

July 2021

Critical aspects for designing and conducting extended one-generation reproductive toxicity (EOGRT) studies under REACH

The extended one-generation reproductive toxicity (EOGRT) study (EU B.56, OECD TG 443) has been the new information requirement for reproductive toxicity under REACH (Annexes IX and X, Section 8.7.3.) since March 2015. When the EOGRT study was introduced into REACH, there was very limited practical experience about the study.

Currently, experts from ECHA and 10 Member States are evaluating the study reports of EOGRT studies for an ongoing *EOGRTS review project* to analyse specific aspects relating to the performance of the study with respect to design, conduct and toxicological findings.

The analysis of EOGRT studies started in May 2021 and, to date, 12 cases have been evaluated. Although the project is at an early stage, critical issues have been identified in terms of how the EOGRT studies were designed and conducted. These issues are considered critical because they compromise data analysis and interpretation of the results, and might raise the question of compliance with legal requirements or result in requests for additional studies if there is a concern.

Industry sponsors and test laboratories can take these critical issues into account to improve ongoing and future EOGRT studies.

1. Dose-level selection must be based on effects on sexual function and fertility, and be sufficiently high

To be compliant and not rejected due to too low dose levels, the highest dose level must induce clear evidence of an adverse effect on sexual function and fertility, but avoid deaths¹ or severe suffering such as persistent pain and distress² in P0 animals.

If there is no clear evidence of an adverse effect on sexual function and fertility, a limit dose of at least 1 000 mg/kg bw/day or the highest possible dose level not causing severe suffering or deaths in P0 must be used as the highest dose level.

A descending sequence of dose levels should be selected to demonstrate any dose-related effect with the aim to establish the lowest dose level as a no-observed-adverse-effect level (NOAEL).

In all cases, the aim must be to demonstrate whether the classification criteria of the most severe hazard category for sexual function and fertility (Repr. 1B; H360F) and developmental

¹ No more than 10 % mortality.

² OECD GD 19, paragraph 18.

toxicity (Repr. 1B; H360D) of the CLP Regulation apply for the substance^{3,4,5}, and whether the substance meets the criteria for a substance of very high concern regarding endocrine disruption according to Article 57(f) of REACH.

Numerical results (*i.e.* incidences and magnitudes) and a description of the severity of effects at all dose levels from dose range-finding studies have to be reported to facilitate the assessment of the dose level selection and the interpretation of the results of the main study.

A justification should be provided that the study results demonstrate that the dose level selection meets the conditions described above.

The highest dose level must demonstrate the highest possible dose level without severe suffering or deaths in P0 animals, or to follow the limit test concept.

2. Histopathological analysis of organs and tissues in Cohort 1B

According to paragraph 67 of OECD TG 443, "*Cohort 1B animals should have the following organs weighed and corresponding tissues processed to the block stage: Vagina (not weighed), uterus with cervix, ovaries, testes (at least one), epididymides, seminal vesicles and coagulating glands, prostate, pituitary, and identified target organs. Histopathology in cohort 1B would be conducted **if results from cohort 1A are equivocal or in cases of suspected reproductive or endocrine toxicants.***"

When Cohort 1B is extended to produce the F2 generation, paragraph 41 of the OECD GD 151 explains, "... Cohort 1B, is included for termination at approximately 14 weeks (if not mated) or 20-25 weeks (if mated) of age and should be subject to gross necropsy with organ weights and tissues processed to block for future analysis, if required."

Therefore, irrespective of whether Cohort 1B is extended, the procedure described in paragraph 67 of OECD TG 443 applies and must be followed.

As such, organs and tissues of Cohort 1B (extended or not) animals processed to the block stage, including those of identified target organs, must be subjected to histopathological investigations if:

- the results from Cohort 1A are equivocal;
- the test substance is a suspected reproductive toxicant; or
- the test substance is a suspected endocrine toxicant.

