Institutet för miljömedicin - IMM

IMM-rapport 1/2013

Risk assessment of developmental neurotoxicity: Evaluation of the OECD TG 426 test guideline and guidance documents

Anna Beronius
Annika Hanberg
Rachel Heimeier
Helen Håkansson

Karolinska Institutet

Risk assessment of developmental neurotoxicity: Evaluation of the OECD TG 426 test guideline and guidance documents

Anna Beronius
Annika Hanberg
Rachel Heimeier
Helen Håkansson

Preface

On behalf of the Swedish Chemicals and Environmental Protection Agencies, the Institute of Environmental Medicine (IMM) at Karolinska Institutet (KI) was asked, in June 2009, to evaluate the OECD TG 426 for developmental neurotoxicity (DNT) testing, as well as the corresponding guidance documents. The evaluation aim was to identify areas of possible improvements of TG 426 and the corresponding guidance documents in terms of their reliability and usability for scientific and regulatory judgements and decisions, which are key processes in chemicals safety and health risk assessment. Specific issues in the evaluation concerned the considerable flexibility of TG 426 in terms of the DNT study design, such as the choice of behavioural tests methods included in the study, the design of those individual tests, the potential variability of test results that may accompany the flexibility of the guideline, as well as the lack of detailed interpretation support and advice in the corresponding guidance documents.

Stockholm 2012-12-05

Anna Beronius*

Annika Hanberg*

Rachel Heimeier

Helen Håkansson

*These authors have contributed equally to this report.

Acknowledgements

The authors would like to thank colleagues at the Swedish Chemicals Agency and the Swedish Environmental Protection Agency (EPA) for fruitful discussions throughout the project. Grants from the European Commission (contract: FOOD-CT-2005-022923), Formas (216-2007-2056), and Ragnar Söderbergs stiftelse (Dnr ERN90/09) have contributed to the financial contract by the Swedish EPA (contract: 802-0172-09).

Table of Contents

Summary

The Institute of Environmental Medicine (IMM) at Karolinska Institutet (KI), has evaluated the OECD TG 426 for developmental neurotoxicity (DNT) testing, as well as the corresponding guidance documents. The evaluation was based on DNT-studies of bisphenol A (BPA), decabromodiphenyl ether (PBDE 209), deltamethrin, 2,2',4,4',5,5'-hexachlorobiphenyl (PCB 153), and perfluorooctane sulfonate (PFOS).

The investigation included detailed comparisons of the experimental design and results, as well as accordance of the individual compound studies to the criteria specified by OECD TG 426. In addition, external reviewers with expertise within the fields of neurotoxicity research/testing and/or safety and health risk assessments were asked to answer specific questions raised during the compound evaluation and to provide comments on a draft of the compound evaluation report. The compound evaluation work and the external reviewer survey both paid special attention to issues such as flexibility of the OECD TG 426 and the potential variability of test results that may accompany this flexibility, the need of research and procedural development, as well as the importance of expert judgement in DNT testing and assessment.

The results of the compound evaluation work and the external reviewer survey were presented and discussed. Briefly, it is clear from the compound evaluation work and the reviewer survey that the areas of neurodevelopment and neurotoxicity are inherently very complex, and, in particular, there are massive gaps in knowledge about normal brain development on the functional, structural and molecular levels, which complicates both DNT testing as well as compound safety and health risk assessment. Also, it makes defining strict criteria for testing, data interpretation, and risk assessment difficult. The external reviewers were in general agreement that TG 426 need to remain flexible in order to enable investigators to design the most sensitive and appropriate study relevant for the exposure and toxicity of the compound under investigation. Decreasing the flexibility of TG 426 in order to make testing more standardized or to facilitate the evaluation of study results is not considered the right way forward. Nevertheless, there were several areas where the reviewer opinions differed markedly e.g. whether the requirements of additional endpoints, such as "anxiety" or "sexual behaviour" should be added to the TG 426 or not, as well as the importance of including historical controls in the study design.

Conclusions and suggestions for how to move forward to ensure that reliable and relevant DNT testing and assessments were made. Briefly, further efforts to improve guidance documents provided for testing laboratories and authorities to ensure reliable and relevant testing and assessment of DNT seems to be an important way forward to ensure that health protection aims of chemicals legislation are fulfilled. Research to decrease knowledge gaps, training to increase interpretation and judgement abilities, as well as procedural efforts to allow for continuous revision of regulatory guidelines and guidance documents as new knowledge becomes available and adopted, are all important contributors to such efforts.

1. Introduction

The Institute of Environmental Medicine (IMM) at Karolinska Institutet (KI) was asked, in June 2009, to evaluate the OECD TG 426 for developmental neurotoxicity (DNT) testing, as well as the corresponding guidance documents. The evaluation aim was to identify areas of possible improvements of TG 426 and the corresponding guidance documents in terms of their reliability and usability for scientific and regulatory judgements and decisions, which are key processes in chemicals safety and health risk assessment.

The scientific background to the evaluation of the OECD TG 426 and the corresponding guidance documents, stems from the growing public and societal concern, during recent years, that exposure to chemicals during gestation and periods of fast development of the brain's morphology and physiology can be particularly harmful and may result in neurotoxic effects that are lasting into childhood, puberty, adulthood, and even into old age. Support for causal relationships is particularly strong for DNT in humans and the exposure to lead, methyl mercury and polychlorinated biphenyls (PCBs) (Grandjean and Landrigan 2006). For these chemicals, there are observational data, which link background exposure levels to DNT in human studies. Furthermore, observational and/or experimental studies of these compounds also demonstrate that DNT occur at doses that are lower than those which affect adult brain function.

Grandjean and Landrigan (2006) also discuss the neurotoxicity of arsenic and toluene, where there is some evidence of DNT-potential in humans, while for a large number of listed neurotoxic chemicals, including 90 pesticides, little is known about their potential to cause DNT in humans. An evaluation of available regulatory DNT studies for close to 70 pesticides was recently conducted by the Office of Pesticide Programs (OPP) at the US EPA (Raffaele *et al.,* 2010). The evaluation concluded that almost half of these regulatory DNT studies (40%) were either used directly as a critical study in an OPP risk assessment, or had the potential of being used as such a critical study for future risk assessments. Furthermore, the evaluation stressed the importance of evaluating a spectrum of endpoints in order to improve the potential for DNT-detection

The test paradigm for regulatory testing for DNT has been the subject of much scientific debate and has been reviewed by many expert groups (Makris *et al.,* 2009, Raffaele *et al.,* 2010). On-going discussions concern matters such as whether or not the DNT test paradigm is sensitive enough to serve as a reliable basis for the risk assessment of DNT in humans.

1.1 History of OECD TG 426

The OECD Test Guideline (TG) 426 (OECD, 2007) for developmental neurotoxicity (DNT) testing was developed mainly based on the already existing US EPA guideline for DNT testing (US EPA, 1998a). The US EPA DNT guideline was first issued in 1991 and was founded on scientific literature within the field of DNT, which first started to appear in the 1960's (Makris *et al*., 2009). It has since been extensively revised on a number of occasions (Fitzpatrick *et al*., 2008; Makris *et al*., 2009) and was for a long time the only DNT guideline available to testing laboratories. Work to develop a DNT guideline to further accommodate the regulatory needs of OECD

countries was initiated in 1995 by an OECD Working Group on Reproduction and Developmental Toxicity (OECD, 2007). The work entailed using the US EPA DNT guideline as a prototype, further improving it by identifying and addressing a number of important issues regarding *e.g.* testing period, dosing regimen and endpoints to be included, as well as criteria for evaluating results. A number of expert consultant meetings and workshops were held over the years to finalize the OECD guideline. OECD TG 426 was then finally adopted in 2007 (OECD, 2007).

1.2 DNT study performance evaluations

Both the US EPA and the OECD guidelines for DNT testing are structured to include investigations of developmental landmarks and behavioural ontogeny, motor activity, motor and sensory function, learning and memory, and neuropathology. For some of these categories several different validated test methods are available and the guidelines are largely flexible regarding which test method to include in the study design.

A number of workshops and expert meetings involving experts from academia, industry, regulatory bodies and public interest groups have taken place over the years to review and evaluate the test methods recommended in the DNT guideline. These efforts include test method development as well as characterizing the sensitivity, reliability and performance of test methods. Makris *et al*. (2009) conducted a review of the outcomes of several such meetings and collaborations between 1960 and 2003 in support of the finalization and implementation of OECD TG 426. The conclusions from this review were that the different validated methods recommended in the guideline are based on sound and solid science conducted in the area of DNT and that these methods have been shown to be sensitive, reliable and relevant in regard to identifying potential adverse effects. In the review it is recognized that the DNT guideline have been criticized both for not including enough endpoints, such as social behaviour or pharmacokinetics, and for being too complex, *i.e.* including too many endpoints, and that continued research is needed within this field. Overall, the review concludes that the OECD TG 426 “represents the best available science for assessing the potential for DNT in human health risk assessment, and data generated by DNT studies are relevant and reliable for this assessment” (Makris *et al*., 2009).

The performance of the US EPA DNT guideline has also been reviewed and evaluated by an International Life Sciences Institute (ILSI) expert working group on neurodevelopmental endpoints consisting of scientists from governmental, academic, industrial and public interest sectors and established in 2004. Several reports are available from this initiative mainly aimed at providing guidance for proper execution, analysis and interpretation of DNT studies. These reports include a practical guide for the selection and use of positive control agents (Crofton *et al*., 2008), a review and recommendations for statistical approaches in the analysis of DNT studies (Holson *et al*., 2008), an operational framework for evaluating variability in DNT study data (Raffaele *et al*., 2008), and guidance on the interpretation of DNT study data (Tyl *et al*., 2008).

1.3 Study design

To address the overall aim, i.e. to identify areas of possible improvements of TG 426 and the corresponding guidance documents in terms of their reliability and usability for scientific and regulatory judgements and decisions, a study was designed by IMM, which analysed in detail, how well available DNT studies for the five selected compounds followed criteria specified in TG 426 in their performance and reporting. Detailed comparisons of the experimental design and results, as well as accordance to the criteria specified in TG 426 were included and tabulated.

The five compounds under study were bisphenol A (BPA), decabromodiphenyl ether (PBDE 209), deltamethrin, 2,2',4,4',5,5'-hexachlorobiphenyl (PCB 153), and perfluorooctane sulfonate (PFOS). These compounds were selected based on societal concern, data availability, and scientific grounds.

In addition to the compound evaluations, this study included an external reviewers survey. Experts within the fields of neurotoxicity research/testing and/or compound safety and health risk assessments were asked to provide written answers to specific questions related to the compound evaluation work, and to provide comments on a draft of the compound evaluation report. The experts and their affiliations are listed in the Appendix, while the survey is presented as "questions for reviewers" and "reviewer comments", respectively, in the discussion section of this report. The compound evaluations and the external reviewer survey both paid special attention to issues such as the considerable flexibility of TG 426 in terms of the study design, such as the choice of behavioural tests methods included in the study, the design of those individual tests, the potential variability of test results that may accompany the flexibility of the guideline, as well as the lack of detailed interpretation support and advice in the corresponding guidance documents. The need of research and procedural development, as well as the importance of expert judgement in DNT testing and assessment were further topics addressed in the external reviewer survey.

