rivm

RIVM Report 340700002/2007

Current innovations in regulatory reproductive toxicity assessment of chemicals

A.H.Piersma
B. Hakkert
J.J.A.Muller

Contact:
A.H. Piersma
GBO
aldert.piersma@rivm.nl

This investigation has been performed by order and for the account of Ministry of Health, Welfare and Sports, within the framework of project 340700: 'Kennisbasis Carcinogenese, Mutagenese en Reproductietoxicologie'.

rivm

rivm

Abstract

Current innovations in regulatory reproductive toxicity assessment of chemicals

Chemicals must be tested for adverse effects on human reproduction. Thanks to innovations, testing will become more efficient, thereby reducing the use of experimental animals. The main developments have been signalled in an overview of current developments in the testing of reproductive toxicity (the adverse effects of chemicals of reproduction) documented in this RIVM report.

Chemicals can adversely affect reproduction, which may manifest itself as reduced fertility or maldevelopment of the fetus. Manufacturers of chemicals are required to assess these adverse effects, for example, in the framework of the European programme for chemical safety called REACH.

Current methods for reproductive toxicity assessment, which rely mainly on experimental animal studies, stem from the early eighties. The high animal use in the testing of reproductive toxicicity is primarily due to the need to study more than 1-generation.

Reductions of animal use are, however, imminent, with developments in the endocrine disruption, animal welfare and REACH having stimulated innovations in this area. Standard protocols are in the process of being amended and there are proposals for novel test systems. These developments also influence testing strategies, which combine individual tests in a tiered approach. These are innovations then that should lead to increased efficiency and reduced animal use.

Key words: reproductive toxicology, test guidelines, alternatives, intergrated testing strategy, REACH, risk assessment, hazard identification, fertility, development

Rapport in het kort

Actuele ontwikkelingen in regelgeving over reproductietoxiciteit van chemische stoffen

Chemische stoffen moeten getest worden op schadelijke effecten op de voortplanting van de mens. Dankzij innovaties zal dit steeds efficiënter kunnen, waardoor ook minder proefdieren gebruikt hoeven te worden.

Dit staat in dit rapport van het RIVM. Het rapport geeft een overzicht van actuele ontwikkelingen in het meten van reproductietoxiciteit, de schadelijke effecten van chemische stoffen op de voortplanting.

Chemische stoffen kunnen schadelijke effecten op de voortplanting teweegbrengen, zoals een verminderde vruchtbaarheid en een verstoorde ontwikkeling van de ongeborene. Onder meer in het kader van het Europese testprogramma REACH moeten producenten van chemische stoffen deze schadelijke effecten vaststellen.

De huidige methoden om reproductietoxiciteit in kaart te brengen stammen uit de jaren tachtig. Deze methoden zijn voornamelijk gebaseerd op studies met knaagdieren. Het gebruik van proefdieren ligt daarbij hoog. Dit komt voornamelijk doordat meer dan één generatie moet worden bestudeerd.

Een verminderd proefdiergebruik is echter in aantocht. Onderzoek aan hormoonverstoring, proefdierwelzijn en het Europese testprogramma REACH hebben innovaties op dit gebied gestimuleerd. Hierdoor worden gestandaardiseerde testen aangepast en nieuwe testmethoden voorgesteld. Dit beïnvloedt tevens teststrategieën, die individuele testen combineren op een getrapte manier. De innovaties moeten leiden tot een verhoogde efficiency en een verminderd proefdiergebruik.

Trefwoorden: reproductive toxicology, test guidelines, alternatives, intergrated testing strategy, REACH, risk assessment, hazard identification, fertility, development

rivm

Contents

rivm

Summary

This report reviews current developments in regulatory reproductive toxicity assessment. Intrinsically complex and continuously relevant in chemical risk assessment, the area is permanently under discussion in national and international regulatory bodies. Developments in politics and society further stimulate research and regulatory initiatives in this area.

The first chapter briefly reviews current practice in regulatory reproductive toxicity assessment, highlighting the function of generation studies and developmental toxicity studies in experimental animals. These guideline-based protocols, which stem from the early eighties, have shown their strengths and weaknesses. They are now ready for review on the basis of over 25 years of experience and data collection. We identify a number of current issues, including the endocrine disrupter issue, the EU REACH programme and animal welfare issues, which have stimulated innovation in the area. The estimation by the European Chemicals bureau that 70% of experimental animals under REACH will be needed for reproductive toxicity testing has stimulated research into refinig testing strategies and reducing animal use. Standardized test protocols are being amended and new protocols are defined, including animal and non-animal tests.

In chapter 2, individual test protocols and their applications are reviewed. The uterotrophic and Hershberger assays for female and male sex hormone activity, the OECD407 subchronic toxicity study updated for endocrine parameters, the draft developmental neurotoxicity study (OECD426), and the proposed extended one-generation study are addressed in detail. Juvenile toxicity testing, alternatives to animal tests, and in silico methods are briefly covered.

In chapter 3, guidance documents and testing strategies which combine individual tests into tiered approaches, are discussed. Both OECD and EU (REACH) have produced guidance documents for reproductive toxicity testing. The issue of rat versus rabbit developmental toxicity testing is reviewed. In the area of classification and labelling, the globally harmonized system (UN-GHS) and the issue of potency are covered. Finally, integration of testing strategies within the wider realm of toxicity testing is under extensive study. This array of developments are ongoing at the national, the EU, OECD and UN levels.

In chapter 4, future developments of main issues are considered and a preliminary minimal testing strategy for reproductive toxicity is discussed. This wealth of activities is aimed towards improved efficiency of safety testing of chemicals with reduced experimental animal use, whilst increasing the quality of basic toxicological information used for risk assessment and classification and labelling of chemicals.

1 Introduction

In recent years reproductive toxicity testing protocols have become subject of renewed discussion. This has lead to changes in existing testing protocols as well as to the development of new testing protocols. As a consequence and in parallel, testing strategies for regulatory safety assessment of chemicals are also being reformulated. This report gives an overview of current activities worldwide, with an outlook towards developments in the future. This section will give a birds eye view of current developments, whereas details of developments in guidelines are discussed in chapter 2, and of guidance documents in chapter 3. A general discussion and outlook follows in chapter 4.

1.1 Reproductive toxicology

Reproductive toxicology is concerned with all possible adverse effects of chemical exposures on any aspect of the reproductive cycle (Figure 1). Classically, fertility and prenatal development have been the two areas of main concern. They constitute a wide variety of mechanisms on the molecular, cellular, tissue and organism level, with different windows of sensitivity in time. Classically, morphological effect assessment and functional integrity of the reproductive system have been used as end points. Novel functional end points of toxicity have received increasing interest, such as developmental neurotoxicity and behaviour, and developmental immune toxicity assessed through immune function tests in offspring at adulthood.

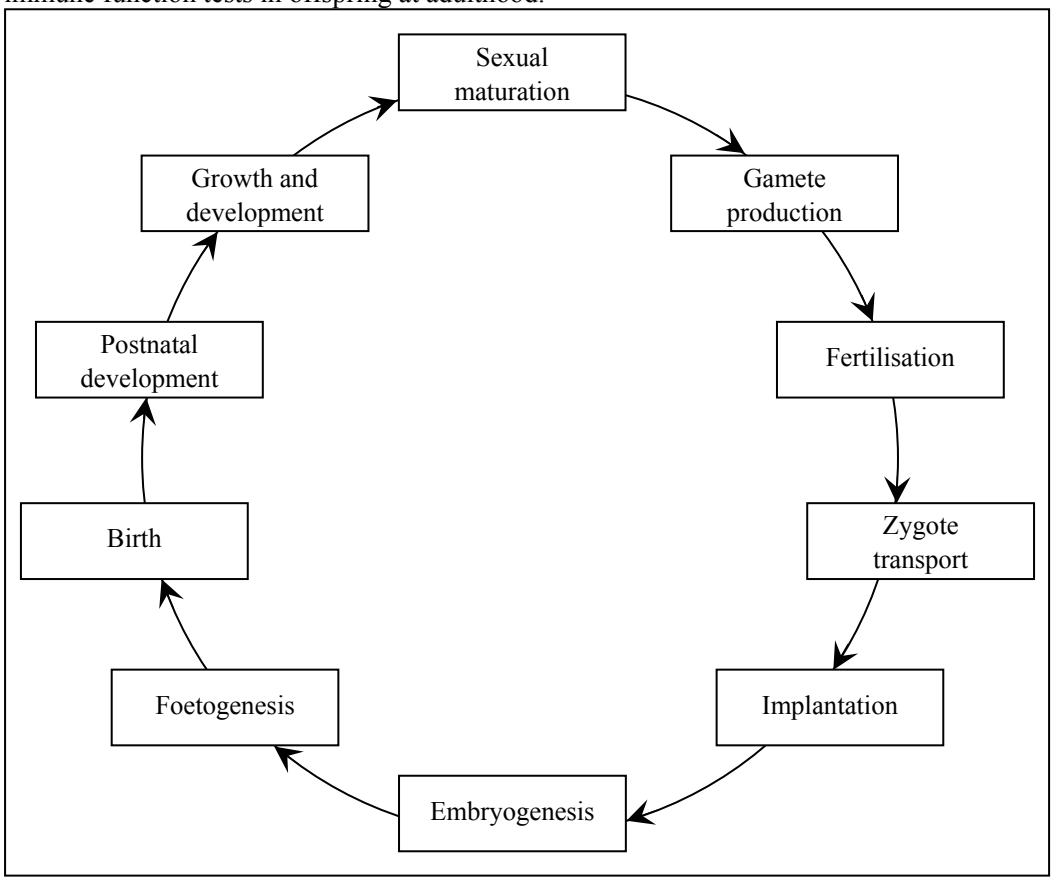

Figure 1. Schematic representation of the reproductive cycle.

1.2 Current regulatory reproductive toxicity tests

Standardized regulatory reproductive toxicity testing dates back to the early eighties of the twentieth century, when OECD protocols were published for the prenatal developmental toxicity study (OECD414), the one-generation reproductive toxicity study (OECD415), and the 2-generation reproductive toxicity study (OECD416). Protocols can be retrieved from the OECD website (1). The underlying justification for the design of this set of three tests was as follows. The OECD414 prescribes prenatal exposure starting after implantation and necropsy one day before expected birth. This design precludes effects on implantation and also excludes the possibility of maternal cannibalism of malformed pups. Thus, developmental effects on implanted embryo-fetuses can be studied on all implants after explantation from the uterus shortly before birth. The OECD415 one-generation study is primarily a study to detect effects on fertility, prescribing parental exposure of both sexes and of dams throughout pregnancy and until weaning of the pups. The OECD416 2-generation study is designed to allow fertility assessment of animals prenatally exposed, and therefore only this study design in principle covers the entire reproductive cycle. In addition to these definitive tests, a relatively quick screening method which can give initial clues about possible fertility and developmental effects of chemicals is the OECD421 screening study adopted in 1995. In this study, exposure of parental animals is for two weeks premating and until postnatal day 6 in dams, which is the day of necropsy. Males are dosed to a total duration of four weeks and necropsied. This protocol prescribes only 8 pregnant dams per group. As noted above this method is designed as screening protocol and is not adequate for reliable conclusions in case of absence of observed toxicity. However, if toxicity is found, this may trigger further studies and/or hazard and risk related measures.

