

RIVM report 340720002/2007

**Alternative methods in reproductive toxicity
testing: state of the art**

M. Luijten, A. de Vries, A. Opperhuizen,
A.H. Piersma

Corresponding author:

M. Luijten
Laboratory of Health Protection Research
Email address: Mirjam.Luijten@rivm.nl

This investigation has been performed by order and for the account of the ministry of Public Health, Welfare and Sports, within the framework of project V/340720, Alternatives to Animal Testing.

Abstract

Alternative methods in reproductive toxicity testing: state of the art

This report overviews the currently existing *in vitro* models for reproductive toxicity and describes their potential use in reproductive toxicity testing, as well as the modular approach of validating these alternative test systems. Here, it is also indicated why these tests have not been implemented yet in any of the guidelines. A limited number of alternative tests for embryotoxicity, which can be applied after optimisation, has been validated. Although none of the *in vitro* methods represents a replacement for current animal tests, these methods can function as a pre-screen in a tiered approach. In this way, the industry can set priorities for *in vivo* testing and thus reduce animal use in reproductive toxicology. Development of *in vitro* methods for reproductive toxicity testing is essential, since testing according to the present guidelines requires a high number of test animals. The numbers of test animals needed range from 560 – 6,000 animals per chemical, dependent on the production volume. Moreover, increased animal use is envisaged as a result of the implementation of REACH, the new European legislation for manufacturers and users of chemicals. REACH requires toxicological information to be submitted for about 30,000 existing chemicals, which will require millions of test animals. Many alternative test methods have already been developed for reproductive toxicity testing, and several of them are in a (pre-)validation process. However, none of these methods offers enough reliability to replace *in vivo* assays currently used. Testing toxic effects on the reproductive cycle is complex, since reproduction is characterised by many different processes and sensitive periods. Due to its complexity, the whole cycle cannot be modelled in one *in vitro* system. Parts of the system need to be studied individually and then integrated into testing strategies.

Key words: reproductive toxicity, alternatives to animal testing, *in vitro* method, validation, embryotoxicity

Rapport in het kort

Alternatieve testmethoden voor toxiciteit op reproductie: stand van zaken

Dit rapport geeft een overzicht van de *in vitro* testmethoden die de afgelopen decennia voor reproductie-toxiciteitonderzoek zijn ontwikkeld. Het validatieproces van deze alternatieven voor dierproeven wordt hier eveneens beschreven. Daarnaast geeft het rapport aan wat de knelpunten zijn om deze methoden in internationale testrichtlijnen te kunnen invoeren. Inmiddels zijn voor embryotoxiciteit een aantal testen gevalideerd, die in de toekomst na verdere optimalisatie toegepast kunnen worden. Daarnaast kunnen individuele *in vitro* testen mogelijk ingezet worden bij het screenen van stoffen. Dit kan leiden tot een aanzienlijke vermindering van het proefdiergebruik voor reproductie-toxiciteitstesten. Hoewel verscheidene alternatieven voor dierproeven voor reproductie-toxiciteitstesten voorhanden zijn, blijken veel van deze methoden nog onvoldoende ontwikkeld en niet gevalideerd te zijn. De huidige testrichtlijnen voor reproductie-toxiciteit vergen een groot aantal proefdieren. De vereiste aantallen variëren van 560 – 6.000 proefdieren per chemische stof, afhankelijk van de hoeveelheid waarin de stof wordt geproduceerd of geïmporteerd. Bovendien dreigt het gebruik van proefdieren sterk toe te nemen vanwege de invoering van REACH, de nieuwe Europese wetgeving voor registratie, evaluatie en autorisatie van chemische stoffen. Binnen REACH zullen naar verwachting miljoenen dieren gebruikt worden om 30.000 bestaande stoffen alsnog te registreren. *In vitro* testmethoden op het gebied van reproductie toxiciteit zijn sterk in ontwikkeling. Ze bevinden zich momenteel veelal in het stadium van validatie, waardoor ze nog niet inzetbaar zijn. Reproductie is een complex samenspel van zeer verschillende processen. Om het effect van chemische stoffen hierop te kunnen toetsen is een reeks van *in vitro* modellen nodig, die vervolgens samengevoegd worden in een teststrategie.

Trefwoorden: reproductie toxiciteit, alternatieven voor dierproeven, *in vitro* testmethode, validatie, embryotoxiciteit

Contents

1. Reproductive toxicity testing.....	6
1.1 Reproductive toxicity.....	6
1.2 Current test guidelines.....	7
1.3 Testing strategy	9
1.4 Classification and labelling.....	10
1.5 Risk assessment	11
2. Alternative methods in reproductive toxicity testing: development of assays.....	13
2.1 In vitro methods for reproductive and developmental toxicity.....	13
2.2 State of the art	14
2.3 Validated alternatives for developmental toxicity	14
3. The validation paradigm: from development to implementation of an alternative test	17
3.1 The modular approach	17
3.2 Test definition	18
3.3 Within-lab variability	18
3.4 Transferability.....	18
3.5 Between-lab variability.....	19
3.6 Predictive capacity.....	19
3.7 Applicability domain	21
3.8 Performance standards.....	21
3.9 Guideline development	22
4. Reduction of animal use: near future?	23
4.1 Current animal use in reproductive toxicology.....	23
4.2 Possibilities for reduction in animal use	23
4.3 REACH	25
5. General conclusions and discussion	26
References.....	27
Appendix 1: Overview of in vitro alternatives for reproductive and developmental toxicity.....	29
Appendix 2: Abbreviations	32

Summary

Reproductive toxicity testing according to the present guidelines requires a high number of test animals. These numbers are envisaged to increase because of the implementation of REACH, the new European legislation for manufacturers and users of chemicals. As a consequence of REACH, toxicological information will have to be submitted for about 30,000 existing chemicals. This will require millions of test animals. Development of alternative methods is therefore essential. The present report gives an overview of the currently existing *in vitro* models for reproductive toxicity, and describes their potential use in reproductive toxicity testing. Many alternative test methods already have been developed. However, none of them offers enough reliability to replace currently used *in vivo* assays.

Although none of the *in vitro* methods represents a replacement for current animal tests, they can function as pre-screen in a tiered approach. It is clear that considerable effort is required to further develop the various existing *in vitro* methods into reliable alternative tests that cover critical aspects of the reproductive cycle. In addition, new tests will have to be developed since the currently available tests are not sufficient. However, they should be developed with great care to make sure that reliable *in vitro* methods are obtained, which can be implemented in testing strategies. Furthermore, new methodologies, such as gene expression profiling in *in vitro* methods, need to be investigated.

Testing of toxic effects on the reproductive cycle is complex, since reproduction comprises many different processes and sensitive periods. Due to its complexity, it is not possible to model the whole cycle in one *in vitro* system. Parts of the system need to be studied individually and then integrated into testing strategies. Both the individual test methods and the testing strategy need to be validated. The modular approach of validation of alternative test systems is described in this report. A limited number of the alternative tests developed so far has been validated or are in the (pre)validation process. Despite successful validation, not a single *in vitro* method has been implemented yet in international test guidelines. This is mainly caused by outstanding issues with regard to the relevance of the various alternative methods. This aspect of the validation process should receive more attention, since it seems to be the bottleneck for implementation.

1. Reproductive toxicity testing

1.1 Reproductive toxicity

Many compounds are known to have adverse, toxic effects on the reproductive cycle. Well-known examples are thalidomide, also known as Softenon, retinoic acid (vitamin A), and diethylstilbestrol (DES). The various toxicological effects that may occur in any phase within the reproductive cycle (see Figure 1) are referred to as reproductive toxicity. This includes effects on fertility, sexual behaviour, embryo implantation, embryonic/foetal development, parturition, postnatal adaptation, and on subsequent growth and development into sexual maturity. As a consequence, an enormous variety of mechanisms on the molecular, cellular and tissue levels cooperate in a concerted and genetically programmed way to regulate normal development. The sensitivity to chemical insults may differ extensively between processes. In addition, different windows of sensitivity in time have been observed for different processes. As an example, neural tube closure occurs early in pregnancy, and most adverse effects on this process can only be detected after exposure during this critical period of time.