It should be noted that it was not the purpose of this report to evaluate the individual studies as to their reliability and relevance for risk assessment. All conclusions regarding the results of the individual studies are those of the study authors. Furthermore, although it has been recognized that developmental neurotoxicity constitutes an area in critical need of testing methods alternative to current in vivo test paradigms e.g. for the screening of large numbers of chemicals (Crofton et al., 2011), this need is not a specific topic of the present investigation. The development of alternative DNT testing methods, as well as their use in chemical regulation, is subject of other on-going discussions within the regulatory and research communities, for example at the recent Third International Conference on Alternatives for Developmental Neurotoxicity (DNT) Testing held in Varese, Italy in May 2011.

2. Guideline and guidance documents for neurotoxicity testing

2.1. OECD Test Guideline 426: Developmental Neurotoxicity Study (2007)

The developmental neurotoxicity study provides information on the potential functional and morphological effects of chemicals on the developing nervous system in offspring exposed in utero and during early postnatal development. The criteria and recommendations stated in TG 426 have been summarized in Table 1.

A developmental neurotoxicity study can be conducted as a separate study, incorporated into a reproductive toxicity and/or adult neurotoxicity study (*e.g.* TG 415, 416, 424) or it may be added to a prenatal developmental study (*e.g.* TG 414). When it is incorporated within or added as part of another study, it is critical to preserve integrity of both study types.

The preferred rodent species is the rat (of common strain), although other rodent species may be used if justified. The test substance should be administered daily to mated females. The route of exposure that is most relevant for human exposure, generally orally, should be used. Administration of the test substance should be conducted from the time of implantation at gestational day (GD) 6 throughout lactation to post natal day (PND) 21. If another species than rat is used the timing of exposure has to be adjusted to cover comparable days of development. At least three dose groups and a concurrent control should be used. Enough pregnant dams should be treated so as to ensure that enough offspring are produced for neurotoxicity evaluation; 20 litters are recommended for each dose level.

Dams should be tested to evaluate any effects during pregnancy and lactation; this information can provide comparative information. On or after PND 4 one pup per sex and litter, should be randomly selected and assigned for each test in the neurotoxicity evaluations. Clinical and functional observations are required for both dams and all offspring, where clinical observations should be performed at least once daily in regards to animal health, with morbidity and mortality to be recorded. In offspring additional required measurements consist of brain weight and neuropathology, which should be conducted on or before PND 22 and at study termination (adulthood). Also, functional and behavioural endpoints should be evaluated in offspring; behavioural ontogeny at least twice during the pre-weaning period; measurements of motor activity at least once pre-weaning and once in young adulthood (recommended PND 60-70); motor and sensory function as well as learning and memory at adolescence (recommended PND 25±2) and in young adulthood (recommended PND 60-70).

Reporting from developmental neurotoxicity studies should include the body weight, food/water consumption; detailed clinical observations, necropsy findings, a detailed description of all behavioural observations, the number of animals at the start and end of the study, absorption and metabolism data (if available) and the toxic response is to be recorded by sex and dose level. The statistical unit of measure should be the litter (or dam) and not the pup.

2.2 Guidance documents

2.2.1 Guidance document for Neurotoxicity Testing: OECD Series on Testing and Assessment number 20 (2004)

This OECD guidance document is stated to be an essential supplement to OECD TG 426 and other test guidelines used to evaluate neurotoxicity. It is intended to "provide guidance on strategies and methods for testing of chemicals for potential neurotoxicity" with the purpose to "ensure that necessary and sufficient data are obtained to enable adequate evaluation of the risk of neurotoxicity arising from exposure to a chemical". The document gives an overview of neurotoxicity assessment and guidance primarily to aid in the selection of methods and design of tests. It includes information on available test methods for the different endpoints, brief information (with references) on how these tests are conducted and what they measure, and in some cases factors that may influence the outcome of the test. Interestingly, testing of motor activity is not mentioned at all. General strengths and weaknesses of behavioural tests are mentioned but very little actual guidance is given on how to interpret test results. It is recommended that at least one scientist who has practical experience of behavioural testing is consulted in both designing tests and interpreting the results.

2.2.2 Guidance document on Mammalian Reproductive Toxicity Testing and Assessment: OECD Series on Testing and Assessment number 43 (2008)

Although not specifically intended as guidance on the testing and assessment of developmental *neurotoxicity*, this document is stated to be an essential supplement to OECD TG 426 as well as other test guidelines that are used to provide data on reproductive toxicity. It is "intended to provide guidance on methodological aspects, interpretation of data and an overall approach for testing of chemicals for potential human and other mammalian reproductive toxicity".

The guidance explains general methodological issues in testing for reproductive toxicity, which are relevant to TG 426, *e.g.* standardization of litter size, route of exposure and time points for behavioural testing. It also specifically describes methodological considerations for conducting the neurobehavioural measures required in OECD TG 426. In addition it provides general guidance on data interpretation, *i.e.* issues to consider when interpreting data, such as the relationship between behavioural measures, the timing of testing in relation to treatment and the plasticity of the nervous system, as well as the validity and sensitivity of neurobehavioural tests. However, there is no guidance for the interpretation of data from specific tests.

This document provides more extensive and detailed guidance for developmental neurotoxicity testing according to OECD TG 426 than guidance document no. 20.

2.2.3 US EPA Guidelines for Neurotoxicity Risk Assessment (1998b)

This is a guidance document used by US EPA in the evaluation of potential neurotoxicity of environmental toxicants. The purpose is to "help develop a sound scientific basis for neurotoxicity risk assessment, promote consistency in the Agency's assessment of toxic effects on the nervous system, and inform others of the approaches used by the Agency in those assessments". This document is intended to provide "comprehensive guidance on the use and interpretation of neurotoxicity data". It also offers guidance on the general procedure and methodology of risk assessment, with a focus on the assessment of potential neurotoxicity, *e.g.* dose-response considerations, adversity of effects and weight of evidence. It is recognised that neurotoxicity hazard characterization must be based on expert judgement to a large extent, both when it comes to evaluating the validity of available data and in judging the adversity of observed effects.

In regard to neurobehavioural testing in animals the guidance aim to provide "an overview of the major types of endpoints that may be evaluated in animal neurotoxicity studies, [describe] the kinds of effects that may be observed and some of the tests used to detect and quantify these effects, and [provide] guidance for interpreting data". Five categories of endpoints are described, namely structural or neuropathological, neurophysiological, neurochemical, behavioural, and developmental. Based on these aims table 2 was constructed in the present report to provide a summary of the guidance actually provided, focusing on guidance for neurotoxicity *testing* and interpretation of results from such tests. The categories of neurophysiological and neurochemical endpoints have been excluded as they are not required in OECD TG 426.

As can be inferred from both the amount and type of information in table 2, examples of methods or testing strategies and guidance for interpreting tests are not provided in any consistent manner for the different endpoint categories. There is no guidance provided for the interpretation of data from specific neurobehavioural tests.

The guidance document notes that there are particular issues of importance in the evaluation of developmental neurotoxicity studies, for example that adverse effects may occur at lower doses after developmental exposure than after exposure in adults and that there are critical windows of functional development. It also mentions that "a pharmacological/physiological challenge may also be valuable in evaluating neurological function and 'unmasking' effects not otherwise detectable". This idea of "silent toxicity", effects which may elude the DNT study paradigm, is not addressed in any of the other literature reviewed here. The US EPA document also states that "in the case of developmental neurotoxicity both monotonic and biphasic dose-response curves are likely". The possibility of non-monotonic (*e.g.* U-shaped or biphasic) dose-response curves in DNT study data are discussed elsewhere (*e.g.* Tyl *et al*., 2008), however, rather from the perspective that extra careful consideration has to be given when interpreting and evaluating such data, and not giving the impression that such dose-response curves are "likely".

The guidance document includes a very brief section on special considerations concerning kinetics of neurotoxic compounds. Mainly it highlights the importance of considering the blood-brain barrier as well as structures in the brain that are served

by blood vessels that lack blood-brain barrier properties. Special considerations regarding the pharmacokinetics of developmental exposure are not discussed.

2.3 Functional Observational Battery

Functional observational batteries (FOB) may be included in general toxicity tests to detect any signs of neurotoxicity. The term "FOB" seems to be generally used by the US EPA while OECD guidelines (*e.g.* in OECD TG 407 and 408) refer to rather less defined "functional observations".

Descriptions of what to include in functional observations in OECD guidelines are not very specific. In general, these guidelines state that sensory reactivity to stimuli of different types, *e.g.* auditory, visual and proprioceptive stimuli, as well as assessment of grip strength and motor activity should be conducted (see Table 1). References are provided for further details of the procedures that could be followed but alternative procedures than those referenced could also be used. Further, functional observations may be omitted if such information is already available from other studies and the daily clinical observations do not reveal any functional deficits or if the animals show toxicity to an extent that would significantly interfere with the functional test performance.

The FOB, as used by the US EPA, typically covers several functional domains, including neuromuscular (*e.g.* weakness, incoordination, gait, and tremor), sensory (*e.g.* audition, vision, and somatosensory), and autonomic (*e.g.* pupil response and salivation) function (US EPA, 1998b).

3. Comparisons of studies according to TG 426 (guideline studies) and other studies on developmental neurotoxicity (non-guideline studies)

For the comparisons of TG 426 and other studies of developmental neurotoxicity five compounds of concern have been selected. These are bisphenol A (BPA) decabromodiphenyl ether (PBDE 209), deltamethrin, 2,2',4,4',5,5'-hexabiphenyl (PCB 153) and perfluorooctane sulfonate (PFOS). For all model compounds but PCB 153 there are one study available performed according to TG 426 and one or more other studies of developmental neurotoxicity. For PCB 153 the comparison is based on two "non-guideline" studies.

In this section studies investigating the developmental neurotoxicity of the five model compounds have been compared to the criteria and recommendations stated in TG 426. Systematic comparisons have been conducted by entering detailed information from the studies into tables. These tables are provided at the end of this report. In addition to criteria stated in TG 426 some further information judged to be of interest in this context has been included for comparison, i.e:

- Any additional behavioural tests not required in the guideline, *e.g.* sexual behaviour, anxiety, etc
- If estrous cyclicity was controlled for in female offspring subjected to behavioural testing

- What effects have been reported by the author and conclusions regarding evidence for developmental neurotoxicity
- Conclusions regarding at what dose effects occur (or NOAEL/LOAEL if stated)

Below follow summaries of comparisons from the tables for each model compound.

3.1 Bisphenol A

3.1.1 Selection of studies

Thirty-seven studies that have investigated some aspect of developmental neurotoxicity of BPA were identified from the published literature. In 31 of these studies (all non-guideline studies) the authors report some effect in DNT endpoints, while in 6 of the studies the authors concluded that the study is negative for DNT. Out of the negative studies one was conducted according to OECD TG 426. The vast data material on BPA necessitated a selection of studies to be included in the evaluation. Consequently, four studies were selected; the study conducted according to TG 426 (Stump *et al*., 2010) as well as the studies by Ema *et al*. (2001), Jones *et al*. (2011) and Negishi *et al*. (2003). The three studies not conducted according to TG 426 were selected based on the following criteria:

- The study should be conducted in rats
- Oral exposure
- There should be more than one dose tested
- Exposure should include the prenatal and lactational period
- The study should not *only* assess behavioural response after a pharmaceutical or physiological challenge, such as additional treatments with amphetamines or nicotine

An overview of the aims, experimental designs and conclusions regarding developmental neurotoxicity from the four studies is provided in Table 3.

