



Report from the BfR expert hearing on practicability of hormonal measurements: recommendations for experimental design of toxicological studies with integrated hormonal end points

Olena Kucheryavenko¹ · Glenn Lurman¹ · Anja Lehmann¹ · Juliana Braz^{1,3} · Lars Niemann¹ · Ibrahim Chahoud⁴ · Alberto Mantovani⁵ · Helen Håkansson⁶ · Steffen Schneider⁷ · Volker Strauss⁷ · Pragati S. Coder⁸ · Alexius Freyberger⁹ · John C. O'Connor¹⁰ · Martina Rauch^{2,14} · Kostja Renko⁴ · Marize L. M. Solano^{1,11} · Niklas Andersson¹² · Olivier Blanck¹³ · Vera Ritz¹ · Roland Solecki¹

Received: 12 March 2019 / Accepted: 14 March 2019 / Published online: 30 March 2019
© The Author(s) 2019

Abstract

This publication summarizes discussions that were held during an international expert hearing organized by the German Federal Institute for Risk Assessment (BfR) in Berlin, Germany, in October 2017. The expert hearing was dedicated to providing practical guidance for the measurement of circulating hormones in regulatory toxicology studies. Adequate measurements of circulating hormones have become more important given the regulatory requirement to assess the potential for endocrine disrupting properties for all substances covered by the plant protection products and biocidal products regulations in the European Union (EU). The main focus was the hypothalamus–pituitary–thyroid axis (HPT) and the hypothalamus–pituitary–gonadal axis (HPG). Insulin, insulin-like growth factor 1 (IGF-1), parathyroid hormone (PTH) and vitamins A and D were also discussed. During the hearing, the experts agreed on specific recommendations for design, conduct and evaluation of acceptability of studies measuring thyroid hormones, thyroid stimulating hormone and reproductive hormones as well as provided some recommendations for insulin and IGF-1. Experts concluded that hormonal measurements as part of the test guidelines (TGs) of the Organisation for Economic Co-operation and Development (OECD) were necessary on the condition that quality criteria to guarantee reliability and reproducibility of measurements are adhered to. Inclusion of the female reproductive hormones in OECD TGs was not recommended unless the design of the study was modified to appropriately measure hormone concentrations. The current report aims at promoting standardization of the experimental designs of hormonal assays to allow their integration in OECD TGs and highlights research needs for better identification of endocrine disruptors using hormone measurements.

Keywords Endocrine disrupting chemicals · Endocrine disruptors/toxicity · Pesticides · Biocides · Risk assessment · Thyroid hormones · Reproductive hormones

Introduction

According to the definition proposed by World Health Organization/International Programme on Chemical Safety in 2002, endocrine disruptor (ED) is an “an exogenous substance or mixture that alters function(s) of the

endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations” (WHO/IPCS 2002). Since 2002, the definition has been endorsed by the Scientific Committee (SC) of the European Food Safety Authority (EFSA 2013). The European Union (EU) legislation for plant protection products (PPP) and biocidal products (BP) stipulates that identification of an active substance as an endocrine disruptor precludes a substance from approval in the EU, unless exposure of or risk to humans is negligible (Reg. (EC) No. 1107/2009 (EP and Council of the EU 2009); Reg. (EU) No. 528/2012 (EP and Council of the EU 2012)). To generate a harmonized framework for assessment of endocrine

Olena Kucheryavenko, Glenn Lurman and Anja Lehmann are equal contributors.

✉ Olena Kucheryavenko
Endocrine.Disruption@bfr.bund.de

Extended author information available on the last page of the article

disrupting properties during registration and re-approval of an active substance, the European Commission (EC) has endorsed scientific criteria for the identification of EDs in the context of BP and for active substances, safeners and synergists in PPPs (EC 2017, 2018).

The ED Guidance Document drafted jointly by the EFSA and the European Chemical Agency (ECHA) is foreseen to guide the implementation of the criteria in regulatory practice. It is designed to implement a hazard-based evaluation principle of the relevant endocrine disrupting properties of a substance and to advise applicants and regulatory authorities on how to assess available data using a weight of evidence approach. The ED Guidance Document (ECHA et al. 2018) emphasizes the relevance of an understanding of the mode of action (MoA) linked to the observed adverse effect and the endocrine activity. Consequently, changes in hormone concentrations gain particular weight in the assessment of possible endocrine disrupting properties of the substance during the pre-approval process. For example, significant changes in concentrations of oestradiol, testosterone and thyroid hormones may be indicative of an endocrine activity of an active substance on the experimental organism, and this should be contextualized in the MoA framework (i.e. dose and temporal concordance of the observed effects). However, because hormone concentrations fluctuate in response to the influence of internal and external factors (e.g. oestrous cycle, stress, circadian rhythm), a detailed description of the experimental conditions of the hormonal assays need to be provided to facilitate generation of meaningful and reliable results that can efficiently contribute to the MoA analysis.

Given the urgent need for practical guidance regarding the measurements of circulating hormones in regulatory toxicology studies, the BfR organized an expert hearing that was held in Berlin, Germany, on 18–19 October 2017. Prior to the expert hearing, the BfR, as a part of the collaborative scientific project with the EFSA, performed a laboratory survey to provide the experts with a detailed overview of the current methods applied in laboratories performing hormone measurement assays. Almost all of the replies from the contacted laboratories described procedures applied for the measurement of hormones circulating in blood. Consequently, the following aspects related to the hormonal measurements were addressed during the hearing:

1. The possibility of incorporating certain hormone measurements into existing OECD testing guidelines (TGs), the inherent methodological constraints pending such inclusion and the overall usefulness of the measurement of hormonal concentrations in the evaluation process.
2. The quality criteria required to assess the reliability of hormone measurement assays.