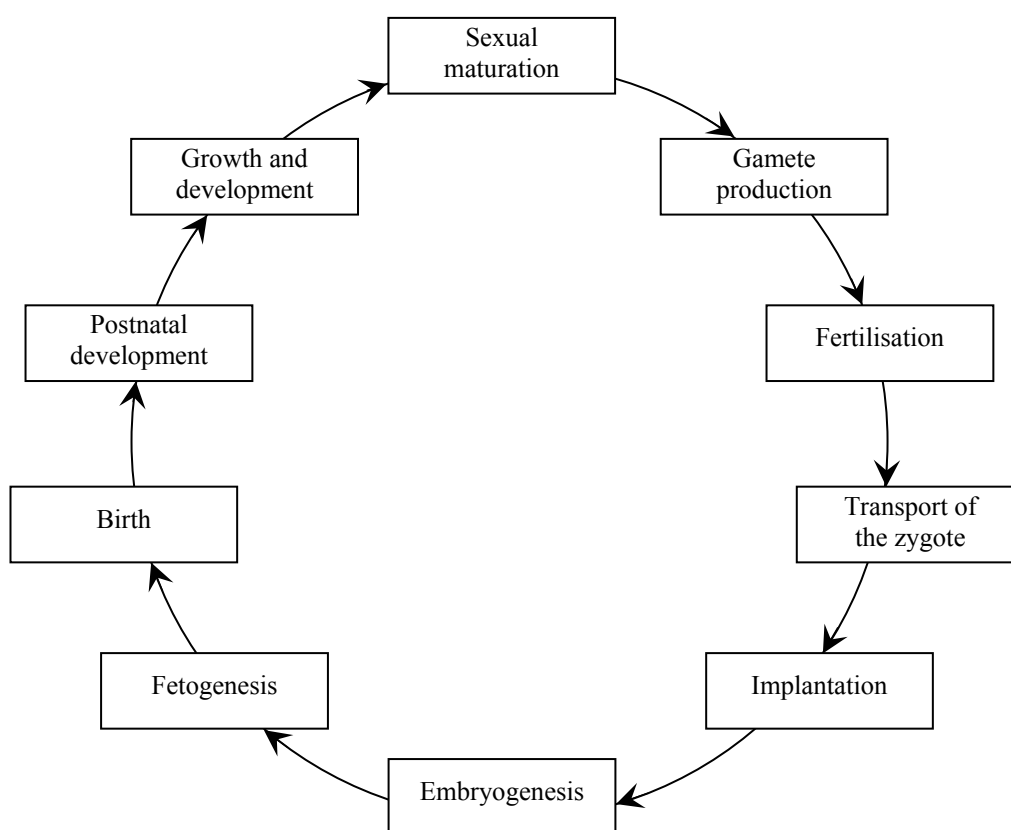


Figure 1. The reproductive cycle.

1.2 Current test guidelines

The European Union (EU) legislation requires the assessment of the potential toxicological effects of new chemicals or drugs on human health, before they are accepted on the market. Currently, a number of key phases and events in the reproductive cycle are covered in standardised animal tests used for regulatory purposes. Three major key phases of the reproductive cycle have been identified:

- (i) maturation, which includes postnatal development, lactation and gamete production and release;
- (ii) mating, which includes fertilisation and implantation of the embryo;
- (iii) gestation, which comprises prenatal development.

These key phases are covered by five OECD (Organisation for Economic Co-operation and Development) test guidelines (TGs): the prenatal developmental toxicity study (OECD TG 414), the one-generation study (OECD TG 415), the two-generation study (OECD TG 416), the reproductive/developmental toxicity screening test (OECD TG 421), and the repeated dose toxicity study combined with the reproductive/developmental toxicity study (OECD TG 422). In addition, the OECD is working on a new guideline, the developmental neurotoxicity study (draft OECD TG 426).

These test guidelines are animal-based tests. They all require high numbers of animals, mainly because of the study design, but also because in each of these guidelines at least three dose levels, besides a control group, are tested. The one- and two-generation studies, TG 415 and TG 416, are used to provide information on fertility, growth and development of the foetus, and, in case of the two-generation study, functional assessment of fertility after complete reproductive cycle exposure. Animals are exposed from 4 weeks prior to mating to weaning of the first and second generation, respectively. Both generation studies start with 25 breeding pairs. Assuming an average litter size of 12 pups, the next generation consists of 300 animals. In total, the one- and two-generation studies require no less than 1,400 and 2,600 animals, respectively. The developmental neurotoxicity (DNT) test, designed to screen for adverse effects of pre- and postnatal exposure on the development and function of the nervous system, is ideally performed in animals generated in the two-generation test.

The reproduction/developmental toxicity screening tests, TG 421 and TG 422, do not provide complete information on all aspects of reproduction and development. However, information on possible effects on male and female reproductive performance such as fertility, pregnancy length and birth, development of the foetus and survival to day 4 post-partum can be obtained from these tests. Ten breeding pairs are exposed from 2 weeks prior to mating to day 4 post-partum. Both tests require a total of 560 animals each. The prenatal developmental toxicity study, 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 foetus. Each group in this study consists of 25 breeding pairs. Pregnant animals are exposed from implantation, i.e. day 5 post mating, to the day before birth. This study requires a total of 1,400 animals. An overview of the guidelines, including the end-points measured, is given in Table 1.

Table 1. Overview of *in vivo* tests for reproductive toxicity testing.

Test	OECD	Exposure period	End-points	Number of animals
Prenatal developmental toxicity	TG 414	From implantation to the day before birth	Embryonic development; fetal growth; morphological variations and malformations	1,400
One-generation	TG 415	Continuously over one generation	Fertility; growth, development and viability; histopathology of sex organs, brain and target organs	1,400
Two-generation	TG 416	Continuously over two generations	Fertility; growth, development and viability; histopathology of sex organs, brain and target organs; functional assessment of fertility after complete reproductive cycle exposure	2,600
Reproduction/developmental toxicity screening test	TG 421/422	From 2 weeks prior to mating to day 4 post-natally	Fertility; pregnancy length and birth; Fetal and pup growth; survival to day 4	560
Developmental neurotoxicity	Draft TG 426	During pregnancy and lactation	Pregnancy length and birth; physical and functional maturation; behavioural changes due to PNS/CNS effects; brain weight and neuropathology	

In order to fully assess the hazardous properties of a compound with respect to reproductive toxicity, the key data requirements are a reproductive/developmental toxicity screening test (TG 421), a two-generation study (TG 416) and a prenatal developmental toxicity study (TG 414). The level of data requirements increases progressively with increasing annual production volumes (see Figure 2). These production volumes are subdivided into tonnage bands of ≥ 1 tonne per year; ≥ 10 tonnes per year; ≥ 100 tonnes per year; and $\geq 1,000$ tonnes per year. Reproductive toxicity testing routinely commences at a production volume of 10 tonnes per year. Below this level, reproductive toxicity testing is not required and general toxicity tests suffice. General toxicity tests include (sub)acute toxicity studies, skin and eye irritation tests, the skin sensitisation test and the Ames mutagenicity assay.

Tests that belong to the basic requirements for reproductive toxicity testing are marked yellow in Figure 2. For production volumes from over 1 to 1,000 tonnes, this is the reproductive/developmental toxicity screening test; for production volumes of 1,000 tonnes or more, a two-generation study is required. Additional testing (marked orange), for both fertility and developmental toxicity, is only required in case of concern on the basis of earlier testing.

Base set 1 tpa	General toxicity tests				Total number animals
Level 1 10 tpa	OECD 421 Screening <i>n</i> = 560	OECD 416 2-Generation <i>n</i> = 2,600	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,960
Level 1 100 tpa	OECD 421 Screening <i>n</i> = 560	OECD 416 2-Generation <i>n</i> = 2,600 + DNT	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,960
Level 2 1000 tpa		OECD 416 2-Generation <i>n</i> = 2,600 + DNT	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,400

Figure 2. Reproductive toxicity testing strategy. Tests marked in yellow belong to basic requirements. Tests marked in orange are only required in case of concern. tpa =tonne(s) per annum; DNT = developmental neurotoxicity.