3.1.2 Comparison of experimental design and methods and agreement with OECD TG 426

The criteria for a developmental neurotoxicity study stated in OECD TG 426, as well as the experimental design and methods used to evaluate developmental neurotoxicity in the four selected BPA studies, have been summarized in Table 4.It should be noted that the study by Ema *et al*. (2001) was conducted according to OECD and US EPA guidelines for reproductive toxicity, *e.g.* OECD TG 416.

3.1.2.1 Experimental animals and administration of test substance

As a result of the selection process all BPA studies included in this evaluation were conducted in rats and using oral exposure, which is in accordance with OECD TG

426. However, the strain and number of animals and vehicle of administration differ between studies.

In the study by Stump *et al*. 24 mated females/dose were included, which is in accordance with the requirements in TG 426. Ema *et al*. included 25 animals/sex/dose in the tests in open field but only 6 rats/sex/dose were used in the tests for learning and memory. Much fewer animals were used per dose group in the non-guideline studies. Jones *et al*. only had 2 or 3 dams at each dose.

The method for assigning offspring to different tests was clearly described by Stump *et al.* and Ema *et al*. also making it clear that the litter was used as the statistical unit in these two studies. However, this information was missing in the studies by Jones *et al.* and Negishi *et al*. but it can be assumed that all pups were subjected to all tests and that the individual pup was used as the statistical unit in these two cases. Also, only Stump *et al.* and Ema *et al.* have reported that pups were individually identified and marked.

BPA was administered slightly differently in the four studies; in feed in the study by Stump *et al*., in oil which the animals were allowed to drink from a syringe in the study by Jones *et al*. and in oil and water, respectively, by gavage in the studies by Negishi *et al.* and Ema *et al*. It should be noted that the use of different vehicles may result in differences in toxicokinetics of the test substance, and potentially on the study outcome, but kinetics was not investigated in these studies. The doses administered differ somewhat. Stump *et al.* and Negishi *et al.* administered doses within the same range, although Stump *et al.* included more dose groups. Jones *et al.* and Ema *et al*. included only low doses, comparable to the lower end of the dose range tested by Stump *et al*. Three or more dose groups were used in all four studies, with appropriate spacing to evaluate a dose-response relationship, which is in agreement with the criteria stated in TG 426. However, the doses chosen by Ema *et al.* were very low and included much lower doses than the other studies. The period of dosing covers (at least partly) the gestation and lactation periods in all studies but differs somewhat. None are in exact agreement with the recommendations of TG 426. No motivation for using a certain exposure period is given, apart from the study by Ema *et al*., which is in accordance with guidelines for reproductive toxicity. There are some differences in exposure duration but primarily postnatal exposure is different in the study by Stump *et al*. compared to the other studies. Because BPA was administered in feed in the study by Stump *et al*. pups were exposed directly via the diet during the last week before weaning. No direct dosing of pups occurred in the other studies where BPA was administered directly to the dam.

3.1.2.2 Endpoints and methods

Apart from the study by Jones *et al*. there is quite a large agreement between the studies in terms of what endpoints have been evaluated. However, the methods differ between studies (see Table 4)

Detailed clinical observations

Clinical observations, mainly body weight measurements, of dams and offspring were conducted in all studies except the study by Jones *et al*. However, the level of detail

in the reporting differs considerably. No treatment related effects were noted during the clinical observations in any of these studies.

Brain weight and neuropathology
Neuropathology was investigated in the studies by Stump *et al*. and Negishi *et al*., 2003. TG 426 states that tissue samples should be collected twice during the study; on PND 22 and at termination. However, Negishi and co-workers only collected brain tissue at the end of the study. No treatment related neuropathological changes were observed in the study by Stump *et al*. Some microscopic alterations in the brain, *i.e.* nests of embryonal cells, were found in a few BPA-treated animals but since these nests were unilateral in nature, there was a lack of a dose response and similar findings had been observed in the historical control study, they were considered incidental. No results from the brain tissue sample analysis were presented by Negishi *et al*.

Behavioural ontogeny
Behavioural ontogeny in pre-weaning offspring was only investigated in the two guideline studies by Stump *et al*. and Ema *et al*. No treatment related effects in these endpoints were observed in either study.

Motor Activity
Motor activity was evaluated in the studies by Stump *et al*., Negishi *et al*. and Ema *et al*. although slightly different apparatuses were used in these studies. Importantly, the evaluations of motor activity were conducted at different ages. Stump *et al.* conducted their measurements using the Kinder Scientific Motor Monitor System on PNDs 13, 17, 21, 61, which is in accordance with the recommendations of TG 426. Each test session was 60 min in duration; total activity, measured as interruption of any photo beam, was recorded. Data were compiled into six 10-min subintervals for tabulation and evaluation of intrasession habituation. No effects on motor activity were observed. In the study by Negishi *et al*. spontaneous motor activity was tested using a Supermex in which the body heat of an animal is detected by infrared detectors. All spontaneous movement was detected and counted. Measurements were collected during 12 h sessions (dark phase) on PND 28-34, which is between the time periods recommended in TG 426. Counts were automatically totalled and recorded at 5 minute intervals. Locomotion during the dark phase was also measured in the Open field test at 8 weeks (PND 56-62), where behaviour was observed and recorded with video camera for 5 minutes. No apparent effect on motor activity was observed in the Supermex or the Open field test. However, detailed analyses from the Supermex test showed that the immobile time was significantly prolonged in female offspring. Ema *et al*. measured ambulation, rearing and grooming (as well as urination and defecation) in the Open field test at 5-6 weeks of age (approximately PND 35-42), also falling between the time points for testing required in TG 426. The animals were allowed to explore freely for 3 minutes. No significant differences between BPA-treated and control animals were observed.

Motor and sensory function
Only the study by Stump *et al*. investigated this endpoint. Auditory startle was measured using the Startle Monitor System on PNDs 20 and 60, in accordance with TG 426. No effects on auditory startle response amplitude or the pattern of habituation were observed.

Learning and memory
Different aspects of learning and memory were evaluated in the studies by Stump *et al*. and Ema *et al*. as compared to the study by Negishi *et al*. However, all three studies used tests that meet the criteria of TG 426. In the study by Stump *et al.* the Biel water maze test was conducted in accordance with TG 426 on PNDs 22 and 62. Evaluations of swimming ability and motivation to escape were performed in a first phase where animals were allowed four consecutive trials in a 4-foot straight channel during day 1 of testing. Phase two evaluated sequential learning (solving the maze in path A and B) on test days 2-6. On day 7 of testing animals were tested for their memory in path A. On PND 62 a significantly higher error rate for path A during trials 1-4 (learning) was observed in males of the lowest dose group compared to controls. However, this difference was attributed to uncharacteristically low error rate in the study control compared to the historical control. Also, males and females combined in the next to lowest dose group showed a higher error rate than controls in one of the path B trials. As observed effects did not occur consistently between or within testing periods, did not show a dose response or were associated with atypical control responses they were considered spurious and not treatment related. Negishi and co-workers studied the effects on active avoidance using the two-way shuttle box at 4 or 8 weeks of age (PND 28-34 or 56-62, respectively), which falls outside the time periods required for testing in TG 426, *i.e.* at post-weaning (PND 25±2) and in young adults (PND 60-70). Also, it is unclear if the criteria for testing memory (*i.e.* retention trial) were fulfilled in this study. Animals were allowed to explore the shuttle box for 5 min before testing and were then submitted to 50 daily trials of avoidance conditioning on 3 consecutive days. Males in the two highest dose groups showed significantly higher response at 4 weeks compared to controls while males in the lowest dose group showed significantly lower response than controls at 8 weeks. No effects were observed in females. Ema *et al.* subjected pups to tests in the T-maze at 6-7 weeks of age (approximately PND (42-49), which falls outside the time periods required for testing in TG 426. No details regarding the nature or number of trials are given. No effects of BPA-treatment were observed (no data were shown).

Other neurobehavioural testing
Anxiety was investigated in the Open field test in the studies by Negishi *et al*. and Ema *et al*. Negishi *et al*. recorded open-field behaviour on PND 56-62 for 5 min during the dark phase, classified into grooming, locomotion, stretching, rearing or “others”. A higher percent of grooming was observed in males in the lowest dose groups. No effects were observed in the higher doses. Ema *et al*. allowed animals at 5 to 6 weeks of age to explore the open field for 3 minutes. Ambulation, rearing, grooming and the occurrence of urination and defecation were recorded. No effects of BPA-treatment were observed (data were not shown).

Jones *et al*. (2011) evaluated the effects on sexual behaviour in adult offspring at between PND 90 and 120. Animals were tested during the first 4 h of the dark phase. Males and females were paired with sexually experienced stimulus animals of the opposite sex. Male sexual performance was evaluated at the beginning (sexual naivete) and end (sexual experience) of a 4-day trial period. The number of anogenital investigations, mounts and intromissions as well as latencies to mount, intromit, ejaculate and the post-ejaculatory interval were recorded and an index of copulatory efficiency was created. Female sexual behaviour was evaluated in two

sessions separated by 4 days (*i.e.* the duration of one estrous cycle). Proceptive behaviours (ear wiggling and hop-darting) were quantified and a lordosis quotient was calculated. Male sexual behaviour was impaired mainly in the two lowest dose groups (5 and 50 µg/kg bw/day, respectively) compared to controls, with the most marked differences observed in the next to lowest dose of 50 µg/kg bw/day. The dose-response relationship was shown to be significantly non-monotonic for many of the parameters of male sexual behaviour. Female sexual behaviour seemed to be unaffected by BPA treatment.

3.1.3 Conclusions

One main difference between the studies investigating developmental neurotoxicity of BPA is the degree of detail in the reporting. The study by Stump *et al*., which was conducted in accordance with TG 426, and also to some extent the study by Ema *et al*, which was conducted in accordance with guidelines for reproductive toxicity, have included more detail in their reporting of study design *e.g.* regarding if/how animals were individually identified or if body weights were measured, than the studies by Jones *et al.* and Negishi *et al*. Further, it is not clear whether the absence of certain information means measurements were not included in the study design or were simply left out of the report. Specifically, information from clinical observations was very poorly reported, if at all, in all studies except the study by Stump *et al*.

The studies differ in several aspects in term of study design, for example the rat strain used, number of animals per dose group, exposure scenario and assignment of pups to different tests. The two guideline studies by Stump *et al*. and Ema *et al*. clearly included a larger number of animals as well as more dose groups and used the litter rather than the individual pup as the statistical unit.

Only the study by Stump *et al*. covered tests for all endpoints required in TG 426. Behavioural ontogeny was only studied in the two guideline studies by Stump *et al*. and Ema *et al*. Motor activity and tests for learning and memory were covered by all studies except the study by Jones *et al*.

Methods for testing of learning and memory differed between the studies. Stump *et al.* and Ema *et al*. conducted tests in spatial mazes to investigate learning and memory while Negishi *et al*. evaluated active avoidance. TG 426 recommends that the choice of tests for evaluation of learning and memory be based on “its demonstrated sensitivity to the class of compound under investigation”. However, this seems to not have been considered in the study by Stump *et al*. In that study the Biel water maze was chosen with reference to the study by Ema *et al*. who had used the T-maze. However, Ema *et al*. did not observe any effects from BPA on learning and memory using the T-maze, making this a poor reference for the choice of such a test.

Also the timing of testing, *e.g.* for motor activity and learning and memory, varied between the studies by Stump *et al.*, Negishi *et al*. and Ema *et al*. Only Stump *et al*. covered the time points required in TG 426.

