Organisation for Economic Co-operation and Development

ENV/JM/MONO(2018)18

Unclassified

English - Or. English

1 March 2019

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

**Cancels & replaces the same document of 7 December 2018**

# FEASIBILITY STUDY FOR MINOR ENHANCEMENTS OF TG 414 (PRENATAL DEVELOPMENTAL TOXICITY STUDY) WITH ENDOCRINE DISRUPTER-RELEVANT ENDPOINTS

SERIES ON TESTING AND ASSESMENT
Number 285

JT03443968

**OECD Environment, Health and Safety Publications**

**Series on Testing and Assessment**

**No. 285**

# FEASIBILITY STUDY FOR MINOR ENHANCEMENTS OF TG 414 (PRENATAL DEVELOPMENTAL TOXICITY STUDY) WITH ENDOCRINE DISRUPTER-RELEVANT ENDPOINTS

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

**Environment Directorate**
**ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT**
Paris 2018

## About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 35 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

*This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.*

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

**This publication is available electronically, at no charge.**

**For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/ehs)**

**or contact:**

**OECD Environment Directorate,**
**Environment, Health and Safety Division**
**2, rue André-Pascal**
**75775 Paris cedex 16**
**France**

**Fax : (33-1) 44 30 61 80**

**E-mail : ehscont@oecd.org**

# *FOREWORD*

This document is a Feasibility Study Report for minor enhancements to TG 414 Prenatal Developmental Toxicity Study to include endocrine disrupter-relevant endpoints. The Feasibility report was prepared by Denmark (lead), and was part of Project 4.100 added to the Test Guidelines Programme work plan in April 2015.

A call for TG 414 data was organised in 2015 and based on analyses of data identified, this report addresses the feasibility, scientific and technical concerns regarding inclusion of additional endpoints related to endocrine disruption measured in the rat fetus and dam.

At the May and October 2017 EDTA meetings, Denmark presented initial findings on the sensitivity of anogenital distance measurements, peripheral testosterone levels and peripheral thyroid hormones (T4, T3, and TSH) in fetuses and thyroid hormones measured in dams. The documents discussed during the October 2017 meeting and synopsis of the discussion are available in the meeting summary record posted on the Clearspace site (https://community.oecd.org/community/edta).

The draft report was revised following input from an expert group on Reproductive and Developmental Toxicity. The expert group provided guidance through a series of teleconferences, the summary records of which can be found on the Clearspace site for Expertise in Test Guidelines (http://community.oecd.org/community.tgeg).

Additionally, the draft Feasibility Report and revised TG 414 were circulated to the WNT and EDTA for two written commenting rounds in October and December 2017. The revised final draft TG 414 is also available ENV/JM/TG (2018)18.

## ACKNOWLEDGEMENTS

The Feasibility study for minor enhancements of TG 414 with Endocrine Disrupter-relevant endpoints was prepared by Sofie Christiansen & Ulla Hass (Division of Diet, Disease prevention and Toxicology, Research Group for Molecular and Reproductive Toxicology, National Food Institute, Technical University of Denmark).

The data was collected and analysed by lead country and discussed with the OECD expert group on developmental and reproductive toxicity (EG). Technical and scientific issues were discussed in the EG and a superior statistical model was developed by US EPA statisticians and reanalyzed by BIAC representatives in the EG. The OECD EG also provided support to the lead country during the development of the feasibility report and the revision of the Test Guideline 414, and held several teleconferences between November 2015 and February 2018.

# Table of contents

## Terms of reference

1. This Feasibility report for minor enhancements of TG 414 (Pre-natal Developmental Toxicity Study) with ED (endocrine disrupter)-relevant endpoints was prepared by the Division of Diet, Disease prevention and Toxicology, Research Group for Molecular and Reproductive Toxicology, National Food Institute, Technical University of Denmark, which is leading the project in OECD. The report has given input for discussions in various groups in OECD. Subsequently, the report has been revised based on input after discussions in the OECD Expert Group on Reproductive and Developmental Toxicity Testing, at EDTA meetings and in WNT commenting rounds in OECD.

## Aim

2. The aim of this project is to do a Feasibility study for minor enhancements of TG 414 (Prenatal Developmental Toxicity Study) with ED-relevant endpoints. This report addresses feasibility, scientific and technical concerns regarding inclusion of these additional ED related endpoints in TG 414. The endpoints considered in this project included anogenital distance (AGD) in all fetuses, testosterone in male fetuses and guidance for genital malformations in all fetuses. Moreover, later in the process also thyroid measurement in fetuses and in the dams has been considered to be included. For these endpoints, the scientific and technical questions considered include:

- Are standardized methods available?
- Is the sensitivity sufficient with the number of litters per group?
- Are the endpoints of relevance for humans?
- Are there animal welfare concerns?
- Is the enhancement possible without changes or with only minor changes in study design?

## Background and expected regulatory need/data requirement that will be met by the proposed outcome of the project

3. A scientific approach will be used to give input to the existing TG 414 (Prenatal Developmental Toxicity Study) in relation to the feasibility of inclusion of sensitive endpoints in all fetuses and dams for detection of chemicals with endocrine disrupting properties.

4. The specific purpose of this project is to consider the relevance and feasibility of enhancement of the OECD 414 (OECD, 2001). The TG 414 provides information on adverse effects on prenatal development and is used in various regulatory frameworks (such as REACH and several pesticide regulations) to generate information for risk assessment of chemicals.

5. OECD TG 414 is included in Level 4 (OECD conceptual framework) as the TG involves repeated dosing of pregnant females and therefore potential exposure of the developing fetus. The assay includes some endpoints that may detect endocrine disruption (OECD, 2018a).

6. However, this update in relation to inclusion of more endpoints in TG 414 in dams or fetuses at the time of caesarean section could be a significant enhancement with regards to detection of effects of endocrine disrupting substances.

7. After the inclusion on the work plan in OECD (as project number 4.100) the possibility to include thyroid hormones in the dams and/or fetuses has been mentioned in the OECD Expert Group on Reproductive and Developmental Toxicity Testing as well as at international meetings in 2017 (DG-Environment, 2017) and EDTA meetings in OECD in May and October 2017.

8. The OECD Expert Group on Reproductive and Developmental Toxicity Testing (EG) from the update on TG 421/422 (Project 4.71 on OECD Workplan) has been convened as the scientific discussion is similar. Additional experts have been invited to the EG via the WNT NCs. These EG and the EDTA meetings in 2017 (May and October) have provided guidance on which endpoints to be considered and the feasibility (e.g. timing and logistics) based on a proposal from the lead.

9. DK has undertaken the examination of available existing data. Data have been received from OECD countries and also peer reviewed scientific relevant papers have been included to make a proposal to the EG on whether or not it is relevant to include the ED related endpoints in a proposal for revision of OECD TG 414.

10. It will also be considered whether certain slight adaptions of the test design of the test guideline may be warranted to include for consideration if other ED related endpoints are suggested by the EG for this project. However, the timing of the OECD TG 414 study cannot be changed as was the case in TGs 421/422 screening studies (which was terminated later due to assessment of Nipple retention).

11. The results of this project may contribute to an improved sensitivity for identification of developmental toxicants in mammalian species at an early stage in the regulatory testing schemes for industrial chemicals (e.g. REACH) as information from TG 414 is already required in such regulatory testing schemes.

12. If these endpoints are implemented in TG 414 it will enhance the international harmonization of hazard assessment with regard to developmental toxicity effects.

13. An important point is that the ability for detection of EDs should be enhanced without increasing the number of experimental animals used.

14. TG 414 is designed to provide general information concerning the effects of prenatal exposure on the pregnant test animal and on the developing organism; this may include assessment of maternal effects as well as death, structural abnormalities, or altered growth in the fetus. The proposed update of TG 414 must not impair the ability to fulfil the purpose of TG 414.