1.3 Current reproductive toxicity testing strategy

The toxicity testing strategy in the European Union for New Substances is primarily based upon production tonnage levels. The system is described in an EU technical guidance document (2) and schematically represented in Figure 2. Reproductive toxicity testing commences classically with a one-generation study at tonnage level 1 to gather some but not comprehensive information both on development and fertility. Dependent on the outcome of this study and on tonnage level, a developmental toxicity study (in case of concern for developmental effects) and/or a 2-generation study (in case of concern for fertility effects) may follow. A developmental toxicity study in a second species (usually the rabbit) can be warranted based on equivocal findings in the first developmental toxicity study (usually the rat). The reasoning for a second species goes back to the thalidomide episode which occurred around 1960 (3,4). Thalidomide, used as a sedative, caused severe limb reduction defects in children of mothers taking the drug in pregnancy. These defects could not be reproduced in the rat, where only some general foetotoxicity was observed. In the rabbit, limited limb reduction defects were observed, although more sporadically and at higher doses. This prompted regulators in some regulatory frameworks to request developmental toxicity testing in two species in order to prevent similar devastating consequences of other chemicals in the future.

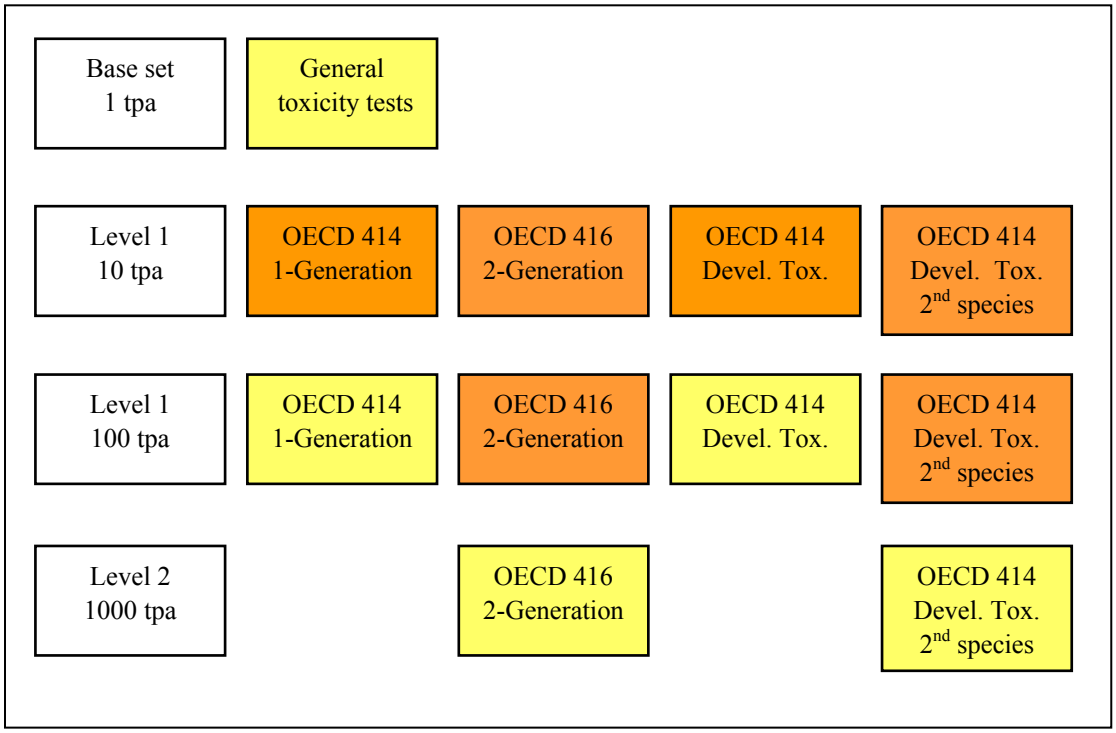

Figure 2. Schematic representation of current EU reproductive toxicity testing strategy.Tests marked in yellow are basic requirements. Tests marked in orange are only required in case of concern.
tpa =tonne(s) per annum.

1.4 Endocrine disruption

The issue of endocrine disription receives attention since the early nineties, when observations on wildlife and human fertility were related to environmental chemicals exposures (5,6). The question was raised whether current test protocols were sufficiently sensitive to detect endocrine disrupting effects. The OECD initiated the Endocrine Disrupter Testing and Assessment (EDTA) task force, which subsequently came up with a wealth of animal and non-animal tests which could detect endocrine activities of chemicals (7). A series of specific in vitro receptor binding and activation assays was listed. In addition, two short term in vivo assays received renewed attention. The uterotrophic assay for (anti-)estrogenicity and the Hershberger assay for (anti-)androgenicity came into consideration for endocrine disruption assessment. In addition, enhancements of existing reproductive toxicity guidelines (OECD414/5/6) were proposed, specifically with reference to hormone level assessments and histopathology of reproductive organs, thyroid gland and pituitary gland. Finally, such enhancements are also proposed for the 28 day subchronic toxicity study (OECD407).

1.5 EU REACH

The European chemicals regulation REACH (Registration, Evaluation and Authorisation of CHemicals) specifically mentions reproductive toxicants together with carcinogenic and mutagenic compounds (8). These compounds need extra scrutiny according to the legislation. As a consequence,

such compounds may need additional testing if the current toxicity profile is incomplete. It has been estimated that 30,000 chemicals will enter the REACH legislation, of which around 5000 chemicals reach tonnage level one and another 5000 chemicals reach higher levels of testing, both requiring reproductive toxicity testing. This testing will require millions of experimental animals, of which no less than an estimated 70% will be needed in reproductive toxicity testing (9). This very large percentage is due to the fact that in reproductive toxicity testing always more than *1-generation* is included in each study, and that foetuses from gestation day 18 onward (rat) and pups count as experimental animals under the animals ethics legislation. This realization has greatly accelerated the attention for simplifying existing animal test protocols in reproductive toxicology and has further stimulated the development, validation, and implementation of animal-free alternatives. In addition, the extended one-generation reproduction study protocol under discussion as a possible replacement for the current 2-generation study is a promising new development in this respect. Non-testing in silico methods such as read-across, group approach and (quantitative) structure-activity relationships (Q)SAR are increasingly being employed to optimize hazard assessment.

1.6 Renewing tests and testing strategies for hazard identification

After around 25 years of experience with the existing system of reproductive toxicity test protocols, and given the high animal use and high cost and time consumed by these tests, efforts are now underway to review the efficiency and necessity of test protocols and current testing strategies in actual practice. In addition, novel end points for endocrine effects, developmental effects on immunity and neurobehavioural effects have received interest as they may represent important parameters that may be affected after prenatal and juvenile exposure which are not yet measured to the extent necessary in current tests and testing strategies. Current major activities in this area are briefly listed here, and decribed in more detail in chapters 2 and 3.

The enhancement of the OECD407 subchronic toxicity study in adult rats with endocrine parameters has resulted in an updated draft guideline published April 2006.

The need for addressing developmental neurotoxicity has resulted in the OECD426 draft Guideline.

An important development is the design of an extended one-generation study, in which basically the OECD415 1-generation study is extended to raise the offspring to adulthood for various additional assessments, but omitting the production of a second generation as in the OECD416 two generation study.

The increased concern for possible specific sensitivity of children to (endocrine acting) chemicals has been translated into using direct exposure of pups before weaning to achieve relevant exposure levels that may not be achieved after lactational exposure via exposed dams.

In reproductive toxicity testing there are several issues related to strategically employing existing tests more efficiently and without redundancy. One issue relates to the added value of the second developmental toxicity study, for which a review of past experience is underway at the National Institute for Public Health and the Environment (RIVM) to underpin decisions on possible eventual alterations in testing strategies.

Furthermore, the question has been put forward to what extent reproductive toxicity end points have an added value in the determination of the overall no-observed-adverse-effect level (NOAEL). A relevant

comparison for addressing this question is that between generation study outcome and subchronic adult toxicity outcome, which is also currently underway at RIVM.

The OECD is finalizing Guidance Document 43 (GD 43), which gives guidance on strategic aspects of mammalian reproductive toxicity testing and assessment.

In 2005, OECD Guidance Document GD34 was adopted, which addresses the validation and international acceptance of new or updated test methods for hazard assessment.

On a higher level of integration, the question can be put forward at what levels and in what order the different animal tests should be performed. In relation to that, the role and place of in silico and in vitro alternatives as well as possible situations in which tests can be waived dependent on outcomes in previous studies should be considered. Within the REACH Implementation Plans, such a strategy has been proposed for reproductive toxicity testing. The same type of question is being addressed for toxicity testing in general, where reproductive toxicity testing is one of a series of classes of end points that have to be assessed in an integrated testing strategy (ITS).

Within the framework of OECD, a globally harmonized system for classification and labelling has been developed including criteria for substances toxic to reproduction, which will likely become effective in the EU in foreseeable time. One open end in this system is the question of whether and how potency should play a role in 1. classification of preparations and 2. classification of individual substances. This is an issue which the EU working group on classification and labelling is currently addressing under Dutch leadership.

2 Test guidelines under construction and revision

This section reviews novel and adapted guidelines which are currently being developed, and discusses their possible application and use in regulatory reproductive toxicity testing.

2.1 The uterotrophic assay

2.1.1 Principle and status

The uterotrophic assay protocol was developed in the thirties. It is based on the increase in uterine weight or uterotrophic response. It evaluates the ability of a chemical to elicit biological activities consistent with agonists or antagonists of natural oestrogens (e.g. 17ß-estradiol), however, its use for antagonist detection is much less common than for agonists. The uterus responds to oestrogens in two ways. An initial response is an increase in weight due to water imbibition. This response is followed by a weight gain due to tissue growth. The uterus responses in rats and mice are qualitatively comparable. The uterotrophic assay is intended to be included in a battery of in vitro and in vivo tests to identify substances with a potential to interact with the endocrine system, ultimately leading to risk assessments for human health or the environment. The recent OECD validation program used both strong and weak estrogen agonists to evaluate the performance of the assay to identify estrogenic compounds (10). Thereby the sensitivity of the test procedure for oestrogen agonists was well demonstrated besides a good intra- and interlaboratory reproducibility. The specificity of the test was not addressed in detail in the validation study and is still an issue for discussion. The OECD working group of national coordinators on test guidelines (WNT) adopted the draft guideline in March 2007. This finalizes the guideline subject to the OECD Joint Meeting approval.

