakeholder no. | General comment (if any) |
|---|---|
| | assay for reproductive toxicity testing. According to the S5 ICH work plan (http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S5/S5_R3__EWG_Work_Plan_08Aug2017.pdf) as well as discussions in the ICH public meetings, specific alternative methods were discussed in detail (e.g. WEC, EST and zebrafish tests) throughout the drafting process and it was suggested that some of them would be described in an appendix.<br><br>It would be beneficial to include some examples of the most promising alternative assays that are currently proving useful to drug developers. Although most of the methods are not accepted as standalone replacement purposes at this stage, they are still valuable for reducing the number of animals used in drug development, i.e. as preliminary screens or to waive the second species. It seems appropriate that discussion of the potential to use alternative methods, which is mentioned in the testing strategies to waive the need for the second species at the end of the document, would include information on what type of alternatives might be considered. Additionally, it should be stated that the examples provided are not an inclusive list of acceptable alternative test methods.<br><br>The SWOT analysis (Barrow, 2016) describes the embryonic stem cell test (EST) and the zebrafish embryo-larval assay as examples of tests that could be included in the guideline. A review article states that these methods are commonly used in practice 'to identify and de-risk potential pharmaceuticals that have a teratogenic or reproductive toxicity liability' (Brannen et al., 2016).<br><br>The EST was fully validated by ECVAM in 2002 and shown to have an overall accuracy of 78% with 20 substances (Genschow et al., 2004). In 2008, Pfizer concluded that the overall performance of the EST was generally good with an accuracy of 75% for 63 chemicals, and that they were confident to use the assay to aid compound-development decisions (Paquette et al., 2008). Improvements have been made recently to increase the applicability and speed of the assay and to account for metabolism.<br><br>Zebrafish have also shown to be good predictive models for screening drug discovery compounds. 'Several studies at various laboratories have found that concordance between mammalian and zebrafish assays is strong and may be higher than 80% with fairly balanced false positive and negative rates' and 'an analysis with data from 214 ToxCast chemicals indicated that the agreement between outcomes in zebrafish and rats or between zebrafish and rabbits was almost as high as the agreement in study outcome between rats and rabbits' (Brannen et al., 2016).<br><br>The guideline only seems to focus on the use of alternatives to address embryo-fetal developmental (EFD) risk and does not provide the option to use alternatives for any of the other reproductive stages. According to the EPA, 'numerous in vitro tests are available and under development to measure or detect chemically induced changes in various aspects of both male and female reproductive systems. These include in vitro fertilisation using isolated gametes, whole organ (e.g. testis, ovary) perfusion, culture of isolated cells from the reproductive organs (e.g. Leydig cells, Sertoli cells, granulosa cells, oviductal or epididymal epithelium), co-culture of several populations of isolated cells, ovaries, quarter testes, seminiferous tubule segments, various receptor binding |

| Stakeholder no. | General comment (if any) |
|---|---|
| | assays on reproductive cells and transfected cell lines, and others' (https://www.epa.gov/sites/production/files/2014-11/documents/guidelines_repro_toxicity.pdf).<br><br>It is inappropriate to limit the use and consideration of alternatives to just EFD strategies and we would therefore like to see some opportunities to use non-animal methods being presented for the other reproductive stages.<br><br>For example, in vitro human spermatogenic models, including sophisticated 3D human testis organoids and in vitro models of the blood-testis barrier, can be used as an alternative strategy to address possible effects on male fertility (Easley et al. 2015, Pendergraft et al., 2017, Siemann et al., 2017). There are also methods that are relevant for female fertility testing, such as EpiVaginal – a commercially available model of the human vagina – and the use of microfluidic platforms that have demonstrated great potential for use in drug discovery and toxicology studies (Ayehunie et al, 2015, Xiao et al., 2017).<br><br>Additionally, sponsors should be encouraged to submit data from human-based in vitro and in silico approaches, even prior to qualification, so that reviewers become more accustomed to seeing and interpreting them. Frequent use of specific methods would help identify which methods are ripe for qualification and broader use.<br><br>**5. Qualification criteria for alternative assays**<br><br>According to the guideline, each alternative assay must be tested against a set of at least 45 ICH reference compounds and demonstrate a sensitivity of at least 80% along with a list of other qualification criteria. While we support the validation of alternative methods, it is important to remember that the traditional animal tests were never formally validated or required to demonstrate the same level of sensitivity with respect to human effects.<br><br>For example, a number of studies have shown that reproductive toxicity tests in animals are only able to detect about 60% of known human reproductive toxicants (Bailey at al., 2005). A 2005 review that examined decades of animal-based teratology data revealed 'significant variability in positive and negative predictability, and high rate of false-positive, false-negatives and equivocal outcomes across twelve species' (Bailey et al., 2005). A recent survey found that out of 12 pharmaceutical candidates submitted by 5 companies, only 1 compound that had shown male reproductive toxicity in experimental animals also demonstrated toxicity in men. The study found that the 'predicted value of experimental findings for adverse effects in men was 8 to 14%' and that 'non-clinical studies appeared to over-predict reproductive toxicity in men'. It concluded that there is 'poor correlation between the male reproductive toxicity produced in non-clinical animal studies and results from human clinical trials' (Scialli et al., 2017).<br><br>We are pleased to see clear information on the criteria required to validate (qualify) an alternative assay. However, we are concerned that alternatives experts such as ECVAM and ICCVAM, who are involved in the validation of alternative assays for reproductive toxicity, do not appear to have been involved in the creation of this guideline. These organisations could provide valuable contribution to the appropriate recommendation |

| Stakeholder no. | General comment (if any) |
|---|---|
| | of alternative assays and help with the development of testing strategies.<br><br>**6. Avoidance of duplicative animal testing**<br><br>According to the guideline, EFD studies are *designed '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)'*. While PPND studies are designed to '*detect adverse effects following exposure of the mother from implantation through weaning on the pregnant or lactating female and development of the offspring […] (Stages C to F)'*.<br><br>It is unclear why two separate studies are warranted when there appears to be an overlap in the stages that are covered i.e. the PPND already investigates Stage C, which is the focus of the EFD studies. These tests should be combined in the interest of avoiding duplicative animal testing.<br><br>It is also unclear under what circumstances a pEFD study would be required. According to table 9-5, it is virtually the same as the definitive EFD study except it uses fewer animals and does not have GLP status. Is there a risk that results from the pEFD studies would not be seen as credible given their lack of GLP status and would therefore likely be followed up by a definitive study anyway? If so, then it would be better to omit the preliminary study in the interest of ensuring that fewer animals are used overall. |