15. Assessment of AGD in both sexes is mandatory in TG 443 and TGs 421/422 and this report will elucidate whether this endpoint could also be included in TG 414 at the day of caesarean section.

16. TG 414 was revised in 2001 but not with regard to inclusion of ED relevant endpoints. It seems relevant to include some ED relevant endpoints in TG 414 as the exposure periods cover some of the sensitive periods for sexual differentiation (prenatal period). The proposed endpoints are described below.

17. The OECD TG 407 (Repeated dose 28- day oral toxicity study in rodents) has been updated in 2008. The assay has been validated for some endocrine endpoints but the

sensitivity of the assay is not sufficient to identify all EATS-mediated EDs. The validation of the assay (OECD, 2006) showed that it identified strong and moderate EDs acting through the ER and AR; and EDs weakly and strongly affecting thyroid function. It was relatively insensitive to weak EDs acting through the ER and AR. This assay also has some optional endpoints such as uterine and ovary weight, vaginal cytology (to provide a marker for physiological estrogenicity that would guide the histological interpretation of the ovary and estrogen-sensitive tissues), histopathologic changes in mammary gland histopathology (mandatory in females optional in males) as well as serum T3, T4, TSH as well as thyroid weight which can be examined if there is additional concern.

18. The extended one-generation reproductive toxicity study (EOGRTS) (OECD, 2012) includes more endpoints sensitive to endocrine disruption than OECD TG 416 (OECD, 2001) and, as it also uses reduced animal numbers if conducted without F2. It is expected that it will often replace OECD TG 416 for mammalian reproductive toxicity testing (OECD GD 150, 2018a). Endpoints sensitive to endocrine disruption include anogenital distance at birth, areola/nipple retention, measurement of thyroid hormones and TSH levels. Effects on the developing nervous and immune systems are also assessed by the DNT and DIT cohorts. These systems may also be sensitive to endocrine influences. This test is also expected to have greater sensitivity than OECD TG 416 as it requires an increased number of pups to be examined. In summary, the new EOGRTS (OECD TG 443) is preferable for detecting endocrine disruption because it provides an evaluation of a number of endocrine endpoints in the juvenile and adult F1, which are not included in the 2-generation study (OECD TG 416) adopted in 2001.

19. In 2015 and 2016, the OECD 421 (Reproduction/Developmental Toxicity Screening Test) and OECD 422 (Combined Repeated Dose Toxicity Study with the Reproduction/ Developmental Toxicity Screening Test) guidelines were revised to include ED relevant endpoints (OECD, 2016a; OECD 2016b; OECD, 2015).

20. In April 2015, OECD launched this feasibility study for the enhancement of OECD 414 with selected parameters intended to increase the detection of EATS disrupting potential. In the autumn 2015 the lead and OECD secretariat requested data (control litter means + SD) from OECD member states to enhance TG 414 with endocrine disrupter relevant endpoints. All conclusions can be found in paragraph 100 of this report.

## Anogenital distance (AGD)

### *Methodology*

21. New-born male rats have no scrotum, and the external genitalia are undeveloped, and only a genital tubercle is apparent for both sexes. The AGD is the distance from the anus to the insertion of this tubercle, the developing genital bud. The AGD is androgen dependent, and studies show that the AGD is normally about twice as long in male as in female rats. Similarly, in new-born humans the AGD measure was about two-fold greater in males than in females (Salazar-Martinez et al. 2004). At caesarean section, the distance between the proximal end of the anus and the genital tubercle of all fetuses must be measured. Anogenital distance is a non-invasive measure of *in utero* androgenicity. The distance will be measured from the base of the genital tubercle to the proximal end of the anal opening using a dissecting microscope with a micrometer eyepiece or another sensitive method. The fetal weight will be measured to derive anogenital index and the measurement need to be included as a covariate in the statistical analysis of AGD (OECD, 2008).

22. In TG 414 the AGD will be measured in both male and female fetuses in all litters at caesarean section one day prior to the expected day of delivery (OECD, 2018b). This report uses data from Wistar rats and Sprague Dawley rats which are relevant species commonly used in prenatal developmental toxicity testing.

### *Data analysis, sensitivity and power*

23. Important parameters when evaluating the sensitivity and power of the data are i) the sample size of the data, ii) the standard deviation SD ($\sigma$) and iii) the target biological difference or ii) the coefficient of variation CoV (the ratio of the standard deviation to the mean of the sample) and iii) the target percent biological difference in the population mean $\mu$.

24. The standard deviation is an expression of how much the value disperses from the population mean ($\mu$). Coefficient of variation is expressed as $\sigma/\mu$ and is often evaluated as a percentage and therefore expresses the standard deviation as the percentage of the population mean, $\mu$. In the feasibility study of the screening studies (TGs 421/422), all statistical analysis was done on the basis of the coefficient of variation (OECD, 2015).

25. In order to take into account the size of the rat when evaluating the AGD, the AGD was divided by the cubic root of the body weight, i.e. (AGD [mm])/$\sqrt[3]{}$(body weight [g]) resulted in the anogenital distance index (hereafter AGDI).

26. The rat grows rapidly, as growth occurs in three dimensions so that body weight can be viewed as a cubic measure. In contrast, AGD is a purely linear metric. Therefore the relationship between AGD and body weight should be more properly evaluated using the cube root of body weight as with AGDI or using the bodyweight as covariate in statistics (OECD, 2015; OECD, 2008).

#### *Data collection*

27. The evaluation of both AGD and plasma testosterone levels as endpoints is based on data from studies at DTU National food institute (only published), external (non-published data from the expert group) and published studies primarily from Saillenfait et al. from The French National Research and Safety Institute for the Prevention of Occupational Accidents and Diseases (INRS) (Saillenfait, et al. 2009; Saillenfait et al, 2011a and 2011b; Saillenfait, et al. 2013). The references for the data from DTU Food data is (Taxvig, et al. 2007; Taxvig, et al. 2008; Taxvig, et al. 2013; Hansen, et al. 2009; Kristensen, et al. 2011; Boberg et al. 2011).

28. Unfortunately, the data sent from the expert group (unpublished) could not be included in the analysis. One dataset missed the information about litter affiliation and thereby group means based on litter means could not be obtained from these data. The other dataset did not give any details on the strain of rodent and could therefore not be compared with others.

29. Furthermore, a literature search was conducted to increase the amount of data. The parameters for this search were based on the dosing period requirements in TG 414. "*Normally, the test substance should be administered daily from implantation (e.g., day 5 post mating) to the day prior to scheduled caesarean section* […] *Females should be killed one day prior to the expected day of delivery*." (OECD, 2001).The dosing period for the rats used in the DTU group (in house data) was GD 7 to 21 and GD 6-21 for the French group. GD 0 is understood as the day of mating and GD 21 is 3 weeks after (our Wistar or SD rats give birth at day 22 or 23).

30. In general, only few published papers satisfied the dosing period requirement in TG 414, either because the administration of the compound was initiated later on (e.g. GD 12-GD 17), or because the caesarean section was conducted before the expected day of delivery e.g. GD 19. Studies that did not sufficiently meet the dosing duration of OECD 414 were excluded from the analysis (Parks et al., 2000; Ema et al., 2003; Fisher et al., 2003; Hotchkiss et al., 2004; Thompson, et al., 2004; Liu et al., 2005; Saillenfait et al., 2016). A reason for this deviation could be the scientific interest in the developmental stages in the embryotic phase or in the testosterone peak levels during the fetal development.

31. Another important criterion was the presence of data in the papers that could actually be used for this analysis. Hence, data presentation in the form of graphs or bar charts was excluded because of inaccuracies when transforming these and therefore only data in the form of tables was included.

### *Results*

32. To evaluate whether a specific endpoint is sensitive enough to be included in a test guideline, it is important to consider the control values. If large variations are present in these samples, the identification of any effect is hampered.