1.3 Testing strategy

To ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimized, a testing strategy has been developed. This strategy includes four general principles: preliminary information, test species, route of

administration, and dose levels. First, all preliminary information available should be gathered. Whatever is known on the physico-chemical, toxicokinetic, and toxicodynamic properties of the test substance, as well as any available relevant information on chemical analogues (i.e. structure activity relationships) and the results of previously conducted toxicity studies, should be taken into consideration. Second, the choice of species for the reproductive toxicity tests must be carefully considered in light of all available information. Normally, the rat is the species of first choice for the two-generation study because of the enormous extent of background data and because this species is routinely used for general toxicity tests. Third, the route of administration used should be the most appropriate one in relation to the likely route(s) of human exposure, the nature and physico-chemical properties of the substance and its toxicity (systemic effects), as well as the practical aspects of conducting tests for reproductive toxicity. Fourth, a test for reproductive toxicity should include a range of dose levels. The highest dose level should normally not exceed 1,000 mg/kg bodyweight/day and should, where possible, give rise to some, but not severe, toxicity in the parent animals without obvious suffering or lethality. The lower dose levels should be selected with the aim of establishing any dose-response relationship and the no observed adverse effect level (NOAEL) for reproductive toxicity.

The first specific reproductive toxicity test to be conducted usually is the reproductive/developmental toxicity screening test. In case of concern, a two-generation study and/or a developmental toxicity study are conducted. The design of the developmental toxicity study should take account of any information derived from previous studies, in particular dose-response relationships and information on maternal toxicity.

1.4 Classification and labelling

Classification and labelling involves an evaluation of the hazard of a substance and a communication of that hazard via the label. This evaluation must be made for any substance or preparation manufactured within or imported into the EU and placed on the EU market, and may result in classification of the substance as dangerous for one or several end-points concerning physical-chemical properties, health or environmental effects. Classification and labelling is therefore used as a tool for risk management of chemicals.

In the EU system reproductive toxicity is divided into effects on fertility and developmental toxicity. In addition, substances that may interfere with lactation or are present in breast milk in significant amounts can be classified as hazardous to breast-fed babies. There are three distinct categories regarding reproductive toxicity:

- Category 1: compounds known to impair fertility in humans and to cause developmental toxicity in humans;
- Category 2: compounds treated as if they impair fertility in humans or cause developmental toxicity in humans;
- Category 3: compounds, which cause concern for human fertility or which cause concern for humans owing to possible developmental toxic effects.

Classification of chemicals as reproductive toxin is intended to be used for chemicals which have an intrinsic or specific property to produce such toxic effects. Chemicals of most concern are those which are toxic exclusively to reproduction at exposure levels which do not produce other signs of toxicity.

1.5 Risk assessment

Reproductive toxicity is one of a series of subdisciplines within toxicology. Risk assessment is based on the complete toxicity profile of a chemical. The lowest no observed adverse effect level (NOAEL) is critical for use in the extrapolation to risk in man.

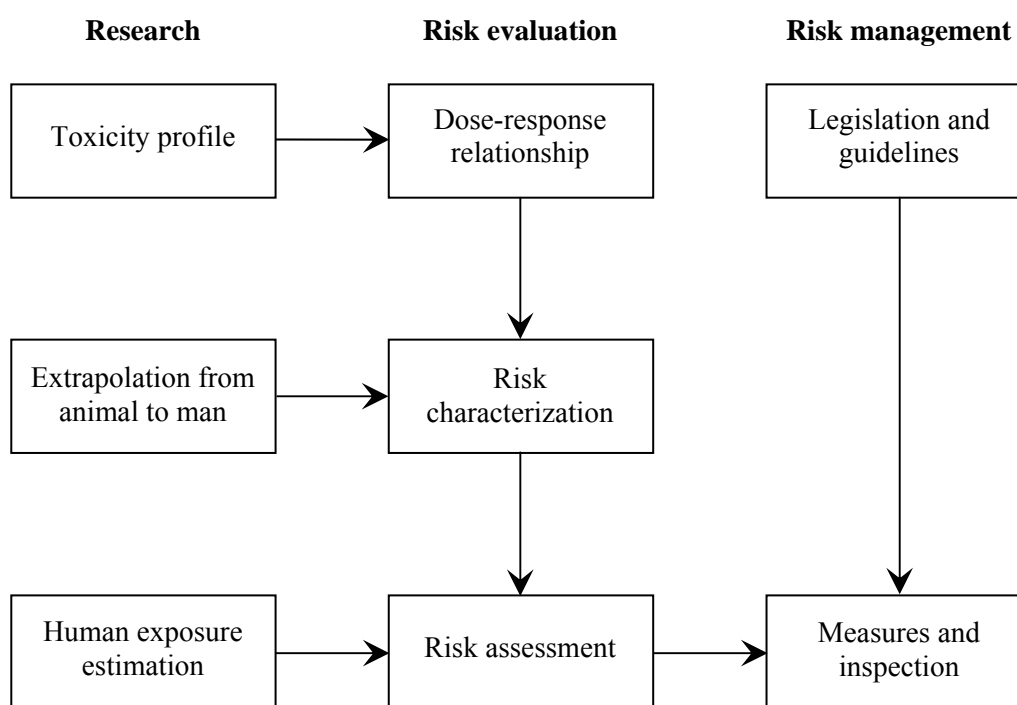


Figure 3. Schematic representation of the risk assessment paradigm.

This overall NOAEL may be derived from a reproductive toxicity end-point or from any other parameter indicating a toxic effect. Figure 3 shows the risk assessment paradigm.

Apart from toxicodynamics the NOAEL takes account of the dose-response relationship, the kinetics of the compound and the human exposure scenario. Uncertainty factors are applied to extrapolate the NOAEL to an Acceptable Daily Intake (ADI). By default, two factors of 10 are used, one for intraspecies and one for interspecies extrapolation, resulting in an overall default uncertainty factor of 100. For some instances regulatory bodies in the USA have introduced an extra uncertainty factor, specifically when the NOAEL was determined by reproductive end-points considered very harmful. However, this is not common use in the European Union. The ADI determines, together with exposure estimates, how a chemical can be used safely. In addition, legislation gives the instrumentation to take measures for e.g.

restricting use in order to keep human exposure below the ADI. Finally, inspection monitors whether measures are implemented and carried out appropriately.

2. Alternative methods in reproductive toxicity testing: development of assays

2.1 In vitro methods for reproductive and developmental toxicity

As described in the previous chapter, reproductive toxicity testing largely relies on *in vivo* assays. Therefore, there is an urgent and continuing need for alternative tests. Over the years, more than thirty culture methods have been developed for the various phases in reproductive and developmental toxicity. Individual *in vitro* models will never be able to cover all aspects of the reproductive cycle, because reproduction requires a complex of integrated functions. However, parts of the process can be mimicked in *in vitro* systems, and it is possible that a panel of well-designed and validated *in vitro* tests could replace a substantial proportion of *in vivo* testing procedures.

For example, *in vitro* alternatives for male reproductive toxicity use cultures of testicular cell types, such as Sertoli-germ cell co-cultures and Leydig cell cultures, to study specific features of testicular toxicity¹. A significant current effort to evaluate, optimize, and prevalidate various *in vitro* methods in reproductive toxicology is the integrated project ReProTect, funded by DG Research in the EU 6th Framework Programme. The overall aim of ReProTect is the integration of existing and newly developed *in vitro* models into a testing strategy that will provide detailed information on the hazard of compounds to the mammalian reproductive cycle. Our laboratory participates in ReProTect with the whole embryo culture.