| 2 | **References:**<br><br>Ayehunie S. et al. (2015) Characterization of a Hormone-Responsive Organotypic Human Vaginal Tissue Model: Morphologic and Immunologic Effects. Reprod Sci. 22(8):980-90.<br><br>Bailey, et al. (2005). The future of teratology research is in vitro. Biogenic Amines 19, 97-145<br><br>Barrow. (2016). Revision of the ICH guideline on detection of toxicity to reproduction for medicinal products: SWOT analysis. Reproductive Toxicology, 64: 57-63.<br><br>Brannen et al. (2016). Alternative models of developmental and reproductive toxicity in pharmaceutical risk assessment and the 3Rs. ILAR Journal, 57(2): 144-156.<br><br>Cappon et al. (2013). Sensitivity of male reproductive endpoints in nonhuman primate toxicity studies: a statistical power analysis. Reproductive Toxicology, 41: 67-72.<br><br>Chellman et al. (2009). Developmental and reproductive toxicology studies in nonhuman primates. Developmental and Reproductive Toxicology, 86(5): 446-462.<br><br>Easley et al. (2015). Assessing reproductive toxicity of two environmental toxicants with a novel in vitro human spermatogenic model. Stem Cell Research, 14: 347-355.<br><br>Faqi. (2012). A critical evaluation of developmental and reproductive toxicology in nonhuman primates. Systems Biology in Reproductive Medicine, 58(1).<br><br>FDA. (2012). Conference Report: Reproductive and developmental toxicity testing: from in vivo to in vitro. ALTEX, 29: 333-339. |

| Stakeholder no. | General comment (if any) |
|---|---|
| | Ford K.A. (2016). Refinement, Reduction, and Replacement of Animal Toxicity Tests by Computational Methods. ILAR J. 2016 Dec;57(2):226-233<br><br>Genschow et al. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim. 2004 Sep;32(3):209-44.<br><br>Paquette et al. (2008). Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. Birth Defects Res. B Dev. Repro. Toxicol. 83, 104-111.<br><br>Pendergraft S.S., et al. (2017) Three-dimensional testicular organoid: a novel tool for the study of human spermatogenesis and gonadotoxicity in vitro. Biol Reprod. 96(3):720-732.<br><br>Scialli et al. (2017). Predictivity of nonclinical male reproductive findings for human effects. Wiley Periodicals, doi: 10.1002/bdr2.1102.<br><br>Siemann D.N. et al. (2017) Zika Virus Infects Human Sertoli Cells and Modulates the Integrity of the In Vitro Blood-Testis Barrier Model. Oct 27;91(22)<br><br>Xiao S. et al. (2017) A microfluidic culture model of the human reproductive tract and 28-day menstrual cycle. Nat Commun. 2017 Mar 28;8:14584. |
| 3 | Guidance provided on dose selection and risk assessment is appreciated along with options to reduce or defer animal use<br><br>Guidance is complex and difficult to follow at times with approaches for small molecules intermingled with options for alternative assays and for cases when no PD suitable species exist, or highly targeted molecules are being assessed.<br><br>Suggest laying out the potential approaches for small molecules first (types of studies [individual or combined], typical species, selection of doses, options for deferring studies) then address how to modify approach outlined for small molecules when dealing with biopharmaceuticals or highly targeted molecules (ie., one vs two species, evaluation in surrogates, genetically modified, or disease models, etc.) and other exceptions (vaccines).<br><br>The use of surrogate molecules, genetically modified animals and disease models appears to be emphasized throughout document.<br><br>Guidance appears to be advocating the development of surrogate molecules, genetically modified animal or disease model when traditional approach for reproductive studies is not feasible (no PD active species for highly targeted molecule; ie, monoclonal antibody). However, these approaches may not be very practical and relevance of findings to human risk is likely to be unknown.<br><br>Suggest consolidate guidance on using these approaches into a single section. Add that the use of these approaches will only be used in cases where the more traditional approach does not fully assess the potential human risk and should be discussed with appropriate Regulatory Agency. |
| 4 | The restriction of the scope described in ll 82-89 is useful when considering the |

| Stakeholder no. | General comment (if any) |
|---|---|
| | application of the Guideline. However, a clear need exists for guidance regarding cellular therapies, gene therapies and tissue-engineered products. |
| 6 | Line 79 (see Section 9.4.3.3) – section does not exist<br><br>Line 154 (see Section 9.4.2.1) – section does not exist<br><br>Line 162 (see Section 9.4) – section does not exist<br><br>Line 213 (see Section 9.5.5) – section does not exist<br><br>Line 219-221: off target effects on EFD or data from a repeat dose toxicity study? In contradiction with 230-233?<br><br>Line 243 (see Section 9.5.5) – section does not exist<br><br>Line 257 (see Section 9.5.5) – section does not exist<br><br>Line 267 (see Section 9.5.2) – section does not exist<br><br>Try to shorten, avoid redundancies |
| 7 | **Introductory comments:**<br><br>Several aspects in the present ICH S5 (R3) draft version are welcomed. This includes aspects related to the current standard testing in place, and the risk assessment process.<br><br>However, there are some general aspects which we find problematic and not in line with our general understanding of what is preferable to include in an ICH guideline. These are outlined further below.<br><br>Regarding structure and readability, the document has improved since an earlier version we have seen. However, it is still challenging to get an understanding of the key messages that are intended to be conveyed, as well as of key details. This is partly due to some remaining repetition throughout the document. Another reason is a lot of details in the Annexes which not always is easy to get a grip on and not always seems to be in full line with the body of the guideline. For instance, the when and how of 1 to 2 species testing scenarios and alternative testing scenarios would be helped by a better integration and cross-referencing.<br><br>We would therefore welcome a more streamlined document. One attractive option would be to make this into two separate (sub) guidelines, e.g. in line with ICH S7. One document could focus on refinement of the current standard testing as well as risk assessment, while all aspects related to alternative testing could be develop within a separate document. That would partly make the individual document(s) more accessible and also make it easier to revise those aspects when more knowledge becomes available (especially as DART alternative testing research is a field in change).<br><br>**General issues**<br><br>Within this document, proposals for testing options are made for which scientific data or sufficient experience are lacking in support of these being reasonable and adequately reassuring for assessing reproductive hazard as intended. This includes the |

| Stakeholder no. | General comment (if any) |
|---|---|