2.1.2 Test protocol

Juvenile or overiectomized young adult rats are exposed with three daily doses of test compound, either orally or subcutaneously (Figure 3). Graduated test substance doses are administered to a minimum of two treatment groups of experimental animals using one dose level per group and a minimum administration period of three consecutive days. The animals are necropsied approximately 24 hours after the last dose. For oestrogen agonists, the mean uterine weight of the treated animal groups relative to the vehicle group is assessed for a statistically significant increase. A statistically significant increase in the mean uterine weight of a test group indicates a positive response in this bioassay.

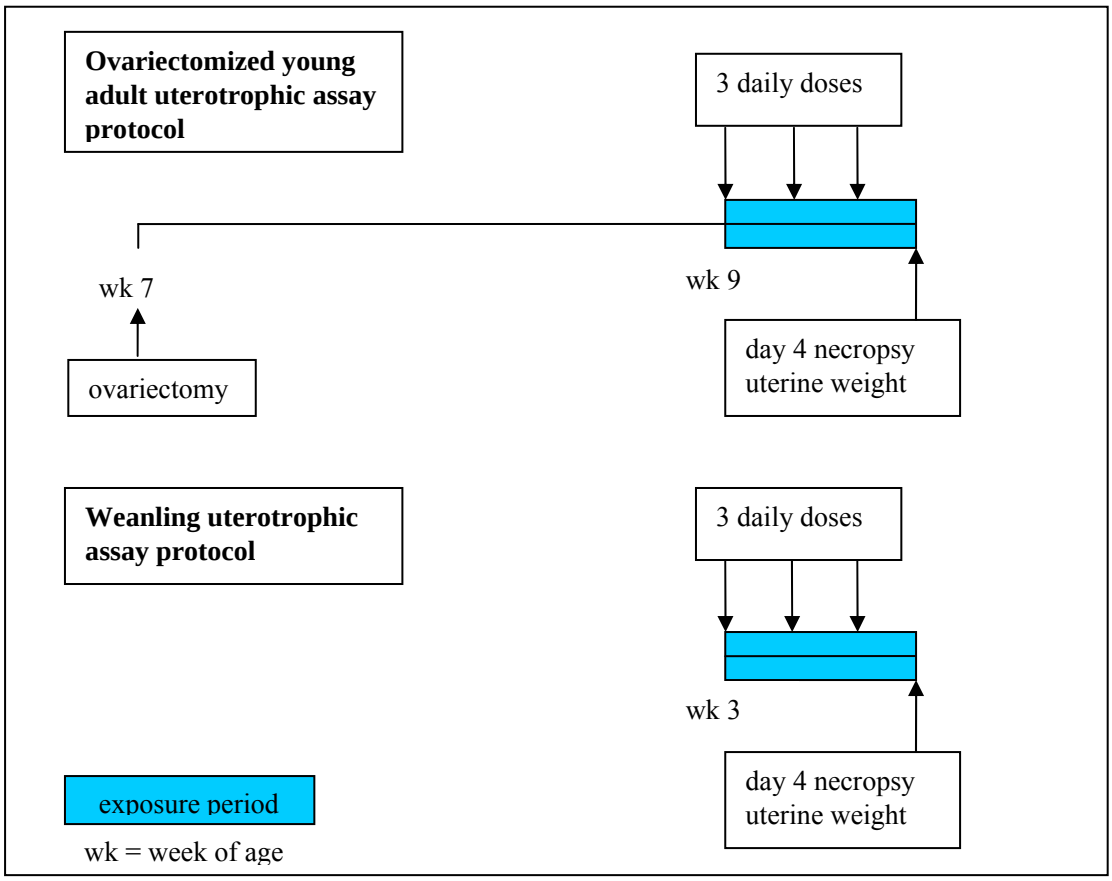

Figure 3. Schematic representation of the uterotrophic assay.

2.1.3 Application in the testing strategy

The uterotrophic assay was revived for use in endocrine disrupter screening. It can detect estrogen receptor alpha agonists, and if coadministered with an agonist, compounds acting as antagonists can also be detected. The discussion on the application of this test focuses on the necessity of using animal experimentation for this specific goal, whereas a wealth of in vitro receptor binding and activation assays are available. The counterargument mostly used is that in an in vivo assay also the kinetics of compounds comes into play, which would make this in vivo test more relevant for the human situation. In response to that, first, the subcutaneous route is less relevant in view of human exposure, and second, the short exposure duration of three daily doses (with usually relatively high doses) is hard to extrapolate to longterm low dose exposures. It is anticipated that at least in the European Union the uterotrophic assay will not become a first choice method for (anti-)estrogenicity testing. Rather, the ongoing development of sophisticated in vitro receptor binding and activation assays will probably reduce the need for the uterotrophic assay. In conjunction with in vivo kinetic studies, in vitro assay results will likely give sufficient relevant information to decide on (anti-)estrogenic potential of test compounds.

2.2 The Hershberger assay

2.2.1 Principle and status

The basis of the Hershberger assay is the absolute requirement for testosterone (made in the testis) and/or dihydrotestosterone (converted from testosterone by 5α-reductase in the testis and other end target organs) for the rapid growth and maturation of the accessory sex organs during puberty in intact males, for their maintenance postpuberty, for their rapid regression and involution after castration (removal of the source of testosterone), and for their rapid regrowth in the castrate administered an exogenous androgen (typically testosterone propionate). The accessory sex organs of interest are predominantly the epididymides (if they are not removed at castration), the prostate (the ventral lobe or ventral plus dorsolateral lobes), and the levator ani plus bulbocavernosus complex (LABC) muscle, as well as the other accessory sex organs (seminal vesicles with coagulating glands, Cowper's glands, and preputial glands). The Hershberger assay has been proposed by both EDSTAC (1998) and OECD (1998) to be validated for use in a comprehensive screen to detect potential endocrine disruptors. An OECD background review paper has been published in draft form December 2006 (11), as a prelude to the development of an OECD test guideline.

2.2.2 Test protocol

The Hershberger assay has been used in various versions (Figure 4). The most important versions include the prepubertal intact male assay, the castrated adult male assay, and the peripubertal castrated male assay. The protocols have different advantages. The intact assay contains the complete intact hypothalamic-pituitary-gonadal axis which makes this an apical assay, and it uses an age window with increased sensitivity to androgens. The concomittant disadvantages include the lack of mechanistic information and the relatively small time window of opportunity, and in this respect the castrated adult male assay is superior. The peripubertal castrated male assay has been advocated for the high sensitivity of accessory sex organs around puberty, in combination with the possibility for mechanistic information in this castrated model (12).

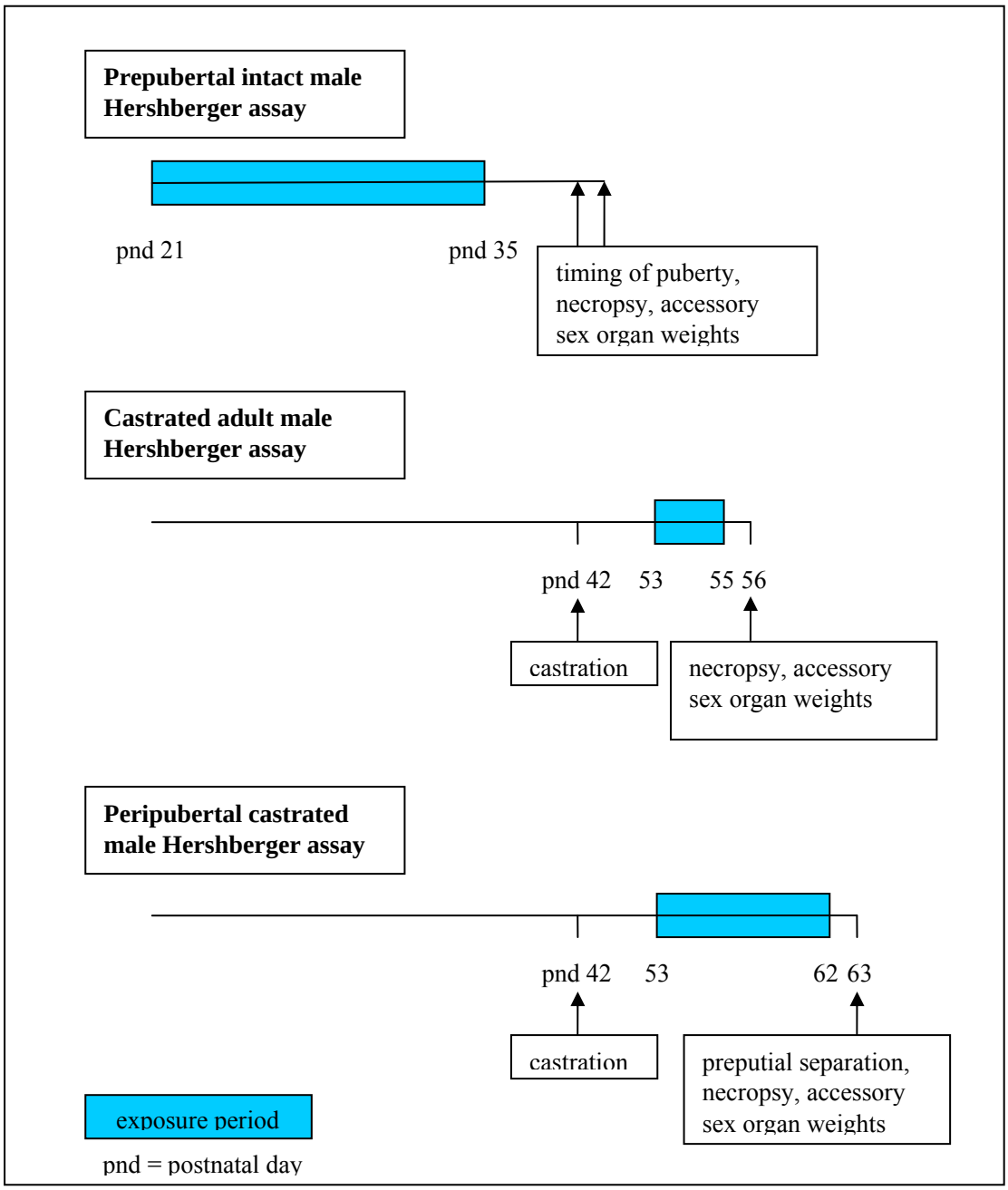

Figure 4. Schematic representation of typical versions of the Hershberger assay.