³ OECD GD 151, paragraph 28: "*Based on weight of evidence and/or specific regulatory authority's requirements, evidence of systemic toxicity or reproductive toxicity may be required at the highest dose level in order to ensure that the test system is optimised to be able to investigate any reproductive toxic property of a substance measured in the test system. [...] It is recognised that some dose levels of the test substance may affect fertility, such that an insufficient number of pups may be produced for assessment of the F1 generation. In situations where fertility is affected, the lower dose levels should therefore be carefully selected to ensure the objectives of the study can be met.*"

⁴ Annex I Section 1.0.1. to REACH: "*the objectives of the human health hazard assessment shall be to determine the classification of a substance in accordance with Regulation (EC) No 1272/2008; and to derive levels of exposure to the substance above which humans should not be exposed*".

⁵ If the substance meets the criteria for classification as toxic for reproduction category 1A or 1B: May damage fertility (H360F), and the available data are adequate to support a robust risk assessment, then no further testing for fertility will be necessary. If a substance meets the criteria for classification as toxic for reproduction category 1A or 1B: May damage the unborn child (H360D), and the available data are adequate to support a robust risk assessment, then no further testing for developmental toxicity will be necessary.

A request to extend Cohort 1B in the ECHA decision always reflects a concern for reproductive toxicity especially in filial generations, irrespective of the triggers. Therefore, histopathological investigations of the organs and tissues of the extended Cohort 1B must be conducted.

3. Organs/tissues of low and mid-dose must be subjected to full histopathology if treatment-related changes are observed in these organs/tissues at high dose

According to paragraph 70 of OECD TG 443, *“full histopathology of the organs listed in paragraphs 63 and 64 is performed for all high-dose and control P animals. Organs demonstrating treatment-related changes should also be examined in all animals at the lower dose groups to aid in determining a NOAEL.”*

According to paragraph 71 of OECD TG 443, *“full histopathology of the organs listed in paragraphs 63 and 64 is performed for all high-dose and control adult cohort 1A animals. All litters should be represented by at least 1 pup per sex. Organs and tissues demonstrating treatment-related changes and all gross lesions should also be examined in all animals in the lower dose groups to aid in determining a NOAEL.”*

Reliable no-observed-adverse-effect level (NOAEL) and limited-observed-adverse-effect level (LOAEL) values cannot be derived based on the findings at the top dose if low and mid-dose groups are not investigated. In such a case, the study is not compliant because no conclusions on risk assessment and classification and labelling can be made.

For treatment-related changes in organs/tissues of high-dose animals, it is important to investigate these organs/tissues in low and mid-dose animals to derive NOAEL and DNEL values for risk assessment and LOAEL values for classification and labelling purposes (see Section 1.0.1 of Annex I, REACH).

4. Investigations on F1 and F2 pups must be identical

According to ECHA Guidance R.7a, Appendix R.7.6–2 *“the registrant is responsible for implementing the overall design of the study as requested, conduct of the study and interpretation of the results in order to meet the regulatory requirements and to insure the scientific integrity of the study in line with the test method.”*

In this respect, having a valid basis for comparative analysis is crucial because the *“extension of Cohort 1B to F2 is considered relevant in the context for classification and labelling and categorisation especially if the effect in P0 parental/F1 offspring is significant but not meeting classification criteria to Repr. 1B and more severe effects are seen in the F1 mating pairs/F2 offspring, thus affecting both P0 parental/F1 offspring and F1 mating pairs/F2 offspring but being more prominent or with a broader/different spectrum in F1 mating pairs/F2 offspring This could lead to a change in the classification from Repr. 2 to Repr. 1B.”*

Special attention should be paid to any target organs identified in the F1 generation to adequately address this concern in F2.

The Annex Table 1.2 of OECD GD 151 clarifies that endpoints and examinations required in F1 litters are identical to F2 up to weaning. This also includes the mandatory measurement of thyroid stimulating hormone (TSH) and thyroxine (T4) at weaning and the optional measurement at postnatal day (PND) 4.

Only by conducting identical investigations in F1 and F2, can an adequate comparative analysis of the first and second filial generation be made.

Therefore, F1 and F2 must be subjected to identical investigations if an extension of Cohort 1B is triggered.

5. Increased statistical power to investigate sexual maturation in F1

Paragraph 12 of OECD GD 151 specifies that a total of 3 pups/sex/litter (*i.e.* a total of 60 animals/sex) must be evaluated for sexual maturation, independently of whether Cohorts 2A/2B and 3 are conducted.