Existing alternative test methods in reproductive toxicology, although not implemented, can be divided into three categories with increasing complexity: established cell lines, primary cell cultures, and mammalian embryos². Continuous cell lines have the exclusive advantage that they are devoid of the use of experimental animals. At the same time, their major disadvantage is their relative simplicity. They allow the study of effects on single mechanisms only, e.g. cellular differentiation³. Examples of alternatives using primary cell cultures are limb bud and neuronal cells from mammalian embryos⁴. Such tests require animal material, but they have the advantage that these cells have not been immortalized which may have possible adverse consequences for responses in test systems. Whole embryo culture systems offer the most complete *in vitro* alternatives for animal testing. They do require animal material and are more laborious but have the advantage of representing embryogenesis in its full complexity from cellular proliferation and differentiation to pattern formation. Test systems range from the Hydra pseudo-embryo system⁵, via frog⁶ and chicken⁷ to mammalian whole embryo culture⁸.

2.2 State of the art

A detailed description of the current *in vitro* alternatives for reproductive and developmental toxicity is given in Appendix 1. Appendix 1 also shows estimates of the time needed to validation. These data have been collected and estimated by the European Centre for Validation of Alternative Methods (ECVAM)⁹.

Reproductive toxicity can be subdivided in toxicity to fertility and developmental toxicity, as was done in the EU Classification & Labelling guideline (see section 1.4). A variety of *in vitro* methods has been developed for toxicity to fertility. Each of them is only a reductionistic representation of the complex combination of physiological processes contributing to fertility. Most of these methods need improvements in protocol and further research, before they can be considered to be ready for prevalidation by testing the effects of chemicals. A few of these *in vitro* models are at a more advanced stage of development. Such models include a Leydig cell model¹⁰, the follicle culture bioassay^{11;12}, and the aromatase test^{13;14}.

Also for developmental toxicity, a variety of *in vitro* tests is available. Again, the majority of these tests need further development and they do not represent complete replacements for animal tests for reproductive toxicity. However, three alternatives using mammalian cells and embryos have been successfully validated by testing a series of chemicals in a double-blind fashion in four independent laboratories: the embryonic stem cell test (EST), the micromass (MM) assay, and whole embryo culture (WEC) test. They will be discussed in more detail in section 2.3. The most promising alternative methods, for both reproductive and developmental toxicity, are given in Table 2.

2.3 Validated alternatives for developmental toxicity

So far, three alternative tests for embryotoxicity have reached the validated status: the embryonic stem cell test (EST), the whole embryo culture test (WEC) and the limb bud micromass (MM) test^{15;16}.

Embryonic stem (ES) cells have the potential to differentiate in culture into any cell type. The EST is based on the determination of inhibition of differentiation and growth. The embryotoxic potential of chemicals is determined by the evaluation of the inhibition of cardiac muscle differentiation of ES cells and the inhibition of growth of ES and 3T3 cells. The EST is performed with permanent cell lines from the mouse. One of the aims of ReProTest is to adapt the murine system to human embryonic stem cells, in order to predict developmental toxic effects in humans more precisely. Furthermore, other routes of differentiation, such as skeletal differentiation, will be studied.

In the WEC, post implantation rat embryos at early stages of organogenesis are cultured. At day 10 of gestation, pregnant rats are killed and embryos are isolated. The deciduous tissue, the Reichert membrane and parietal yolk sac of each embryo are carefully removed (Figure 4). Embryos are cultured for 48 hours in culture vessels and subsequently scored.

Table 2. Most promising in vitro tests for reproductive and developmental toxicity.

In vitro model	End-point	Applicability	Validation status
Reproductive Toxicity			
Follicle culture bioassay	Folliculogenesis, oogenesis, steroidogenesis	Chemicals interfering with ovarian function	Research & Development
Leydig tumour cell lines	Cytotoxicity, proliferation, apoptosis, steroidogenesis	Chemicals interfering with endocrine function (endocrine disruptors)	Prevalidation
H295R cells	Steroidogenesis	Chemicals interfering with endocrine function (endocrine disruptors)	Prevalidation
Aromatase assay	Aromatase inhibition	Chemicals interfering with aromatase activity	Research & Development
Developmental toxicity			
Embryonic stem cell test	Cell differentiation, viability and growth	Direct acting embryotoxic/teratogenic compounds	Validated
Limb bud micromass	Cell differentiation, viability and growth	Direct acting embryotoxic/teratogenic compounds	Validated
Whole embryo culture	Morphology, functionality, growth	Direct acting embryotoxic/teratogenic compounds	Validated

This allows the identification of compounds that induce embryotoxicity and malformations. One of the innovations considered to be valuable to be applied to the WEC system is the introduction of a metabolic system. This is currently investigated by our laboratory within the frame of the ReProTect project.

The MM assay is a simple cell culture system, in which development and differentiation of embryonic limb buds cells are studied. Single cell suspensions are prepared from limb buds isolated from 13-days-old embryos. Undifferentiated mesenchyme cells of limb buds will form differentiating foci of chondrocytes in micromass culture. Teratogenic compounds inhibit the formation of foci and can therefore be detected by a reduced number of foci, or a reduced number of cells within foci.

The three embryotoxicity tests have been validated after a validation program funded by ECVAM. Our laboratory served as lead laboratory in the formal validation study on the WEC¹⁷. For each of these three assays, a prediction model has been developed to classify the chemicals into three classes of embryotoxicity: non, weak, and strong. The scientific validity of the three methods was endorsed by the ECVAM Scientific Advisory Committee in October 2001: the EST and the WEC test were considered to be scientifically valid for distinguishing between non, weak/moderate and strong embryotoxins^{18;19}, whereas the MM test was considered scientifically valid for identifying strongly embryotoxic chemicals²⁰. However, the ESAC also recognized that these methods do not represent replacements for current animal tests. These tests should rather be used as screens to set priorities for *in vivo*

testing for regulatory purposes and as such provide suitable means for reducing and/or refining the use of animal testing in the context of the testing strategy.

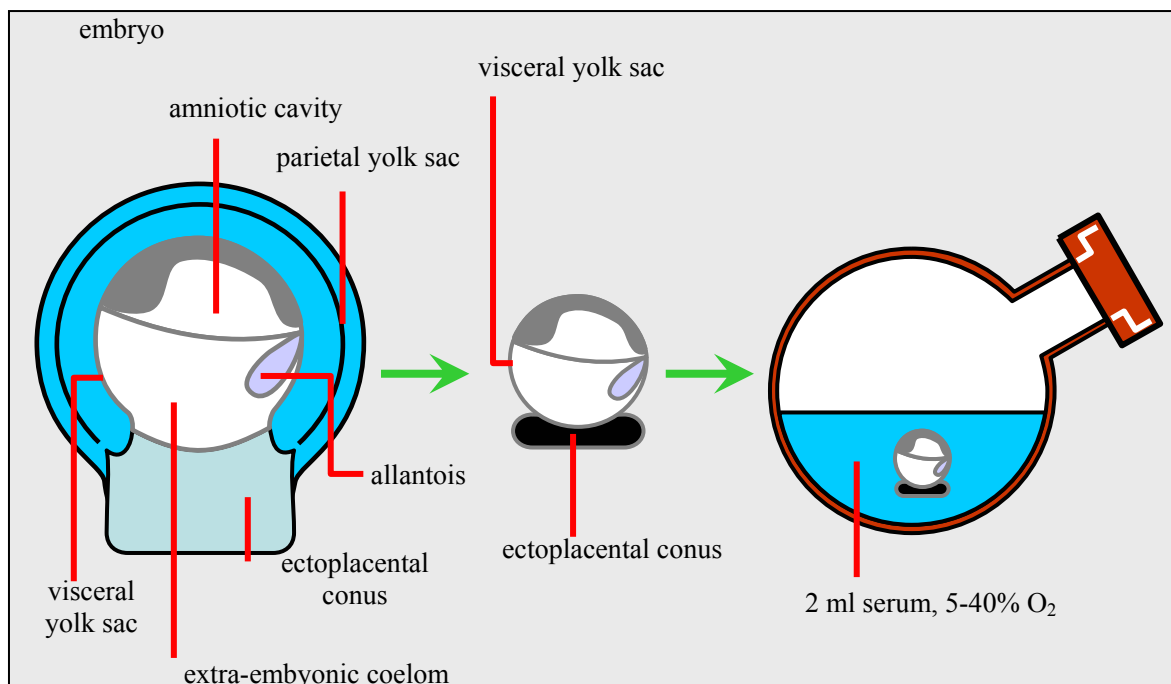


Figure 4. The post implantation whole embryo culture.