3. Specific recommendations for the hormonal assays, in particular with regard to animal housing, handling, sampling and detection methodologies.

Twenty-two participants representing academia, industry and risk assessment agencies from the EU, the USA and Brazil took part in the discussions in three parallel breakout groups. Representatives of the EC and the EFSA participated in the expert hearing as observers. Each group of experts focused on one of the hormonal pathways and corresponding hormones:

1. Hypothalamus–pituitary–thyroid (HPT) axis (triiodothyronine (T3), thyroxine (T4), thyroid-stimulating hormone (TSH)).
2. Hypothalamus–pituitary–gonadal (HPG) axis (oestradiol, progesterone, luteinizing hormone (LH), follicle-stimulating hormone (FSH), testosterone).
3. Other hormones referred further as non-EATS (IGF, insulin, PTH, vitamin D and A).

One of the direct outcomes of the expert hearing and the laboratory survey performed by the BfR was a compilation of a concise list of practical recommendations for the experimental design of hormonal assays relevant to thyroid, thyroid-stimulating and reproductive hormones. The list was subsequently provided to the EFSA and formed the basis for the Appendix B of the ED Guidance document (ECHA and EFSA et al. 2018). The current report contains the agreed recommendations as well as experts' views on inclusion of hormonal measurements in the OECD TGs, scientific justification for selecting of certain hormones for assessment, hormonal measurement at different life stages and sources of scientific uncertainty that could influence regulatory decision-making.

Results

Inclusion of hormonal end points in the OECD TGs

Considering the importance of hormonal measurements as *in vivo* mechanistic parameters for the assessment of EDs, the question of incorporating hormone measurements in the OECD TGs was addressed during the expert hearing. Since 1997, the OECD Advisory Group on Endocrine Disrupters Testing and Assessment (OECD EDTA AG) has implemented endocrine-sensitive parameters in TGs, including measurements of thyroid hormones, TSH, LH, FSH and testosterone. The experts supported OECD EDTA AG activities on inclusion of measurements of thyroid hormones and TSH during the updating of the TGs. The experts also agreed that

currently technical feasibility does not prevent the inclusion of HPT and HPG hormones in the OECD test guidelines.

Thyroid axis hormones (HPT axis)

The measurements of (total) T4, TSH and T3 were recommended as a preferred option to facilitate reliable data interpretation and avoid false assumptions that can arise from measurement of a single hormone. The assessment of all thyroid function-related end points including T3, T4 and TSH, thyroid weight as well as liver weight, thyroid, liver and pituitary histology should be favoured over the measurement of T4 alone. In general, the measurement of T4 and TSH is recommended as a minimum for assessment in post-pubertal animals, in weanlings and on postnatal day 4 (PND4). T3 could be measured as an optional end point at the same time points as T4 and TSH, if assessments are not limited by blood/plasma sample volumes. Measurement of free fractions of T3 and T4 represents a more complex technical challenge and may be investigated on a case-by-case basis in specifically designed mechanistic studies.

Alignment of the time points for mandatory and optional hormonal measurements in OECD TGs 421/422 and 443 should be considered during the next revision of these guidelines. For example, the TGs 421 and 422 hormone measurements on PND 13 are mandatory, but PND 4 is an optional time point. In contrast, the TG 443 mandates measurements on PND 4, PND 21 and in adults. The existing discrepancy in the time points at which measurements are performed complicates the comparison of results obtained from two study types and a proposed alignment of the TGs would greatly improve data interpretation. It is noteworthy that TGs 421 and 422 are not standard data requirements for biocides and PPPs in the EU.

It was also noted that the measurement of thyroid hormones in fetuses is not currently required by the EU regulations; however if this requirement is to be introduced in the future, it should not compromise measurement of other end points such as intrauterine growth, survival and foetal morphology. Instead, a separate study such as the comparative thyroid assay (US EPA 2005) could offer a more appropriate possibility for investigating hormonal changes in the fetuses. It was also noted that the US Environmental Protection Agency (US EPA) has elaborated screening tests for thyroid-active EDs in peripubertal animals, yet these protocols are not currently candidates for OECD TG development (US EPA 2009a, US EPA 2009b). Evidence from other studies, performed according to the TGs 407, 408, 422, 443, 451 and 453 (OECD 2008, 2016, 2018a, c, d, e), could trigger the need for this additional study if gestational effects are anticipated. To overcome the technical limitation of the low blood volume in fetuses, pooling of blood samples within

litters was suggested, but implementation of a more sensitive method would be more advantageous.