**Table 1**. Overview of male AGD in control data GD 21, mean ± SD. Averages for all control group means, control group standard deviations (calculated using litter means) and coefficients of variation (of litter means) for the control groups are given. In relation to all calculated averages, the standard deviation (of litter means) is depicted. The total number of studies and the range of the number of litters for the control groups are also given. Importantly, the controls are also split with regards to the rat strain used. The data AGDI mean is a group mean (control) based of litter means in each study.

| | AGDI mean [mm/g] | AGDI standard deviation [mm/g] | AGDI Coefficient of variation [%] | Control experiments (range of litters) |
|---|---|---|---|---|
| All control experiments | 2.00 ± 0.36 | 0.08 ± 0.04 | 4.03 ± 1.67 | 17 (3-25) |
| Control experiments with Wistar rats | 2.31 ± 0.14 | 0.10 ± 0.04 | 4.49 ± 1.69 | 9 (3-18) |
| Control experiments with Sprague-Dawley rats | 1.72 ± 0.21 | 0.07 ± 0.02 | 3.99 ± 1.09 | 8 (6-25) |

33. An overview of all control data of AGD at GD 21 including the overall mean, standard deviation of litter means (i.e. between litter variation), and coefficient of variation of litter means are given in Table 1.

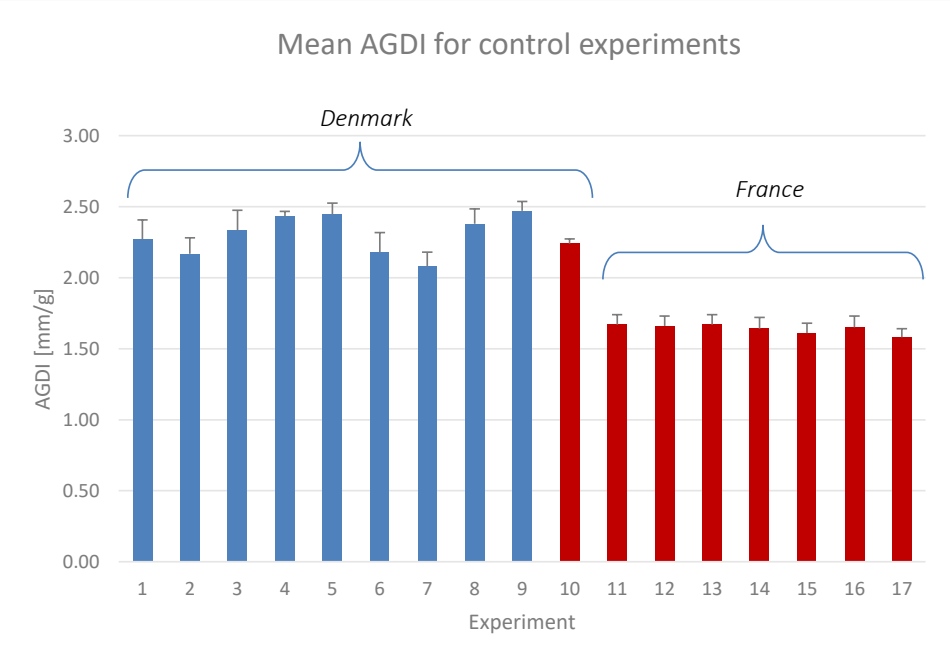


**Figure 1**. Mean AGDI +SD at GD 21 male fetuses, i.e. the mean with corresponding SD for each control group for the studies included. Blue bars indicate Wistar rats (N=3-18 litters) and red bars indicate Sprague-Dawley rats (6-25 litters). The DTU group in general uses Wistar rats (only one study with Sprague Dawley), whereas the French research group uses Sprague-Dawley rats only.

34. As seen in table 1 the combined Coefficient of Variation (of litter means) in male control fetuses is quite low (overall 4.03±1.67) even though most studies have much fewer litters than used in TG 414. However, it is also observed (Table 1, Figure 1) that Sprague-Dawley rats in the French studies have a lower mean AGDI, standard deviation and coefficient of variation compared to Wistar rats.

**Table 2.** Comparison of coefficients of variation for AGDI measured prenatally and postnatally. The given coefficients of variation are group means (control) based on litter means of all control group coefficients, mean ± SD. The total number of studies/experiments and the range of number of litters for the control groups are given. Only data on Wistar rats (from table 1) is included from DTU, DK to assure that data is from same laboratory.

| | AGDI Coefficient of variation | Control experiments (range of litters) |
|---|---|---|
| Prenatal AGDI | 4.49 ± 1.69 | 9 (3-18) |
| Postnatal AGDI | 4.00 ± 1.50 | 23 (3-21) |

35. To evaluate whether the prenatal AGD measurement is sensitive enough it is also relevant to compare to postnatal data. The endpoint AGD is not feasible to include if the prenatal measurements has a much higher variation than postnatal measurement. The measurements must be consistent with standard deviations and coefficients of variation not greatly influenced by the gestation day compared to postnatal measurements. Table 2 and

figure 2 outlines such a comparison between coefficients of variation for AGDI measured pre- and postnatally.

36. As seen in table 2 the CoV for prenatal AGDI is 4.49 (9 studies) whereas in new born males it is 4.00 (23 studies). The given coefficients of variation are group means (control) based on litter means of all control group coefficients, mean ± SD. This indicates that the sensitivity/power for detecting effect on AGD is quite similar in GD 21 fetuses and new-born male pups. For information the sample size in TG 414 studies is much larger than the studies included in this report (as number of litters in TG 414 is at least 20).

37. As described above the analysis for tables 1 and 2 and figures 1 and 2 were calculated from group means based on litter means. In the expert group it was decided that a more superior approach and statistical model should be developed by US EPA statisticians to take into account the size of the studies (e.g. the number of litters included). They provided a SAS® code with a mixed model approach to facilitate the reanalysis. BIAC representatives in the expert group ran the reanalysis of the AGD data. The reanalysis included only a subset of the data included in this feasibility report (tables 1 and 2; figures 1 and 2). These nine studies from the lead (DTU Food) were included for which there was individual animal data. The reanalysis resulted in an overall CoV (based on the total variance within study = between litters variance + within litter variance) of 7.0% and a between litter CoV (based on the between litters variance) of 4.18% was very close to the 4.49% reported in table 1 and 2 above (original statistics).

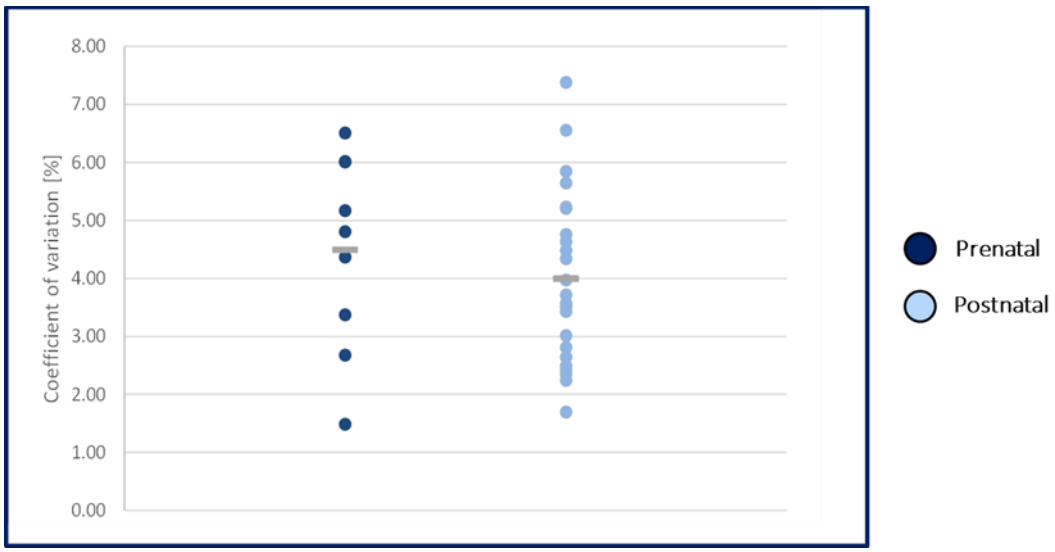


**Figure 2**. Mean coefficient of variation (group means based on litter means in each study) for prenatal studies: 4.49 ±1.69 % (9) and postnatal studies: 4.00 ±1.50 % (23 studies). For the prenatal studies 2 CoVs have the same value and therefore it looks like there are only 8 studies represented