These tests have not been implemented yet in any of the guidelines. Their application as valid alternatives to animal testing is still under discussion. Questions remain as to the biological relevance and applicability of the mathematically derived prediction models, the applicability domain of the test systems in terms of classes of chemicals as well as the effect extrapolation, and the possibilities to test chemicals that need metabolic activation for expressing their embryotoxic properties. Clearly, validation is a key part of development of alternative assays. In the next chapter, the validation process will be discussed in more detail.

3. The validation paradigm: from development to implementation of an alternative test

3.1 The modular approach

Validation is the process by which reliability and relevance of a test are established for a particular purpose. ECVAM has proposed a modular approach for the validation of alternative test systems²¹. The three embryotoxicity tests described in the previous chapter were validated according to this validation paradigm. The approach is schematically represented in Figure 5. The presumption of the system is that the seven modules on the left should be adequately completed as judged by the validation management group before the test can enter the independent peer review. In this chapter this approach will be explained and illustrated with the embryonic stem cell test (EST). In addition, the post-validation trajectory towards guideline development and implementation of alternatives in international testing strategies will be discussed.

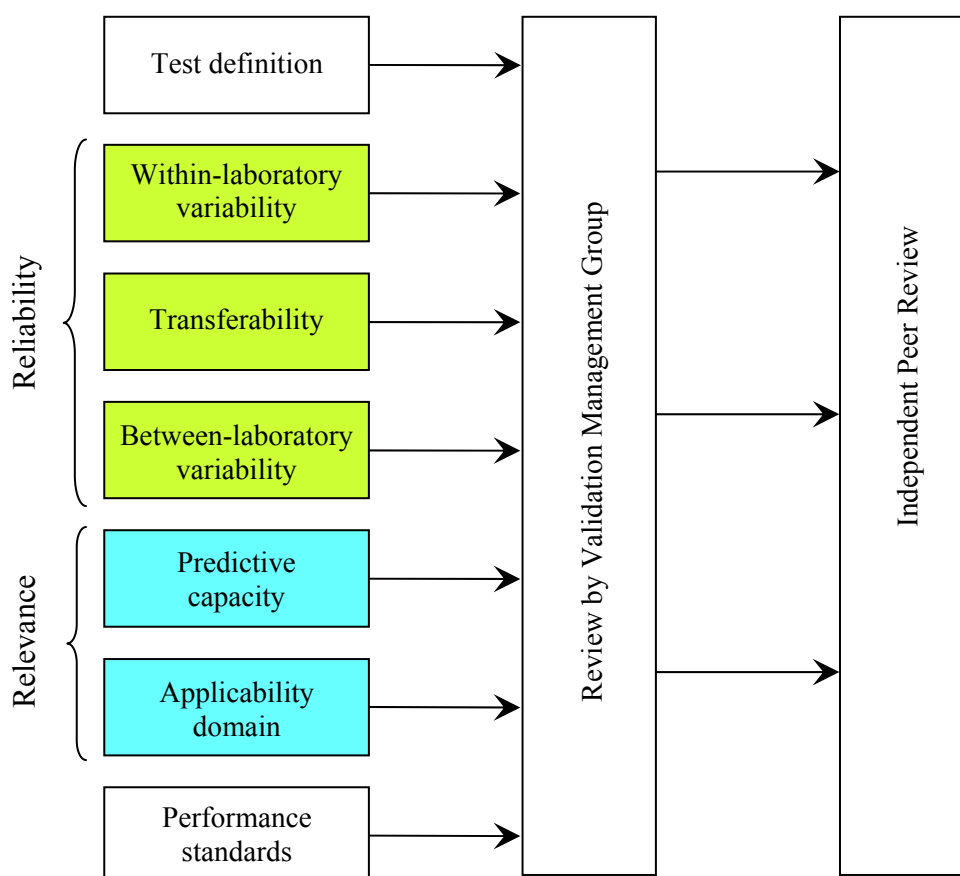


Figure 5. Modular approach for applying ECVAM principles on test validity. Modified from Hartung et al.²¹

3.2 Test definition

The first step in the development of an alternative test is the establishment of an *in vitro* model for an *in vivo* physiologic process of interest. A crucial component of an *in vitro* model is the biological end-point parameter. The end-point parameter should preferably enable quantitative assessment by allowing dose-response analysis. Second, it should ideally represent a mechanistically explained read-out of the effect of tested compounds. The test definition culminates in the description of a standard operating procedure (SOP) describing all details of the method in such a way that other laboratories can copy the methodology successfully. The SOP should be rigid in the essential characteristics of the test, and at the same time allow differences in aspects that are not essential. For example, different serum batches, different culture incubators, and different culture plastics may be allowable, facilitating the transfer of the SOP to other laboratories.

The EST, as an example, is a well-defined test of which the protocol is based on over a decade of dedicated research. An optimized *INVITTOX* protocol is available in the ECVAM database of validated methods (see <http://ecvam.jrc.it/index.htm>).

3.3 Within-lab variability

Together with the modules ‘transferability’ and ‘between-lab variability’ (see below), the module ‘within-lab variability’ represents the reliability of an alternative method (indicated in green in Figure 5). The laboratory which establishes the *in vitro* model will initially test whether the model suffices for model compounds. In addition, it is important to establish whether different operators obtain similar results, and whether results are stable over time. This is a critical prerequisite before the methodology can be transferred to other laboratories. The EST has gone through testing of within-laboratory variability in a series of laboratories, showing that the method is reasonably robust. Possible sources of variability include primarily the quality of undifferentiated stem cells, of which culturing requires specific technical expertise. In addition, culture medium composition is critical for the successful performance of the test. However, these variables can be kept within bounds that make routine use of the test possible.

3.4 Transferability

Starting from a model that has been standardised within a laboratory, the next step is to transfer the methodology to other laboratories in order to test its robustness between laboratories. The robustness is usually tested with a limited set of positively effective as well as negative/ineffective compounds. This procedure can be referred to as prevalidation. The prevalidation exercise may lead to an improved test protocol. In addition, it may also result in the establishment of a prediction model. The prediction model is necessary to translate the *in vitro* effects into a statement on the likelihood of human exposure leading to adverse effects.

The EST has been transferred between laboratories both on an ad hoc basis as well as in the controlled ECVAM validation study²² and in the integrated project ReProTect²³.

3.5 Between-lab variability

In the validation study, the between-laboratory variability of a test is usually evaluated in three or four independent laboratories with a larger set of chemicals. The test is performed in a double-blind design according to the predefined standard protocol. The number and choice of chemicals is an important issue, which depends e.g. on the anticipated applicability domain and on the *in vivo* database available for compounds to be tested. In addition, the number of replicate tests and the methodology of dose-response analysis are part of the design of the validation study.

The EST has been validated in an ECVAM coordinated program: twenty coded test chemicals, classified as non-embryotoxic, weakly embryotoxic or strongly embryotoxic, were tested in four laboratories. The results obtained in the different laboratories turned out to be highly comparable.

3.6 Predictive capacity

The modules ‘predictive capacity’ and ‘applicability domain’ stand together for the relevance of the alternative test (indicated in blue in Figure 5). The predictive capacity of an alternative test demonstrates how well the test can predict the reference standard when using the prediction model. The outcome of the validation study results in statements on the predictability of the test, including percentages of false negatives and false positives. Data obtained in the validation study are ideally analyzed by an independent laboratory and reviewed by an independent review committee, in order to avoid any bias in the analysis.

Predictions of embryotoxicity obtained in the two lead laboratories of the validation studies on the EST and the WEC are shown in Table 3. Both the EST and the WEC provided a good correlation between *in vitro* and *in vivo* data, both in the lead laboratories and the other participating laboratories. The overall predictive capacity of the EST has been calculated as 100%, 70%, and 72% for strongly, weak and non-embryotoxic chemicals, respectively, on the basis of twenty chemicals tested²². Analysis of the results obtained in the WEC test showed 100% predictability for strongly embryotoxic chemicals, 76% predictability for weakly embryotoxic chemicals, and 70% predictability for non-embryotoxic chemicals¹⁷.