| | introduction of the pEFD, and the enhanced eEFD, as well as the options for alternative testing. In our view, this is in contrast to the general way ICH guidelines are built; where recommendations made are based on data / experience which have demonstrated that the recommendations made are fit for purpose. We are therefore questioning putting so much emphasis on these different options in the draft document.<br><br>**Specific issues - step 2b**<br><br>**Main points**<br><br>1. **Dissenting view**. The idea of preliminary EFD (pEFD) and enhanced pEFD (epEFD) studies is introduced (e.g. section 3.3.3.) in a manner that makes them separate from traditional dose-range finding studies. According to Table 9-5, the experimental group size of pEFD is 6 dams, for epEFD it is 8+ dams while the traditional minimum (in definitive studies) is 16 and more often closer to 20+ dams. It is questionable that this would overall reduce the number of animals used, since there still is a need for two definitive studies for marketing approval, and when taking into account the number of animals used for ex-vivo or in-vivo alternative tests or alternative tests that require serum from animals etc.. Furthermore, no convincing scientific arguments have been provided that the pEFD or epEFD designs would actually in most cases allow one to draw conclusions about human developmental hazard based on these studies only. Optimally, one would want to see that a randomized reduction of number of dams (to 6 or 8 dams) in a data set of N substances would reasonably often give the same NOAEL and LOAEL as generated with the full set of dams and that there would be no great loss in the doses where malformations are identified.<br>2. **Issue of feasibility**: Regarding the use of reference compounds in alternative assays. The present document requires "historical background data" from at least 45 compounds in order to validate the assay. This strategy is based on having a universal approach (i.e. one assay should cover as many substance classes as possible). The purpose of a universal reference set is that "all classes should be tested" (page 4, row 1241). On page 45 (rows 1317-1321), it is recommended that a 'robust' predictivity and sensitivity can be estimated based on that no more than 15% of the reference compounds (plus possible additional "self-choice" substances) should be used for the training set, leaving ≥85% for the test set. It is also noted that the ICH reference list is not supposed to be useful for chemically similar substances (page 6, row 1355). Of course, ultimately, as long as the applicant proves that the approach works, this should be enough. Also, on the plus side, the use of a universal reference substance list where one a priori knows which substances generate a teratogenic effect in animals removes the need for conducting additional in-vivo studies in order to classify the substances.<br>That being said, we have some concerns about whether the universal approach is realistic and has support from the predictive toxicology literature.<br>a. The premise is that the validation should cover as big a chemical space as possible. But is this a reasonable goal both from a biological and a |

| Stakeholder no. | General comment (if any) |
|---|---|
| | regulatory DART context? It is true that the reference substance set used to develop and train a predictive test system and model is generally not sufficient to adequately gage the predictivity of a test system (this is seen when using new batteries of test substances for the same test systems and has also been noted in the QSAR field which generally is based on less complex endpoints). But does this addition of additional nominally non-similar chemical substances up to n=45 (with at least 2-3 substances per class) provide a good test set? How likely is it that any given alternative test system for such a complex outcome remains even in the variability of its responses across chemical classes (i.e. remains with a good predictivity despite an increasing inclusion of different chemical classes)? Is it not far more likely that some alternative test systems are better than others (e.g. in some cases, EST cells are better than Zebrafish and vice versa) for certain chemical classes and that compulsively introducing a broad chemical space beyond the class of interest will inadvertently reduce the usefulness of the methods/increase uncertainty. Academic experience for larger independent alternative assay studies (i.e. based on different test substance batteries) often do not get similar levels of prediction and sensitivity (sometimes more dissimilar than similar), indicating that the composition and homogeneity of the reference substance batteries matters for the assessment of a given new substance in a given test system or set of test systems. This might even give a situation where the development of one or more potentially useful alternative test systems based on a chemical space of pharmaceutical interest would lead to consistently bad/diluted predictivity due to always subsequently adding the same greater chemical variation of the remaining universal list (the "≥85%"). Furthermore, the companies would not be stimulated to send in their data for assessment if the regulatory agencies are very likely to decide that the predictivity is too low.<br>b. An option would be to build on the recognition of the prediction/sensitivity differences between training and independent/test substance batteries but to attempt to reduce the chemical space – i.e. to focus on fewer "universal" positive and negative reference control substances (which increases variation) and more on a strategy where a combination of 2-3 assays (ITS approach) is validated for 1-3 similar chemical classes (with n much greater than the 2-3 positive and negative compounds per class in the universal approach). Of course, this option assumes that there are both positive and negative control substances within each chemical class. The applicant would also have to argue for the relevance of the chemical class representatives included (e.g. number of compounds, reasons they should belong to the same class, in-vivo correlation etc.).This approach could provide a long term approach where information is collected on about which chemical classes and reasonably validated alternative test systems are most compatible. |

| Stakeholder no. | General comment (if any) |
|---|---|
| | c. Finally, some considerations on how or even if the suggested approaches will stimulate the usage of alternative testing and knowledge generation and/or application. The universal approach may be more likely to inhibit than stimulate the use/development of alternative test methods for regulatory assessment (the required logistics for the setup for validation runs of 45 reference substances of various properties in different sets of test systems is likely to be unpopular) if the validity status is likely to be low for something that in the end still needs in-vivo testing. The alternative of the above suggested class approach is also not without logistical problems. The companies are possibly also more likely to have a greater range of members of the same nominal chemical class within their drug development/lead optimization pipeline which they are interested in testing both in-vitro and in-vivo. On the minus side, for completely novel substances, i.e. not part of a list of established teratogens, additional in-vivo studies in order to set the teratogen status of the class members would become necessary.<br>3. **Dissenting view**. The document states in rows 856-859 (page 25) that "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". This notion that prenatal weight reduction is a "minor manifestation"/"minimal concern from a risk assessment perspective" is then compounded in rows 862-864 where the level of adversity in fetal growth retardation is made equal to transient inhibition of spermatogenesis. While there is a tradition and pragmatic logic in toxicological pathology to put less emphasis on adult organ weight changes in the absence of histopathological signs, this approach cannot simply be applied to developmental processes. There is an extensive amount of literature that indicates that restricted growth rate and reduced birth weight can have far reaching consequences into juvenile and adult life for both animal models and humans (e.g. increased chance for still birth, poor neurodevelopmental outcome, and increased likelihood for metabolic disorders). Exposures where there was a clear growth reduction at doses below the maternal LOAEL should be of clear interest in a risk assessment.<br><br>Regarding alternative assays and pharmacokinetics/toxicokinetics. It can be noted that some cell and ex-vivo tests (e.g. Whole Embryo Culture) use serum albumin in the culture medium while other do not. Other tests such as zebrafish tests only use water. Any kinetic in-vitro to in-vivo extrapolations should accommodate for such factors (comment in the appendix to the guideline?) |