## *Human relevance*

38. In rats, both AGD and nipple retention have been shown to be highly predictive of or correlated to adverse effects of the male reproductive system including increased incidence of genital malformations (e.g. hypospadias) or dysgenesis, and or altered reproductive organ weight changes e.g. levator ani-bulbocavernosus (LABC) muscle complex (Bowman et al. 2003, Christiansen et al. 2008, van den Driesche et al. 2011, Welsh et al. 2008). In female rats, AGDs can in some cases also be affected by exposure to

endocrine disruptors. In most studies of anti-androgenic exposure, female AGDs have not been affected (Hass et al. 2007, Christiansen et al. 2009); however, androgen exposure has been shown to increase female AGD (Hotchkiss et al. 2007) and prochloraz exposure in utero has increased female AGD in several studies (Laier et al., 2006; Melching-Kollmuss et al., 2017). Exposure to oestrogenic agents like ethinyl estradiol (EE2) and genistein have been shown to increase or decrease female AGDs, depending on study design (Levy et al. 1995, Casanova et al. 1999, Delclos et al. 2009, Mandrup et al. 2013). In two novel studies slight reductions in female pup AGD following butylparaben exposure (Boberg et al. 2016) and bisphenol A exposure (Christiansen et al., 2014) have been reported. Therefore AGD in both sexes is included in this TG.

39. A shorter anogenital distance in humans has been shown to be associated with phthalate exposure (Bornehag et al. 2015). Recent studies reported that male infants and boys with adverse effects such as hypospadias or undescended testis also had reduced AGD (Hsieh et al. 2012; Hsieh et al. 2008; Jain and Singal 2013; Thankamony et al. 2013). Moreover, a shorter AGD in adult men has been related to decreased fertility (Eisenberg et al. 2011), impaired semen quality (Mendiola et al. 2011) and decreased serum testosterone levels (Eisenberg et al. 2012). Shortened AGD has also been suggested as a biomarker of testicular dysgenesis syndrome (Sharpe 2005).

40. As AGD in both sexes is included as endpoint in OECD TG 443 and TG 421/422 it can therefore be considered as an endpoint evaluated to be of human relevance. Moreover an updated ECHA guidance document stated that: “The findings in AGD, nipple retention and fetal T, suggest an anti-androgenic mode of action (androgen deficiency) and may be considered as relevant findings and predictors of potential adverse effect during human development.” (ECHA 2013). The OECD GD 43 and GD 151 states “A statistically significant change in AGD that cannot be explained by the size of the animal indicates effects of the exposure and should be considered in setting the NOAEL” (OECD 2008; OECD 2013). As the NOAEL can be used as the point of departure for setting safe exposure levels for humans this further supports that effects on AGD are of human relevance. Moreover, as observations of similar effects is seen in experimental animals and in humans support that AGD changes in experimental animals are relevant for humans.

### *Animal welfare*

41. An important point to remember is that this TG can be enhanced with the ability for detection of EDs without increasing the number of experimental animals used.

42. Assessment of AGD in fetus at caesarean section does not increase the number of experimental animals used, but requires marginally more handling of the fetuses. This assessment can be done very gently just before humane killing and is therefore not expected to lead to any animal welfare concerns.

### *Inclusion of AGD in TG 414*

43. There are standardized OECD test methods for assessing AGD and AGD measured at birth (e.g. PD 1-4) have been included in several OECD TGs. Moreover, several studies (see table 1 and 2) have shown ability to measure AGD at caesarean section at GD 20 or 21 (one day prior to the expected day of delivery). These studies have used a dissecting microscope with a micrometer eyepiece.

44. The current report has shown that the power for assessment of AGD is almost equal in GD 21 fetuses versus in new-borns as the analysis showed similar CoV in these two time

points. Therefore this endpoint can be included in TG 414 at GD 21 without any modification of the overall test design. Several studies have also measured AGD with success in rats at GD 19 and 20.

45. At the TC in EG statistical experts from the USEPA noted that it is important to ensure the CoV results analysed are statistically interpretable. The lead have therefore sent detailed data on 9 studies in order for USEPA to calculate the CoVs by using a mixed effects model to distinguish between variability between litters from variability between studies and to better address the sensitivity of the AGD endpoint. These results will be discussed in EG early 2018.

46. AGD is an endpoint of high human relevance and there are no concerns for animal welfare related to the assessment of this endpoint (OECD, 2015).

47. This all supports the assessment of AGD in all fetuses can be included in TG 414.

## Testosterone levels in male fetuses

### *Method*

48. Changes in testosterone levels in sensitive prenatal time windows can cause permanent/long lasting reproductive changes in laboratory rats. Studies have shown that exposure during gestation to e.g. some phthalates, can show effects on testosterone synthesis, fetal testicular content or *ex vivo* testicular testosterone production in males fetuses leading to effects observed postnatally e.g. anogenital distance and reproductive organ weight changes (Borch et al. 2004, Saillenfait et al. 2013, Borch et al 2006).

49. In general, only few other published papers satisfied the dosing period used in TG 414, either because the administration of the compound was initiated later than in TG 414 on (e.g. GD 12), or because the caesarean section was conducted before the expected day of delivery, most often at GD 19. Studies that did not sufficiently meet the dosing duration of OECD 414 were therefore excluded from the analysis (Parks et al., 2000; Fisher et al., 2003; Hotchkiss et al., 2004; Lehmann et al., 2004; Thompson, Ross and Gaido, 2004; Saillenfait et al., 2016). A reason for this deviance could be the scientific interest in the developmental stages in the embryotic phase or in the testosterone peak levels during the fetal development.

50. In relation to the feasibility of inclusion of testosterone hormone measurements the lead and expert group have discussed whether serum testosterone levels, testosterone synthesis (testicular) or ex vivo testosterone production in fetal testes would be the most suitable method.

51. It was agreed that only measurement of testosterone hormones in the serum would be possible as other methods might impair the ability to fulfil the purpose of TG 414 that is designed to provide general information concerning the effects of prenatal exposure on the pregnant test animal and on the developing organism; this may include assessment of maternal effects as well as death, structural abnormalities, or altered growth in the fetus.

52. Blood sampling in TG 414 (also other hormones) could be done by:

- Cardiac puncture, can only be done for fetuses selected for skeletal assessment
- Decapitation, will impact all fetal assessments

- The umbilical cord, will not impact fetal assessments but only gives a very small amount of blood (so blood of most fetuses/litter should be pooled)

    - This latter method is not relevant for testosterone as this should be only pooled by sex and the amount of blood might be too small.

53. Fetal measurements of T4, T3 and TSH can be pooled from both sexes, however testosterone should not be pooled.

### *Data analysis, sensitivity/power*

54. Weisz and Ward (1980) reported significantly higher serum testosterone levels in male rat fetuses compared to female fetuses with a peak at GD 18 and this finding was confirmed by Lichtensteiger and Schlumpf (1981, 1985). Due to this testosterone peak in male fetuses at GD18, this stage has been examined when the focus is on sexual dimorphism of sex steroid regulation. As already stated this TG 414 performs caesarean section the day before expected delivery and therefore it is not feasible to assess testosterone level in serum from other time-windows.