The studies by Jones *et al*. and Negishi *et al*. reported adverse effects from perinatal BPA-exposure on neurobehavioural endpoints that are not included in TG 426. Jones

et al. observed adverse effects on adult male sexual behaviour after perinatal exposure to BPA. Effects on performance in the Open field test, indicating increased anxiety in BPA-treated animals, were observed in the Study by Negishi *et al*. However, Ema *et al*. also investigated open-field behaviour but did not report any effects indicating increased anxiety in the animals.

In the two studies by Jones *et al*. and Negishi *et al*. that were not conducted according to guidelines the authors concluded that perinatal exposure to BPA adversely affects neurobehavioural development in rats while the authors of the two guideline studies by Stump *et al*. and Ema *et al*. concluded that BPA is not a developmental neurotoxicant. It should be noted that in the studies by Jones *et al*. and Negishi *et al*. observed differences between BPA-treated animals and controls were considered treatment related even though they did not show a clear dose response or consistency between or within testing periods. In the study by Stump *et al*. such observations were considered spurious and not treatment related. No effects on behavioural endpoints were reported by Ema *et al*. however, no data was shown from behavioural tests in this study.

3.2 PBDE 209

3.2.1 Selection of studies

A developmental neurotoxicity study of PBDE 209, conducted according to OECD TG 426, was available from the Swedish Chemicals Agency (Beck 2009, later published as Biesemeier *et al*. 2011). Three studies on neurotoxicity after exposure to PBDE 209 during the prenatal and/or early postnatal period were found in the open literature (Viberg *et al.,* 2003 and 2007, Rice *et al.,* 2007 and 2009). These studies were not conducted according to OECD TG 426 (non-guideline studies). The four studies were evaluated and compared to the OECD TG 426.

3.2.2 Comparison of experimental design and methods and agreement with OECD TG 426

The criteria for a developmental neurotoxicity study stated in OECD TG 426, as well as the experimental design and methods used to evaluate developmental neurotoxicity in these four studies, have been summarized in Table 5 (see end of report).

3.2.2.1 Experimental animals and administration of test substance

The four PBDE 209 studies used different species. The guideline study was conducted in rats (Sprague-Dawley), which is in accordance with TG 426, and the non-guideline studies were conducted in rats (Sprague-Dawley, Viberg *et al.,* 2007) and mice (NMRI, Viberg *et al.,* 2003, or C57BL6/J, Rice *et al.,* 2007 and 2009).

TG 426 stated that 20 litters are recommended at each dose level. The guideline study included 35 mated females per dose group, while the Viberg *et al*. studies

included 20 (rats) and 10 (mice) pups from 3-5 litters/group. The Rice *et al*. studies included 11-13 mice per sex/dose (1+1 from each litter).

The doses administered differ between the studies. In the guideline study the dams were administered PBDE 209 at dose levels 0, 1, 10, 100 and 1000 mg/kg/d, which is one more dose group than required in the TG 426. In the Viberg *et al*. non-guideline studies two dose levels were used (0, 6.7 and 20.1 mg/kg for rats and 0, 2.22 and 20.1 mg/kg for mice). In the Rice *et al*. non-guideline study also two dose levels were tested (0, 6 and 20 mg/kg/d). The period and method of dosing differ between the studies. In the guideline study PBDE 209 was administered by gavage to the dams and covered the gestation and lactation periods. In the Viberg *et al*. non-guideline studies the test substance was administered as a single dose directly to the pup via gavage on PND 3 (also on PND 10 and 19 for mice). Viberg *et al*. 2003 refer to the sensitive period of "the brain growth spurt" that in rodents occur early postnatally. The test substance PBDE 209 is persistent, why a single dose is sufficient to cover a longer postnatal period. They show that the concentration of PBDE 209 in the brain in fact increases over a 7-day period after dosing. In the Rice *et al*. non-guideline study pups were exposed directly to PBDE 209 in the mouth with a micropipette daily during PND 2-15. The vehicle was corn oil in the guideline study and a 20% fat emulsion in all the non-guideline studies. In the guideline study an additional study measured the levels of PBDE 209 in maternal plasma and milk as well as in plasma of offspring as evidence of postnatal pup exposure.

3.2.2.2 Endpoints and methods

The endpoints and methods used in the studies are summarized in Table 5.

Detailed clinical observations

The guideline study is performed in accordance with the requirements in the TG 426. No effects on clinical observations or body weight were observed. Rice *et al*. (2007) studied clinical observations based on a FOB (referring to US-EPA Guidelines for neurotoxicity risk assessment) and it was conducted every other day between PND 2 and 20. The palpebral reflex, forelimb grip (delayed acquisition) and struggling behaviour during handling were affected by PBDE 209. No information is given in the non-guideline studies by Viberg *et al*. on how clinical observations were evaluated. The authors state that "there were no clinical signs of toxicity". Body weights have been measured and it was reported that there were no differences between groups (all studies).

Brain weight and neuropathology

Brain weight and neuropathology was investigated in the guideline study, but not in the non-guideline studies. In the guideline study, there were no significant effects on these parameters except for increased brain-to-body weight ratio in male offspring of the 100 mg/kg/d group at PND 21 (not at PND 72).

Behavioural ontogeny

Behavioural ontogeny was investigated in the guideline study (motor activity). No treatment-related effect of these endpoints was reported. In the Rice *et al*. (2007) study several behaviours were followed during the lactational period, *e.g.* acquisition of the righting reflex, rearing, grip strength etc. Motor activity was also evaluated, but

not using an automated device. A delay in showing an effective grip with the forelimbs was observed in the high dose group (statistically significant in males, not when males and females were combined!). Behavioural ontogeny was not evaluated in the Viberg *et al*. studies.

Motor Activity
Motor activity (including habituation) was evaluated in the guideline study and the Viberg *et al*. studies, with an additional nicotine-challenge session. In the guideline study locomotor activity was studied at PND 13, 17, 21, 61, 120 and 180 using a validated Kinder Scientific Motor Monitor System with background white noise. Each session was 60 minutes and data was reported as 10-minute subintervals. In the highest dose group (1000 mg/kg/d) there was a statistically significant effect at PND 180, both in the first (however within the laboratory's historical control range) and the nicotine-induced session (males and females were combined). In the non-guideline studies by Viberg *et al*. motor activity was studied at 2 months of age (for mice also at 4 and 6 months). These studies used a Rat-O-Matic automated device with infrared beams. Each session was 60 minutes and data was reported for 20-minute subintervals. A decrease in initial activity and a decreased habituation was reported for both rats and mice and a decreased response to nicotine was reported for mice (rats not tested for nicotine challenge). In the non-guideline study of Rice *et al*. (2007) motor activity was evaluated on PND 70 and at one-year of age using an automated device from Coulbourn with infrared beams. Naïve mice were tested for two hours during the dark cycle. The only treatment-related effect was an increase in motor activity in young high dose males during the first 1.5 hours.

Motor and sensory function
In the guideline study grip strength was evaluated at PND 21, 35, 45 and 60 and auditory startle response was studied on PND 20 and 60. No treatment-related effects were reported for these endpoints. In the Rice *et al*. study (2007) motor and sensory function was evaluated during PND 2-20, but not after weaning. No tests on motor and sensory function were reported in the Viberg *et al*. non-guideline studies.

Learning and memory
Learning and memory were evaluated in the Biel Maze swimming trial at PND 22 and 62 in the guideline study for 7 consecutive days (day 7 for memory test). No treatment-related effects on swimming ability, learning or memory was reported (an improved performance in two trials in the 100 and 1000 mg/kg/d groups was not considered substance-related). In Rice *et al*. (2009) young adults and aging mice were tested for a series of operant procedures including a fixed-ratio (FR) schedule of reinforcement, a fixed-interval (FI) schedule and a light-dark visual discrimination. Mice were trained for these tests when after reaching adulthood (approx. 87 days old, young adults) and at about 16 months of age (ageing), respectively. Performance was evaluated after the mice had successfully learnt the different tasks. For the FR-test, no effects were observed in the young mice, but the high dose mice of the ageing cohort showed initially better performance as they earned more food pellets. In the FI test the high dose group of the ageing cohort showed a higher overall response rate and in the visual discrimination test they had a higher number of total errors. In this test the young mice of the high dose group also had more errors. Learning and memory were not evaluated in the non-guideline studies by Viberg *et al*.

Other neurobehavioural tests
Anxiety was studied in mice in a later publication from the same laboratory as the non-guideline studies (Johansson *et al.* 2008). The test was performed using an elevated plus maze. No significant dose-response relationship was observed for either entries or time spent in the open arms of the maze suggesting that anxiety is not the reason for reduced motor activity. Anxiety was not studied in the other studies.

3.2.3 Conclusions

The guideline study agrees very well with the OECD TG 426. There are large differences between the guideline and the non-guideline studies. There is also a difference in the degree of detail in the reporting. In some parts it is unclear if the absence of some information is because it was not investigated or was left out of the report.

There are large differences between the studies regarding the number of animals/litters per dose group. In the guideline study and in the study by Rice *et al*. the litter constituted the statistical unit (only one pup/sex/litter was tested). More than one pup/litter were studied in the non-guideline studies by Viberg *et al*. and only male pups were tested.

The guideline study agrees with the exposure period stated in the TG 426, *i.e.* GD 6 to PND 21 (includes indirect exposure of pups via milk, levels of test substance in milk and pup plasma shown), while the non-guideline studies exposed the pups directly and postnatally using a single or repeated doses.

Much fewer tests were performed in the non-guideline study by Viberg *et al*. (*i.e.* only motor activity including habituation with nicotine-challenge) compared to the guideline study. The Rice study included more tests but did not agree fully to the recommendations of the TG 426.

The timing of testing in the guideline study agreed with TG 426. In the non-guideline studies by Viberg *et al*. the rats and mice were not tested until adulthood (from 2 months of age). In the Rice *et al*. study the testing of motor and sensory function and learning did not cover all time-points recommended in the guideline.

In the guideline study the authors concluded that no evidence of DNT was observed as no effects were considered treatment-related although many tests were performed and doses up to 1000 mg/kg/d were used. Occasional statistically significant effects were in most cases within the range of the laboratory's historical controls. In the non-guideline studies treatment-related effects were reported for developmental ontogeny, motor activity and different operant behaviour tests in both mice and rats and in the low mg/kg range.

3.3. Deltamethrin

3.3.1 Selection of studies

A developmental neurotoxicity study of deltamethrin, conducted according to OECD TG 426 (draft TG 426 of 2003 and the US-EPA OPPTS 870.6300), was unpublished but available from the Swedish Chemicals Agency (Gilmore *et al.,* 2006). Only one study on neurotoxicity after exposure to deltamethrin during the prenatal and/or early postnatal period was found in the open literature (Eriksson and Fredriksson 1991). This study was not conducted according to OECD TG 426 (non-guideline study). These two studies were evaluated and compared to the OECD TG 426.

3.3.2 Comparison of experimental design and methods and agreement with OECD TG 426

The criteria for a developmental neurotoxicity study stated in OECD TG 426, as well as the experimental design and methods used to evaluate developmental neurotoxicity in these two studies, have been summarized in Table 6 (see end of report).

3.3.2.1 Experimental animals and administration of test substance

The two deltamethrin studies used different species. The guideline study was conducted in rats (Wistar), which is in accordance with TG 426, and the non-guideline study was conducted in mice (NMRI).

TG 426 stated that 20 litters are recommended at each dose level. The guideline study included 30 mated females (resulting in 23 litters) per dose group. However, only 16 rats/sex/dose were used for behavioural testing compared to the 20/sex/dose required in the TG 426. The non-guideline study included 12 pups from 3 litters/group.