Although the predictive capacity of both methods is quite high, the biological relevance and applicability of their mathematically derived prediction models is still questioned. For the EST, for example, a prediction model exists that was derived by an independent statistician. However, this prediction model has been challenged for its lack of biological meaning. The end-points feeding into the model are the inhibition of differentiation of embryonic stem (ES) cells and the inhibition of growth of both ES and 3T3 cells. Mathematically, inclusion of the data obtained with 3T3 fibroblasts resulted in a better prediction model for the set of

compounds tested than when the results for 3T3 were not incorporated. The biological reasoning for incorporating data obtained with 3T3 cells is however unclear. Furthermore, the model discriminates between non-, weak, and strong embryotoxins. The question arises how the limits between these categories should be defined. The ECVAM validation study has shown that most of the misclassifications occurred between non- and weak embryotoxins. Current personal communications with industrial laboratories which try to apply the EST indicate that whereas the EST results themselves may be considered very informative for their chemicals of interest, the prediction model often does not result in the conclusion expected on the basis of the effects observed. This stresses the notion that the current EST prediction model is less than optimal. Therefore, one of the aims of the ReProTect project is to improve the prediction model for the EST.

Table 3. Predictions of embryotoxicity from the EST and the WEC, compared to in vivo data

	EST ¹		WEC ²	<i>In vivo</i>
Hydroxyurea	3	3	3	3
6-aminonicotinamide	3	3	3	3
All-trans-retinoic acid	3	3	3	3
Methotrexate	3	3	3	3
5-bromo-2'-deoxyuridine	3	3	3	3
Methylmercury chloride	1	1	3	3
Boric acid	2	2	2	2
Salicylic acid sodium salt	2	2	1	2
Pentyl-4-yn-valproic acid	2	2	1	2
Valproic acid	2	2	2	2
Lithium chloride	2	2	2	2
Methoxyacetic acid	2	2	2	2
Dimethadione	1	2	1	2
Acrylamide	1	2	1	1
Dimethyl phthalate	2	1	1	1
Diphenhydramine hydrochloride	2	2	1	1
Saccharin sodium hydrate	1	1	1	1
Isobutyl-ethyl-valproic acid	1	1	1	1
D-(+)-Camphor	1	1	1	1
Penicillin G sodium salt	1	1	1	1

¹ Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET), Berlin

² Laboratory of Health Protection Research, RIVM, Bilthoven

1 = Non-embryotoxic; 2 = weakly embryotoxic; 3 = strongly embryotoxic

Predictions that differ from *in vivo* are indicated in bold.

3.7 Applicability domain

The particular purpose for which a test can be applied is the applicability domain. The applicability domain may contain two components: i) a specific (set of) toxicological end-point(s), and ii) a class (or classes) of chemicals for which the method is considered applicable. The first component can be defined as predictive for a certain *in vivo* effect. Most *in vitro* tests, especially those in developmental toxicology, are a very reductionistic derivative of the *in vivo* process which they are supposed to mimic. The second component can be defined as predictive for the effects of a defined class(es) of chemicals. Standard lists of chemicals for validation of alternative tests in developmental toxicity have been developed and reviewed over the past quarter of a century^{16,24}. These lists are limited by the availability of a sufficient database of *in vivo* effects with which the *in vitro* test data could be compared in the validation study. They contain a wide variety of chemical domains, which enables a general idea of the performance of the test. However, this is not very informative if one wants to define the chemical classes for which the test is predictive. In addition, the extrapolation of the results for about twenty chemicals to the world of chemicals requires prediction over more than three orders of magnitude in number of chemicals only. Therefore, validation will need more specific approaches to successfully define applicability domains. For instance, a set of chemicals from the same chemical class will need to be analyzed to be able to determine the predictability of the test for that class of chemicals. This is currently the phase of development in which the most promising alternatives in developmental toxicity testing are situated.

The EST has been validated for twenty selected compounds with very different embryotoxic potentials¹⁶. The extrapolation of these findings to the realm of chemicals is still a matter of dispute. Therefore, as yet the applicability domain of EST is not sufficiently defined. Furthermore, the EST is a murine system. To avoid the need for interspecies extrapolation, human ES cells should be employed. This may allow the prediction of developmental toxic effects in humans more precisely. The applicability of human ES cell differentiation will also be assessed in ReProTect.

3.8 Performance standards

ECVAM proposes that reference chemicals are defined which enable the demonstration of the equivalence in performance between a new test and a previously validated test. In reproductive toxicology this is anticipation rather than current practice, as there are hardly any previously validated test systems to compare with. The current state of the art is that the *in vitro* results are compared with known *in vivo* effects of tested chemicals to determine validity of the *in vitro* test. In addition, equivalence can only be meaningfully tested when both tests have the same applicability domain.

The SOP for the EST describes reference chemicals for transfer between laboratories. Comparisons with existing validated alternative tests have not been done as the latter are

nonexistent so far. However, these reference chemicals can be used for validation of the human EST variant and of various modifications of the EST, involving neuronal or skeletal differentiation rather than the standard heart muscle cell differentiation.

3.9 Guideline development

Once the applicability domain of a test system has sufficiently been defined according to the above procedure, the protocol can be forwarded to international bodies such as OECD for formal worldwide acceptance as a test guideline. This procedure usually takes several years in view of the careful review process, the input of national experts worldwide, and the need to achieve consensus due to the OECD structure.

None of the three embryotoxicity tests has entered the guideline development stage. This is mainly caused by outstanding issues around validation and validity, prediction model performance, and applicability domain.

4. Reduction of animal use: near future?

4.1 Current animal use in reproductive toxicology

As described in section 1.2, the numbers of animals involved in reproductive toxicity testing are very high. The level of data requirements for the assessment of the potential toxicological effects of chemicals increases progressively with increasing annual production volumes (see Figure 2). These production volumes are subdivided into tonnage bands up of ≥ 1 tonne per year; ≥ 10 tonnes per year; ≥ 100 tonnes per year; and $\geq 1,000$ tonnes per year. For each production volume, various animal tests are involved. The basic test for production volumes from over 1 to 1,000 tonnes is the reproductive/developmental toxicity screening test (TG 421); for production volumes of 1,000 tonnes or more, a two-generation study (TG 416) is required. Additional testing, for both fertility and developmental toxicity, is only required in case of concern on the basis of earlier testing.

In the reproductive/developmental toxicity screening test (TG 421) about 560 animals are used. The prenatal developmental toxicity test (TG 414) requires 1,400 animals, and in the two-generation study (TG 416) no less than 2,600 animals are used. Therefore, at least 560 animals are used for reproductive toxicity testing for chemicals with an annual production volume at level 1 (10 tpa). At level 1 (100 tpa), the minimal number of animals used is 1,960. Chemicals that reach level 2 (1,000 tpa) require at least 4,000 animals. If maximal reproductive toxicity testing is required, these numbers may amount to almost 6,000 animals per chemical for each production volume.

4.2 Possibilities for reduction in animal use

What is the reduction in animal use that could be achieved within several years using the current alternative methods for reproductive toxicity?

First of all, the validated alternative tests for embryotoxicity need optimisation with regard to performance of prediction model and definition of applicability domain. Once reliable tests with an appropriate applicability domain are available, they still do not present full replacements for current animal tests. However, they may be used as building blocks in testing strategies: either as part of a battery approach, or as pre-screen in a tiered approach. In a battery approach, all segments of the reproductive cycle would be represented by their own *in vitro* test. This concept could lead to the replacement of an *in vivo* test with the *in vitro* battery, or could reduce *in vivo* testing to only those compounds for which the *in vitro* battery gives equivocal results. However, chemicals with an effect detectable only long after the critical exposure period will not be picked up by such a battery. In the field of reproductive toxicology, with its variety of mechanisms, the design of a predictive battery of alternative tests is a significant challenge.