2.2.3 Application in the testing strategy

The Hershberger assay has been revived and further developed in its several forms for endocrine disrupter detection, much in a similar sequence as for the uterotrophic assay. OECD is in the process of development of a test guideline, which lags somewhat behind the development of the uterotrophic assay into a guideline protocol. The argumentation relevant for the application of the Hershberger assay in a testing strategy is also very similar to the uterotrophic assay. The question of justification of the use of experimental animals for what is basically a receptor-activation assay which could be mimicked to a great extent by in vitro assays is also pertinent here. The exposure routes can be oral or

subcutaneous, the latter should be considered less relevant for actual human exposures, although this route may be used for achieving higher internal exposures necessary for the detection of weak (ant-)agonists. It is anticipated that at least in the European Union the Hershberger assay will not become a first choice method for (anti-)androgenicity testing. Rather, the ongoing development of sophisticated in vitro receptor binding and activation assays will probably reduce the need for the Hershberger assay. In conjunction with in vivo kinetic studies, in vitro assay results will likely give sufficient relevant information in the long run to decide on (anti-) estrogenic potential of test compounds.

2.3 OECD407 updated subchronic toxicity study

2.3.1 Principle and status

In view of the endocrine disrupter issue, an initiative was taken in 1998 to update the existing OECD guideline for 'repeated dose 28-day oral toxicity study in rodents' (TG 407) by parameters suitable to detect endocrine activity of test substances. This procedure underwent an extensive international program to test for the relevance and practicability of the additional parameters, the performance of these parameters for chemicals with (anti)oestrogenic, (anti)androgenic, and (anti)thyroid activity, the intra- and interlaboratory reproducibility, and the interference of the new parameters with those required by the prior TG 407. The updated TG 407 allows to put endocrine mediated effects into context to other toxicological effects. These activities lead to a 'draft updated guideline 407' which is available on the web for public comments until 20 August 2007. A detailed description of the procedure and outcome of the project is given in Gelbke et al., (2007) (13).

2.3.2 Test protocol

Male and female rats, 5 per sex per dose group, are exposed for 28 days starting at age 7-9 weeks (Figure 5). At the end of the exposure period, extensive necropsy is performed, which is enhanced for endocrine parameters as compared to the existing OECD407 guideline. Enhancements are listed in Table 1. The additional parameters were selected from a larger series that was included in the validation studies. From those studies it was concluded that most hormonal parameters such as sex steroids and LH, FSH were too variable to be usefully included. Only thyroid hormone levels were considered informative in the protocol. In addition it was concluded that the protocol is able to detect high and medium potency endocrine compounds, whereas low potency compounds are unlikely to be detected in the system. An optional extended post-dosing observational period of two weeks is suggested for observation of reversibility, persistence and late occurrence of effects, specifically for the control and high dose groups, which may increase sensitivity of the test. Furthermore, it is stated that the protocol is not performed in a life-stage that is most sensitive to endocrine disruption. Extension of the exposure period from 28 up to 90 days was considered unlikely to improve the chance of detection of endocrine activity.

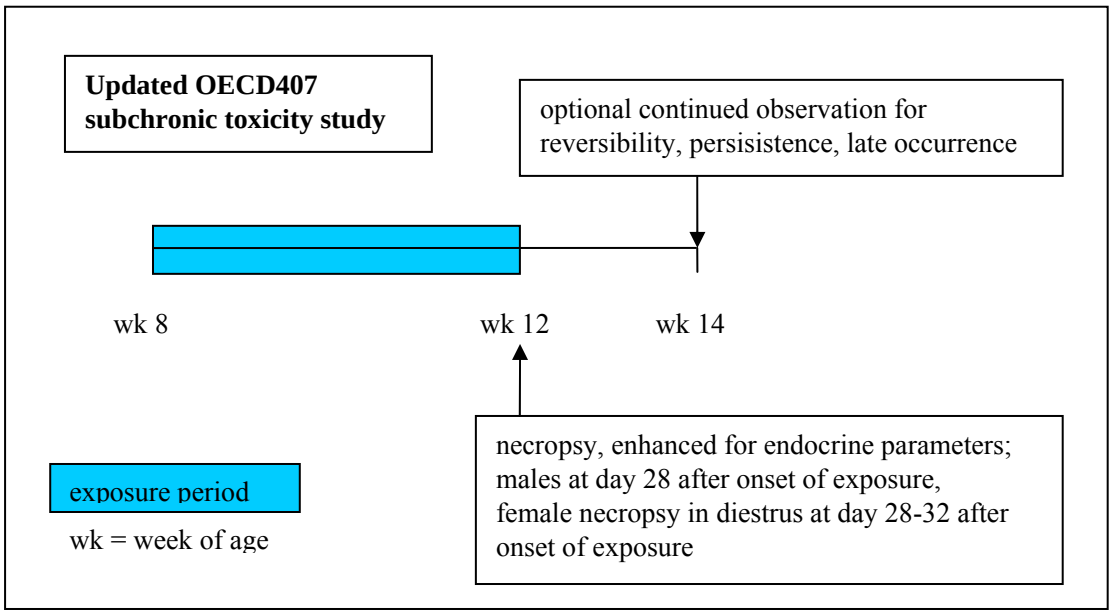

Figure 5 . Schematic representation of the updated OECD407 subchronic toxicity test protocol.

Table 1. Proposed enhancement endpoints to the OECD407 subchronic toxicity study.

Organ/tissue weights
1. Testes (each weighed separately)
2. Seminal vesicles + coagulating glands
3. Prostate (possible dissection and separate weights for ventral and dorsolateral prostate), ovaries
4. Thyroid
5. Uterus
Histopathology
1. Pituitary
2. Vagina
3. Epididymides, seminal vesicles + coagulation glands
4. Mammary gland
Thyroid hormones
1. Circulating levels of T_3 and T_4
2. Circulating levels of TSH
Spermatology
1. Epididymal sperm number

2.3.3 Application in the testing strategy

The updated OECD407 guideline is useful at the base set tonnage level in the EU system for toxicological hazard identification of New Substances and REACH. The enhancements proposed do not cost extra animals, except for the optional extension of observation post-dosing in the control and high dose group. In the absence of dedicated reproductive toxicity testing at the base set level, the information from this protocol may yield unique and important information that may direct further dedicated testing for reproductive toxicity. On the other hand it should be realized that, as the draft guideline rightly states, the age of the animals in this test, young adult, is likely not the most sensitive for detection of adverse effects on the endocrine system. Therefore, the absence of findings on the

endocrine system can not be taken as proof of absence of endocrine activity of the compound tested. The usefulness of the updated OECD407 should be weighed against that of in vitro and in vivo screening assays for sex and thyroid hormone (ant)agonistic properties. In vitro hormone receptor binding and activation assays are animal-free and more sensitive but lack in vivo kinetics, and in vivo screens such as uterotrophic and Hershberger assays are more sensitive, but have limited kinetic relevance and do use additional animals for a single end point test.

2.4 OECD426 developmental neurotoxicity assay

2.4.1 Principle and status

Neurodevelopmental toxicity became of interest in view of increasing awareness of behavioural abnormalities in children such as attention deficit hyperactivity disorder (ADHD), and findings in experimental studies that several compounds affected behaviour after prenatal exposure in otherwise unaffected animals (14). It was felt that important developmental parameters of brain development and functionality were not adequately addressed in existing safety testing of chemicals, which warranted this additional protocol. Developmental neurotoxicity studies are designed to provide data, including dose-response characterizations, on the potential functional and morphological effects on the developing nervous system of the offspring that may arise from exposure *in utero* and during early life. A developmental neurotoxicity study can be either conducted as a separate study, or incorporated into a reproductive toxicity and/or adult neurotoxicity study (*e.g.,* OECD415, 416) or added onto a prenatal developmental toxicity study (*e.g.,* OECD414). The history of the development of the OECD425 protocol dates back to 1995, when a first expert meeting was held in Copenhagen. At present, the protocol has the status of an OECD draft test guideline.

2.4.2 Test protocol

The test substance is administered to animals during gestation and lactation (Figure 6). Dams are tested to assess effects in pregnant and lactating females and to provide comparative information (dams versus offspring). Offspring are randomly selected from within litters for neurotoxicity evaluation. The evaluation consists of observations to detect gross neurologic and behavioural abnormalities, including the assessment of physical development, behavioural ontogeny, motor activity, motor and sensory function, and learning and memory; and the evaluation of brain weights and neuropathology during postnatal development and adulthood.

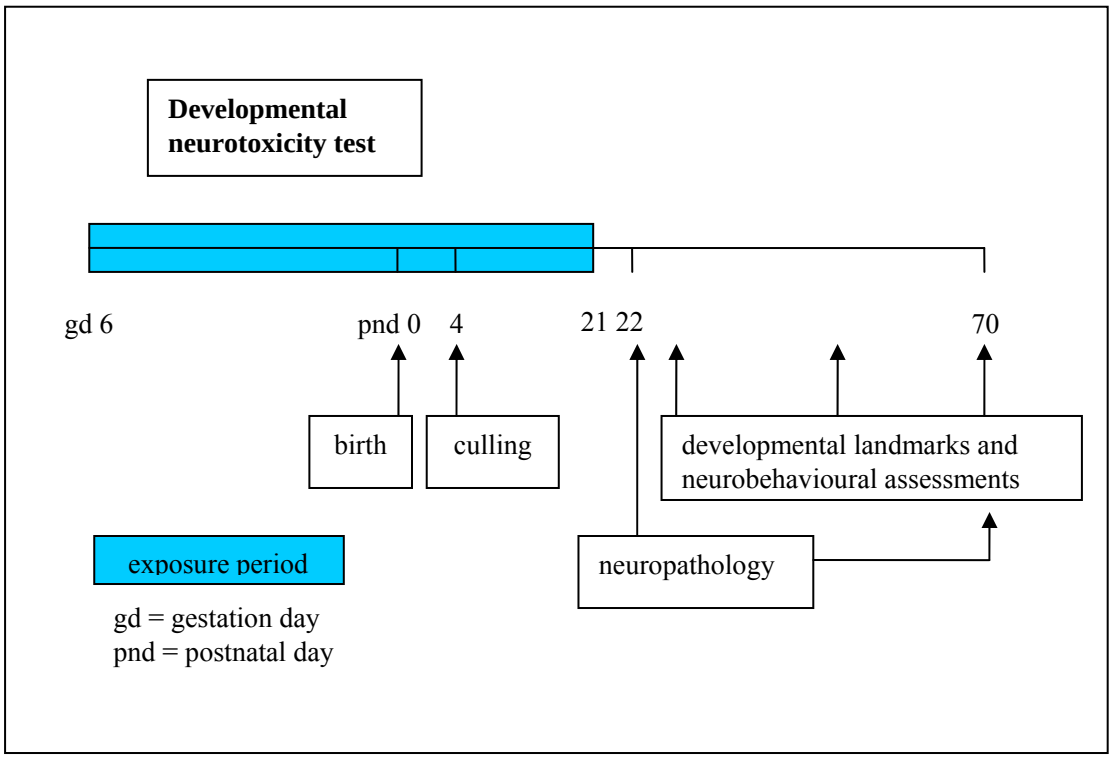

Figure 6. Schematic representation of a typical stand alone version of the developmental neurotoxicity test.