55. As stated above, plasma testosterone levels have also been found to be affected by chemicals at subsequent fetal and neonatal stages, and have been related to developmental disturbances. This is why this project originally suggested determining plasma testosterone at around GD 21 in TG414.

56. In many studies intratesticular testosterone levels rather than plasma levels were reported (Parks et al., 2000; Fisher et al., 2003; Hotchkiss et al., 2004; Lehmann et al., 2004; Thompson, Ross and Gaido, 2004). This endpoint is however not feasible to include in TG 414 due to the purpose of the TG 414 (see paragraph 49).

57. In Table 3, the overall mean and standard deviation for the individually measured mean plasma testosterone levels, the standard deviations and the coefficient of variation for each control experiment are presented. Furthermore, the number of control experiments included in the calculations and the range of the numbers of litters in these are stated.

58. It is seen at table 3 that a large overall coefficient of variation of around 43% in controls is obtained indicating that the possibility for detecting effects is low. However, it is also seen that the number of litters included is relatively low compared the number of litters in TG 414, i.e. 3-9 litters per group compared to 20 litters per group in TG 414. Therefore the power for detecting effects is expected to be clearly higher in the TG 414 than in those studies.

59. Moreover, in two of the 7 studies (see figure 3), the coefficient of variation is actually around 20% in spite of the low number of litters per group. This indicates that the power for detecting effect with around 20 litters per group may be sufficient.

60. Nevertheless, as also mentioned in this report several studies on EDs have shown significant effects on testosterone measurement in males fetuses.

**Table 3.** Overall mean ± SD for the mean level of plasma testosterone at GD 21, the standard deviation and the derived coefficient of variation for each study (7). Additionally, the number of studies/experiments and the range of number of litters in these are included.

| | Plasma testosterone Mean [nM] | Plasma testosterone Standard deviation [nM] | Plasma testosterone Coefficient of variation [%] | Control experiments (range of litters) |
|---|---|---|---|---|
| Control experiments with Wistar rats | 0.30 ± 0.10 | 0.13 ± 0.08 | 42.99 ± 21.14 | 7 (3-9) |

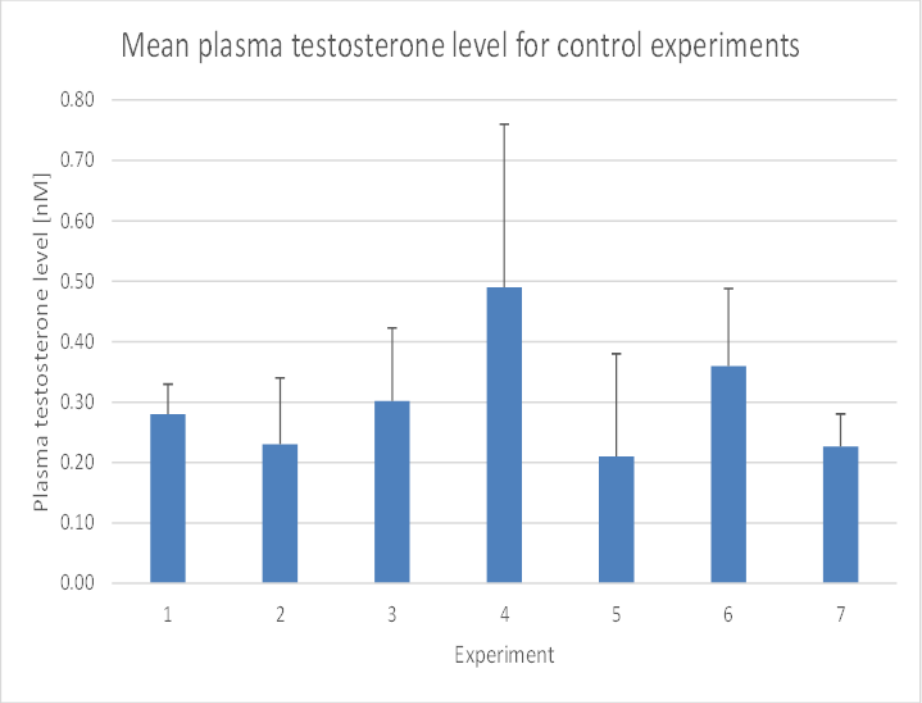


**Figure 3.** The mean plasma (based on litter means) testosterone level +SD in each control group for the 7 studies included in this feasibility report. These studies included fewer animals pr. litter than in the TG 414.

### *Human relevance*

61. Testosterone and dihydrotestosterone are two of the key hormone players in sex differentiation of a fetus and are both classified as sex steroid hormones (Melmed et al. 2015). Therefore by measuring testosterone in TG 414 as an indication of endocrine disruption could indeed be relevant for human risk assessment.

### *Animal welfare*

62. Blood samples for assessment of testosterone or other steroid hormones in fetus could be taken at caesarean section by termination of the study. This will lead to no concern for animal welfare, as the blood samples will be collected at the time of sacrifice.

### *Inclusion of testosterone in TG 414*

63. There are standardized test methods for assessing testosterone hormones in serum. The performed power analysis showed coefficient of variation of around 43% in controls indicating that the possibility for detecting effects is low.

64. Nevertheless, several studies on EDs have shown effects on testosterone measurement in male fetuses.

65. The feasibility study demonstrates large variations in sample measurements and the variation in controls for plasma testosterone is high. These data fail to demonstrate that this measurement is sensitive enough to be included in a test guideline.

66. The inclusion of hormone measurement (testosterone and thyroid hormones) in fetuses was discussed in the EG after 1st commenting round. A representative from BIAC described recent research evaluating technical feasibility of collecting the additional endpoints (AGD, blood samples) from fetuses. To preserve the integrity of other endpoints in TG 414, blood samples were collected by cardiac puncture of fetuses selected for skeletal examination. However, this added about 20-25 minutes per litter and fetuses were alive for up to 35 minutes between Caesarean section and necropsy, raising concerns on variation in processing time for animals within a litter which could lead to high variability in data. BIAC noted that in many cases, blood volumes collected from fetuses were limited and needed to be pooled for analyses.

67. The EG recognised that there is large CoV in fetal testosterone in this report due to the gestational surge which occurs before GD 20, and thus this measurement has limited reliability. Overall, the group felt that there is not enough confidence in fetal testosterone to support including the testosterone measure in the revised TG.

## Thyroid hormones

### *Method*

68. Thyroid hormones were included in the TGs 421/422 as blood samples from the day 13 pups and assessments of thyroid hormones (T4) in the adult males was required (OECD, 2015). In contrast, further assessment of T4 in blood samples from the dams and day 4 pups is to be done if relevant. Moreover the TGs also include an option for other hormones.

69. Due to the circadian rhythm of thyroid hormones, sample collection should occur at approximately the same time of day and be randomized across dosage groups, preferably in the morning hours at which time basal values should be present (Döhler et al., 1979).

70. Blood samples for evaluation of triiodothyronine (T3), thyroxine (T4), and TSH should be collected immediately following sacrifice.

71. Hormonal analyses should be conducted on dams only in this TG 414.

72. Prior to sacrifice, every effort should be made to avoid inducing stress that could affect hormone concentrations (Döhler et al., 1979).

73. When the inclusion of thyroid hormones in the dams was discussed in the EG it was suggested that the revised guideline specify that blood should be collected from dams

within a short timeframe (i.e. two hours) on the morning of the day of necropsy to reduce variability in thyroid hormone levels. Fasting is not necessary for the blood sampling.

74. It was also noted that only non-pregnant dams should be excluded from analyses, and though all blood samples should be collected.

### *Data analysis, sensitivity/power*

75. The number of animals included in TG 414 is similar to the number of animals in TG 443, where assessment of thyroid hormones is included. Thus, specific data analysis of power related to number of animals is not needed here.