In the pre-screening situation, an alternative test may be used at a tonnage level in which reproductive toxicity in animals is not yet mandatory: a production level of up to 1 tonne per year (see Figure 6). The result of the alternative as a pre-screen may be used to either prioritise further testing in case of a positive result, and may possibly defer testing to a higher tonnage level in case of negative results. In case of the availability of a reliable test with an appropriate applicability domain defined, such an approach should be feasible. Implementation of *in vitro* methods for reproductive toxicity testing in the base set of toxicity tests would result in reduction of animal use, as illustrated by the following hypothetical situation.

A chemicals manufacturer is developing a new product and various candidate substances are available. Based on parameters other than toxicity, 90% of the chemicals drop out. For the remaining chemicals general toxicity tests are performed. Suppose that for three substances favourable results are achieved. These three substances are subsequently tested for reproductive toxicity in the reproductive/developmental toxicity screening test (TG 421). One of the three chemicals appears to have a negative effect on reproductive performance and is abandoned. In the above example, the use of optimized validated *in vitro* methods for embryotoxicity as screen would have enabled the industry to set priorities for *in vivo* testing, thereby reducing animal use.

Base set 1 tpa	General toxicity tests	<i>In vitro</i> tests repr. toxicity			Total number animals
Level 1 10 tpa	OECD 421 Screening <i>n</i> = 560	OECD 416 2-Generation <i>n</i> = 2,600	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,960
Level 1 100 tpa	OECD 421 Screening <i>n</i> = 560	OECD 416 2-Generation <i>n</i> = 2,600 + DNT	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,960
Level 2 1000 tpa		OECD 416 2-Generation <i>n</i> = 2,600 + DNT	OECD 414 Develop. Tox <i>n</i> = 1,400	OECD 414 Develop. Tox 2 nd species	5,400

Figure 6. Suggested implementation of alternative methods in reproductive toxicity testing strategy. Tests marked in yellow belong to basic requirements. Tests marked in orange are only required in case of concern. tpa =tonne(s) per annum; DNT = developmental neurotoxicity.

4.3 REACH

The European Commission (EC) adopted a new regulatory framework for the Registration, Evaluation, Authorisation and restriction of CHemicals (REACH) in 2003²⁵. The REACH initiative aims, amongst others, at a more complete toxicity database for chemicals already in use, based on toxicological information provided by industrial manufacturers and users of chemicals. Enterprises that manufacture or import more than one tonne of a chemical substance per year will be required to register it in the central database. As such, they need to perform safety testing.

It has been estimated that toxicological information will have to be submitted for about 30,000 existing chemicals. The number of test animals that will be required as a consequence of REACH is immense. Estimates vary from about 4 million animals (EC's European Chemicals Bureau²⁶) up to 20 million animals (industry, personal communications). No less than 70% of these animals will be needed for collecting toxicological information in the specific area of developmental toxicology, as well as for reproductive toxicology in general. Of the 30,000 existing chemicals that need to be evaluated in REACH, estimates by ECVAM show that about 7,500 chemicals have production level 1 (10 tpa and 100 tpa together) and about 2,700 chemicals have production level 2. This has major consequences for animal usage for reproductive toxicity testing. Assuming that 10 percent of the chemicals at level 1 raise concern, this means that approximately 3,500 chemicals need to be tested in a two-generation study. With 2,600 animals required per test, this equals around 9 million test animals. To put these numbers into perspective: a total of no more than 100 two-generations studies have been performed over the past 25 years. This calculation also shows that the estimation of van der Jagt *et al.*²⁶ that around 4 million experimental animals will be needed for REACH in total may be an underestimation. On the other hand, the unofficial calculations from industry, with numbers as high as 20 million animals, may be overestimations. In any case, these estimates emphasize once more the urgency of the implementation of alternative methods in this area in order to reduce animal use.

5. General conclusions and discussion

The present report gives an overview of the currently existing *in vitro* models for reproductive toxicology. Testing of toxic effects on the reproductive cycle is complex, since reproduction comprises many different processes and sensitive periods. Due to its complexity, it is not possible to model the whole cycle in one *in vitro* system. Parts of the system need to be studied individually and then integrated into testing strategies. Many alternatives for animal testing already have been developed, but currently there are no alternative tests available for reproductive toxicity testing, which offer enough reliability to be integrated into testing strategies.

A limited number of the alternative tests have been validated or are in the (pre)validation process. Despite successful validation, not a single *in vitro* method has been implemented yet in international test guidelines. This is mainly caused by outstanding issues with regard to the relevance of the various alternative methods. Optimisation is needed for both prediction model and applicability domain. These aspects of the validation process should receive more attention, since they seem to be the bottleneck for implementation. Subsequently, testing strategies need to be developed and the potential role of the various *in vitro* tests needs to be discussed. It should be pointed out that validation of single tests does not imply validation of the corresponding testing strategy. Therefore, extra time will be necessary to analyse the relevance and reliability of testing strategies.

The high numbers of animals involved in reproductive toxicity testing clearly indicate that considerable effort is required to further develop the various existing *in vitro* methods into reliable alternative tests that cover critical aspects of the reproductive cycle. Existing tests will also have to be expanded to include further end-points and to introduce metabolising capacities. In addition, new tests will have to be developed since the currently available tests are not sufficient. However, they should be developed with great care to make sure that reliable *in vitro* methods are obtained, which can be implemented in testing strategies. Furthermore, new methodologies, such as gene expression profiling in *in vitro* methods, need to be investigated.

The current extensive attention for alternative methods as a consequence of the introduction of REACH has given new momentum to this search for alternatives. The potential for alternatives for reducing animal use is theoretically the highest in reproductive toxicology as compared to other areas of toxicology. This is a direct consequence of the relatively very high animal use in this area, i.e. 70% of all animal use in REACH. If we suppose that the prenatal developmental toxicity test (TG 414) could be waived on the basis of informative *in vitro* testing, this could reduce reproductive toxicity testing for that compound with around 50%, and would reduce the animal use for all toxicity testing of that compound with 35%. Assuming that the prenatal developmental toxicity test could be waived on the basis of *in vitro* testing for even only 11% of chemicals in REACH, this would already result in a reduction of animal use amounting to one million experimental animals. Such a straightforward calculation shows the potential impact of alternatives in reproductive toxicology and should stimulate and increase current and future efforts towards the development and implementation of alternatives.