2.4.3 Application in the testing strategy

At this moment in time there is no absolute requirement for this test at any tonnage level in the European Union. However, the use of the test has been extensively discussed in the expert group which developed a guidance document for reproductive toxicity testing under REACH. As the test has no legal status yet both in terms of an OECD guideline and in terms of current EU regulations, no mandatory requirement could be decided upon. Discussions in the group also considered the relatively high animal use and the practicality of the labour-intensive protocol, in addition to the limited past experience with the test system. The majority feeling was however that the test could probably be optimally used as an adjunct to the 2-generation study (OECD416) at the time when the latter would be required in cases triggered by findings in earlier studies pointing to a possible effect on the central nervous system.

2.5 Extended 1-generation study

2.5.1 Principle and status

Various initiatives worldwide are currently exploring the possibility of replacing the existing OECD416 2-generation reproductive toxicity study with an extended 1-generation study. Proposals for protocols have been generated independently by US and Japanese groups and in Germany and in the Netherlands activities along the same lines are ongoing (15). The rationale for these activities are

varifold. The usefulness in terms of informative yield for risk assessment and classification & labelling of effects observed in the second generation of the 2-generation study has become an issue of detailed study. The possible reduction of animal use and the efficiency gain are important additional aspects. The critical issue is whether testing functional fertility in a prenatally exposed generation is essential for reproductive hazard and risk assessment. This aspect is included in the 2-generation study but not in the extended one-generation study.
RIVM recently performed an analysis of 176 multi-generation studies to assess potential differences between the first and the second generation, both in terms of the types of effects observed and in terms of the effective doses (16). All substances classified as reproductive toxicants by the Directive 92/32/EEC or considered as toxic to fertility by the California EPA for which a multi-generation study was found were included (n=58 studies). The rest of the studies (n=118) related to substances that are not classified as reproductive toxicants by the mentioned institutions. The second generation in the 2-generation studies considered affected neither the overall NOAEL nor the critical effect. Therefore, it had no impact on the ensuing risk assessment nor on classification and labeling. These results clearly support the proposal of replacing the current 2-generation study by a 1-generation study with a more extensive assessment of parameters at F_1 adulthood.
The OECD Validation Management Group – Mammalian has identified the extended one-generation study as an important issue for further consideration in upcoming meetings. In the EU, in view of the burden of reproductive toxicity testing on the animal use in the REACH program, this subject is absolutely timely. Nevertheless, it will probably take some time before a decision on possible changes of current guidelines and testing requirements will be made. The quality of data for hazard identification is at stake, which warrants careful scrutiny of basic principles of reproductive toxicity testing and of existing data gathered over two-and-a-half decade since the institution of the OECD416 2-generation study in 1981. In addition, the place of the test in the testing strategy should be considered.

2.5.2 Test protocol

Cooper et al. (15) published a proposal for an extended one-generation study which is given here as an example (figure 7). The male and female P-generation is exposed for 4 and 2 weeks, respectively, before mating. Males are exposed up to 6 weeks postmating (total male exposure at least 10 weeks), and females are exposed throughout pregnancy and weaning. F1 animals are exposed up to postnatal day 70 and developmental landmarks are monitored. At postnatal day 70 three sets of assessments are listed: 1. clinical pathology and developmental neurotoxicity, 2. immunotoxicity, and 3. estrus cyclicity and reproductive toxicity parameters. Necropsy of F1 generation animals at adulthood allows detailed pathological assessment of organs relevant for sex and thyroid hormone homeostasis, which, together with the data on parental reproduction, is suggested to give sufficient insight into the possible reproductive toxicity of the test compound.

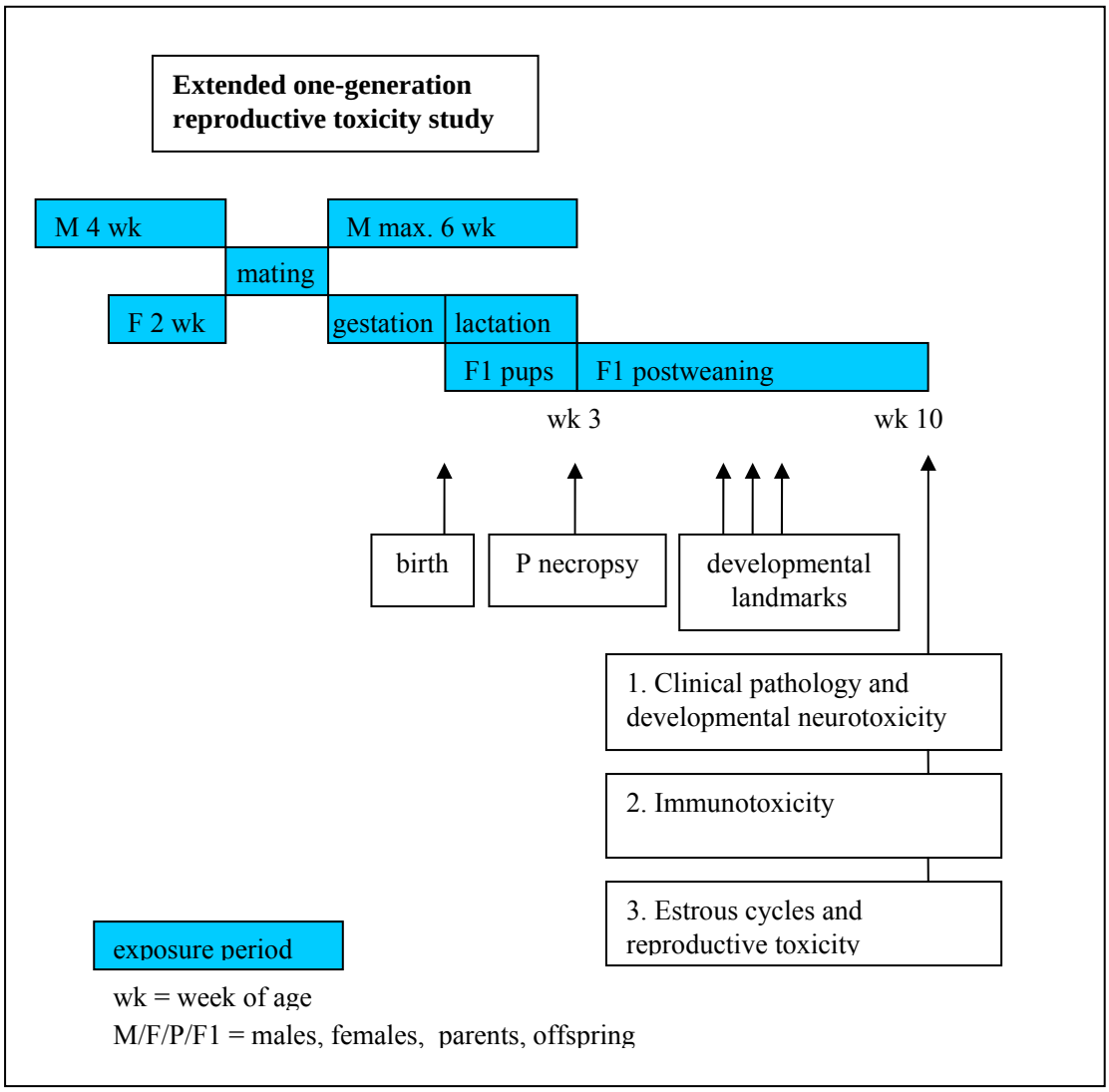

Figure 7. Schematic representation of an extended one-generation reproduction study.

2.5.3 Application in the testing strategy

The extended one-generation study may ultimately replace the OECD416 2-generation study as the definitive study on fertility and reproduction effects on which hazard and risk assessment for these end points is based. At present this development appears as a promising one, both on the basis of the retrospective study performed at RIVM, and in view of foreseen reduction of animal use. As it is critical not to miss essential end points, a thorough retrospective assessment of all experience so far has to indicate the feasibility of replacing the existing two generation guideline with an extended one-generation study. Final assessment however awaits further consideration as currently anticipated in the OECD VMG-Mammalian working group.

2.6 Direct pup exposure for juvenile toxicity testing

The need for juvenile toxicity testing in rats via direct oral exposure of suckling pups is a subject of discussion primarily in the world of pharmaceuticals testing. Although it has not been a prime issue in the chemicals domain sofar, it may be of relevance there as well, and the subject is therefore briefly touched upon. The idea is that assessment of the safety of drugs for pediatric use meets with considerable difficulty when based on adult toxicity studies only. In addition, perinatal toxicity studies which use maternal exposures only (gavage or feed exposure) are dependent to a large extent on lactational transfer of compounds for pup exposure during the first two to three weeks of life. To fill this exposure gap, it may in some cases be warranted to use direct pup exposure of sucklings in order to achieve doses relevant for risk assessment. In addition, exposure may be continued throughout growth and development until adulthood. In 2005, the European Medicines Agency's Committee for Human Medicinal products (EMEA-CHMP) produced a draft guideline for testing human pharmaceuticals for paediatric indications in juvenile animals (17). Whereas most of this design is covered in the chemicals domain by the OECD416 study and the proposed extended one-generation study (see section 2.5), the direct exposure of pups is an aspect that is not covered. Pup exposure may however be warranted for chemicals as well in view of developing parameters related to e.g. sexual maturation, brain function and immune competence.

2.7 Alternative test systems for hazard identification

A wealth of alternatives have been developed in the area of reproductive toxicology, boosted by the relatively high animal use in reproductive toxicology. By their nature, such tests represent a reductionistic approach to reproduction and development, which poses important questions about their predictivity. Rat whole embryo culture and embryonic stem cell differentiation are currently the most promising systems which are being extensively scrutinized (18,19). The validity and applicability domain of these test systems for chemical screening are currently issues for research and debate. A comprehensive treatise of the subject has been made in a recent RIVM report (20), and the reader is referred to that report for further details. Suffice it here to state that alternatives, although some have been formally validated, still have not reached a stage where they can reduce animal testing. Issues of concern are e.g. the limited kinetic information and the virtual lack of metabolism in in vitro tests. In a prescreening situation for prioritizing in vivo testing, these tests may prove helpful in the future, but such developments await further definition of applicability domain and predictive capacity.
RIVM is investing in improving existing developmental toxicity alternatives by using gene expression analysis as effect parameter. The basic hypothesis of this approach is that early chemical-induced gene expression changes will be predictive for adverse effects occurring later at the level of the cell, tissue and organism. Thus, if end points outside the period of testing in in vitro tests may be predicted by early gene expression changes in the assays, this would improve the predictive capacity of the in vitro tests. The critical search is for a set or sets of genes that can be taken as a biomarker for developmental toxicity. Most likely, this set will differ to an extent for different classes of compounds and effects. This approach is still in its infancy, but may potentially provide an important step forward in hazard identification and reduction of experimental animal use.