The OECD TGs pre-define a minimum number of animals to be allocated to treatment and control groups, while statistical power analysis should be applied to calculate the group sizes actually needed. Experts strongly argued that statistical power is a critical point for defining the limits of the selected study type to detect changes in hormone concentrations under the chosen experimental conditions with the necessary sensitivity. For example, a study performed according to OECD TG 408 (OECD 2018a) with ten rats per group has sufficient power to detect a 1.5-fold increase in TSH and a 1.35-fold (35%) decrease in T3 or T4 concentration in the treated group vs control, assuming that the coefficient of variation in the control and treatment group is about 25% for T3 or T4 and 35% for TSH (Wilcoxon-test, two sided, power 75%, $p < 0.05$, NQUERY software used for the calculation). In contrast, the expected significant changes between control and treatment groups occurs only at a twofold increase of TSH and a 1.73-fold decrease of T3 or T4 when the group sizes are five rats as in the OECD TG 407 (OECD 2008). Changes below these levels will not be statistically significant. Indeed, if a dedicated satellite group for the hormonal assessment is considered, specific considerations on the number of animals should be carefully evaluated and justified.

Reproductive hormones (HPG axis)

It was agreed that measurement of reproductive hormones is a useful parameter for the identification of a substance with endocrine disrupting properties. To facilitate data interpretation, the measurement in the same study of a battery of hormones should be preferred over the measurement of a single hormonal endpoint. An informed decision regarding the exact hormones to be measured in the study should be made based on the available data from previous studies (e.g. histopathology, organ weight and reproductive parameters). In principle, certain OECD TGs can be extended by incorporation of reproductive hormone measurements. However suitability of each TG should be considered independently and appropriate modifications to the experimental design should be proposed if necessary, for example to consider the inclusion of a dedicated satellite group for the hormonal assessment or change in the number of animals. One important consideration highlighted by the experts was the insufficient number of female animals to account for the variations caused by the oestrous cycle.

Non-EATS pathways

Common for all non-EATS hormones and hormone-like substances addressed at the meeting is that they are not

incorporated in standard toxicology studies and there is no guidance on their use or interpretation. Disturbances of the retinoid system (vitamin A) that can lead to modifications of gene expression programmes of relevance for health later in life were discussed by the experts as an example of a project that is required to bring the non-EATS modalities to regulatory attention (Project 4.97, (OECD 2018f)). It was noted however that for vitamins A and D, additional evidence should be obtained before their inclusion as a standard parameter in the OECD TGs. Experts proposed that the measurements of insulin and IGF-1 could potentially be assessed in the context of evidence of altered growth during the post-natal development or indications of metabolic disease (e.g. diabetes, obesity), but should not be included as an end point in the guideline studies. Overall, experts have encouraged further activities to be initiated by the OECD EDTA AG, on the development of testing strategies to assess non-EATS modalities and guidance on their use or interpretation.

Quality and performance criteria for thyroid hormones, TSH, reproductive hormones and non-EATS pathways

Factors influencing quality and reliability of results, as well as those that lead to variability in the data were scrutinized by the experts. The results of the laboratory survey preceding the expert hearing formed the basis for the discussion. The survey highlighted, on one hand, the broad range of experimental conditions and detection methods reported by various laboratories, and, on the other, advances in the development and implementation of standard operating procedures by laboratories conducting the studies.

Considering the current practice and the fact that advances in technology will likely improve the sensitivity beyond the current standards, the experts suggested defining the range of quality criteria that are expected to be met, independent of the assay methodology in use (Table 1). Meeting these recommended quality criteria would enable greater consistency and reliability in the data, as well as reduced variability. Such an assay-independent approach requires considerable assay validation. Assistance in the validation of assays is in part provided by the EMA guideline on bioanalytic method validation (EMA 2011), and some recommendations on chromatographic method validation are listed in the report from the 3rd AAPS/FDA Bioanalytical Workshop (Viswanathan et al. 2007).

At the plenary session, different opinions were expressed by experts regarding the reporting of the assay validation results, whether it should be summarized in the corresponding study report or be submitted to competent regulatory agency/authority upon request as a validation report. However, experts agreed that independent of the

form of reporting, proof of robust assay validation is critical and should be available for an independent evaluation of study data.

General recommendations for assessing circulating hormones (HPT, HPG and non-EATS)

During the expert hearing, the best practice for hormonal measurements was discussed and consensus was reached regarding multiple parameters of study design, sampling, pre-analytical considerations and data analysis within the breakout groups. Some of the recommendations, however, appeared to be common for measurement of circulating hormones in toxicological studies and these are summarized below:

Considerations relevant to the study design

Number of animals Experts concluded that a minimum number of required animals should be established depending on the normal physiological inter-individual variability and sensitivity of the assay and method for determination of the point of departure, i.e. whether the derivation of a no-observed-adverse-effect level/lowest-observed-adverse-effect level (NOAEL/LOAEL) or a benchmark dose (BMD) approach is taken. It was recommended to perform statistical power analysis to establish the optimal group size. For the thyroid hormones and TSH, see the exemplary calculation above and in Table A.1 of Appendix B in ED Guidance document (ECHA and EFSA et al. 2018). For male and female reproductive hormones, power analysis calculations can be found elsewhere (Andersson et al. 2013; Stanislaus et al. 2012).