The administration of the test substance differs between the studies. In the guideline study the dams were administered deltamethrin orally at dose levels 0, 20, 80 and 200 ppm in the feed (corresponding to 0, 1.64, 6.78 and 16.1 mg/kg/d). The number of dose levels is the same as required in the TG 426. In the non-guideline study only one dose level was used (0 and 0.7 mg/kg/d). The period of dosing covers the gestation and lactation periods in the guideline study, but in the non-guideline study pups were exposed once daily during PND 10-16. In the guideline study the test substance was administered in the diet, while in the non-guideline study it was administered orally with a PVC tube and the vehicle was a 20% fat emulsion.

3.3.2.2 Endpoints and methods

The endpoints and methods used in the studies are summarized in Table 6.

Detailed clinical observations
The guideline study is performed in accordance with the requirements in the TG 426.Both dams and offspring showed reduced body weight gain in the highest dose group. The highest dose males also showed increased incidence of vocalization when handling on PND 4. No information is given in the non-guideline study regarding observations. The authors report “no clinical signs of pyrethroid poisoning”. Body weights have been measured at 4 months of age and no treatment-related effects were reported.

Brain weight and neuropathology
Brain weight and neuropathology was investigated in the guideline study, but not in the non-guideline study. No treatment-related effects were reported.

Behavioural ontogeny
Behavioural ontogeny was investigated (motor activity) in the guideline study, but not reported in the non-guideline study. No effects were observed on motor activity or vaginal opening. A delay in balanopreputial separation was observed for the high dose males.

Motor activity
Motor activity was evaluated in both studies. In the guideline study locomotor activity was studied using a Figure 8 maze for 60 minutes at PND 13, 17, 21, 60 and 120 (including habituation). No treatment-related effects were reported for motor activity or habituation. In the non-guideline study motor activity (including habituation) was studied on PND 17 and at 4 months of age using a Rat-O-Matic automated device with infrared beams. Each session was 60 minutes and data was reported for 20-minute subintervals. There was no effect on motor activity at PND 17, but a decreased habituation at 4-months of age.

Motor and sensory function
In the guideline study auditory startle reflex habituation was evaluated on PND 22 and 60 and reported to show no treatment-related effects. No tests on motor and sensory function were reported in the non-guideline study.

Learning and memory
Learning and memory were evaluated in a water maze at PND 60 in the guideline study; animals demonstrating acquisition were tested for retention on PND 67. In addition acquisition was tested in a passive avoidance test on PND 22, testing retention on PND 29. Neither of these tests showed any treatment-related effects. Learning and memory functions were not evaluated in the non-guideline study.

Other neurobehavioural tests
No additional (non-required) tests were included in any of the studies.

Neurochemistry
High- and low-affinity muscarinic receptor density was studied in different parts of the brain in the non-guideline study. Studies on neurochemistry were not required by TG 426 and was not studied in the guideline study.

3.3.3 Conclusions

The guideline study agrees very well with the OECD TG 426, despite a slightly lower number of offspring tested for behavioural endpoints. There are large differences between the guideline and the non-guideline study. There is a difference in the degree of detail in the reporting. In some parts it is unclear if the absence of some information is because it was not investigated or was left out of the report.

There are differences regarding the species studied and the number of animals/litters per dose group. In the guideline study the litter constituted the statistical unit (only one pup/sex/litter was tested). More than one pup/litter were studied in the non-guideline study and only male pups were tested.

The guideline study agrees with the exposure period stated in the TG 426, *i.e.* GD 6 to PND 21. The pups were thus indirectly exposed via milk as well as directly from feed during the third week of lactation. A limited study on brain levels of deltamethrin in pup brains shows some evidence of exposure. However, the level of indirect postnatal exposure via milk may not reflect a scenario with direct exposure of deltamethrin. In the non-guideline study, pups were exposed directly once daily on PND 10-16.

Much fewer tests were performed in the non-guideline study (*i.e.* only motor activity including habituation and density of brain muscarinic receptors) compared to the guideline study.

The timing of testing in the guideline study agreed with TG 426. In the non-guideline study the mice were tested on PND17, which agrees with TG 426, and at 4 months, which is slightly later than recommended.

In the guideline study effects on body weight, balanopreputial separation and vocalization were observed at the highest dose level (16.1 mg/kg/d). No effects were observed on motor activity or learning and memory function. However, in the non-guideline study effects on motor activity (habituation) were observed at 0.7 mg/kg/d.

3.4. PCB 153

3.4.1 Selection of studies

Three studies investigating developmental neurotoxicity of PCB 153 were identified in the open literature (Piedrafita *et al*., 2008, Fischer *et al*., 2008 and Schantz *et al*., 1995). No developmental neurotoxicity studies conducted according to regulatory guidelines were identified. The three selected studies were evaluated and compared to the OECD TG 426.

3.4.2 Comparison of experimental design and methods and agreement with OECD TG 426

The criteria for a developmental neurotoxicity study stated in OECD TG 426, as well as the experimental design and methods used to evaluate developmental neurotoxicity in these three studies, have been summarized in Table 7 (see end of report).

3.4.2.1 Experimental animals and administration of test substance

The studies used different species; the studies by Piedrafita *et al*. and Schantz *et al*. were conducted in Wistar and Sprague-Dawley rats, respectively, which is in accordance with TG 426. The study by Fischer *et al*. was conducted in mice (NMRI). The study by Fisher *et al*. included investigations of male mice offspring. The sensitivity of this mouse strain (NMRI) to effects on motor activity and habituation has been reported to be comparable to that of Sprague-Dawley rats in another study (study on PBDE 209 by Viberg *et al.,* 2007).

The study by Piedrafita *et al*. used 6 rats of each sex from 3 different litters per dose at each test of learning ability. In the study by Fischer *et al*. 8 male mice from 3-4 litters per dose group were used for behavioural testing. Thus, the individual pup was used as the statistical unit in both these studies. However, Fischer *et al*. stated that they have previously “established that studies using mice randomly selected from at least 3 different litters yield the same statistical effect and power as using litter-based studies". In the study by Schantz *et al.* one rat/sex from 6-9 different litters per treatment group were assigned to tests of learning and memory, *i.e.* the litter was the statistical unit in this study.

Only one dose level of PCB 153 was included in the studies by Piedrafita *et al*. and Fischer *et al*. and two dose levels were included in the study by Schantz *et al*. The dose levels selected were quite similar in the studies by Piedrafita *et al*. and Fischer *et al*; 1 mg/kg bw/day and 0.51 mg/kg bw/day, respectively, but were decidedly higher (16 and 64 mg/kg bw/day) in the study by Schantz *et al*.

The timing and duration of exposure differed significantly between the studies. Piedrafita *et al*. exposed dams to PCB 153 mixed in sweet jelly from GD 7 to PND 21, thus offspring were indirectly exposed *in utero* and via lactation (pups may have consumed jelly during the last week of lactation). In the study by Schantz *et al*. pregnant dams were exposed via gavage to PCB 153 in corn oil from GD 10 to 16. In the study by Fischer *et al*. pups were directly exposed to a single dose of the test substance via gavage on PND 10.

3.4.2.2 Endpoints and methods

The endpoints and methods used in the studies are summarized in Table 7.

Detailed clinical observations

No detailed clinical observations seem to have been conducted in any of the studies. Fischer *et al*. briefly stated that no overt signs of clinical toxicity occurred during the experimental period, but it is not clear which parameters were evaluated.

Brain weight and neuropathology
Brain weight and neuropathology were not investigated in any of the studies.

Behavioural ontogeny
Behavioural ontogeny was not investigated in any of the studies.

Motor activity
Motor activity, including habituation, was studied by Fischer *et al*. in male offspring at 2 and 4 months of age using an automated device (Rat-O-Matic). Slight reductions in motor activity were reported for PCB exposed mice at both 2 and 4 months of age without any apparent effect on habituation. Motor activity was not investigated in the studies by Piedrafita *et al* and Schantz et al.

Motor and sensory function
No measurements of motor and sensory function were included in any of the studies.

Learning and memory
Learning ability at 3 and 7 months (different animals were used) was investigated in the study by Piedrafita *et al*. using a wooden Y-shaped maze with food pellets as rewards. Learning was impaired by PCB 153 in young, but not in adult rats, similarly in males and females. Piedrafita *et al*. concluded that the normally higher learning ability in young rats compared to adult rats was not observed in the PCB 153 treated rats. Memory function was not studied by Piedrafita el al. Schantz *et al*. investigated spatial learning and memory in adult rats using two different test methods. Working and reference memory was evaluated using the radial arm maze. Tests were conducted daily (Monday – Friday) for 7 consecutive weeks starting at 90 days of age. Delayed spatial alternation (DSA) was tested in the T-maze for 3 consecutive weeks starting at approximately PND 165. Learning and memory was not investigated in the study by Fisher *et al*.

Other neurobehavioural tests
No additional (non-required) tests were included in any of the studies.

Neurochemistry
The function of the glutamate-nitric oxide-cGMP pathway was studied by Piedrafita *et al*. by in vivo brain microdialysis. This function was decreased in PCB 153 treated rats at 3 months and the effect correlated with the impaired learning ability.

3.4.3 Conclusions

None of the three studies identified for PCB 153 was conducted according to standardized test guidelines.

There are large differences between the studies regarding the species studied and the timing and duration of exposure. In the study by Piedrafita *et al*. the animal species and strain (Wistar rat) and exposure duration (GD 7 to PND 21) is in accordance with TG 426. The study by Schantz *et al.* also used an appropriate

species and strain (Sprague-Dawley rats) but exposure was restricted to GD 10 – 16. In the study by Fischer *et al*. male mice were used and offspring were exposed postnatally to the test substance on PND 10.

None of the studies covered all endpoints required for study in the OECD TG 426. Detailed clinical observations seem not to have been conducted, or the results from such observations were not reported. Further, only effects in adult animals (two months and older) were evaluated in these studies, thus there was no testing at adolescence, as required in TG 426. The study by Piedrafita *et al*. included one behavioural test for learning ability, but it is not clear whether memory was also evaluated. In the study by Schantz *et al*. two different tests, evaluating different aspects of learning and memory, were conducted. In the study by Fischer *et al*. only motor activity and habituation were measured. The pup was used as the statistical unit in the studies by Piedrafita *et al*. and Fischer *et al.* while the litter was the statistical unit in the study by Schantz *et al*.

Piedrafita *et al*. and Schantz *et al*. reported effects of developmental PCB 153 treatment on learning and memory in rats at adult age. However, in the study by Schantz *et al*. effects were only observed in one of the tests, *i.e.* the T-maze DSA task and not in the radial arm maze. Further, effects were only observed in females in this study. Fischer *et al*. reported that exposure to PCB 153 caused slight reductions in motor activity in adult male mice.

3.5. PFOS

3.5.1 Selection of studies

One developmental neurotoxicity study of PFOS conducted according to OECD TG 426 and US EPA OPPTS 870.6300 was available in the open literature (Butenhoff *et al.* 2009). Two additional studies on neurotoxicity after exposure to PFOS and PFOA during the early postnatal period were found in the open literature (Johansson *et al*., 2008; Onishchenko *et al*., 2011). These studies were not conducted according to OECD TG 426. The studies were evaluated and compared to the OECD TG 426.