## 2. Specific comments on text

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 12-14 | 2 | **Comments:**<br><br>The original title has been accidentally retained. |
| 44-55 | 6 | **Comments:**<br><br>You may want to relate these stages to stages/definitions as commonly used in labels, guidelines incl. this guideline, i.e. fertility, early embryonic development (FEED), pre-and postnatal development (PPND). |
| 78-81 | 6 | **Comments:**<br><br>duplication<br><br>**Proposed change:**<br><br>Reductions in animal use can also be achieved by deferring or replacing definitive EFD studies (see Section 9.4.3.3) until later in pharmaceutical development (see below). |
| 94-95 | 6 | **Comments:**<br><br>Sentence "However, high quality scientific standards should be applied, with data collection records readily available" is redundant and/or should be combined with previous arguments<br><br>**Proposed change:**<br><br>shorten whole paragraph. |
| 96 | 2 | **Comments:**<br><br>It should be clearly stated at the beginning of the guideline that the use of alternatives is encouraged to avoid and minimise animal use in accordance with the 3Rs principles. The proposed change should be added to the end of the first paragraph in the introduction.<br><br>**Proposed change:**<br><br>In accordance with the 3Rs (replacement, reduction and refinement of animal use), data from qualified alternative approaches, such as in vitro and in silico approaches, should be considered before animal tests are conducted to assess nonclinical reproductive toxicity. |
| 97-99 | 2 | *'To asses 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'.*<br><br>**Comments:**<br><br>The request for adult animals explicitly calls for animal data and sets the stage for the rest of the document to prioritise animal tests. It should be reworded to provide flexibility for evolving science. What is important is not |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | "exposure of adult animals" per se but the impact on stages of development.<br><br>**Proposed change:**<br><br>'To asses a human pharmaceutical's effects on reproduction and development, the information should generally include **tests that address** ~~exposure of adult animals and~~ the impact on all stages of development from conception to sexual maturity'. |
| 99-100 | 5 | **Proposed change:**<br><br>Suggest bold or italicized text for emphasis for the sentence "No guideline can provide sufficient information to cover all possible cases, and flexibility in testing strategy is warranted." |
| 101-107 | 2 | '*Key factors to consider when developing an overall integrated 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, and similarity to other compounds in structure or activity and selection of specific studies, test species/test system and dose levels'.*<br><br>**Comments:**<br><br>An extra bullet point should be included in between the consideration of existing data and the selection of animal studies that mentions the use of non-animal alternatives.<br><br>**Proposed change:**<br><br>Add bullet point on non-animal alternatives (e.g. in vitro, in silico). |
| 104-105 | 6 | **Comments:**<br><br>please clarify this example since men can father also when elderly. Or: Why are these stages relevant for pharmaceuticals for males only at any age? Please define elderly male. |
| 106 | 7 | **Comments:**<br><br>*similarity to other compounds in structure* should be taken into account in an overall strategy. We are a bit hesitant raising this as a specific point for developing a testing strategy, as we are not convinced this being a key factor to consider when developing a testing strategy for a specific compound. |
| 108-112 | 7 | **Proposed change:**<br><br>Row can be taken out due to repetition |
| 109 | 6 | **Comments:**<br><br>Short term treatment can occur in men, WOCBP or pregnant women; how would a controlled setting reduce a reproductive risk? |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 110 | 2 | *'[…] to perform only the those studies […].'*<br><br>**Comments:**<br><br>The word 'the' is duplicative of the word 'those'.<br><br>**Proposed change:**<br><br>'[…] to perform only those studies […].' |
| 111 | 4 | **Comments:**<br><br>typo …ito perform only ~~the~~ those studies… |
| 113-116 | 3 | **Comments:**<br><br>Appears to omit stage A (premating to conception in F0 generation).<br><br>**Proposed change:**<br><br>(i.e., from ~~conception~~ premating in one generation through conception in the following generation) |
| 115-116 | 3 | **Comments:**<br><br>The definition of GD 0 seems out of place here since this term is not used in associated text (text only references conception.)<br><br>**Proposed change:**<br><br>Eliminate sentence here or clarify how this relates to conception. |
| 117-121 | 6 | **Comments:**<br><br>Unclear, even the examples given for when testing can be warranted; what could be off-target effects when we know that the expected pharmacologic effects on reproductive endpoints are non-adverse? |
| 127 | 6 | Examples of such modifications? |
| 129-131 | 3 | **Comments:**<br><br>Clarify that the risk to all stages of development should be assessed.<br><br>**Proposed change:**<br><br>The stages covered in individual studies are left to the discretion of the Sponsor; however, the risk to all stages should be assessed unless not relevant to the intended patient population. The timing of studies within… |
| 142-146 | 3 | **Comments:**<br><br>Extraneous information that would be better suited elsewhere in the document.<br><br>**Proposed change:**<br><br>Consider deleting |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 144 | 6 | **Comments:**<br><br>Reference is made to timing schedule in ICH S6(R1). Should that be applied also for non biotechnology-derived pharmaceuticals, for instance small molecules?<br><br>**Proposed change:**<br><br>add clarification |
| 146-157 | 6 | **Comments:**<br><br>Is section 3.1.5 needed?<br><br>**Proposed change:**<br><br>Information should be added where appropriate and this section should be deleted. |
| 147 | 7 | **Comments:**<br><br>This sentence is nearly read as a key aim of a testing strategy is to reduce the use of animals, while the primary aim is to gather adequate data for human risk assessment. We suggest that the latter is further stressed in section 3.1. |
| 147-152 | 3 | **Comments:**<br><br>Reorder to improve flow<br><br>**Proposed change:**<br><br>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. 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). 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). |
| 151 | 5 | **Proposed change:**<br><br>The two words "where as" should be corrected to the unified linker word "whereas." |
| 155-164 | 4 | **Comments:**<br><br>The flexibility around GLP standard studies is to be welcomed. |
| 156-158 | 5 | **Proposed change:**<br><br>Suggest bold or italicized text for emphasis for the sentence "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." |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 158-160 | 3 | **Comments:**<br><br>This sentence appears to be out of place. Paragraph provides general guidance on need for GLP and/or high quality scientific standards with data records readily available. Specific information on pEFD addressed later.<br><br>**Proposed change:**<br><br>Consider deleting |