76. Moreover, the feasibility study from TGs 421/422 (OECD, 2015) included intensive power simulations for TH measurements with even fewer animals.

77. However, blood samples will be taken in dams in TG 414 compared to non-pregnant adult animals in TG 443 and TG 421/422 (males). This may affect the sensitivity.

### *Human relevance*

78. Thyroid hormones (TH) are needed for proper nerve cell differentiation and proliferation, and normal status of these hormones during early development is therefore crucial. In humans even moderate and transient reductions in maternal T4 levels during pregnancy, may adversely affect the child's neurological development (OECD, 2015; DG-environment, 2017).

79. This indicates that by measuring thyroid hormones (T4, T3 and TSH) in dams in TG 414 as an indication of thyroid disruption is relevant for human risk assessment as also described in the feasibility study for TGs 421/422 (OECD, 2015).

### *Animal welfare*

80. Blood samples for assessment of thyroid hormones could be taken at caesarean section by termination of the study. This leads to no concern for animal welfare, as the blood samples will be collected at the time of sacrifice.

81. Blood samples will be collected from all dams at termination for mandatory assessment of thyroid hormones T4 and T3 or TSH (within a short timeframe (i.e. two hours) on the morning of the day of necropsy).

### *Inclusion of thyroid hormones in TG 414*

82. There are standardized OECD test methods for assessing thyroid hormones. The performed power analysis in TGs 421/422 made in the feasibility study supported that assessment of thyroid hormones could also be included in these TGs.

83. Thyroid hormone measurements (T3, T4 and TSH) will be included as mandatory endpoints in the updated TG 408 (90 day study).

84. Therefore the assessment of thyroid hormones in TG 414 dams is sufficiently sensitive to provide relevant data.

85. Due to the adverse effects seen in humans after developmental hypothryoxinemia, this endpoint is of high human relevance and there are no concerns for animal welfare related to the assessment of this endpoint as long as blood sampling is done in animals that are being sacrificed anyway.

86. The revised TG 414 will include mandatory measures of T4, T3, and TSH from the dams. As agreed by EG the revised guideline will specify that blood should be collected from dams within a short timeframe (i.e. two hours) on the morning of the day of necropsy to reduce variability in thyroid hormone levels. Fasting is not necessary for the blood sampling.

87. The EG discussed proposed inclusion of optional T4, T3, and TSH measures in male and female fetal blood in the revised TG 414 after 1st commenting round (see paragraphs 66-67). The EG discussed possible criteria to trigger the optional endpoints, but no considerations were agreed upon by the group. Several experts noted that if other indications of potential thyroid impairment were observed (e.g. decrease T4 among dams in TG 414), it may be more helpful to request data from guidelines that include more specific information and adverse responses (e.g. neurodevelopmental endpoints). There was no clear support for including optional blood samples in TG 414 revision without considerations for when these should be included. Moreover the fetal blood samples would add extra processing time for animals within a litter which could lead to high variability in data (see paragraphs 66-67). It was then decided by the EG to not include any serum hormone measurements from fetuses in TG 414.

## Abnormalities of external genital organs

### *Method*

88. In TG 414 the reproductive tract is examined for signs of altered development. The SPSF in this project have described guidance on evaluation of abnormalities of external genitalia in fetuses such as hypospadias (Hsieh et al. 2007). However, Hsieh et al. (2007) used histopathological examination of the genital tubercle to detect hypospadias.

89. In the TG 414 it is mentioned (paragraph 32) *Particular attention should be paid to the reproductive tract which should be examined for signs of altered development*. This project has also included the following text: *External fetal sex (as determined by gross examination) should be compared with internal (gonadal) sex in all fetuses (examined for both skeletal and soft tissue malformations). In addition, indication of incomplete testicular descent/cryptorchidism should be noted in male fetuses*.

### *Data analysis, sensitivity/power*

90. The feasibility study for TG 421/422 (OECD, 2015) included a calculation of the effect size needed for finding significant effect for abnormalities/malformations early postnatally.

91. This limited sensitivity for detecting significant effects on rare adverse outcomes is generally recognized for malformations or dysgenesis. Thus, the occurrence of a few similar rare genital malformations may generally be considered toxicologically relevant although the finding is not statistically significant (OECD, 2015). The new text on cryptorchidism in TG 414 is therefore supported.

### *Human relevance*

92. Exposure during critical developmental phases such as *in utero* and in the early postnatal period may lead to adverse effects on reproductive development. The fact that many of the basic mechanisms underlying this developmental process are similar in many known species of mammals indicates that chemicals that have adverse effects on

reproductive development in rodents should be considered as potential human reproductive toxicants as well (Gray 1992).

### *Animal welfare*

93. Assessment of abnormalities of external genital organs requires slightly more handling of fetuses. This assessment can be done very gently and is therefore not expected to lead to any animal welfare concerns. However, as the assessment of abnormalities of external genital organs is done after termination of the fetuses in TG 414, there will obviously be no concern for animal welfare.

### *Inclusion abnormalities of external genital organs in TG 414*

94. Assessment of abnormalities is already included in TG 414 (paragraph 32). The proposed text to be added in the revised TG 414 in relation to external fetal sex (as determined by gross examination) and internal (gonadal) sex in all fetuses and indication of incomplete testicular descent/cryptorchidism is now added in new paragraph 30 in TG 414. This inclusion is agreed by EG and is already common practice according to BIAC.

## Overall discussion and conclusions

95. The aim of this project was to do a feasibility study for minor enhancements of TG 414 with ED-relevant endpoints. The endpoints considered for inclusion in the SPSF were AGD, testosterone levels in fetuses, thyroid hormones in the dams and fetuses and more guidance on abnormalities of external genital organs.

96. The EG have decided not to include any serum hormone measurements in fetuses because of too large CoVs and moreover these measurements would be time consuming and thereby impair the ability to fulfil the purpose of TG 414.

97. For all included endpoints, OECD test methods are available for assessing these however this is the first TG to include thyroid hormones in pregnant dams. Calculation of CoVs for AGD has shown that it is a sufficiently sensitive endpoint to get relevant data with the number of litters per group in TG 414. All endpoints are of relevance for humans as described in this report. Inclusion of the additional endpoints in TG 414 does not trigger any animal welfare concerns.

98. The Test Guideline has been updated with specific text proposals. No changes in study design and only few text changes are necessary to include the assessment of anogenital distance (AGD) in all fetuses, thyroid hormone measurements (in the dams only) and text on a text on external and internal fetal sex (see paragraph 32; OECD, 2018b).

99. In conclusion, it is feasible to make the proposed minor enhancements of TG 414 with ED-relevant endpoints: anogenital distance (AGD), thyroid hormones (in dams) and text on external and internal fetal sex and indication of incomplete testicular descent/cryptorchidism.

100. Following discussions with EG and two WNT written commenting rounds the final proposed revision for the updated TG 414 (OECD, 2018b) includes:

1. Mandatory AGD measurement in all fetuses (paragraph 30)
2. Mandatory T4, T3 and TSH mandatory measurement in dams; other hormone measurements if relevant (paragraph 35)
3. Observation of external fetal sex (as determined by gross examination) and internal (gonadal) sex in all fetuses (paragraph 32)
4. Indication of incomplete testicular descent/cryptorchidism in male fetuses (paragraph 32).

## References

Boberg, J (2011) Screening study on phthalate-like effects of an alternative plasticizer. Report for FVST DK.

Boberg J, Axelstad M, Svingen T, Mandrup K, Christiansen S, Vinggaard AM, Hass U (2016). Multiple Endocrine Disrupting Effects in Rats Perinatally Exposed to Butylparaben. Toxicol Sci. 2016 Jul;152(1):244-56.