Considerations relevant to stress minimization and animal welfare

Levels of thyroid (e.g. T3, TSH) as well as some reproductive hormones (e.g. prolactin) can alter rapidly in response to stress (Balcombe et al. 2004). Thus, implementation of stress reduction/minimization is of particular importance in studies where hormonal endpoints are planned to be assessed. This is especially important in adult animals. Animal care and housing should meet the EU and national legislation requirements (EP and Council of the EU 2010). Standard group housing with three to five rats per cage of a suitable size is not expected to affect thyroid hormone concentrations, while isolated housing was considered to be more stressful. To reduce stress associated with the procedure of blood sampling, sampling time should not exceed 3 min per animal with and 1 min without anaesthesia,

Table 1 Quality criteria parameters to be considered for robust assay validation when designing and reporting on the hormonal measurement assay independent of the detection method applied

Hormonal axis	HPT	HPG	Non-EATS	Comments
Parameter				
Reproducibility across control groups	+		Pre-defined performance criteria should be met (e.g. EMA guideline as guidance)	Coefficients of variation (CV) should be calculated. For example, for control group, CV for T3/T4 below 25% and for TSH below 35% are required by OECD TG 407 (2008) and OECD TG 408 (2018). If a higher variation is observed in isolated cases, an explanation should be provided
Validation study with a positive control	+	+		Positive control study is essential to establish proficiency of the laboratory to perform hormonal assays, but inclusion of the positive control group in each assay was not considered as necessary. However, it may be dispensable for other species (e.g. dogs, monkeys) for animal welfare reasons and limited knowledge
Accuracy	+	+		EMA Guideline on bioanalytical method validation may serve as a reference. A comparison with international (animal) standards is needed when available (e.g. for rat TSH, NIH-RP2)
Sensitivity, limits of quantification	+	+		Limit of detection (LOD), lower (LLOQ) and upper limit of quantification (ULOQ) should be established by using sample matrix
Measurement range	+	+		Linearity should be proved of the whole measurement range and the upper limit should be set
Precision	+	+		EMA Guideline on bioanalytical method validation and existing recommendations for hormone measurements should serve as a guide
Matrix effects	+	+		If a different matrix is used (e.g. serum, plasma), an assay should be tested with each matrix independently. Lack of effect of matrix on the linear sample dilution should be proved as well
Cross-reactivity of antibodies (only for immunoassays)	+	+		Validation on the level of the reagent supplier can be acceptable
Quality control (QC) samples	+	+		The type of applied QC samples should be recorded: e.g. QC samples provided by kit manufacturer, spiked samples, biological control samples, reference range, etc.
Sample stability	+	+		The storage conditions should be validated
Species effect	+	+		For a given species the suitability of the chosen assay should be proven*
Lot-to-lot validation	+	+		Necessary upon availability of several lots of antibodies or upon change from one lot to the other. This is important if different kit lots are used in one study or results are compared to the normal range established in the laboratory

*From literature and communication with experts, the cross-reactivity of antibodies can be utilized for thyroid hormone measurement. For example, use of human ELISA kits for toxicological purposes for rat samples was described to achieve higher sensitivity

independent of the method used and sampling should be performed by a trained technician. If available, a separate room may be used for survival sampling. If animals are moved to another location (e.g. procedure rooms), at least 30 min should be given for acclimatization, but acclimatization should not necessarily be extended to 24 h. The required preparatory procedures (e.g. shaving prior to jugular vein bleeding) should be performed 1 day prior to sampling. The selection of the sampling technique should be based on the previous experience of the laboratory, considering the importance of reducing the stress associated with the procedure to minimum. For example, in adult animals, tail vein sampling should be avoided due to the stress associated with restraint, while cardiac puncture should not be used as survival sampling method due to animal welfare reasons. For pups, cardiac puncture or alternatively decapitation followed by collection of trunk blood was recommended as blood sampling methods of choice. For foetuses (not required in the EU legislative framework), collection of umbilical cord blood or alternatively decapitation followed by collection of trunk blood was recommended. For survival sampling, the maximum volume of collected blood should be determined according to the relevant EU and national welfare regulations.

Anaesthesia should be given prior to exsanguination for the purposes of euthanasia of adult animals or pups for terminal sampling. To this end, the use of isoflurane followed by exsanguination is recommended. The general consensus was that isoflurane is the anaesthetic of choice due to its rapid and smooth induction. Carbon dioxide gas should be avoided due to animal welfare reasons, namely induction often takes longer and animals exhibit restlessness and hyperactivity (Deckardt et al. 2007). Using isoflurane alone for euthanasia should be avoided in adults and ether for euthanasia of pups should be avoided due to laboratory personnel safety reasons.

Considerations relevant to circadian rhythm

Staggering and stratification/randomization during terminal killing Randomization/stratification of animals to control for the differences in the timing of the sampling is highly recommended when sampling. Certain recommendations regarding the length of the sampling procedure for hormones of HPG and HPT axes are provided below. If more time is required (e.g. terminal sampling), then staggering of animals can be implemented (e.g. parturition staggering); however, the same number of animals from the control and all treated groups should be sampled on the same day.

Pre-analytical parameters

Sample matrix and storage Considering that both plasma and serum can be used in the subsequent steps of hormone detection, the selection of the fraction can be made based on a protocol previously established in the laboratory. To obtain serum, whole blood should be left to clot for at least 30 min at room temperature and for further sample processing serum-separation tubes can be used. For plasma preparation, whole blood can be collected in heparin- and EDTA-treated tubes, while care should be taken to avoid usage of sodium-citrate-treated tubes. Upon separation, plasma or serum samples can be aliquoted and stored frozen for future processing. It is recommended that the validation of sample stability be performed under all circumstances including validation of sample storage conditions (e.g. temperature, length, freeze–thaw stability).