| 165-174 | 4 | **Comments:**<br><br>Given Art. 33 (Clinical trials on pregnant or breastfeeding women) of CT Regn. 536 and similar provisions within 21 CFR, perhaps this section needs to be expanded to address the situation of a drug being developed specifically to manage a pregnancy-emergent condition, and so will be administered intentionally and possibly repeatedly to a pregnant woman. |
| 168-170 and 174-176 | 6 | **Comments:**<br><br>combine and shorten<br><br>**Proposed change:**<br><br>A comprehensive histopathological examination of the reproductive organs of the reproductive organs from the repeat-dose toxicity studies is a sensitive<br><br>method of detecting the majority of effects on male and female fertility, provided animals are sexually mature. (Note 1). |
| 170 | 4 | **Comments:**<br><br>Regarding: "A pharmaceutical for use in an elderly male does not warrant conduct of studies to evaluate stages E and F." Comment: Male only exposure can produce EF toxicity. But is this possibile in other ways than genotoxicity? If a pharmaceutical is not genotoxic, the exposure of a pharmaceutical to males only, can be expected to result in no more than a minimal exposure to WOCBP (via semen). This makes reprotoxicity unlikely. Would reprotox studies stages C-D normally not be waived as well? Especially when exposures in semes can be expecte to be much lower than plasma? |
| 170-171 | 7 | **Comments:**<br><br>it is stated that indications in elderly males do not warrant segment III studies. It is unclear why, by implication, segment III studies are warranted to support treatment of young men, and/or why there is a difference in risk between young and old. |
| 174 | 7 | **Comments:**<br><br>it is stated that "*Short-term therapies under highly controlled settings*" influences the extent of reproductive toxicity testing. It is unclear what this actually means/which scenarios this indicates. There can be several examples where it is warranted to undertake such testing also in these therapeutic |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | settings. |
| 183-184 | 6 | **Comments:**<br><br>Do we have to understand that when effects on reproductive organs are (only) assessed by histopathology in the repeat-dose toxicity studies, a dedicated fertility study serves to check for e.g., impact on fertility parameters (which are not assessed by histopathology of reproductive organs). |
| 187-188 | 6 | **Comments:**<br><br>sentence is redundant " Likewise, a fertility study is not warranted for pharmaceuticals that will not be used in subjects of reproductive age."<br><br>**Proposed change:**<br><br>delete as said before already |
| 192-193 | 3 | **Comments:**<br><br>The intent of this sentence is unclear. |
| 193 | 7 | **Proposed change:**<br><br>[…] *exposure under conditions of use. This should be addressed and the consequences for reproductive risk assessment should be discussed*. |
| 194-207 | 4 | **Comments:**<br><br>In the case of a drug being developed to treat pregnancy-emergent conditions not seen in non-pregnant women, the Guideline should stipulate that the complete reproductive toxicology package should be complete before the first administration to humans. There seems no ethical basis for proceeding with any human studies until the drug is known to be sufficiently safe to be assessed in the target population. |
| 197-203 | 6 | **Comments:**<br><br>In this introduction it lacks a reference to the Annex with detailed description of the in vivo study designs (11.2.3). Propose to include a reference to this chapter. |
| 201-203 | 6 | **Comments:**<br><br>argumentation should be more straight-forward, preferably PD active species, any positive result (EFD tox, not only TEFL): no further testing, wouldn't PD active species also allow to assess off-target EFD tox? only in some case 2nd species needed. |
| 204-246 | 6 | **Comments:**<br><br>Since this section refers to Figure 3-1, it would be appropriate to describe in the text the decisions step by step along to the figure and not to start with the PD activity in routine species. |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 204 f | 6 | **Comments:**<br><br>Under which circumstances are in vivo studies in two pharmacologically not relevant species required (two off target studies)? |
| 209 | 6 | **Comments:**<br><br>What is meant here with "these data"? |
| 210 | 6 | **Comments:**<br><br>Is the wording "it can be appropriate" here correct? Is an EFD study in (at least) one routine species not always needed for non-highly targeted molecules?<br><br>**Proposed change:**<br><br>adapt text to avoid such uncertainties |
| 211-217 | 6 | **Comments:**<br><br>The approach is sufficiently described in Lines 111-116. |
| 212 | 6 | **Comments:**<br><br>Table 3-1 is not part of this section.<br><br>Proposed change: Refer to section with this table (currently section 3.3.3) |
| 213-215 | 5 | **Proposed change:**<br><br>(additions in bold): When designing a pre- and post-natal development (PPND) or ePPND study, thought should be given to the value **of including** juvenile animal **toxicity** endpoints for supporting the safety of pediatric use (see Section 215 9.4.2.1). |
| 229-231 | 5 | **Proposed change:**<br><br>Consider clarifying text: If NHPs are to be used to assess effects on fertility, this is based on histologic evaluation of reproductive tissues from ~~there should be a sufficient number of~~ sexually mature animals of both sexes ~~should be~~ at study termination. |
| 230-231 | 2 | '*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 the effects on fertility, there should be a sufficient number of sexually mature animals at study termination'*.<br><br>**Comments:**<br><br>See general comment 1. We don't see a need to make a separate distinction for NHPs, surely there should be a sufficient number also if dogs or mini pigs are used?<br><br>**Proposed change:**<br><br>If dogs, **NHPs** and minipigs used in long-term repeat-dose studies **are used** |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | **to assess effects on fertility then a sufficient number** 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~~. |
| 235-237 | 5 | **Proposed change:**<br><br>Consider adding text "…except in the female dog due to the infrequent and relatively unpredictable progression of estrous cycles" |
| 237-239 | 5 | **Comments:**<br><br>Suggest move: "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 toxicity studies." to line 228 for clarity of section.<br><br>**Comments:**<br><br>consider adding text (bold)<br><br>**Proposed change:**<br><br>This information will later be supplemented with similar evaluations in the subchronic toxicity studies ***and if there is a concern, sexually mature animals could be included.*** |
| 263 | 5 | **Comments:**<br><br>Greater clarity should be included regarding the surrogate molecules<br><br>**Proposed change:**<br><br>Species-specific surrogate molecules may be useful for understanding some issues but should not be required. In those cases where surrogate molecules are used, there should be sufficient characterization to ensure pharmacologic relevance. |
| 271-273 | 7 | **Comments:**<br><br>repetition also later in the paragraph. |
| 288-290 | 5 | **Comments:**<br><br>The phrase "however, the approach used should be justified" has no meaning if a relevant model as listed in the first portion of the sentence cannot be identified.<br><br>**Proposed change:**<br><br>Amend the wording of the phrase to read: "however, the decision not to perform in vivo reproductive toxicity testing due to the absence of a relevant nonclinical model should be provided." |