Borch J, Ladefoged O, Hass U, Vinggaard AM. Steroidogenesis in fetal male rats is reduced by DEHP and DINP, but endocrine effects of DEHP are not modulated by DEHA in fetal, prepubertal and adult male rats. Reprod. Toxicol. 2004;18(1):53-61

Borch J, Axelstad M, Vinggaard AM, Dalgaard M. Diisobutyl phthalate has comparable anti-androgenic effects to di-n-butyl phthalate in fetal rat testis. Toxicol. Let. 2006, 163,183-90.

Bornehag CG1, Carlstedt F, Jönsson BA, Lindh CH, Jensen TK, Bodin A, Jonsson C, Janson S, Swan SH. Prenatal Phthalate Exposures and Anogenital Distance in Swedish Boys. Environ Health Perspect. 2015 Jan;123(1):101-7.

Bowman CJ, Barlow NJ, Turner KJ, Wallace DG, Foster PMD. 2003. Effects of in utero exposure to finasteride on androgen-dependent reproductive development in the male rat. Toxicological Sciences 74:393-406.

Casanova M, You L, Gaido KW, Archibeque-Engle S, Janszen DB & Heck HA 1999 Developmental effects of dietary phytoestrogens in Sprague–Dawley rats and interactions of genistein and daidzein with rat estrogen receptors a and b in vitro. Toxicological Sciences 51 236–244. (doi:10.1093/toxsci/51.2.236)

Christiansen S, Scholze M, Axelstad M, Boberg J, Kortenkamp A, Hass U. Combined exposure to anti-androgens causes markedly increased frequencies of hypospadias in the rat. International Journal of Andrology 2008;31:241–8.

Christiansen S, Scholze M, Dalgaard M, Vinggaard AM, Axelstad M, Kortenkamp A & Hass U 2009 Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. Environmental Health Perspectives 117 1839–1846. (doi:10.1289/ehp.0900689)

Christiansen S, Axelstad M, Boberg J, Vinggaard AM, Pedersen GA, Hass U (2014). Low-dose effects of bisphenol A on early sexual development in male and female rats. Reproduction 147, 477-487.

Delclos KB, Weis CC, Bucci TJ, Olson G, Mellick P, Sadovova N, Latendresse JR, Thorn B & Newbold RR 2009 Overlapping but distinct effects of genistein and ethinyl estradiol (EE2) in female Sprague–Dawley rats in multigenerational reproductive and chronic toxicity studies. Reproductive Toxicology 27 117–132. (doi:10.1016/j.reprotox.2008.12.005)

DG-Environment, October 2017. Supporting the organization of a workshop on thyroid disruption – final report. ISBN 978-92-79-73427-4. https://publications.europa.eu/en/publication-detail/-/publication/472d2c88-a8b1-11e7-837e-01aa75ed71a1/language-en

Döhler, K. D., Wong, C. C., & von zur Mühlen, A. (1979). The rat as model for the study of drug effects on thyroid function: consideration of methodological problems. Pharmacology & Therapeutics. Part B: General & Systematic Pharmacology, 5(1-3), 305–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/386373

ECHA (European Chemicals Agency). 2013. Evaluation of new scientific evidence concerning DINP and DIDP in relation to entry 52 of annex xvii to reach regulation (EC) no 1907/2006. Available: http://echa.europa.eu/documents/10162/31b4067e-de40-4044-93e8-9c9ff1960715 [accessed 18 November 2014]

Eisenberg ML, Hsieh MH, Walters RC, Krasnow R, Lipshultz LI. 2011. The relationship between anogenital distance, fatherhood, and fertility in adult men. PLoS One 6(5):e18973.

Eisenberg ML, Jensen TK, Walters RC, Skakkebaek NE, Lipshultz LI. 2012. The relationship between anogenital distance and reproductive hormone levels in adult men. J Urol 187(2):594-598.

Ema, M., Miyawaki, E., Hirose, A. and Kamata, E. (2003) Decreased anogenital distance and increased incidence of undescended testes in fetuses of rats given monobenzyl phthalate, a major metabolite of butyl benzyl phthalate, Reproductive Toxicology, 17, pp. 407–412

Fisher, J. S., Macpherson, S., Marchetti, N. and Sharpe, R. M. (2003) Human "testicular dysgenesis syndrome": a possible model using in-utero exposure of the rat to dibutyl phthalate, Human Reproduction, 18(7), pp. 1383–1394.

Gray LE. 1992. Chemical-induced alterations of sexual differentiation: A review of effects in humans and rodents. In: Chemically induced alterations in sexual and functional development: the wildlife/human connection (Colborn T, Clement C, eds). Princeton, NJ:Princeton Scientific Publishing,203-230.

Hansen PR, Taxvig C, Christiansen S, Axelstad M, Boberg J, Kiersgaard MK, Nellemann C, Hass U. Evaluation of endocrine disrupting effects of nitrate after in utero exposure in rats and of nitrate and nitrite in the H295R and T-screen assay.Toxicol Sci. 2009 Apr;108(2):437-44. doi: 10.1093/toxsci/kfp023. Epub 2009 Feb 3.

Hotchkiss, A. K., Parks-Saldutti, L. G., Ostby, J. S., Lambright, C., Furr, J., Vandenbergh, J. G. and Gray, L. E. (2004) A mixture of the "antiandrogens" linuron and butyl benzyl phthalate alters sexual differentiation of the male rat in a cumulative fashion, Biology of Reproduction, 71, pp. 1852–1861.

Hass U, Scholze M, Christiansen S, Dalgaard M, Vinggaard AM, Axelstad M, Metzdorff SB & Kortenkamp A 2007 Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environmental Health Perspectives 115 (Suppl 1) 122–128. (doi:10.1289/ehp.9360)

Hotchkiss A K., Lambright C S., Ostby J S., Parks-Saldutti L., Vandenbergh J G., and Gray Jr L E. (2007). Prenatal Testosterone Exposure Permanently Masculinizes Anogenital Distance, Nipple Development, and Reproductive Tract Morphology in Female Sprague-Dawley Rats Reproductive Toxicological Sciences 96(2), 335–345, doi:10.1093/toxsci/kfm002

Hsieh,MH, Grantham,EC, Liu,B, Macapagal,R, Willingham E, Baskin LS. 2007. In utero exposure to benzophenone-2 causes hypospadias through an estrogen receptor dependent mechanism. Journal of Urology 178, 1637-1642.

Hsieh MH, Breyer BN, Eisenberg ML & Baskin LS 2008 Associations among hypospadias, cryptorchidism, anogenital distance, and endocrine disruption. Current Urology Reports 9 132–142. (doi:10.1007/s11934- 008-0025-0)

Hsieh MH, Eisenberg ML, Hittelman AB, Wilson JM, Tasian GE, Baskin LS. 2012. Caucasian male infants and boys with hypospadias exhibit reduced anogenital distance. Human reproduction 27(6):1577-1580.

Jain VG. Singal AK. 2013. Shorter anogenital distance correlates with undescended testis: a detailed genital anthropometric analysis in human newborns. Human Reproduction 28(9): 2343–2349.

Lehmann, K. P., Phillips, S., Sar, M., Foster, P. M. D. and Gaido, K. W. 2004. Dose-dependent alterations in gene expression and testosterone synthesis in the fetal testes of male rats exposed to di(n-butyl) phthalate, Toxicological Sciences, 81, pp. 60–68.

Levy JR, Faber KA, Ayyash L, Hughes CL Jr 1995The effect of prenatal exposure to the phytoestrogen genistein on sexual differentiation in rats. Proceedings of the Society for Experimental Biology and Medicine 2008 60–66.

Lichtensteiger W, Schlumpf M. 1981. Steroids and neurotransmitter mechanisms in the prenatal period. In: Steroid Hormone Regulation of the Brain, Wenner Gren Symposium Series Vol. 34 (K. Fuxe, J. A. Gustafsson and L. Wetterberg, Eds.), Pergamon Press, Oxford, pp. 161-172.