Data analysis and interpretation

Statistical analysis of data The statistical analysis of hormonal concentrations does not require application of any special test, but a suitable statistical method of analysis should be applied. No consensus was reached regarding the statistical treatment of outliers, with some experts arguing strongly against the removing of outliers as it may mask the biologically determined variability between responders and non-responders to the treatment. High variability of the data even if all precautions to limit analytical and physiological variation were taken still does not justify the removal of outliers. Nonetheless, exclusion of data points may be justified for physiological reasons [e.g. non-pregnant rats in a teratogenicity study (OECD TG 414, (OECD 2018b))] or because of technical failures (e.g. strong haemolytic sample, etc.).

Use of historical control data In regulatory toxicology studies, the effect size is normally determined in comparison to the concomitant control group, thus historical control data are rarely needed. Indeed inter-laboratory comparisons are even less probable if comparison to the intra-laboratory reference range for the purpose of result interpretation is not expected to be performed. Yet, each laboratory conducting hormone measurement assay(s) is expected to develop its own reference range(s) for quality control purposes. The reference range should be formed from data generated using the same method (including sampling time), performed in untreated animals of the same sex, strain and age kept under standardized housing/dietary/environmental conditions with hormone detected with the same analytical assay.

Specific recommendations for the HPT axis (T3, T4, TSH)

As evident from the survey performed prior to the expert hearing, thyroid hormones and TSH were found to be routinely measured in contract research laboratories, industrial and academic research centres, indicating that sufficient experience and expertise already exist. The dedicated breakout group reached a consensus on the number of parameters relevant for the measurements of T3, T4 and TSH in the context of the nonclinical toxicological studies.

Considerations relevant to the study design

Free vs bound T4/T3 hormone It was concluded that the measurement of (total) thyroid hormones as part of regulatory toxicology studies was technically feasible. At the same time, the measurement of free fractions of T3 and T4 represents a more complex technical challenge, which in addition requires a high volume of serum (up to 200 µl), and thus could be recommended as a follow-up mechanistic study. An indication for the mechanistic study would be de-regulation of T4 in only one species or discordance between T4 and TSH responses. When measurement of the free hormone is foreseen, the sample would require pre-treatment via ultracentrifugation or dialysis, followed by chromatography or a comparably sensitive detection technique.

Test animals The current recommendations were formulated for rats, thus the assay parameters would need to be adjusted according to the species examined, e.g. dogs for regulatory toxicology studies or sheep for research studies. Experts agreed that T4 and TSH should be measured in post-pubertal animals (around 12 weeks of age), at weaning and at PND 4 and T3 could be measured optionally at the same time points. The measurement of thyroid hormones can be performed in both male and female animals, while preliminary synchronization of females was not considered to be necessary. In general, eight to ten animals per group should be sufficient to detect statistically significant differences between the control and treated groups. A preliminary power analysis is recommended to predict the minimum effect size that is likely to be detected if a lower number of animals are used, for example, as in the OECD TG 407.

Terminal sampling Terminal sampling is expected to provide adequate information for hazard identification without compromising other study end points. However, the interpretation of data may be difficult if terminal sampling is performed in senescent animals.

Circadian rhythm Blood samples should be collected in the morning, between 8 a.m. and noon for the measurement of thyroid hormones. Samples must be taken in a randomized manner, but within the shortest possible time frame (within 2 h) to minimize impact of the circadian variation on the outcome of the measurement. Experts concluded that up to 40 animals can be sampled within a given time frame of 2 h for survival sampling, while for terminal sampling staggering would be possible.

Analytical considerations

To quantify the hormone, any detection method that has undergone a robust validation as described above is suitable for the measurement of hormonal concentrations. The laboratory survey demonstrated that laboratories use radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), multiplexing and mass spectrometry (MS/MS) technologies to measure TSH, T3 and T4.

Data interpretation

As already mentioned, measurement of a battery of hormones (i.e. at least T4, TSH and T3 for thyroid assessment) was recommended for meaningful data interpretation. As the concurrent control group is used as a reference point, measurement of the hormonal levels at the baseline was not necessarily recommended. In the occasion of non-concordant changes in hormones' concentrations, a mechanistic follow-up study is to be considered. However, it should as well be noted that in cases of a non-concordance, attention should be paid to the biological plausibility of the changes, along with the measured variability. In those cases, when the effects on thyroid are investigated in juvenile rats, a histopathology atlas such as the Atlas of Histopathology of Juvenile Rat (Parker 2016) can be consulted to facilitate interpretation of the results obtained in the studies with juvenile rats.

Specific recommendations for the HPG axis (oestradiol, progesterone, LH, FSH and testosterone)

The measurements of oestradiol, progesterone, LH, FSH and testosterone provide important input to the identification of a substance with endocrine disrupting properties. Among survey respondents reporting measurement of the HPG hormones, 12, 13, 10 and 15 respondents reported measurement of oestrogen, progesterone, FSH and LH, while 16 reported measurement of testosterone. The following recommendations are given to assist in design of reproductive hormone measurements and were developed for oestradiol, progesterone LH, FSH and testosterone.

Considerations relevant to the study design

When reproductive hormone investigation was necessary, experts once again agreed that assessment of the battery of hormones (i.e. at least LH, FSH, oestradiol and progesterone for female hormonal reproductive status) would be preferential to the measurement of a single hormone. Selection of particular hormones should be done based on the results of preceding toxicological studies or any relevant information that would facilitate generation of the hypothesis driving the investigation. It might be necessary to adapt the study to accommodate specific pattern of hormone changes in response to oestrous cycle or circadian rhythm, as some hormones are more prone to the variation in response to stress or circadian rhythm (e.g. prolactin) than others (e.g. LH, testosterone). For more specific recommendations on the design of male and female reproductive hormone measurements, reviews by the Scientific and Regulatory Policy Committee might be helpful (Andersson et al. 2013; Chapin and Creasy 2012).