| 295 - Figure | 3 | **Comments:**<br><br>If yes to first decision point (adverse events of intended pharmacol…), goes to |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 3.1 | | "sufficient to communicate risk; no studies warranted". However, does not address what to do if information is insufficient to communicate risk. Presumably would move to "highly targeted molecule" decision point<br><br>**Proposed change:**<br><br>Consider revising figure to include path when insufficient to communicate risk. |
| 295 - Figure 3.1, Table 3-1, section 9.5.5. (appendix) | 7 | **Comments:**<br><br>The relations between testing strategies in Figure 3-1 (page 10), Table 3-1 (page 12) and the recommendations in appendix section 9.5.5 (page 59+) need to be clarified. It remains unclear under what conditions the different paths should be followed. For instance, the appendix seems to make it clear that there are three scenarios with alternative testing, but this is not clear in the main text. Should Table 3-1 be considered a continuation from the "2 relevant species scenarios" path in Fig 9-1? How do the different Table 3-1 approaches (A, B & C) integrate with the appendix scenarios? The reference to 8.5.5.1 & 8.5.5.2 in Fig 9-1 seems to be wrong and should likely refer to 9.5.5.1 & 9.5.5.2? |
| 263-265 | 3 | **Comments:**<br><br>Unclear what "these data" is referring to<br><br>**Proposed change:**<br><br>If it is a highly-targeted pharmaceutical, data from GM animals or surrogate can be sufficient. |
| 304-307 | 2 | *'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.'*<br><br>**Comments:**<br><br>We are disappointed that the document does not go further in terms of the scenarios whereby alternative assays could be used to replace animal tests. We therefore request that flexibility be added to the document such that as alternative assays become available and qualified, regions can accept them in a wider range of scenarios.<br><br>**Proposed change:**<br><br>'Use of qualified alternative assays is appropriate for risk assessment under certain circumstances where they are interpreted in conjunction with in vivo |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | 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. **As experience with alternative assays grows and they become qualified and/or externally validated, regions may decide to extend the scenarios for which they may be suitable to replace in vivo testing.** |
| 320-333 | 6 | **Comments:**<br><br>In this introduction it lacks a reference to the Annex with detailed description of the in vivo study designs (11.2.2). Propose to include a reference to this chapter. |
| 326 | 7 | **Comments:**<br><br>Why should GLP conditions only be aimed for if the assay did not identify a hazard? This can also not be known on beforehand, and planning should include adherence to GLP /GLP like conditions. Propose to reword. |
| 333 | 6 | **Comments:**<br><br>see comment line 144 (meant for all molecule types?) as ICH S6 refers only to biotech products<br><br>**Proposed change:**<br><br>add clarification |
| 337 | 6 | **Comments:**<br><br>In case a study is negative, and only one species was testing it would be reassuring that the foetus was exposed and that the outcome is true negative. The translation to human foetal data is not warranted since these data are not available, the information is relevant for the interpretation of the study. |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 359-361 | 3 | **Comments:**<br><br>Appears to be distinguishing between the pEFD and enhanced pEFD.<br><br>**Proposed change:**<br><br>One such approach is the use of an enhanced pEFD study for one of the species. ~~In this case~~For the enhanced pEFD, 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 examination. |
| 362 – Table 3-1 | 3 | **Comments:**<br><br>Propose reordering approaches in Table<br><br>**Proposed change:**<br><br>Approach C (becomes Approach A) first since this is the approach described in ICH M3. Move Approach A to end (becomes Approach C) since currently qualified alternative assays are not readily available.<br><br>**Comments:**<br><br>Table on its own is difficult to fully understand. Additional text clarifying the 3 different approaches would be helpful.<br><br>**Proposed change:**<br><br>For example<br><br>Approach A: conducting an enhanced pEFD or definitive EFD along with a qualified alternative assay will support unlimited inclusion of WOCBP up to start of Phase 3. A definitive EFD in a second species should be conducted to support unlimited of WOCBP in Phase 3 trials.<br><br>Approach B: Unlimited inclusion of WOCBP up to start of Phase 3 can also be supported with pEFD study in one species and an enhanced pEFD or definitive EFD in a second species. A definitive EFD in the first species should be conducted to support unlimited inclusion of WOCBP in Phase 3 trials.<br><br>Approach C: pEFD studies in two species (ICH M3(R2)) can support clinical trials with limited inclusion of WOCBP (up to 150 WOCBP) for short duration (up to 3 months). Definitive EFD studies in both species should be conducted to support unlimited inclusion of WOCBP in larger Phase 2 and 3 trials.<br><br>In all cases, definitive EFD studies in two species, when appropriate, should be conducted to support marketing. |
| 362 – Table 3-1 | 7 | **Comments:**<br><br>It appears that in Table 3-1, the A, B, C scenario differ from the A, B, C scenario outlined in the Appendix 9.5.5.1/2. We propose that it is further clarified how the information in the body of the document and the appendices is related. |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 364-372 | 6 | **Comments:**<br><br>the criteria for the selection of the species seems to be stated at a very high level; in contrast the rabbit is recommended as the second species (section 4.1.2) for which some data (criteria) are very often not available (e.g. pharmacologic target, metabolic profile); that could trigger more animals testing in rabbits; until now the missing data in rabbits have been accepted by the regulatory authorities |
| 369-371 | 2 | '*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'*.<br><br>**Comments:**<br><br>If additional animal tests (preliminary PPND in this case) are not considered absolutely necessary and can be avoided, then firmer recommendation should be made to discourage their use as per the 3Rs.<br><br>**Proposed change:**<br><br>'Although 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, in most cases this **would be considered unnecessary**, because the appropriate information is generally available from prior studies to design the definitive study'. |
| 376-387 | 4 | **Comments:**<br><br>Given the extensive changes of ADME seen in human pregnancy, toxicokinetics should address whether relevant levels of exposure have been attained, or, alternatively, significantly exceeded. In the event that human data shows PKs in pregnancy are significantly different, the Guideline should address whether reproductive toxicology studies shown be repeated using the new information. |
| 390 | 6 | **Comments:**<br><br>preventive vaccines are often given only during childhood and then no reprotox studies are warranted<br><br>**Proposed change:**<br><br>refer to vaccines guidelines, e.g. WHO and possible acceptable lack of reprotox studies , see lines 648/649 (this argument of lines 648/649 should come earlier) |
| 395-398 | 2 | '*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 NHPs where proven mothers can be an advantage for ePPND studies'*. |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | **Comments:**<br><br>The above example of an NHP study has been presented in the section on 'routine test species'. This is inappropriate as the NHP is not considered a routine species and any discussion of NHP tests should be limited to the 'non-routine' species section so as to avoid any unnecessary confusion.<br><br>**Proposed change:**<br><br>delete any mention of NHPs in the routine test species section. |
| 407-431 | 6 | **Comments:**<br><br>this paragraph is redundant as information was given already before |