Lichtensteiger W, Schlumpf M. 1985. Prenatal nicotine affects fetal testosterone and sexual dimorphism of saccharin preference. Pharmacol. Biochem. Behav. 23: 439-444.

Liu K., Lehmann KP, Sar M, Young SS, Gaido KW. 2005. Gene expression profiling following in utero exposure to phthalate esters reveals new gene targets in the Etiology of testicular dysgenesis. Biology of Reproduction, 73, pp. 180–192.

Laier, P., Metzdorff, S. B., Borch, J., Hagen, M. L., Hass, U., Christiansen, S., Axelstad, M., Kledal, T., Dalgaard, M., McKinnell, C., et al. (2006). Mechanisms of action underlying the antiandrogenic effects of the fungicide prochloraz. Toxicol. Appl. Pharmacol. 213, 160–171.

Mandrup, K. R., Jacobsen, P. R., Isling, L. K., Axelstad, M., Dreisig, K., Hadrup, N., Vinggaard, A. M., Hass, U., and Boberg, J. (2013). Effects of perinatal ethinyl estradiol exposure in male and female Wistar rats. Reprod. Toxicol. 42, 180–191.

Melching-Kollmuss, S., Fussell, K. C., Schneider, S., Buesen, R., Groeters, S., Strauss, V., and van Ravenzwaay, B. (2017). Comparing effect levels of regulatory studies with endpoints derived in targeted anti-androgenic studies: Example prochloraz. Arch. Toxicol.

Melmed S, Polonsky KS, Larsen PR, Kronenberg HM (2015). Williams Textbook of Endocrinology. Elsevier Health Sciences. pp. 711–. ISBN 978-0-323-29738-7.

Mendiola J, Stahlhut RW, Jørgensen N, Liu F, Swan SH. 2011. Shorter anogenital distance predicts poorer semen quality in young men in rochester, new york. Environ Health Perspect 119(7):958.

OECD (2001) Prenatal Developmental Toxicity Study, Guideline for the Testing of Chemicals (no.414). Organisation for Economic Cooperation and Development, Paris.

OECD (2006). Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats. No.59.

OECD (2008). Guidance document on mammalian reproductive toxicity testing and assessment. OECD Series on Testing and Assessment no. 43. Organisation for Economic Cooperation and Development, Paris. 88 pp

OECD (2012), Test No. 443: Extended One-Generation Reproductive Toxicity Study, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264185371-en

OECD (2013). Guidance document in support of the test guideline on the extended one generation reproductive toxicity study No. 151. Organisation for Economic Cooperation and Development, Paris.

OECD (2015). Feasibility Study for Minor Enhancements of TG 421/422 with ED Relevant Endpoints. Environment, Health and Safety Publications, Series on Testing and Assessment (No. 217), Organisation for Economic Cooperation and Development, Paris.

OECD (2016a). OECD guideline for testing of chemicals No. 421. Reproduction/developmental toxicity screening test. 27-July-1995, 28-July-2015 (former) and 29-July-2016 (updated) http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

OECD (2016b). OECD guideline for testing of chemicals No. 422. Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test. 22-March-1996, 28-July-2015 (former) and 29-July-2016 (updated) http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

OECD (2018a). Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment No. 150, *ENV/JM/MONO(2012)22* (revised pre-publication)

OECD (2018b). OECD guideline for testing of chemicals No. 414. Prenatal Developmental Toxicity Study. July 208 (updated).

Parks, L. G., Ostby, J. S., Lambright, C. R., Abbott, B. D., Klinefelter, G. R., Barlow, N. J. and Gray, L. E. (2000) The plasticizer diethylhexyl phthalate induces malformations by decreasing fetal testosterone synthesis during sexual differentiation in the male rat, Toxicological Sciences, 58, pp. 339–349.

Saillenfait, A., Gallissot, F. and Sabaté, J. (2009) Differential developmental toxicities of di-n-hexyl phthalate and dicyclohexyl phthalate administered orally to rats, Journal of Applied Toxicology, 29, pp. 510–521.

Saillenfait, A., Roudot, A. and Gallissot, F. (2011a) Developmental toxic potential of di-n-propyl phthalate administered orally to rats, 31, pp. 36–44.

Saillenfait, A., Roudot, A., Gallissot, F. and Sabaté, J. (2011b) Prenatal developmental toxicity studies on di-n-heptyl and di-n-octyl phthalates in Sprague-Dawley rats, Reproductive Toxicology, 32, pp. 268–276.

Saillenfait, A.-M., Sabaté, J.-P., Robert, A., Rouiller-Fabre, V., Roudot, A.-C., Moison, D. and Denis, F. (2013).Dose-dependent alterations in gene expression and testosterone production in fetal rat testis after exposure to di-n-hexyl phthalate. J. Appl. Toxicol., 33: 1027–1035.

Saillenfait, A.-M., Ndiaye, D., Sabaté, J.-P., Denis, F., Antoine, G., Robert, A., Rouiller-Fabre, V. and Moison, D. (2016) Evaluation of the effects of deltamethrin on the fetal rat testis, Journal of Applied Toxicology, 36, pp. 1505–1515.

Salazar-Martinez E, Romano-Riquer P, Yanez-Marquez E, Longnecker M, Hernandez-Avila M. 2004. Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. Environmental Health: A Global Access Science Source 3:8.

Sharpe RM. 2005. Phthalate exposure during pregnancy and lower anogenital index in boys: Wider implications for the general population? Environ Health Perspect 113(8):A504-5

Taxvig C, Hass U, Axelstad M, Dalgaard M, Boberg J, Andeasen HR, Vinggaard AM.Endocrine-disrupting activities in vivo of the fungicides tebuconazole and epoxiconazole.Toxicol Sci. 2007 Dec;100(2):464-73. Epub 2007 Sep 4.

Taxvig, C., Vinggaard, A. M., Hass, U., Axelstad, M., Metzdorff, S. and Nellemann, C. (2008) Endocrine-disrupting properties in vivo of widely used azole fungicides, International Journal of Andrology, 31, pp. 170–177.

Taxvig C, Hadrup N, Boberg J, Axelstad M, Bossi R, Bonefeld-Jørgensen EC, Vinggaard AM. In vitro-in vivo correlations for endocrine activity of a mixture of currently used pesticides. Toxicol Appl Pharmacol. 2013 Nov 1;272(3):757-66. doi: 10.1016/j.taap.2013.07.028. Epub 2013 Aug 14.

Thankamony A, Lek N, Carroll D, Williams M, Dunger DB, Acerini CL et al. 2013. Anogenital distance and penile length in infants with hypospadias or cryptorchidism: Comparison with normative data. Environ Health Perspect http://dx.doi.org/10.1289/ehp.1307178.
Thyroid workshop report, 2017

Thompson, C. J., Ross, S. M. and Gaido, K. W. (2004) Di(n-butyl) phthalate impairs cholesterol transport and steroidogenesis in the fetal rat testis through a rapid and reversible mechanism, Endocrinology, 145(3), pp. 1227–1237.

van den Driesche S, Scott HM, MacLeod DJ, Fisken M, Walker M, Sharpe RM. 2011 Relative importance of prenatal and postnatal androgen action in determining growth of the penis and anogenital distance in the rat before, during and after puberty. Int J Androl. Dec;34(6 Pt 2):e578-86. doi: 10.1111/j.1365-2605.2011.01175.x. Epub 2011 Jun 2.

Weisz J, Ward IL (1980). Plasma testosterone and progesterone titers of pregnant rats, their male and female fetuses, and neonatal offspring. Endocrinology 106: 306-316.

Welsh M, Saunders P, Fisken M, Scott HM, Hutchison GR, Smith L & Sharpe RM 2008 Identification in rats of a programming window for reproductive tract masculinization, disruption of which leads to hypospadias and cryptorchidism. J.Clin.Invest. 118 1479-1490.