2.8 In silico non-testing methods for hazard identification

Besides in vivo and in vitro test methods, there is increasing use worldwide of (Q)SAR, grouping and read-across techniques in regulatory testing programs for chemical safety, concurrent with the development and validation of such methods. These methods are based on chemical and physical properties of compounds, which are compared to predict toxicity across defined classes of compounds. In brief, the toxicity of an untested compound is predicted on the basis of knowledge about analogues within the group. The EU (REACH) legislation clearly includes such in silico methods to be applied where appropriate to optimize hazard evaluation. However, their application in various areas of toxicity is still under development and discussion. Especially with regard to (Q)SAR sufficient validation and documentation of the methods are mentioned as prerequisites. Pedersen et al. (ECB, 2003) (26) estimate that the acceptance of (Q)SAR methods for reproductive and developmental toxicity end points is 10% and 25% respectively, which is far below any other end point of toxicity. An RIVM study showed that the vast majority of classified reproductive toxicants was not recognized by two existing (Q)SAR models (27). These findings are not surprising in view of the limited database available to derive (Q)SAR models for reproductive toxicity, and is also understandable in view of the limited mechanistic knowledge in this area with its complex variety of mechanisms and windows of sensitivity in time throughout the reproductive cycle. Significant steps forward in methodology, development and validation are necessary before these methods can be usefully employed in the area of reproductive toxicology hazard and risk assessment.

3 Guidance documents and testing strategies

At the strategic level the challenge is to combine the individual test protocols in such a way that information is gathered sufficiently and efficiently for hazard and risk assessment purposes. This section will review novel and adapted guidance documents and testing strategies which are currently being developed.

3.1 Testing strategy OECD Guidance Document No.43

The OECD Working Group on Reproduction and Developmental Toxicity met in Copenhagen, Denmark in June 1995. The meeting recognised that there was a need for a Guidance Document on developmental and reproduction toxicity testing, covering testing strategies, approaches for tier testing, relationship with neurotoxicity testing and data interpretation procedures. The 7th WNT meeting approved to develop the Guidance Document in October 1996 and the US EPA volunteered to take the lead in this activity. The currently available draft guidance document is dated November 10, 2004. This document is presently being finalized for adoption as an OECD Guidance Document (1).
The document is intended to provide guidance on methodological aspects, interpretation of data and strategy for testing of chemicals for potential reproductive toxicity. The primary objective is to ensure that necessary and sufficient data are obtained to enable adequate evaluation of the risk of reproductive toxicity arising from exposure to a chemical. A stepwise assessment/testing strategy is recommended. To minimise animal usage and optimise
allocation of resources, data should be assessed following each step of testing to decide if they are adequate for the evaluation of the risk arising from the intended use of the chemical, or if further testing is needed. The guidance document presents a Conceptual Framework, in which tests are listed in four levels according to increasing complexity (Table 2). Testing could start at level 1 and move through the ensuing levels dependent on the outcome at lower levels.

Table 2 : OECD Conceptual Framework for the Regulatory Hazard Assessment of Chemicals With respect to Mammalian Reproduction, based on increasing levels of information provided.

Level 1 **Profiling and alerts based upon existing information**	Expected human & environmental exposure and use patterns - hazard, *e.g.,* available toxicological data (enhanced TG 407) - QSARs for blood / testis barrier - QSARs for blood / placenta barrier - QSARs for blood / breast milk barrier - QSARs for blood / brain barrier -- physical & chemical properties, *e.g.,* MW, reactivity, volatility, biodegradability	
Level 2 ***In vitro* assays providing mechanistic data and focus on target-aimed tests**	- ER, AR, TR receptor binding affinity - transcriptional activation - aromatase and steroidogenesis *in vitro* - aryl hydrocarbon receptor recognition / binding	- embryonic stem cell tests - *ex vivo* sperm test - *ex vivo* oocyte test - *in vitro* fertilization - target cell toxicity - Leydig cell viability - fetal oocyte viability - *in vitro* genetic toxicity
Level 3 ***In vivo* assays providing data about single mechanisms related to fertility/reproduction**	- Uterotrophic bioassay (estrogenic related) - Hershberger bioassay - non-receptor mediated hormone function	copulation behavior
Level 4 ***In vivo* assays providing data about multiple mechanisms related to fertility/reproduction**	- male and female pubertal assays - adult intact male assay - reproductive screening test (TG 421) - combined TG 421 / TG 407 - Segment I, II, III studies	- 1-generation assay (TG 415) - 2-generation assay (TG 416) - prenatal development (TG 414) - developmental neurobehavior (TG 426)

The approach taken is reminiscent of the EU-REACH testing strategy (see below), in that the OECD421 protocol is given as a screening test which can give initial information, but may be waived if more definitive testing is anticipated. Also similar to EU-REACH, the minimum data requirement in order to have an assessment of the various endpoints covered by the term reproductive toxicity is 2-generation study (TG 416) and developmental toxicity studies in two species (TG 414), the need for the second species being dependent on the outcome of the study in the first species.

3.2 REACH reproductive toxicity testing strategy

During the calendar year 2006 the testing strategy under REACH for reproductive toxicity was extensively debated in an EU expert group consisting of experts from academia, government and industry. A majority proposal was formulated which can be summarized as follows. Chemicals may not need to be tested in case they are classified as reproductive toxicant or genotoxic carcinogen or germ cell mutagen and sufficient robust data for risk assessment are already present, as well as in case of low

toxicological activity and negligible systemic absorption and negligible exposure. In line with and within the framework of existing EU regulations, testing is triggered by tonnage level. Figure 8 gives an overview of requirements and options. The OECD421 screening test is the basic initial reproductive toxicity test to be performed at tonnage level 1, followed by optional 2-generation and developmental toxicity studies at the 10 tpa level if triggered by the OECD421 test and other existing data. At the 100 tpa level, in addition to the 10 tpa level requirements, the developmental toxicity study is mandatory for all substances, and a developmental neurotoxicity study (OECD draft 426) may be added as an adjunct to the 2-generation study when triggered by existing data. At tonnage level 2 both a 2-generation study and a developmental toxicity study are mandatory and therefore preclude the need for a screening study. At levels 1 and 2, a developmental toxicity study in a second species may be triggered by equivocal findings in the first developmental toxicity study. The second developmental toxicity study is done in an alternative species, usually the rabbit to complement the first study which is usually done in the rat.
The discussion about the application of the developmental neurotoxicity study (OECD draft 426) was focussed on what findings should trigger such a study, on its practicality in terms of labour-intensiveness, on the difficulties in interpretation of the data, and on the fact that it is still a draft guideline and not a formal requirement yet.
There has been ample discussion about the utility of alternative test methods in this testing strategy. In agreement with current EU regulation, these tests can give rise to useful information that can trigger and direct further dedicated testing. On the other hand, also in line with current EU regulation, results from alternatives cannot be used on their own for final decisions on classification & labelling and risk assessment.

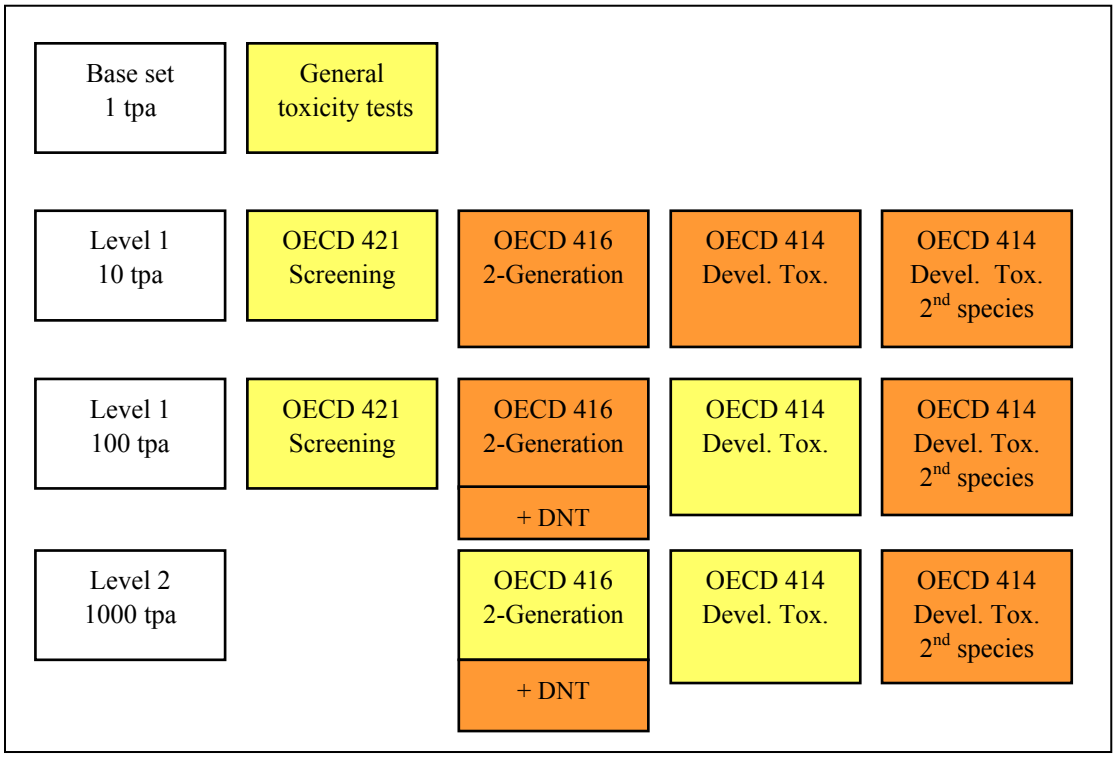

Figure 8. Schematic representation of suggested REACH reproductive toxicity testing strategy. Tests marked in yellow are basic requirements. Tests marked in orange are only required in case of concern. tpa =tonne(s) per annum; DNT = developmental neurotoxicity

3.3 Rabbit versus rat developmental toxicity study

Current strategies for reproductive toxicity testing prescribe the rat as the experimental species for the first developmental toxicity study. This may be followed by a similar study in the rabbit as a non-rodent species, dependent on the outcome of the rat study. Especially equivocal findings in the rat study may trigger a rabbit study. The question has been put forward whether the rabbit study has additional value over a similar study in the rat. The thalidomide experience (see above) has shown that the rabbit may show effects different from those in the rat, although in both species embryotoxic effects were observed that might have triggered specific risk management. It is pertinent therefore to retrospectively compare findings in rats and in rabbits for the same chemicals tested. The data gathered during the last 2-3 decades using the standardized OECD414 developmental toxicity test design make this possible. RIVM is currently carrying out such a comparison, in order to be able to judge whether the conduct of a rabbit study is needed if a rat study according to modern standards has been performed. A critical appraisal is necessary before any well-informed discussion on simplifying the testing strategy can be possible. The reduction of animal use, cost and time will have to be carefully weighed against any possible increased risk of missing effects of chemicals to avoid dramatic consequences.