References

1. Yu X, Sidhu JS, Hong S, Faustman EM. Essential role of extracellular matrix (ECM) overlay in establishing the functional integrity of primary neonatal rat Sertoli cell/gonocyte co-cultures: an improved in vitro model for assessment of male reproductive toxicity. *Toxicol Sci* 2005; 84(2):378-393.
2. Piersma AH. Alternative methods for developmental toxicity testing. *Basic Clin Pharmacol Toxicol* 2006; 98(5):427-431.
3. Mummery CL, van den Brink CE, van der Saag PT, de Laat SW. A short-term screening test for teratogens using differentiating neuroblastoma cells in vitro. *Teratology* 1984; 29(2):271-279.
4. Faustman EM. Short-term tests for teratogens. *Mutat Res* 1988; 205(1-4):355-384.
5. Johnson EM. A subvertebrate system for rapid determination of potential teratogenic hazards. *J Environ Pathol Toxicol* 1980; 4(5-6):153-156.
6. Bantle JA, Burton DT, Dawson DA, Dumont JN, Finch RA, Fort DJ et al. Initial interlaboratory validation study of FETAX: phase I testing. *J Appl Toxicol* 1994; 14(3):213-223.
7. Jelinek R. Use of chick embryo in screening for embryotoxicity. *Teratog Carcinog Mutagen* 1982; 2(3-4):255-261.
8. New DA, Coppola PT, Cockroft DL. Improved development of head-fold rat embryos in culture resulting from low oxygen and modifications of the culture serum. *J Reprod Fertil* 1976; 48(1):219-222.
9. Bremer S, Cortvrindt R, Daston G, Eletti B, Mantovani A, Maranghi F et al. Reproductive and developmental toxicity. *Altern Lab Anim* 2005; 33 Suppl 1:183-209.
10. Cooke BA. In vitro models for the investigation of reproductive toxicology in the testis. *Adv Exp Med Biol* 1998; 444:95-102.
11. Cortvrindt RG, Smitz JE. Follicle culture in reproductive toxicology: a tool for in-vitro testing of ovarian function? *Hum Reprod Update* 2002; 8(3):243-254.
12. Sun F, Betzendahl I, Shen Y, Cortvrindt R, Smitz J, Eichenlaub-Ritter U. Preantral follicle culture as a novel in vitro assay in reproductive toxicology testing in mammalian oocytes. *Mutagenesis* 2004; 19(1):13-25.
13. Brueggemeier RW, Katlic NE. Aromatase inhibition by an enzyme-activated irreversible inhibitor in human carcinoma cell cultures. *Cancer Res* 1990; 50(12):3652-3656.
14. Letcher RJ, van H, I, Drenth HJ, Norstrom RJ, Bergman A, Safe S et al. Cytotoxicity and aromatase (CYP19) activity modulation by organochlorines in human placental JEG-3 and JAR choriocarcinoma cells. *Toxicol Appl Pharmacol* 1999; 160(1):10-20.
15. Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A et al. The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. *European Centre for the Validation of Alternative Methods. Altern Lab Anim* 2002; 30(2):151-176.
16. Brown NA. Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. *European Centre for the Validation of Alternative Methods. Altern Lab Anim* 2002; 30(2):177-198.
17. Piersma AH, Genschow E, Verhoef A, Spanjersberg MQ, Brown NA, Brady M et al. Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 2004; 32(3):275-307.
18. Balls M, Hellsten E. Statement of the scientific validity of the embryonic stem cell test (EST) -- an in Vitro test for embryotoxicity. *Altern Lab Anim* 2002; 30(3):265-268.
19. Balls M, Hellsten E. Statement on the scientific validity of the postimplantation rat whole-embryo culture assay -- an in vitro test for embryotoxicity. *Altern Lab Anim* 2002; 30(3):271-273.
20. Balls M, Hellsten E. Statement on the scientific validity of the micromass test -- an in Vitro for embryotoxicity. *Altern Lab Anim* 2002; 30(3):268-270.

21. Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S et al. A modular approach to the ECVAM principles on test validity. *Altern Lab Anim* 2004; 32(5):467-472.
22. Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N et al. Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 2004; 32(3):209-244.
23. Hareng L, Pellizzer C, Bremer S, Schwarz M, Hartung T. The integrated project ReProTect: a novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* 2005; 20(3):441-452.
24. Hardin BD, Schuler RL, Burg JR, Booth GM, Hazelden KP, MacKenzie KM et al. Evaluation of 60 chemicals in a preliminary developmental toxicity test. *Teratog Carcinog Mutagen* 1987; 7(1):29-48.
25. Proposal Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). 2003;
26. 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.

Appendix 1: Overview of in vitro alternatives for reproductive and developmental toxicity

Current endpoints addressed in animal test	Alternative tests available	Endpoints measured	Purpose	Area(s) of application	Validation status	Regulatory acceptance	Comments	Estimated time to have the method validated (ESAC endorsement) ^a
Developmental toxicity (OECD TGs 414, 421)	EST	Inhibition of differentiation of ES cells into cardiomyocytes (ID50) (IC50) 3T3 cells (IC50) ID3 cells	Partial replacement: tiered strategy and/or test battery	Testing structures analogous to developmental compounds; prioritisation	Validated (ECVAM ZEBET)		The test can only be used in a test strategy that covers main signs of development toxicity; it can only be used for a narrow range of chemicals, and its performance with other chemical classes has to be determined	0 years
Developmental toxicity (OECD TGs 414, 421)	MM test	Inhibition of differentiation	Partial replacement: tiered strategy and/or test battery	Testing structures analogous to developmental compounds; prioritisation	Validated (ECVAM ZEBET)		The test can only be used in a test strategy that covers main signs of development toxicity	0 years
Developmental toxicity (OECD TGs 414, 421)	WEC test	Various	Partial replacement: tiered strategy and/or test battery	Testing structures analogous to developmental compounds; prioritisation	Validated (ECVAM ZEBET)		The test can only be used in a test strategy that covers main signs of development toxicity	0 years
Developmental toxicity (OECD TGs 414, 421)	Chick embryo retina cell cultures	Various	Partial replacement: tiered strategy and/or test battery	Testing structures analogous to developmental compounds; prioritisation	R&D		The test can only be used in a test strategy that covers main signs of development toxicity	6 years
Developmental toxicity (OECD TGs 414, 421)	Fetal gonad cultures	Various	Partial replacement: tiered strategy and/or test battery	Testing structures analogous to developmental compounds; prioritisation	R&D		The test can only be used in a test strategy that covers main signs of development toxicity	> 10 years

^aThis table estimates the time needed to achieve ESAC endorsement for individual alternative tests, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. 'Optimal conditions' means that all necessary resources, e.g. technical, financial, human and coordination are met at all times in the process, and that the studies undertaken have successful outcomes.

Current endpoints addressed in animal test	Alternative tests available	Endpoints measured	Purpose	Area(s) of application	Validation status	Regulatory acceptance	Comments	Estimated time to have the method validated (ESAC endorsement) ^a
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	Sliced testis assay	Testosterone production; Leydig cell count	Partial replacement: tiered strategy and/or test battery	Interference with testosterone production	Prevalidation (US EPA EDMVS)		The test can only be used in a test strategy that covers main features of the reproductive cycle	6 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	H295R cells	Steroid production	Partial replacement: tiered strategy and/or test battery	Interference with steroidogenesis	Prevalidation (US EPA EDMVS)		The test can only be used in a test strategy that covers main features of the reproductive cycle	6 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	R2C cells	Steroid production	Partial replacement: tiered strategy and/or test battery	Interference with steroidogenesis	R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	H 540 cells	Steroid production	Partial replacement: tiered strategy and/or test battery	Interference with steroidogenesis	R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	MLTC-1 cells	Steroid production	Partial replacement: tiered strategy and/or test battery	Interference with steroidogenesis	R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	TM 3 cells	Various	Partial replacement: tiered strategy and/or test battery	Interference with steroidogenesis	R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years

^aThis table estimates the time needed to achieve ESAC endorsement for individual alternative tests, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. 'Optimal conditions' means that all necessary resources, e.g. technical, financial, human and coordination are met at all times in the process, and that the studies undertaken have successful outcomes.

Current endpoints addressed in animal test	Alternative tests available	Endpoints measured	Purpose	Area(s) of application	Validation status	Regulatory acceptance	Comments	Estimated time to have the method validated (ESAC endorsement) ^a
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	Seminiferous tubule culture assay	Various	Partial replacement: tiered strategy and/or test battery	Testicular toxicity	R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	Sperm activation assay	Various	Partial replacement: tiered strategy and/or test battery	Effects on sperm function	Under R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	<i>In vitro</i> fertilisation oocyte toxicity test	Various	Partial replacement: tiered strategy and/or test battery	Oocyte and preimplantation toxicity	Under R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	Oogonia proliferation test	Various	Partial replacement: tiered strategy and/or test battery	Interference with oocyte meiotic process	Under R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years
One and two generation reproduction toxicity study (OECD TGs 415, 416, 421)	Sertoli cell lines test	Various	Partial replacement: tiered strategy and/or test battery	Effects on the blood-testis barrier; information on whether chemicals are able to cross the blood-testis barrier and can effect the sensitive germ cells	Under R&D		The test can only be used in a test strategy that covers main features of the reproductive cycle	> 10 years

^aThis table estimates the time needed to achieve ESAC endorsement for individual alternative tests, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. 'Optimal conditions' means that all necessary resources, e.g. technical, financial, human and coordination are met at all times in the process, and that the studies undertaken have successful outcomes.

Appendix 2: Abbreviations

DNT	Developmental neurotoxicity
EC	European Commission
ECVAM	European centre for validation of alternative methods
ES	Embryonic stem cell
EST	Embryonic stem cell test
EU	European Union
MM	Limb bud micromass test
NOAEL	No observed adverse effect level
OECD	Organisation for economic co-operation and development
REACH	Registration, evaluation and authorisation of chemicals
tpa	tonne per annum
WEC	Whole embryo culture