3.5.2 Comparison of experimental design and methods and agreement with OECD TG 426

The criteria for a developmental neurotoxicity study stated in OECD TG 426, as well as the experimental design and methods used to evaluate developmental neurotoxicity in these three studies have been summarized in Table 8.

3.5.2.1 Experimental animals and administration of test substance

The studies used different species; the study by Butenhoff *et al*. was conducted in rats (Sprague-Dawley), which is in accordance with TG 426, and the studies by Johansson *et al*. and Onishchenko *et al*. were conducted in mice (NMRI and

C57BL/6/Bkl, respectively). The study by Johansson *et al*. included investigations of male offspring only. The sensitivity of the NMRI mouse strain to effects on motor activity and habituation has been reported to be comparable to that of Sprague-Dawley rats in another study (study on PBDE 209 by Viberg *et al.,* 2007).

TG 426 states that 20 litters are recommended at each dose level. The study by Butenhoff *et al*. included 25 mated females (no information given on final number of litters) per dose group. The Johansson study included 16 male offspring picked at random from 3-5 litters per dose group. Onishchenko *et al*. included 6 pregnant dams in the treatment group and 10 pregnant dams in the control group.

The dose levels administered differ between the studies. In the study by Butenhoff *et al*. dams were administered PFOS orally by gavage at dose levels of 0, 0.1, 0.3 and 1.0 mg/kg/day from GD 0 to PND 20. The period of dosing thus covers the gestation and lactation periods and the number of doses is the same as required in the TG 426. In the study by Johansson *et al*. two doses, 0.75 or 11.3 mg/kg bw, of PFOS were administered via gavage to male pups once on PND 10. Onishchenko *et al*. applied the test substance, dissolved in ethanol, on “palatable food” in a volume adjusted according to individual body weights to reach the exposure of 0.3 mg/kg. Pregnant dams were exposed from GD 1 throughout pregnancy, thus only covering prenatal exposure.

3.5.2.2 Endpoints and methods

The endpoints and methods used in the studies are summarized in Table 8.

Detailed clinical observations

The study by Butenhoff *et al*. includes information on daily clinical observations and measurements of bodyweight in both dams and pups. An extensive FOB was conducted in pups on PND 4, 11, 21, 35, 45 and 60. A statistically significant effect in hind limb grip strength was observed in the male pups of the highest dose group on PND 21. However, since the grip strength mean of this group was within the lab’s historical control mean, the effect was only observed at an isolated time point and there were no effects in correlated measures, such as forelimb grip strength and gait, it was concluded that this observation was not treatment related. No information is given regarding clinical observations in the studies by Johansson *et al*. and Onishchenko *et al*. However, Johansson *et al*. included body weight measurements on PNDs 10 and 28 and state that no overt signs of clinical toxicity occurred during experimental period.

Brain weight and neuropathology

Brain weight and neuropathology was investigated in the study by Butenhoff *et al*. and brain weight was investigated in the study by Onishchenko *et al*. No effects were observed in these endpoints. Brain weight or neuropathology was not investigated in the study by Johansson *et al*.

Behavioural ontogeny

In the study by Butenhoff *et al*. locomotor activity was investigated on PND 13, 17 and 21 and righting reflex on PND 21 as part of the FOB. No effects on righting reflex were observed and all animals showed the expected pattern of lower activity levels

on PND 13, highest levels on PND 17 and slightly lower levels again on PND 21. Increased total cumulative activities were observed in some groups at isolated time points (see below “Motor activity”). No investigations of behavioural ontogeny were reported in the studies by Johansson *et al*. or Onishchenko *et al*.

Motor activity
Motor activity was investigated in all three studies. In the study by Butenhoff *et al*. locomotor activity, including habituation, was studied on PND 13, 17, 21 and 61 using an automated system (SDI Photobeam Activity System). Increased total cumulative activities were observed for male pups in the two highest dose groups on PND 17 and for females in the highest dose group on PND 21. However only the effect in the males of the highest dose group on PND 17 were considered treatment related, as they were accompanied by the absence of habituation. Habituation in the lower dose group and in high-dose females on PND 21 was not different from concurrent controls. In the study by Johansson *et al*. locomotion, including habituation, rearing and total activity was studied at 2 and 4 months of age using a Rat-O-Matic, an automated test device measuring interruptions of infra-red beams. In addition, all types of vibration within the cage are registered by a needle mounted on a horizontal arm with a counter weight and connected to the test cage. Motor activity (locomotion, rearing and total activity) was measured during a 60-minute session divided into three 20-minute intervals. At both 2 and 4 months of age, in the high dose group, locomotion, rearing and total activity were significantly decreased during the first two time intervals of the testing session and increased in the last 20-minute interval, compared to controls. Habituation capability was shown to decrease significantly with age (2 months vs. 4 months). Onishchenko *et al*. measured locomotor activity at the age of 5-8 weeks. Data on walk distance was collected for each animal in 30-minute sessions, divided into 5-minute intervals, using an automated video tracking system. The distance walked was significantly lower in PFOS-treated males compared to controls. No effects were seen in females.

Motor and sensory function
In the study by Butenhoff *et al*. acoustic startle response was studied On PND 20 and 60 using the SR-Lab Startle Response System. No effects were observed. No tests on motor and sensory function were reported in the study by Johansson *et al*. In the study by Onishchenko *et al*. muscle strength and motor coordination were evaluated at 3-4 months of age using the hanging wire test and accelerating rotarod test, respectively. PFOS-exposed males had significantly shorter fall latencies on the hanging wire test and exhibited a reduced ability to stay on the rotarod. There were no statistically significant effects in these endpoints in females.

Learning and memory
In the study by Butenhoff *et al*. learning and memory was evaluated in the Biel water maze for 7 consecutive days starting on PND 22. On day one an evaluation of swimming ability and motivation to escape from the maze was conducted. Sequential learning was evaluated on days 2-6 where rats were allowed two trials per day for two days to solve path A and two trials per day for three days to solve path B. On the last day (PND 28) two trials probing for memory were conducted for each rat. No effects on learning ability or memory were observed. Learning and memory functions were not investigated in the studies by Johansson *et al*. and Onishchenko *et al*.

Other neurobehavioural testing
Johansson *et al*. studied anxiety-related behaviour at 4 months using the elevated plus maze and nicotine-induced behaviour using the Rat-O-Matic. The behavioural tests in the Rat-O-Matic followed the same principles as previously described, *i.e.* motor activity (locomotion, rearing and total activity) was measured during a 60-minute session divided into three 20-minute intervals. No effects on anxiety-related behaviour were observed in the elevated plus maze. In the high dose group the mice responded to nicotine with a decrease in activity in the first and second time intervals, rather than the increase observed in controls. They also showed failure to habituate. In the study by Onishchenko *et al*. circadian activity using the TraffiCage system as well as anxiety related behaviour in the elevated plus maze and depression-like behaviour in the forced swimming test was evaluated at the age of 5-8 weeks. It is not clear whether the same animals were subjected to all tests. In the TraffiCage system the home cages are placed on platforms with 5 embedded circular antennas, which detect the presence of transponders injected into the animals at weaning. Activity is measured as the crossing of the animal from one antenna to another and "resting time" as the total duration of inactive periods. Activity was monitored for 48 consecutive hours. Since home cages were moved to an experimental room for testing the first 3 hours were analysed separately as adaptation to the novel environment. PFOS-exposed animals showed decreased activity in the TraffiCage during the first 3 hours of adapting to the novel environment, with a more pronounced effect in males. In addition, prenatal exposure to PFOS was associated with an increase in the total number of inactive periods during the dark phase in both males and females. No significant effects were observed for anxiety-related behaviour or depression-like behaviour.

3.5.3 Conclusions

The guideline study by Butenhoff *et al*. agrees well with the OECD TG 426. There are large differences between this study and the non-guideline studies by Johansson *et al*. and Onishchenko *et al*. Also, there is a difference in the degree of detail in the reporting.

The studies differ in terms of the species studied and the number of animals/litters per dose group. In the study by Butenhoff *et al*. the litter constituted the statistical unit (only one pup/sex/litter was tested). More than one pup/litter were studied in the study by Johansson *et al.* and only male pups were tested. In the study by Onishchenko *et al*. it is stated that one or two offspring from each litter were randomly selected for testing. However, it is not entirely clear whether, in fact, 1-2 pups *per sex* from each litter were selected. Given that PFOS-treated and control groups are stated to consist of 8 offspring/sex/group, and there were only 6 dams (*i.e.* max 6 litters) in each group, this seems plausible.

Dosing schedules differ between the studies and none adheres exactly to the requirements in TG 426. In the guideline study by Butenhoff *et al*. dams were exposed from GD 0 to PND 21, without justification for starting exposure earlier than what is stated in TG 426 (*i.e.* GD 6). Johansson *et al*. exposed male pups directly,

and only once, on PND 10. In the study by Onishchenko *et al*. dams were exposed from GD1 and throughout pregnancy, *i.e.* offspring were only exposed prenatally.

Out of the tests required in TG 426 only motor activity, including habituation, were investigated in the studies by Johansson *et al*. and Onishchenko *et al*. However, these studies also included tests of nicotine-induced behaviour, circadian activity, anxiety-related behaviour and depression-like behaviour, which are not required tests as stated in TG 426.

The timing of testing in the study by Butenhoff *et al*. is in accordance with TG 426. In the study by Johansson *et al*. the mice were tested only at adulthood. Onishchenko *et al*. conducted tests for locomotor and circadian activity at the age of 5-8 weeks (*i.e.* between PND 35-56) and tests for muscle strength and motor coordination at 3-4 months of age.

All three studies concluded that PFOS can cause disturbances in neurobehavioural development.

4. Discussion

The conclusions from the comparisons of the fifteen studies of the five model compounds are summarized in Table 9. In total five guideline studies and eleven non-guideline studies were compared.

Accordance of studies to TG 426

In general, the guideline studies followed the TG 426. However, the Ema study followed TG 416 for reproductive toxicity, but included also behavioural testing of offspring. For the non-guideline studies published in the open literature there was naturally less detail in the reporting and it is not clear whether the absence of certain information means measurements were not included in the study design or were only left out of the report. The non-guideline studies were more limited in the design as to the number of animals, number of dose levels, number of tests performed as well as the number of time points (age) for testing. When the main conclusion of the study authors was compared, regarding evidence of developmental neurotoxicity, all non-guideline studies and one of the five guideline studies were positive.

Differences in sensitivity between species, strains and sexes

In TG 426 rat is the preferred species of testing and use of other test species shall be justified. Both sexes shall be tested. Rats have been used in all guideline studies as well as four out of eleven non-guideline studies. Different strains of rats have, however, been used in these studies; Sprague-Dawley and Wistar rats in the guideline studies and Wistar, Long-Evans, Sprague-Dawley and F344 rats in the non-guideline studies. In the mouse studies NMRI and C57BL/6 mice were used. In the studies of NMRI mice only males were studied. In separate publications the sensitivity of the male NMRI mice to effects (on spontaneous behaviour and habituation) has been shown to be comparable to the sensitivity of male and female

C57BL/6 mice (compared for exposure to PBDE 99, Viberg *et al.,* 2004) and male Sprague-Dawley rats (compared for exposure to PBDE 209, Viberg *et al.,* 2007).

Question for reviewers: Are there any known, general or specific, differences in sensitivity to DNT between sexes, strains or species?