3.4 Reproductive toxicity in integrated testing strategies

Reproductive toxicity represents one group of toxicity end points among a wealth of others such as acute and chronic toxicity, carcinogenicity, neurotoxicity and immune toxicity, to mention just a few examples. When designing testing strategies it can be argued that reproductive toxicity should be regarded in relation to other toxic properties of the chemical under study. For instance, systemic genotoxic carcinogens will be reproductive toxicants as well, and therefore extensive reproductive toxicity testing for such compounds may not be necessary. In addition, techniques for predicting toxic properties such as (qualitative) structure-activity relationships ((Q)SAR), read-across, and grouping of chemicals may be taken into account. Integrated testing strategies (ITS) are being designed to improve toxicity testing efficiency by optimizing a tiered approach of toxicity testing integrating all toxicity end points aimed at the removal of any redundant testing. RIVM plays a pivotal role in the development of ITS in the EU Framework-6 project OSIRIS (21) which kicked off in 2007.
In conjunction with the above, RIVM has recently compared the sensitivity of the 2-generation study versus the 90-day subchronic toxicity study (22). A retrospective analysis was done of toxicity data of chemicals for which both a 2-generation study and a 90-day subchronic toxicity study were available. Overall No-Observed-Effect Levels (NOEL) were compared as well as the critical effect parameters. It was concluded that in terms of the overall NOEL the generation study was not more sensitive than the 90-day study, although the critical effects could be different in both studies. When redesigning testing strategies, the differences in data requirement for overall risk assessment (RA) versus classification and labelling (C&L) should be taken into account. For RA, a well-based overall NOAEL suffices, whereas for C&L specific information on toxicity to fertility and toxicity to development as well as information on lactational effects is required. Therefore, although for RA the generation study did not contribute additional info for the chemicals studied, the additional information on reproductive end points was important for C&L for reproductive toxicity. These findings are of utmost importance under the REACH regulation where several essential decisions are made on basis of C&L . These findings will also be included in ongoing discussions on ITS.

3.5 Globally harmonized system for classification and labelling

A globally harmonized system for classification and labelling of chemicals (GHS) was finalized in 2003, based on the mandate from the 1992 United Nations Conference on Environment and Development (UNCED) (23). Given the reality of the extensive global trade in chemicals, and the need to develop national programs to ensure their safe use, transport, and disposal, it was recognized that an internationally harmonized approach to classification and labelling would provide the foundation for such programs. Once countries have consistent and appropriate information on the chemicals they import or produce in their own countries, the infrastructure to control chemical exposures and protect people and the environment can be established in a comprehensive manner. The European Union is in the process of adopting and implementing this guideline (24). A stakeholders consultation period was open during the fall of 2006. The Commission will adopt the proposal as soon as possible. After adoption the proposal will be sent to Parliament and Council for co-decision. This regulation will replace the current Classification & Labelling practice in the EU, based on Directive 67/548/EEG, after a transitional period.
In daily practice in the EU, the GHS is not expected to bring forward major changes as it is largely similar to the current EU regulations. However, finding consensus for adaptations of the GHS will be even more time-consuming than for an EU regulation, in view of the larger and more diverse group of stakeholders. In actuality, the current discussions within the EU about potency considerations within C&L (see below) will meet with this situation.

3.6 Potency considerations in classification and labelling

The criteria for some of the classifications in the current EU C&L system and in the EU-GHS are based on the minimal dose of a substance necessary to exert a certain effect (potency) or on the weight of evidence that a substance has a certain effect. Examples of endpoints for which classification is mainly based on the weight of evidence are carcinogenicity, mutagenicity and reproductive toxicity (CMR). The CMR properties of mixtures are normally not tested. The classification and labelling of mixtures for CMR endpoints is therefore based on the classification of the ingredients and the percentage of each ingredient in the mixture. The current European system and the EU-GHS system contain default percentages for classification of mixtures with CMR properties but allow the use of specific concentration limits (SCL) based on the potency. Methods for the determination of SCL are used in the current European system for carcinogenicity and reproductive toxicity by the Technical Committee on Classification and Labelling (TC-C&L). The T25 concept for carcinogenicity with additional considerations is used as a measure for potency and a guidance document has been established in order to assist in establishing SCLs for carcinogens. However, the development of a potency scheme for reproductive toxicity is still in its infancy and a discussion of the scientific basis for a method has still to take place.
Also the UN-GHS contains default concentrations limits for the classification of mixtures for reproductive toxicity and allows the setting of SCLs also for self-classification. However, a method for deriving SCL is neither included in the UN-GHS nor in the EU-GHS. Therefore a method for the determination of SCL for reproductive toxic substances is needed. This method can be incorporated in the Guidance on classification and labelling (EU REACH Implementation Plan RIP3.6) of the EU.
The use of reproductive toxicity potency for the setting of SCL was discussed at the OECD level. However, it was concluded that the available scientific knowledge on the complex issue of reproductive toxicity potency does not allow a general revision of the classification criteria, and only a case by case approach, on the basis of GHS section 1.3.3.2, is possible. However, it was considered that the establishment of a list of chemicals with potency ranking and information on the distribution of

NOAELs/LOAELs for critical endpoints would be useful in providing information for increasing background knowledge for decisions on reproductive toxicity potency.
Classification of a substance as a reproductive toxicant category 1A/B or 2 (EU-GHS system) has several legislative consequences in Europe independent on the potency. Therefore some parties advocate to include potency considerations in the classification system.
A current study by a subgroup of the TC-C&L under Dutch (RIVM) leadership aims at the development of a scientifically robust methodology for the determination of potency of substances that are toxic to reproduction (25). This methodology could subsequently be used to underpin the development of guidance on the setting of specific concentration limits. Also the scientific underpinning of the general concentration limits in 1999/45/EC and EU-GHS could be investigated when a method is available. As a second step, and when and/or if considered appropriate, the concept of potency could be discussed to be one factor to be considered when discriminating substances in different categories for classification.

3.7 Validation of novel test systems: OECD Guidance Document No.34

With the expanding development of new and updated test guidelines, the need to harmonize the assessment of their validity also emerged. In August 2005, OECD published a Guidance Document on the validation and international acceptance of new or updated test methods for hazard assessment, after a process that was initiated at an OECD Workshop in Solna, Sweden in 1996. The purpose of the document is to provide guidance on issues related to the validation of new or updated test methods consistent with current approaches. The document provides a synopsis of the current state of test method validation. Test method validation is a process based on scientifically sound principles by which the reliability and relevance of a particular test, approach, method, or process are established for a specific purpose. Reliability is defined as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardised protocol. The relevance of a test method describes the relationship between the test and the effect in the target species and whether the test method is meaningful and useful for a defined purpose, with the limitations identified. In brief, it is the extent to which the test method correctly measures or predicts the (biological) effect of interest, as appropriate. Regulatory need, usefulness and limitations of the test method are aspects of its relevance. New and updated test methods need to be both reliable and relevant, *i.e.*, validated.
The GD34 procedures of validation are internationally agreed and scientifically sound. They provide an important safeguard for test system quality. For in vitro alternative methods the procedures can be met in practice although they are laborious. However, validation of in vivo methods according to GD34 has met with difficulty. Examples are the enhancement of the OECD407 and the uterotrophic assay, for which validation studies were carried out which were individually of good quality but with a very limited total number of chemicals which limited definitive statements on the validity of the assays. This is due partly to difficulties in funding expensive animal studies that are ‘only’ aimed at method validation. In addition, it can be argued that in vivo methods would not need to be validated to the extent necessary for reductionistic in vitro methods for human hazard characterization, because (mammalian) whole-animal methods are by default the most complete approximation of the human situation available. Using large numbers of animals just for validation purposes may meet with ethical objections. Rather, for in vivo methods, retrospective validation may be preferable after sufficient data have been gathered by applying the method. In line with this notion, GD34 states that for test methods in which there are published data available, the first effort should be to conduct a thorough examination of the peer-reviewed scientific literature, and any other relevant and credible reports and publications

for information about the performance of the test method. GD34 continues to state that in cases when data gaps need to be filled as part of the process to achieve adequate validation for a specific proposed use, attempts should be made to limit the extent of new animal testing as far as possible, and this is indeed a point of concern as indicated. GD34 will have implications for OECD test development and implementation, to an extent that will become apparent in the future.

4 Discussion and conclusions

Regulatory reproductive toxicity testing is currently undergoing a period of extensive revision and renewal. This is stimulated by developments such as the endocrine disrupter issue and the EU REACH legislation, and by the realization that animal use in regulatory reproductive toxicity testing is relatively high. These issues have become of interest in a period when the existing testing paradigms have been in place for more than two decades, which in itself warrants a retrospective evaluation of their performance with the aims of improvement and refinement.

As regards updating of existing guidelines, several activities are ongoing. Alternative assays which reduce animal use will hopefully be useful in screening situations preceding in vivo testing. The updated OECD407 subchronic toxicity study carried out at base set will increase the knowledge of adverse effects on adult reproductive organs. The OECD421 reproductive screening assay will be able to give insight at tonnage level 1, 10 tpa of reproductive performance without the immediate need to perform a complete generation study in case of no adverse effect findings. The extended one-generation study may in the long run replace the 2-generation study and thereby reduce animal use, and time and cost of testing. Within this protocol, neurobehavioural testing as now defined in the draft OECD426 protocol can be incorporated as necessary. In addition, specific direct juvenile exposure could be incorporated in this protocol as appropriate. Finally, if the second species developmental toxicity study (OECD414) can be omitted, this will also result in a considerable reduction of testing. It should be made clear that most of these changes are still subject to investigation, and will rely heavily on the retrospective analysis of existing data. Such analyses are ongoing, with RIVM playing an important role. If for the moment we speculate that all these changes will take effect, this would result in the testing strategy under REACH depicted in Figure 9, which provides a considerable simplification of the current practice given in Figure 2 above.