In the review of the EOGRT studies, however, some cases were encountered where only 2 pups/sex/litter had been evaluated for sexual maturation. This deviation from the OECD GD 151 text results in lower statistical power.

Paragraph 44 of the OECD GD 151 clarifies that *"the litter mean values should still be considered in the analysis of the data and a statistical method based on data from all investigated pups should be used."*

The results should not be analysed separately by cohort. Instead, the results of these measurements should be statistically analysed by combining the results of all the F1 pups of the same dose group to achieve higher statistical power. A total of 3 pups/sex/litter must be evaluated.

6. Immunotoxicity testing in P0 and Cohort 1A animals is mandatory

Some test laboratories omitted splenic lymphocyte subpopulation analysis.

ECHA emphasises, however, that immunotoxicity testing in P0 and Cohort 1 animals in the EOGRTS is mandatory as long as not clearly described as optional in the protocol.

Investigations stated in paragraph 66 of the OECD TG 443 are mandatory because neither the test guideline nor its guidance document (OECD GD 151) indicate that these investigations are optional in any circumstances. In other words, the following investigations are mandatory in Cohort 1A:

"For the investigation of pre- and postnatally induced immunotoxic effects, 10 male and 10 female cohort 1A animals from each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) will be subject to the following at termination:

- *weighing of the lymph nodes associated with and distant from the route of exposure (in addition to the weight of the adrenal glands, the thymus and the spleen, already performed in all cohort 1A animals)*
- *splenic lymphocyte subpopulation analysis (CD4+ and CD8+ T lymphocytes, B lymphocytes, and natural killer cells) using one half of the spleen, the other half of the spleen being preserved for histopathological evaluation,*

Analysis of splenic lymphocyte subpopulations in non-immunized (cohort 1A) animals will determine if exposure is related to a shift in the immunological steady state distribution of "helper" (CD4+) or cytotoxic (CD8+) thymus-derived lymphocytes or natural killer (NK) cells (rapid responses to neoplastic cells and pathogens)."

Missing investigations without scientific reasons might raise the question of compliance, or requests for further studies if there is a concern. Therefore, it is important that the test guideline is followed with respect to mandatory investigations.

The splenic lymphocyte subpopulation analysis is a mandatory investigation and must always be conducted irrespective of whether Cohort 3 is triggered or not.

7. Methods need to be described in sufficient detail

Apparent inconsistencies in results may be due to different methodologies used to measure parameters rather than inconsistencies in the results themselves.

To allow independent evaluation of the data reported, used methods need to be reported in the study report and robust study summary with sufficient details. In brief, the method section should provide sufficient documentation on how the investigations were conducted, the methods used and reasons for choosing the methods.

When commercially available kits or devices are used, these should be identified, as well as the possible computational software used. Furthermore, data supporting the reliability and sensitivity of the test method (*i.e.* positive and historical control data) needs to be reported.

Deficiencies that have been observed in conducting/reporting and affecting interpretation and acceptance of the results include:

- T-cell dependent antibody response (TDAR) (number of responsive animals in positive control vs test animals);
- auditory startle response (ASR) (no historical controls, no habituation in controls);
- morphometry of brain areas (historical controls, clarity in measured areas);
- areolae/nipple retention (control values do not reflect background variation);
- post-implantation loss/stillbirths (clarity on how this is calculated);
- Follicle counts (method not well explained, high variation); and
- T4 and TSH measurements (laboratory validation of the method lacking? No historical control data).

For example, the percentage coefficient of variation (%CV) in T4 or TSH assays has been unacceptably high, and many values have been below the detection level. Therefore, these results are inconclusive.

For a proper analysis and interpretation of results, it is necessary that the methodology used to derive the result is adequate.

7.1. Considerations for interpretation

As a general rule, effects on reproductive toxicity are relevant for classification even in combination with other toxic effects. In some cases, effects on reproductive toxicity are considered by the registrants/in the study report to be a secondary, non-specific consequence of other toxic effects, or as a consequence of specific maternally mediated mechanism.

Sometimes only a hypothesis or assumptions are presented with no reliable evidence (*e.g.* mechanistic information). An unequivocal demonstration⁶, covering all possible mechanisms and leaving no open questions, is needed on a case-by-case basis to conclude that classification is not warranted.

⁶ CLP Regulation 3.7.2.4.2