| 414-415 | 3 | Comment: Implies selection of pharmacologically relevant species is particularly important for highly targeted molecules. However, throughout the document it seems to apply having a pharmacologically relevant species is important in all cases and this appears to be emphasized throughout document.<br><br>**Proposed change:**<br><br>Consider deleting as extraneous sentence.<br><br>Line 412 "c" already indicates pharmacological activity should be considered when selecting the test species. Suggest to incorporate the reference to highly targeted molecules here. |
| 416-421 | 3 | **Comments:**<br><br>Implies rat should be used even when mouse is the primary rodent tox species.<br><br>**Proposed change:**<br><br>Consider clarifying |
| 416-427 | 7 | **Comments:**<br><br>The text in sections 4.1.1 and 4.1.2 can easily be reduced into one or two sentences (e.g. "*The rat and the rabbit are the most commonly used animal model species in human pharmaceutical EFD testing and are consequently considered to be routine animal model species*"). |
| 420-431 | 6 | **Comments:**<br><br>Paragraph refer to testing strategy and not to species selection. |
| 432-435 | 6 | **Comments:**<br><br>The content of this paragraph does not fit to the title of the section, and it is redundant.<br><br>**Proposed change:**<br><br>delete paragraph |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 435-437 | 3 | **Comments:**<br><br>This appears to be extraneous information.<br><br>**Proposed change:**<br><br>Consider deleting |
| 463 | 3 | **Comments:**<br><br>"some" appears extraneous<br><br>**Proposed change:**<br><br>For ~~some~~ therapeutic modalities that lack orthologous target engagement in useful reproductive toxicology species and also have anticipated off-target effects,…" |
| 465-469 | 6 | **Comments:**<br><br>Move information to chapter 3.1.2 |
| 471-491 | 2 | **Comments:**<br><br>We are concerned about the inclusion of a new section on 'disease models' which was not in the previous version of the guideline. The recommendation of additional animal tests that are not strictly required should be avoided at all costs. Especially the use of disease models, which have a poor track record, are highly basic in nature and are not relevant to the human condition due to unavoidable species differences. |
| 495-497 | 3 | **Comments:**<br><br>It is unclear what the relevance of this example is or how it applies to genetically modified models or surrogates. Isn't this true any time nonclinical data is used to make a risk assessment? Is the intent to indicate these models are best used for hazard identification?<br><br>**Proposed change:**<br><br>Consider deleting or clarifying |
| 503-505 | 3 | **Comments:**<br><br>This sentence can be confusing. Is the thought both a genetically modified model and surrogate molecule would be used at same time? If off target effects are expected shouldn't a routine test species be used with the clinical molecule in addition to the genetically modified model or surrogate molecule?<br><br>**Proposed change:**<br><br>Consider clarifying |
| 510 | 6 | **Comments:**<br><br>In case only small multiples of the clinical exposure are achieved in absence of maternal toxicity, how could a change of the route of administration improve |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | this situation? It is assumed that the systemic exposure would be increased, as would the maternal toxicity. |
| 512-513 | 3 | **Comments:**<br>Is intent in routine test species?<br>**Proposed change:**<br>"… evaluation of off-target reproductive toxicity using the clinical candidate in a routine test species is warranted." |
| 541-545 | 4 | **Comments:**<br>The statement regarding TK confirmation in pregnant animals is to be welcomed. |
| 570-572 | 6 | **Comments:**<br>refers to species selection, could be removed here. |
| 577-580 | 6 | **Comments:**<br>re-word; content is unclear – what is difference between some pharmacodynamic activity and pharmacodynamic activity? |
| 581-584 | 6 | **Comments:**<br>contradict line 524 (equally appropriate) |
| 606-622 | 3 | **Comments:**<br>The text in this section is not clear. Examples are straight forward (Lines 616 -622) but lack of understanding in Line 610-612 "alternatively, the fraction unbound can be used regardless…"<br>**Proposed change:**<br>Consider clarifying |
| 635-636 | 6 | **Comments:**<br>How can a dose level resulting in an exposure below the clinical exposure inform on risks for patients? |
| 648-649 | 6 | **Comments:**<br>this argument should come earlier , e.g. lines 390 |
| 702-704 | 3 | **Comments:**<br>This would be better addressed in Section 3.2 than here; it currently seems to be out of place.<br>**Proposed change:**<br>consider deleting |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 702-704 | 7 | **Comments:**<br><br>remove text, the same is said later in the paragraph. |
| 707-713 | 3 | **Comments:**<br><br>No example of study design in Annex 9.4.4 provided unlike other options in Section 6.2<br><br>**Proposed change:**<br><br>Consider adding study design for this approach to Annex 9.4.4 |
| 707-722 | 7 | **Comments:**<br><br>Sections 6.3 (Two study design) and 6.2 (single study design) should change place since section 6.1 discusses three separate studies design. |
| 730 | 5 | **Comments:**<br><br>Assessing mating and fertility after 26 weeks of dosing is not recommended, suggest delete text |
| 739 | 7 | **Comments:**<br><br>[…] used for this assessment can come from any repeat-dose study *of at least two weeks duration*. |
| 754-759 | 1 | **Comments:**<br><br>As a general principle fetal observations should be recorded as individual findings i.e. "grouping" should not be carried out during the examination. However, many laboratories use computer based collection systems which also have a reporting function. It is common, especially in the "Summary of observations" table to group findings. This can be a simple as grouping left/right/bilateral findings or can include different observations e.g. incomplete ossification of skull bones. The individual table, however, should not include such groupings and should reflect the original observations made during the examinations. I have seen "Individual" tables that are merely listings of the grouped observations per fetus and reference had to be made to the original database in order to break the groupings down.<br><br>**Proposed change:**<br><br>Ensure that individual tables reflect the record of individual fetal observation findings. |
| 754-771 | 7 | **Comments:**<br><br>Under section 6.5.1, one can question if the text on data presentation is necessary in the main text (as opposed to an appendix). |
| 760-763 | 1 | **Comments:**<br><br>Would it be more pertinent to have all structural change data, which is considered to be related to administration of the test compound, tabulated as |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | the primary listing. If there are none this should be stated up front. |
| 770-771 | 1 | **Comments:**<br><br>The industry standard reference for terminology is Makris et al (2009) the Terminology of Developmental Abnormalities in Common Laboratory Mammals (Version 2) – why not be specific in the guidance about its use? Alternatively mandate naming the source document used for terminology in the report text/reference section of the study report? |
| 770-771 | 7 | **Comments:**<br><br>Under section 6.5.1, the text regarding terminology, it may be reasonable to refer to the teratological terminology harmonization project (so far 8 workshops, the latest reported by Solecki R et al., (2015) Continuing harmonization of terminology and innovations for methodologies in developmental toxicology: Report of the 8th Berlin Workshop on Developmental Toxicity, 14–16 May 2014 Reprod Toxicol 57: 140–146). |