Test animals Differences between males and females should be addressed in the study design, as information collected from both sexes may be helpful in the assessment of overall changes in reproductive hormones. The stage of the oestrous cycle should be considered in females at the time point of blood sampling, while synchronization was not considered as absolutely necessary. However, it should be noted that in a study with 10–15 females per group when divided by four stages of the cycle, only 3–4 females will be at the same stage of the cycle, per se reducing the power of the experiment (Biegel et al. 1998; Stanislaus et al. 2012). Due to the fact that concentrations of the hormones are influenced by the stage of the cycle which in turn is strictly regulated by circadian rhythm, sampling in the intact females should be properly timed. In such cases, it might be more helpful to design an additional mechanistic study where female rats are sampled according to the stage of the oestrus cycle and/or time of the day (e.g. early evening for decreased prolactin and oestradiol) (Andersson et al. 2013).

Circadian rhythm In general, to measure reproductive hormones, blood sampling should take place in the morning (unless another time of the day is preferential) within 3 h. The stage of the oestrous cycle should be considered when stratifying females.

Consideration relevant to stress minimization It was recommended to use anaesthesia prior to blood sampling to mitigate the stress. However, any direct impacts of the anaesthetic on the hormone level must be considered (Deckardt et al. 2007; Nazian 1988).

Data interpretation

Some of the experts suggested that the results should be interpreted by considering any potential interaction with mechanisms resulting from excessive toxicity; for this reason, changes observed only at the high dose (at or above the maximum tolerated dose) should be evaluated with caution.

Specific recommendations for non-EATS hormones and hormone-like substances

Compared to the EATS-related hormones, only a limited number of responses were collected for non-EATS-related hormones and hormone-like substances (2, 3, 5 and 5 for IGF, vitamin D, insulin and PTH, respectively). These molecules and the retinoic acid pathway were discussed by experts in the dedicated breakout group. The following practical recommendations were generated mainly for measurements of insulin and IGF-1. Some considerations were expressed by experts relevant to PTH and vitamin D, while no specific recommendations were developed for retinoid (vitamin A) analyses in the blood.

Insulin and insulin-like growth factor-1

The measurement of insulin in non-clinical toxicological studies is a helpful indicator of metabolic changes or of more profound alterations induced by the treatment with the test substance leading to the development of the metabolic disease. Measurements of IGF-1 (and IGF-2) are recommended as indicators of altered growth (e.g. body composition, fat pads, bone length, e.g. femoral–tibial length) observed in the extended one-generation reproductive toxicity study (OECD TG 443, (OECD 2018c)) or in any other repeated-dose toxicity test.

Study design The rat and mouse were considered acceptable species for investigations of the effects of treatment on the IGF-1 and insulin levels. Measurement of IGF-2 during foetal and IGF-1 during post-natal development was suggested. Both male and female animals can be used for the measurement of IGF-1 and insulin; however, sex-specific differences in IGF-1 concentrations (e.g. during puberty) may exist according to some of the experts. If repeated sampling is required, e.g. for IGFs and insulin measurements, the use of satellite groups would be necessary, while the amount of blood collected during repetitive sampling should meet national and EU animal welfare regulations.

Stress minimization With regard to animal handling, stress minimization was mentioned as a desirable, but probably not critical factor when IGF-1 and insulin are measured.

Blood sampling Insulin should be measured in fasting animals.

Analytical considerations ELISA is considered an applicable detection method for the quantification of IGF-1 and insulin when relevant calibration of the instrument is guaranteed and the assay fulfils the performance criteria (e.g. EMA guideline as guidance (EMA 2011)).

Parathyroid hormone

No specific recommendations were provided for the PTH. The measurement of the hormone is technically feasible; however, other parameters might be more informative than PTH to assess changes in skeletal phenotype, for example serum calcium or measurement of bone mass.

Vitamin A (retinol and retinoic acid)

No specific recommendations were developed for the measurement of retinoids. It was noted that there might be a future need for analytical capacity in this area to match current regulatory awareness and possible future demands. While biologically inactive retinol (vitamin A) can be readily detected in serum (Arnold et al. 2012), its relevant metabolite retinoic acid might require establishment of sensitive LC–MS/MS methods. In any case, measurements of retinoic acid should be interpreted in the context of histopathological changes in the liver, eyes and other tissues/organs.

Conclusions

Overall, the expert hearing led to the establishment of quality criteria and gave practical recommendations for study design, study conduct and the evaluation of T3, T4, TSH and reproductive hormones, as well as for some non-EATS hormones and hormone-like substances. The principle conclusions of the meeting can be summarized as follows:

1. An extensive body of expertise and knowledge in laboratories routinely conducting measurement of T3, T4 and TSH and reproductive hormones (oestradiol, progesterone LH, FSH, testosterone) already exists.
2. Different methods (e.g. hormone quantification, blood sampling procedures) may be used in different laboratories, but it is essential to justify their use and critical to provide evidence of their validation.
3. Study designs should be tailored to hormonal physiology and with special consideration towards stress reduction

measures. Animal and sampling randomization/stratification methodologies should be used.