Reviewer comments:
The reviewers were generally in agreement that differences in sensitivity to DNT between animal species, strains and sexes exist. For example, differences in toxicokinetics and toxicodynamics, as well as in attainment of developmental landmarks in different species and strains may contribute to differences in sensitivity. Reviewers also commented that it is important to keep in mind that endpoints may differ significantly across species and strains, so a comparison in terms of sensitivity to DNT may be misleading. It was pointed out by a few reviewers that the Fischer 344 rat strain would be inappropriate for DNT testing because of poor fecundity of the strain and offspring maturing slightly later than Sprague-Dawley and Wistar rats, which would require adjustment of the timing of behavioural tests.

Concerning sex-differences in sensitivity, these are considered biologically plausible (in sexually mature animals) because of the sexually dimorphic development of the brain, although it seems that the mechanisms behind these differences are not entirely elucidated. However, sex-differences in sensitivity are unexpected prior to sexual maturation.

While some examples were given by some reviewers of where differences in sensitivity have been reported one comment is also that, although there are anecdotal reports, there are so far no well conducted studies that provide empirical evidence of differences in sensitivity between sexes and species.

Using litter or pups as the statistical unit
When discussing the assignment of offspring to the different tests TG 426 states that "Selection of pups should be performed so that to the extent possible both sexes from each litter in each dose group are equally represented in all tests." A minimum of 1 pup/sex/litter is recommended for most tests (except for clinical observations, body weight measurements and optional developmental landmarks where the recommendation is to include all animals). It thus seems acceptable to include more animals per litter for testing. However, the guideline clearly states that "The statistical unit of measure should be the litter (or dam) and not the pup", meaning that appropriate statistical models should be used to account for litter effects, *e.g.* by nesting litters in groups in a factorial ANOVA. Little guidance on the use of statistics in the analysis of DNT studies is given in the OECD and US EPA guidance documents. Holson *et al*. (2008) have provided a comprehensive review of key considerations in the analysis of DNT studies, with recommendations for statistical approaches and reporting of data, stressing that litter effects must be considered throughout the entire DNT study.

All guideline studies included in this report were performed in accordance with the requirement to use the litter as the statistical unit. In most, but not all, non-guideline studies the pup, however, seems to be the statistical unit as more pups per litter and

fewer litters are used. Although the studies using direct exposure to pups do not involve litter effects due to maternal toxicity, other litter effects may be possible and the pups from the same litter may not be "independent" in all aspects. In addition, in total fewer animals were studied in the non-guideline studies as compared to the guideline studies, where at least 20 litters per dose group are required.

Question for reviewers: Which are the consequences for the DNT study if the pup is used as the statistical unit?

Reviewer comments:
There was general agreement that when exposure is via the mother the statistical unit should be the litter and not the individual pup. Using the pup as the statistical unit in this case would lead to an inflated N, consequently increasing the alpha level and the potential for false positive results. Alternatively, using the pup and not litter as statistical unit could potentially also lead to decreased sensitivity of the study because the variation due to litter effects would be included in the random variation.

Question for reviewers: When pups are only exposed directly (and not in utero via the mother), and later studied when being (young) adult, is it still relevant to consider litter effects or would it be OK to use the pup as the statistical unit in this specific case?

Reviewer comments:
Opinions diverge on this question. Some reviewers argued that it is OK to treat individual littermates as independent measures if pups are only exposed via direct administration. Others were of the strong opinion that litter effects (*i.e.* correlations across littermates due to *e.g.* genetics or stress during gestation and lactation) exist also when offspring become adult.

Several reviewers suggested that the litter effect can be minimized in this case if pups within the same litter are randomly chosen to different dose-groups and/or by analyzing the data for a litter x treatment interaction so that maternal-/litter-based influences may be ruled out.

Dosing and exposure pre- and postnatally
Exposure within the DNT study should adequately model the exposure of children (Tyl *et al*., 2008). This includes considering sensitive windows for exposure and differences between species in the timing of birth and brain development. TG 426 recommends exposure at least daily during the period from GD6 to PND21. However, the frequency and duration of exposure may be adjusted to better model human exposure and dosing durations *should* be adjusted for other species to ensure exposure during all early periods of brain development. TG 426 also states that administration of the test substance to the dams may start already on GD 0 but in such cases consideration has to be given to the potential for the substance to cause pre-implantation loss. Also, direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation. Further, the actual postnatal pup exposure shall be considered. Thus, in cases where there is a lack of evidence of continued exposure to offspring, *e.g.* from pharmacokinetic studies, direct dosing of preweaning pups should be considered. Tyl *et al*. (2008) point out

that in cases where the test substance is administered in the feed pups will be directly exposed via feed consumption during the third week of lactation and possibly also via exposure to maternal feces and dermal contact. Information of developmental pharmacokinetics of the substance can therefore be invaluable in data interpretation.

All guideline studies reviewed here followed the requirements for exposure from GD6 (or GD 0) to PND21. In the studies by Butenhoff *et al.* (2009) and Stump *et al.* (2010) exposure was started on GD 0. Three of the guideline studies administered the dose via gavage and two via the feed. Only two guideline studies have measured pup exposure postnatally. In addition, the guideline studies using exposure via feed comment on the direct exposure of pups when they start to consume the feed in their third week of life. Four non-guideline studies administered the dose to dams via gavage and in these cases the offspring was not directly exposed to the test substance. The five non-guideline studies performed at Per Eriksson's laboratory expose the pups directly on PND 3, 10 or 10-16 (authors refer to the "brain growth spurt" that occurs during this period). These mouse pups were thus directly exposed during a sensitive time period, however, not exposed prenatally or through the entire lactation period. Also Rice *et al*. exposed the pups directly postnatally. Onishchenko *et al*. and Schantz *et al*. exposed the dams during pregnancy only.

The Viberg *et al.* studies on PBDE 209 considered the long half-life of the substance and thus administered a single dose (on PND 3) to cover a longer period. If and how the kinetic profile of the test substance was considered in the other studies included here is not clear. Similar to the Viberg studies administration of a single dose postnatally was also conducted in the studies by Fischer *et al*. (2008), Eriksson and Fredriksson (1991) and Johansson *et al*. (2008). In the other studies the test substance was administered repeatedly (daily) or continuously (in feed).

Question for reviewers: Which are the consequences for the sensitivity of the study with deviation from the recommended exposure period (small deviations of days or lack of prenatal or postnatal exposure)?

Reviewer comments:
Reviewers generally agreed that exposure should include both gestation and lactation and that deviation from this exposure paradigm may decrease the sensitivity of the study.

The argument often given was that different brain areas, modulating different cerebral functions, develop and differentiate at different periods of development. Different exposure periods may cover only parts of all events and the critical time point for exposure may be missed. For example, effect on morphological development may be missed if gestational exposure is excluded and functional changes may go unidentified if postnatal exposure is eliminated. If the mechanism of action of the substance and the sensitive period of development is known then exposure during a certain time period during development may be justified. However, since mechanism of action is often unknown, and determining if and when a critical time point for onset of adverse neruodevelopmental effects of a certain compound may manifest itself is extremely difficult and requires a lot of effort and expense, it is prudent to administer the substance from (at least) implantation to weaning.

However, one reviewer observed that “F1 pups often show much greater sensitivity to the test item when administered the compound directly than through indirect exposure via the milk, moreso than can adequately be explained by the difference in route alone.”

Question for reviewers: Is the “brain growth spurt” accepted as a very sensitive period, or even the most sensitive period in the brain development?

Reviewer comments:
There was general agreement that there are several brain growth spurts and that these are sensitive periods but not the only sensitive periods. It was also pointed out that brain growth spurts do not occur on a set schedule and that they often differ between individuals.

Question for reviewers: Which are the consequences for the sensitivity of the study if there is no or lower exposure postnatally (due to low indirect exposure via milk)?

Reviewer comments:
The general message conveyed by the reviewers was that this must be considered on a case-by-case basis. Sensitivity of the study will be reduced if there is no or less exposure during a developmental period when children would be expected to be exposed. However, if the critical time point for onset of effects is passed then it not necessarily mean a loss of sensitivity. It was also noted that rat and human offspring are born at different developmental stages, *i.e.* that rats at birth are at a developmental stage comparable to the human fetus around month 6 and at around 10 days old development of the brain in rat offspring is comparable to a new-born human. Consequently, “no or clearly lower exposure of rats postnatally means that there is a lack of data regarding brain development during the last trimester of human development.”

Question for reviewers: Would it be feasible to include requirements for investigating kinetics in TG 426?

Reviewer comments:
Most reviewers commented that even though it would be useful (sometimes even critical) to have information about the kinetics of the compound before the DNT study is conducted, this would require that more animals are included in the study.

Some reviewers argued that kinetic studies would not sufficiently increase the value of the study, that it is outside the scope of the study, and that more effort should be put into studying the mechanism of action of the compound. It was often stated that kinetics is already studied in other toxicity tests. However, it must then be mentioned that not all compounds can be expected to go through rigorous regulatory testing.

One reviewer commented that kinetics is not sufficiently covered in TG 426 or any of the relevant guidance documents.

Number and choice of tests

In Table 9 the tests performed on the model compounds for motor activity, learning and memory function and other neurobehavioural test in the studies reviewed are summarized. The guidance documents (OECD, 2004 and 2008; US EPA, 1998b) provide descriptions of available methods for evaluating neurotoxicity and, to some extent, also provide guidance on interpretation of results. However, the design of tests and selection of methods to include must be based on the expert judgement of the investigator. For example, the OECD guidance document for neurotoxicity testing states that the selection of the most appropriate methods should be done "on a case-by-case basis and be guided by all of the available information of the chemical" (OECD, 2004). It is important to acknowledge that the flexibility in the guideline, in regard to the choice and design of the behavioural test, may be a source of extrinsic variability in study data (Raffaele *et al*., 2008).

In all guideline studies, except for the reproductive study by Ema *et al*., both motor activity and learning and memory function were tested according to TG 426. In the non-guideline studies only a few covered both these endpoints. Both motor activity and learning and memory function were evaluated using several different devices. TG 426 states that the test of learning and memory should be chosen on the basis of its demonstrated sensitivity to the class of compound under investigation, if such information is available in the literature. Importantly, different studies (of the same model compound, *e.g.* for BPA and PCB 153) have included different methods for testing cognitive function, such as learning and memory. Some of these show effects and some do not. It is known that different tests for cognitive function, *e.g.*, passive avoidance, active avoidance or spatial mazes, assess different forms of learning and/or memory (OECD 2004). It is therefore reasonable to expect that any one chemical may show different effects on cognitive function depending on which test method is being used.

In addition, anxiety was tested in five of the 16 studies and sexual behaviour was tested in one study. Testing for anxiety and sexual behaviour is not required in TG 426.

Question for reviewers: How can the choice of test method to study *e.g.* motor activity and learning and memory function influence study outcome?

Reviewer comments:

The reviewers agreed that the choice of test method can influence the study outcome, especially for learning and memory tests.

Different learning and memory tests evaluate different aspects of learning and memory which may be differently influenced by different compounds. Therefore there is no test that is the most sensitive or appropriate for all compounds.

Different methods for measuring motor activity were discussed and that many factors influence these measurements. Some reviewers commented that when motor activity is measured in an open field, the activity will be influenced by stress and anxiety-like behaviour which can be reduced if measurements are made in a home environment. However, a few reviewers stated that different methods for motor activity are equally robust and that "as long as appropriate controls are tested simultaneously, the

paradigm and apparatus are validated, and the outcome has been replicated within/between labs, the results should be reliable", and therefore the choice of method for this endpoint is not likely to influence study outcome.