| 790 | 5 | **Proposed change:**<br><br>Suggest bold or italicized text for emphasis for the sentence "Therefore, cesarean and fetal data should be calculated for the litter as the unit of measure." |
| 804 | 7 | **Comments:**<br><br>add *dose – response trends* |
| 813 | 7 | **Proposed change:**<br><br>*Therapeutic benefit considerations can influence the assessment of the human risk management strategy*. |
| 818 | 4 | **Comments:**<br><br>The statement that "Definitive human data will supersede nonclinical data" is helpful. However, establishing that a drug has caused a teratogenic effect is one of the most difficult avenues we face. |
| 822-823 | 7 | **Comments:**<br><br>Rows 822-823 can be moved after the first paragraph in the same section (after row 812). |
| 859-861 | 1 | **Comments:**<br><br>e.g. higher incidences of palatal rugae changes (Ikemi, 2001) |
| 861 | 7 | Add '*presence or absence of maternal toxicity*' within the parenthesis. |
| 862 | 5 | **Comments:**<br><br>Section 7.1: "transient inhibition of spermatogenesis". Not clear what is intended for this pathology endpoint, suggest clarify or use alternate example. |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| 882-892 | 3 | **Comments:**<br><br>Risk assessment for lactation section states that animal lactation studies not relevant to human risk assessment; therefore, this section does not seem relevant. In addition, the key information is already addressed earlier in document (Line 383-385).<br><br>**Proposed change:**<br><br>Delete this section as lactation data from nonclinical species is generally not relevant for risk assessment. |
| 897 | 5 | **Comments:**<br><br>Quantitative assessment of spermatogenesis<br><br>**Proposed change:**<br><br>Consider revising the text as indicated, with reference:<br><br>A quantitative analysis of spermatic stages (i.e. staging) is not ~~generally~~ recommended, ***nor useful***. [Creasy DM (1997). Evaluation of testicular toxicity in safety evaluation studies: The appropriate use of spermatogenic staging. Toxicol Pathol 25: 119—131]. |
| 907-922 | 3 | **Comments:**<br><br>Note 3: It is unusual to have guidance soliciting data. No information provided for where to submit data. |
| 954 | 4 | **Comments:**<br><br>It would be useful to add a table of agents which have shown evidence of reproductive toxicity an animals, but have never been reported to express similar effects is humans. |
| 973-975 | 3 | **Comments:**<br><br>Does highly targeted or highly selective pharmaceutical/therapeutics apply to antivirals or antibacterials that are specific for viral or bacterial (non-mammalian) targets?<br><br>**Proposed change:**<br><br>Provide clarification and modify definition as appropriate |
| 1040 – Figure 9-1 | 7 | **Comments:**<br><br>Consider writing Cynomolgus monkey instead of Cyno<br><br>**Comments:**<br><br>The box "contributes to mechanistic understanding" is a dead end. Please consider editing. |
| 1042 – Mini- | 5 | **Comments:**<br><br>3-5 month sexual maturity for mini pigs seems too young. i.e. 4 to 7 months |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| pigs | | for sexually mature mini pig females (Tox Pathol 2016; 44(3):482-5) |
| 1042 – Guinea pig | 5 | **Comments:**<br><br>Consider: An additional disadvantage is that neonates in this species are markedly precocious relative to altricial species, so are a poor model for maternal care. |
| 1047 | 7 | **Proposed change:**<br><br>Delete *Numbers* |
| 1047-1050 | 3 | **Comments:**<br><br>Clarify if addressing number of animals per group or number of groups on a study or both.<br><br>**Proposed change:**<br><br>Suggest there are situations when there is value to adjust either the number of groups or number of animals per group. |
| 1058 | 2 | **Comments:**<br><br>Studies with 'two breeding generations' are mentioned without any further explanation. What are the circumstances where two generations are needed? Can only one generation be recommended as common practice? (e.g. EOGRTS). |
| 1065 | 2 | **Comments:**<br><br>The use of alternative methods to determine possible effects on human male and female fertility (described in general comments, point 4) are possible and should be recommended, prior to any fertility assessment in rodents. |
| 1074 | 1 | **Comments:**<br><br>I appreciate that unless there are indications from general toxicity studies of an effect on male/female reproductive organs a combined male/female FEED study is commonly used. However, studies of this design may be potentially problematic since a positive outcome will not tell you if the effect is male or female mediated? Reiterate statements made in paragraphs 211/212 and 1211/ 1214 regarding use of animals (males) on repeat dose general toxicity studies. |
| 1134-1193 | 6 | **Comments:**<br><br>Why is the section about PPND study designs (11.2.2) before the section to EFD studies (11.2.3)?<br><br>**Proposed change:**<br><br>Change order (EFD first, PPND second) |
| 1137-1147 | 7 | **Comments:**<br><br>In section 9.4.2.1 there is a discussion about the extension of segment III |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | studies (maternal exposure based) with juvenile toxicity (direct offspring exposure based) studies (e.g. antiviral treatment). The toxicological literature shows that the outcomes from prenatal-maternal exposure and direct postnatal exposure in rodents (the main animal models) can sometimes be substantially different in type and/or sensitivity. Consequently, if for instance the purpose is to evaluate safety starting by direct exposure before the weaning of infants (within the range of segment III studies), a maternal exposure (e.g. the standard setup until PND21 in segment III studies) would be directly misrepresentative. |
| 1148-1159 | 2 | **Comments:**<br><br>An enhanced PPND study in NHPs is listed and described here, which states that >16 pregnant animals are needed along with a 'sufficient number of infants (6-8 per group). According to the literature, a large number of female and male NHPs are required for mating in order to obtain the necessary pregnancies per group (Faqi, 2012) due to their low fertility and high rate of miscarriage/stillbirths. In the interest of animal welfare, these tests should be discouraged and replaced with testing strategies made up of alternative methods and tests in other species. |
| 1182 – Table 9-5 | 3 | **Comments:**<br><br>pEFD TK - If pEFD or enhanced pEFD is used to defer the definitive EFD studies as described in Table 3-1, shouldn't TK be conducted to assist in assessing potential human risk unless TK data in pregnant animals already exist?<br><br>**Proposed change:**<br><br>Consider adding footnote to address this.<br><br>**Comments:**<br><br>Footnote "g" implies that for pEFD studies there is no requirement to have a minimum number of litters for evaluation. Is this still the case when these studies are used to defer the definitive EFD studies?<br><br>**Proposed change:**<br><br>Suggest to clarify. |
| 1190 | 1 | **Comments:**<br><br>Extremely important to ensure dam necropsy is sufficiently detailed and pay particular attention to any maternal congenital anomalies observed during the post mortem (maternal necropsy can often be "cursory" since the dams are very often regarded as the carriers of the fetuses only....) |
| 1207 | 3 | **Comments:**<br><br>[Section 11.2.4.1] FEFD = fertility and embryo-fetal development<br><br>**Proposed change:** |

| Line no. | Stakeholder no. | Comment and rationale; proposed changes |
|---|---|---|
| | | Change title to reflect acronym<br><br>Fertility and ~~Embryonic~~ Embryo-Fetal Development |
| 1277 | 7 | **Comments:**<br><br>What is the basis for testing 45 compounds? |
| 1281 | 7 | **Comments:**<br><br>At least 2 or at least 3? |
| 1290 | 7 | **Comments:**<br><br>It is an absolute requirement that inter-laboratory reproducibility is established for all assays as part of the assay qualification. |