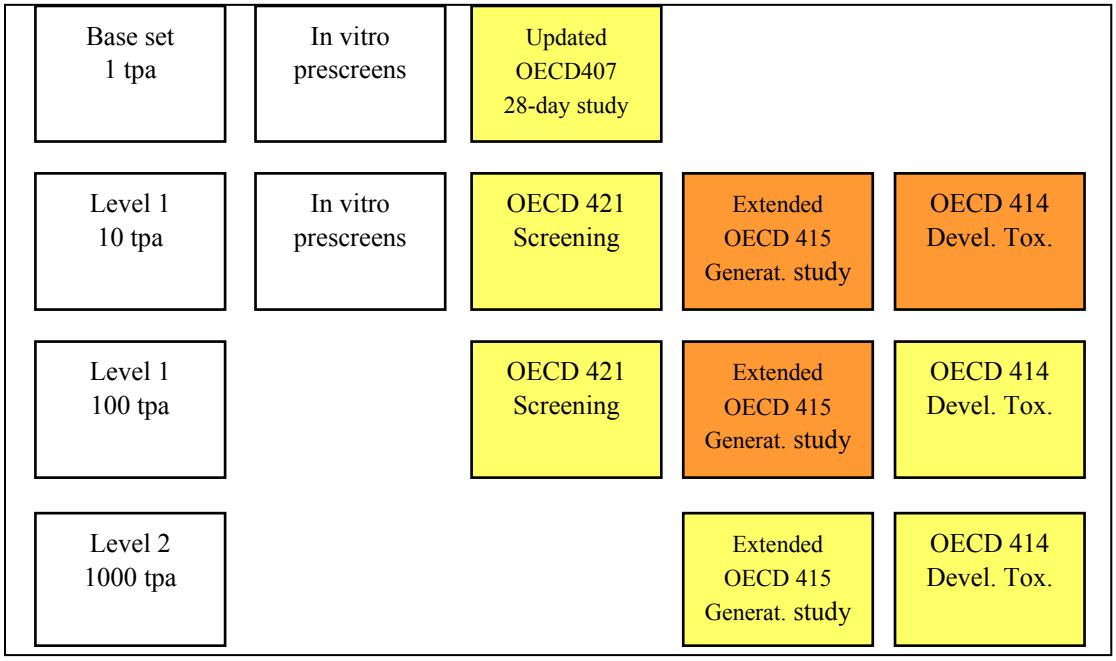

Figure 9. Schematic representation of possible future minimal reproductive toxicity testing requirements under REACH, based on speculative extrapolation from currently ongoing discussions. Tests marked in yellow are basic requirements. Tests marked in orange are only required in case of concern. tpa =tonne(s) per annum; DNT = developmental neurotoxicity.

With respect to novel test guidelines, animal and non-animal tests can be distinguished. The new OECD426 neurodevelopmental toxicity study and the extended one-generation study may eventually be merged as indicated above. Other novel test systems have largely emerged downstream from the endocrine disrupter issue. They are mostly single end point tests and their application will largely be in prescreening and prioritization of compounds before final testing is done using definitive studies for risk assessment and classification and labelling purposes. The uterotrophic and Hershberger assay, being single end point in vivo assays, may be replaced by in vitro receptor binding and activation assays in conjunction with kinetic information from e.g. subchronic toxicity studies. This would save animal use probably without losing critical information at the screening level. In vitro alternatives such as the rat postimplantation whole embryo culture (WEC) and the embryonic stem cell test (EST) can be very useful in providing developmental toxicity and mechanistic data which can be used in understanding modes of action and in prioritizing in vivo testing. RIVM is heavily involved in the further development of these assays which make use of genomic approaches to improve their predictivity. Their reductionistic nature makes it difficult to extrapolate data to the complete pregnant mammal, especially in cases when no effects are found. Similar considerations are relevant for receptor binding and activation assays such as the calux assays, which may inform about possible effects on fertility. Also these can be useful in a screening situation, but will usually not give results on which definitive risk and hazard assessment can be based. In silico methods such as gouping, read-across, and (Q)SAR require extensive further development before they can usefully be applied in reproductive toxicology. Both basic principles as well as available data are limited at the present time.

In the area of Classification & Labelling the immanent adoption of the Globally Harmonized System by the EU will be instrumental in converging C&L practice worldwide. At the same time, the discussion about the hazard basis of classification remains a matter of dispute. In Europe, potency is currently considered for possible inclusion in the derivation of specific concentration limits of chemicals in preparations. The incorporation of potency considerations for classification is a more far-reaching idea, because it would leave the basic concept that classification is strictly hazard based and include some elements of risk. The latter point has been raised in view of the wide array of downstream regulations in the EU which basically take risk management decisions on the basis of hazard-based classification. This practice omits the crucial intermediate stage of risk assessment which also takes exposure estimates and potency considerations into account. It should be stated, however, that possibly overconservative downstream legal consequences provide insufficient arguments for changing the classification system itself. Rather, the consequences of classifications may need some retrospective analysis. In this respect, countries outside the EU have sometimes very different downstream regulations, although in the future the same GHS for classification will be implemented.

Integrated testing strategies take a step away from individual tests and toxicology domains in that they use testing schemes developed within specific areas of toxicology such as reproductive toxicology and attempt to integrate them on the level of toxicology as a whole. The aim is to achieve increased efficiency of testing through waiving of redundant tests based on results gathered in earlier stages of a tiered approach. Also, combination of more end points in one test can improve efficiency, an example of which is the updated OECD407 subchronic toxicity test with added parameters on endocrine organs. Alternatives can play a role at the earliest stages of testing serving to prioritize and direct further testing. The EU FP6 OSIRIS integrated project, in which RIVM participates, aims at developing such an integrated testing strategy. Its design should be critically based on retrospective analysis of existing data gathered over the last decades. In the coming years these activities are expected to lead to new evidence-based proposals for integrated testing strategies. This will hopefully result in significant efficiency gain and reduction and refinement of animal testing whilst retaining the current high standards of basic information requirements which form the indispensable basis for classification & labelling and risk assessment of chemicals.

rivm

References

1. OECD Guidelines for the Testing of Chemicals – Section 4 – Health effects
http://www.oecd.org/document/55/0,2340,en_2649_34377_2349687_1_1_1_1,00.html, June 2007

2. EU Technical Guidance Document on Risk assessment:
http://ecb.jrc.it/documents/TECHNICAL_GUIDANCE_DOCUMENT/EDITION_2/tgdpart1_2ed.pdf,
June 2007

3. McBride, W. G.: Thalidomide and congenital abnormalities, Lancet 2:1358 (Dec. 16) 1961.

4. Lenz, W.: Kindliche Missbildungen nach Medikament wahrend der Graviditat, Deutsch Med Wschr 86:2555-2556 (Dec. 29) 1961.

5. Carlsen E., Giwercman A., Keiding N., Skakkebaek N.E. Declining semen quality and increasing incidence of testicular cancer: is there a common cause?
Environ Health Perspect. 1995 Oct;103 Suppl 7:137-9.

6. Our Stolen Future.Theo Colburn, Dianne Dumanoski, and John Peterson Myers. Dutton Publishing 1996

7. OECD Chemicals Testing – Guidelines - Endocrine Disrupter Testing and Assessment
http://www.oecd.org/document/62/0,2340,en_2649_34377_2348606_1_1_1_1,00.html, June 2007

8. European Chemicals Bureau – REACH http://ecb.jrc.it/REACH/, June 2007

9. van der Jagt K., Munn S., Tørsløv J., de Bruijn J. Alternative approaches can reduce the use of test animals under REACH. Addendum to Assessment of additional testing needs under REACH Effects of (Q)SARS, risk based testing and voluntary industry initiatives. 2004;JRC Report EUR 21405 EN.

10. OECD Report of the Initial Work Towards the Validation of the Rodent Uterotrophic Assay - Phase 1; and OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2 - Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories OECD Review documents 65 and 66,
http://www.oecd.org/document/30/0,2340,en_2649_34377_1916638_1_1_1_1,00.html, June 2007

11. OECD Draft Hershberger background review document, december 2006,
http://www.oecd.org/dataoecd/18/57/37880949.pdf

12. Gray, L.E. Jr., J. Furr, and J.S. Ostby (2005). Unit 16.9, Hershberger assay to investigate the effects of endocrine-disrupting compounds with androgenic or anti-androgenic activity in castrate-immature male rats. In: Current Protocols in Toxicology, Chapter 16: Male reproductive toxicology, John Wiley and Sons, NY, pp. 16.9.1-16.9.15.

13. Gelbke HP, Hofmann A, Owens JW, Freyberger A. The enhancement of the subacute repeat dose toxicity test OECD TG 407 for the detection of endocrine active chemicals: comparison with toxicity tests of longer duration. Arch. Toxicol. 81(4):227-250, 2007.

14. Hass U. The need for developmental neurotoxicity studies in risk assessment for developmental toxicity. Reprod. Toxicol. 22: 148-56, 2006.

15. Cooper RL, Lamb JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, Eisenbrandt DL, Fenner-Crisp PA, Hines RN, Irvine LF, Kimmel CA, Koeter H, Li AA, Makris SL, Sheets LP, Speijers G, Whitby KE. A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol. 2006 Jan;36(1):69-98. Review.

16. Janer G., Piersma A., Slob W., Vermeire T., Hakkert B. "A retrospective analysis of the 2-generation study: What is the added value of the second generation?" Reproductive Toxicology, 2007, accepted for publication.

17. European Medicines Agency – CHMP – Draft Guideline on the need for non-clinical testing in juvenile animals on human pharmaceuticals for paediatric indications, september 2005.

18. Piersma AH, Genschow E, Verhoef A, Spanjersberg MQI, Brown NA, Brady M, Burns A, Clemann N, Seiler A, Spielmann H. Validation of the postimplantation rat whole embryo culture test in the International ECVAM Validation study on three in vitro embryotoxicity tests. ATLA 32: 275-307, 2004.

19. Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K. Validation of the embryonic stem cell test in the international validation study on three in vitro embryotoxicity tests. ATLA 32: 209-244, 2004.

20. M. Luijten, A. de Vries, A. Opperhuizen, A.H. Piersma. Alternative methods in reproductive toxicity testing: state of the art. RIVM report 340720002/2006.

21. EU-FP6 project OSIRIS: http://www.ufz.de/index.php?en=11601

22. Janer G., Hakkert B.C., Piersma A.H., Vermeire T., Slob W.
A retrospective analysis of the added value of the rat 2-generation reproductive toxicity study versus the rat subchronic toxicity study. Reproductive Toxicology, 2007, accepted for publication.

23. UNECE Globally Harmonized System of Classification and Labelling of Chemicals (GHS) http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html, June 2007

24. EU E&I Stakeholder Consultation In Implementation of the Globally Harmonised System of Classification and Labelling of Chemicals (GHS) in Community Legislation http://ec.europa.eu/enterprise/reach/ghs_consultation_en.htm

25. Consideration of potency in classification for toxicity to reproduction. ecb.jrc.it/classlab/3004a1_NL_reprotoxicity_potency.doc

26. Pedersen F, de Bruijn J, Munn S, van Leeuwen K. Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. EU-JRC-ECB, 2003

27. Maslankiewicz L, Hulzebos EM, Vermeire TG, Muller JJA, Piersma AH. Can chemical structure predict reproductive toxicity? RIVM Report no. 601200005/2005.