4. The measurement of a battery of hormones relevant to a given axis is preferred to measurement of single hormones (i.e. at least T3, T4 and TSH for thyroid assessment and at least LH, FSH, oestradiol and progesterone for female hormonal reproductive status).
5. For reproductive hormones, the selection of hormones should be based on information from prior toxicological studies with the substance.
6. If incorporation of hormonal evaluation in routine regulatory toxicology studies is not feasible, then a specific mechanistic study may be conducted and this strategy is preferred to address the MOA evaluation specifically.
7. Changes in hormone levels cannot be used as “stand-alone” determinants, rather they should contribute to the weight of evidence for the identification of endocrine disrupting properties to complement such toxicological information as growth, reproductive performance, organ weights and histopathology.
8. The lack of experience in measuring non-EATS hormones in the toxicological context represents an important gap for the risk assessment of endocrine active substances; therefore, a cross-disciplinary collaboration between pharmacologists, endocrinologists and toxicologists is strongly recommended.

Many assays for the exploration of endocrine MoAs already exist; however validation at an OECD level is either ongoing or has not yet been initiated (Solecki et al. 2017), thus delaying their applicability in the toxicological studies. Currently, the situation is even more complex for non-EATS modalities due to the lack of adequate testing strategies and our understanding of the interpretation of the results. Targeted research is needed as a consequence. It is expected that a more efficient assessment of substances with endocrine disrupting properties could be achieved by closing knowledge gaps and including further validated assay/test methods into the OECD Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (OECD 2018g), to support the interpretation of the results and reduce uncertainties.

Acknowledgements This publication is based on discussions held during the international expert hearing organized by the German Federal Institute for Risk Assessment (BfR) in Berlin in October 2017, Germany. The authors would like to express their gratitude to the responders of the survey that provided background for the discussions, to participants of the expert hearing for their time and valuable input, to Andrea Terron, the European Food Safety Authority (EFSA), and Sharon Munn, the Joint Research Centre (JRC) for their input into the project and to the team of the BfR Academy for their support with organization of the expert hearing. The expert hearing was part of the grant collaboration GP/EFSA/PRAS/2017/01.

Compliance with ethical standards

Conflict of interest The authors affiliated with the BfR, ANVISA, Charité, Universitätsmedizin Berlin, Istituto Superiore di Sanità, Karolinska Institutet, BASF SE, CRL, Bayer AG (Pharmaceuticals), DuPont Haskell Global Center for Health & Environmental Sciences, CTGB, ECHA and Bayer SAS (Bayer CropScience) declare no conflicts of interests.

Disclaimer This publication has been approved by the German Federal Institute for Risk Assessment; however, the opinions expressed in the publication do not represent regulatory position of the institute or organizations, whose representatives participated in the expert hearing.





Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Andersson H, Rehm S, Stanislaus D, Wood CE (2013) Scientific and regulatory policy committee (SRPC) paper: assessment of circulating hormones in nonclinical toxicity studies III. female reproductive hormones. *Toxicol Pathol* 41(6):921–934. <https://doi.org/10.1177/0192623312466959>
- Arnold SL, Amory JK, Walsh TJ, Isoherranen N (2012) A sensitive and specific method for measurement of multiple retinoids in human serum with UHPLC-MS/MS. *J Lipid Res* 53(3):587–598. <https://doi.org/10.1194/jlr.D019745>
- Balcombe JP, Barnard ND, Sandusky C (2004) Laboratory routines cause animal stress. *Contemp Top Lab Anim Sci* 43(6):42–51
- Biegel LB, Cook JC, Hurtt ME, O'Connor JC (1998) Effects of 17 beta-estradiol on serum hormone concentrations and estrous cycle in female Crl:CD BR rats: effects on parental and first generation rats. *Toxicol Sci* 44(2):143–154. <https://doi.org/10.1006/toxs.1998.2469>
- Chapin RE, Creasy DM (2012) Assessment of circulating hormones in regulatory toxicity studies II. Male reproductive hormones. *Toxicol Pathol* 40(7):1063–1078. <https://doi.org/10.1177/0192623312443321>
- Deckardt K, Weber I, Kaspers U, Hellwig J, Tennekens H, van Ravenzwaay B (2007) The effects of inhalation anaesthetics on common clinical pathology parameters in laboratory rats. *Food Chem Toxicol* 45(9):1709–1718. <https://doi.org/10.1016/j.fct.2007.03.005>
- European Commission (EC) (2017) Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. *Off J Eur Union* L301:1–5
- European Commission (EC) (2018) Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. *Off J Eur Union* L101:33–36
- European Chemicals Agency (ECHA) and European Food Safety Authority (EFSA) with the technical support of the Joint Research Centre (JRC), Andersson N, Arena M, Auteri D, Barmaz S, Grignard E, Kienzler A, Lepper P, Lostia AM, Munn S, Parra Morte JM, Pellizzato F, Tarazona J, Terron A, Van der Linden S (2018) Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. *EFSA J* 16(6):e05311. <https://doi.org/10.2903/j.efsa.2018.5311> (ECHA-18-G-01-EN)
- European Food Safety Authority (EFSA) (2013) Scientific opinion on the hazard assessment of endocrine disruptors: Scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. *EFSA J* 11(3):3132. <https://doi.org/10.2903/j.efsa.2013.3132>
- European Medicines Agency (EMA) (2011) Guideline on bioanalytical method validation. EMEA/CHMP/EWP/192217/2009 Rev. 1 Corr. 2, 21 July 2011, EMA, London, p 23
- European Parliament and Council of the European Union (EP and Council of the EU) (2009) Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. *Off J Eur Union* L309:1–50
- European Parliament and Council of the European Union (EP and Council of the EU) (2010) Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Off J Eur Union* L276:33–79
- European Parliament and Council of the European Union (EP and Council of the EU) (2012) Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. *Off J Eur Union* L167:1–123
- Nazian SJ (1988) Serum concentrations of reproductive hormones after administration of various anesthetics to immature and young adult male rats. *Proc Soc Exp Biol Med* 187(4):482–487
- Organisation for Economic Co-operation and Development (OECD) (2008) Test No. 407: Repeated dose 28-day oral toxicity study in rodents, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2016) Test No. 422: Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018a) Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents, OECD Guidelines for the Testing of Chemicals, Sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018b) Test No. 414: prenatal developmental toxicity study, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018c) Test No. 443: extended one-generation reproductive toxicity study, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018d) Test No. 451: carcinogenicity studies, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018e) Test No. 453: combined chronic toxicity/carcinogenicity studies, OECD guidelines for the testing of chemicals, sect. 4. OECD Publishing, Paris
- Organisation for Economic Co-operation and Development (OECD) (2018f) Work plan for the test guidelines programme. September 2018. <http://www.oecd.org/env/ehs/oecd-guidelines-testing-chemicals-related-documents.htm>. Accessed 7 Mar 2019
- Organisation for Economic Co-operation and Development (OECD) (2018g) Revised guidance document 150 on standardised test

- guidelines for evaluating chemicals for endocrine disruption, OECD series on testing and assessment, No. 150. OECD Publishing, Paris. <https://doi.org/10.1787/9789264304741-en>
- Parker G, Picut C (2016) Atlas of histology of the juvenile rat, 1st edn. Elsevier Academic Press, Boston
- Solecki R, Kortenkamp A, Bergman A et al (2017) Scientific principles for the identification of endocrine-disrupting chemicals: a consensus statement. *Arch Toxicol* 91(2):1001–1006. <https://doi.org/10.1007/s00204-016-1866-9>
- Stanislaus D, Andersson H, Chapin R et al (2012) Society of toxicologic pathology position paper: review series: assessment of circulating hormones in nonclinical toxicity studies: general concepts and considerations. *Toxicol Pathol* 40(6):943–950. <https://doi.org/10.1177/0192623312444622>
- United States Environmental Protection Agency (US EPA) (2005) Guidance for thyroid assays in pregnant animals, fetuses and postnatal animals, and adult animals, 24 October 2005. Office of Pesticide Programs, Health Effects Division, Washington DC. <https://www.epa.gov/pesticide-registration/guidance-thyroid-assays-pregnant-animals-fetuses-and-postnatal-animals-and>. Accessed 7 Mar 2019
- United States Environmental Protection Agency (US EPA) (2009a) Endocrine disruptor screening program test guidelines OPPTS 890.1450: pubertal development and thyroid function in intact juvenile/peripubertal female rats. US EPA, Cincinnati, OH
- United States Environmental Protection Agency (US EPA) (2009b) Endocrine disruptor screening program test guidelines OPPTS 890.1500: pubertal development and thyroid function in intact juvenile/peripubertal male rats. US EPA, Cincinnati, OH
- Viswanathan CT, Bansal S, Booth B et al (2007) Quantitative bio-analytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm Res* 24(10):1962–1973. <https://doi.org/10.1007/s11095-007-9291-7>
- World Health Organization/International Programme on Chemical Safety (WHO/IPCS) (2002) Global assessment of the state-of-the-science of endocrine disruptors. World Health Organization, Geneva
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Olena Kucheryavenko¹  · Glenn Lurman¹ · Anja Lehmann¹ · Juliana Braz^{1,3} · Lars Niemann¹ · Ibrahim Chahoud⁴ · Alberto Mantovani⁵  · Helen Håkansson⁶  · Steffen Schneider⁷ · Volker Strauss⁷ · Pragati S. Coder⁸ · Alexius Freyberger⁹ · John C. O'Connor¹⁰ · Martina Rauch^{2,14} · Kostja Renko⁴ · Marize L. M. Solano^{1,11} · Niklas Andersson¹²  · Olivier Blanck¹³ · Vera Ritz¹ · Roland Solecki¹

¹ German Federal Institute for Risk Assessment, Berlin, Germany

² European Food Safety Authority (EFSA), Parma, Italy

³ Present Address: Brazilian Health Regulatory Agency (ANVISA), Brasília, Brazil

⁴ Institute of Experimental Endocrinology, Charité-Universitätsmedizin Berlin, Berlin, Germany

⁵ Istituto Superiore di Sanità, Rome, Italy

⁶ Karolinska Institutet, Stockholm, Sweden

⁷ BASF SE, Ludwigshafen, Germany

⁸ Charles River Laboratories, Ashland, OH, USA

⁹ Bayer AG, Pharmaceuticals, Berlin, Germany

¹⁰ DuPont Haskell Global Center for Health & Environmental Sciences, Newark, Delaware, USA

¹¹ Present Address: CTGB, Ede, The Netherlands

¹² European Chemicals Agency (ECHA), Helsinki, Finland

¹³ Bayer S.A.S., Bayer CropScience, Valbonne, France

¹⁴ Present Address: German Federal Institute for Risk Assessment, Berlin, Germany