Several reviewers pointed out that different laboratories have developed expertise to conduct only certain tests so the choice of test is also dependent on what the lab offers as well as on other practical and logistical issues, such as the availability of positive and historical controls, the design of the study, how many animals are tested, etc.

Question for reviewers: What device/method is most reliable for evaluating motor activity? What is the level of expertise needed to perform and interpret these tests?

Reviewer comments:
The opinions of the reviewers differed in regard to this question. Some reviewers expressed a decided preference for either automated tests (automated actimeters), which can simultaneously record a lot of different information regarding a large number of motor activity related parameters, or video monitoring, because it could be more informative than automated devises. Other reviewers argued that different methods may work better in different situations, or that several different tests should be included in order to be able to draw sound conclusions. Some emphasized the importance of historical and positive controls to characterize normal response and demonstrate the ability of the equipment and methods rather than the test itself.

Training and expertise were stated to be important to perform and interpret tests, *e.g.* in handling animals, setting up the equipment, standardize test conditions, verify test system suitability (positive controls) and evaluating data as well as more specific issues, such as distinguishing between exploratory activity and spontaneous motor activity.

Some important issues to consider when measuring motor activity were brought up, e.g:

- Having appropriate environmental controls to control for the time of day and conditions within the testing room.
- Counterbalancing of animals during data collection.
- The variance in control animals' responses.
- The number of animals needed to generate a representative response.
- The influence of stress and anxiety on motor activity.

Question for reviewers: Can motor function also be evaluated in tests for motor activity, or are separate motor function tests needed to fulfil the requirements of TG 426?

Reviewer comments:
Reviewers expressed different opinions regarding this question. It seems it may depend on the type of motor function evaluated and the type of device used, as some devices can only detect large movements, some both large and small movements and some can even distinguish between the two. One reviewer commented that automated devises to measure motor activity are especially unsuitable for measuring motor function.

Some reviewers stated that the motor activity test is *not* suitable for evaluating motor function because motor activity is an apical test and changes can be caused by a number of underlying causes, including a change in motor function. It was specifically noted by a couple of the reviewers that motor coordination might not be detected in traditional motor activity tests. One reviewer noted that motor function should be tested separately and that this should be included in the TG.

A few reviewers argued that motor function can indeed be measured in tests for motor activity, if the equipment is well calibrated and there is sufficient historical and positive control data available at the test facility.

Some comments were made that several tests, *e.g.* the motor activity assessment in combination with detailed clinical observations (DCO), a functional observational battery (FOB) or separate observational or functional tests (*e.g.* grip strength or rotarod), are needed to sufficiently evaluate motor function.

Question for reviewers: What should be the basis for choosing the most appropriate test for learning and memory function? What is the level of expertise needed to perform and interpret these tests?

Reviewer comments:
It was indicated by several reviewers that testing of learning and memory may be the most challenging aspect of the DNT test paradigm. Reviewers were in agreement that a high degree of expertise (even higher than for the other tests in TG 426) is needed for performing and evaluating such tests.

Some reviewers highlighted the issue that knowledge about the mechanism of DNT of the compound is needed to choose the right learning and memory test since different tests evaluate different types of learning and memory, modulated by different mechanisms and involving different neuronal circuits. It was also noted, however, that such information is rarely available. One suggestion was that results from motor function and motor activity tests should be considered in the choice of learning and memory test.

Some reviewers state that the choice of test is usually based on practical and logistical issues, such as the expertise and experience of the testing facility and their availability of historical and positive control data. Specific comments were also made by a couple reviewers that only validated tests should be included.

A few reviewers suggested that ideally a battery of different learning and memory tests should be conducted to cover the different mechanisms by which this parameter may be affected.

Question for reviewers: Are the functional and behavioural tests recommended in TG 426 appropriate for testing young animals (for example differences in motivation between young and adult animals)?

Reviewer comments:
The reviewers generally agreed that the tests in TG 426 are appropriate, given that appropriate positive and negative controls are performed, that the evaluation scale and equipment are adapted accordingly and that performance is compared to age-matched controls.

A couple of the reviewers expressed a wish to include a wider battery of tests. Another comment was, however, that the TG leaves it up to the test facility/registrant to identify if any additional endpoints should be included.

Question for reviewers: Are any endpoints missing in TG 426 that would possibly increase the sensitivity of the study? For example, shall tests of anxiety and sexual behaviour be recommended for certain groups of compounds (e.g. sexual behaviour when studying EDCs)?

Reviewer comments:
Reviewers disagreed in regard to this question. Some stated that the TG is complex enough as it is and that it is up to the test facility/registrant to identify if any additional endpoints should be included. This needs to be done on a case-by-case basis. It was also stated that any test that is added should be validated and sufficient historical and positive control data should be available.

Other reviewers mentioned endpoints that they feel are missing from the TG or that could be useful in certain cases, e.g:

- Activity in wheel (rewarding)
- Circadian rhythms of activity
- Anxiety
- Preference for specific conditions or drugs
- Sex-related behaviour
- Social behaviour

Question for reviewers: May additional pharmacological/physiological challenges (e.g. additional exposure to amphetamines or nicotine) be useful in "unmasking" effects of DNT that may not otherwise be detected using the TG 426 DNT study design?

Reviewer comments:
Different opinions were expressed by the reviewers. Some stated that including this type of challenge in addition to other tests may be useful for some types of compounds. One reviewer expressed a strong opinion that it should be included and that the selection of such a test should be based on what is known/predicted about the compound's mechanism of action. Another issue that was raised was that this type of challenge may be especially important when evaluating low-dose effects.

The majority of the reviewers commented that there is no evidence that such methods would reveal anything else about the compound that is not likely to be discovered in the tests already included in the TG. Some remarked that it may help characterize the nature of the toxicity but will not necessarily add to the sensitivity of the study. It was also argued that such challenges introduce parameters that are not easy to control, that they can produce responses that are unrelated to the

neurotoxicity of the compound and that results may be difficult to interpret. Further comments were that if such a challenge is added there should be a very clear test hypothesis and ideally positive and negative controls.

Timing of testing

In TG 426 recommendations are given to at what time points certain testing should be performed. This refers both to testing of developmental ontogeny as well as testing in young adults. The guideline studies were generally conducted in accordance with these recommendations. In most non-guideline studies included in this review, fewer time points than required in the TG 426 were used for testing. However, the non-guideline studies all reported an effect on developmental neurotoxicity suggesting that they have covered a sensitive (although maybe not the most sensitive) period for testing.

Question for reviewers: Can differences in time points (age of animals) for testing influence the sensitivity of the study?

Reviewer comments:

The reviewers generally agreed that the timing of testing has a large influence on the ability to identify effects of the compound and that it is a factor that needs to be considered when analysing and interpreting results. Comments were made that this is especially true for developing animals but when they reach adulthood age at testing becomes of less concern. It was also discussed that if tests are conducted at different time points in different studies there will be limited comparability between studies, this also means, however, that adjustments may have to be made for different rat strains because of differences in growth rate.

Some issues that were raised in regard to this question were, e.g:

- You cannot test a function that has not yet been fully developed.
- Some effects become apparent only at later ages when the ability of the organism to compensate impairments and maintain homeostasis is hampered.
- Exposure to low doses can take longer before they are detectable
- The behaviour defect may change with time, *e.g.* hyperactive condition in young but hypoactive when adult; therefore tests could be conducted during a time point when no effect would be noted.

Factors influencing behavioural test outcomes

A large number of varied factors that can influence the outcome of behavioural testing are mentioned in TG 426 and relevant guidance documents. These factors include, *e.g.* animal species/strain, time of day when tests are conducted (considering circadian rhythm), handling, stress and environmental factors, such as sound level, size and shape of the test cage, temperature, relative humidity, light conditions and odours and environmental distractions as well as rearing conditions and previous stress experiences (OECD 2004; Raffaele *et al*., 2008; Tyl *et al*., 2008; US EPA, 1998b). Further, it is important to consider that concurrent systemic toxicity and/or impaired motor and sensory function may affect performance in tests for motor activity, learning and memory. The above-mentioned potentially confounding factors and variables must be controlled for in the test design and/or statistical analyses.

Interestingly, some of the model compound studies in this investigation also included some type of control for estrous cyclicity in females when conducting behavioural testing as it can significantly affect female behaviour and performance in such tests. However, controlling for estrous cyclicity is not required in TG 426. Further, it is not mentioned in any of the guidance documents reviewed here. The ILSI report on the variability in DNT (Raffaele *et al.,* 2008), however, discuss that stage of estrous may greatly influence the performance of female rats. Particularly, they may have a significantly increased level of motor activity, which may also affect performance in tests for learning where animals are required to perform physical activity. Since it may be difficult or impractical to control or compensate for estrous cyclicity Raffaele *et al*. suggest that the estrous cycle could be monitored in order to allow characterization of any effects on test performance.

Question for reviewers: Should estrous cyclicity be controlled for in testing of neurotoxicity in adult females? What would be a feasible way of doing this?

Reviewer comments:
The reviewers disagreed concerning the importance of controlling for estrous cyclicity as well as the feasibility of including such controls.

Some stated that it should be controlled as it affects behaviour. Although, it was also recognized that it may be difficult to do in practice without disturbing the animals. A few suggestions were made of tracking the estrous cycle to help in data interpretation. However, one reviewer commented that cytological monitoring of estrous cycle may not be very reliable as it does not always give similar results with histological evaluation.

Other reviewers argued that controlling for estrous cyclicity is neither warranted nor feasible, highlighting that more animals would be needed; that it would complicate the study design and potentially introduce new confounders. It was further discussed that controlling for estrous cyclicity would have little benefit as all groups are treated equally and the stage of estrous would be expected to be approximately equivalent across all groups, *i.e.* if a sufficient number of animals are used in the study there is no need for controlling for estrous cyclicity.

Question for reviewers: Can male behaviour be affected by the estrous stage of nearby females? Could the sensitivity of behavioural testing be increased by keeping males and females in separate rooms?

Reviewer comments:
The reviewers generally acknowledged the fact that male behaviour may be affected by nearby females, however none of the reviewers suggested that it is warranted or feasible to separate the sexes.

The arguments given were:

- Separation of males and females may be difficult to manage, it would increase costs, there is no evidence that separating them would be better
- A separation of sexes would introduce new problems as females may become acyclic

- It would be possible to test males and females on different days. However, if males and females undergo separate data collection there is no possibility of later combining data from both sexes (to increase sensitivity if no sex-differences are detected)
- Separation of the sexes is not needed since males are continuously housed in an environment where approximately 20-25% of the females would be in estrous
- Keeping males and females on separate racks and cleaning out test equipment between trials (standard lab practice) is enough to reduce the effects of female presence
- The endpoints of any test should be relevant and robust enough not to require such a level of control and effort to standardize the testing environment

Question for reviewers: Is consistency in results between males and females important when evaluating results or may differences between sexes be expected in certain cases?

Reviewer comments:
The majority of the reviewers agreed that for some behaviours/endpoints sex differences are very common, *e.g.* activity, flavour preference, certain types of operant behaviours, reproductive endpoints, haematology, and that therefore consistency between the sexes is not necessarily expected. Comments were made that consistency in results must be considered on a case-by-case basis and that "Consistency is only important if there is a biologically based rationale for expecting consistency."

Non-monotonic dose-response curves
The relevance of observed non-monotonic, *i.e.* (inverted) U-shaped, dose-response curves has been discussed in the area of endocrine disrupting compounds. In this investigation non-monotonic dose-response curves were observed for several aspects of male sexual behaviour in one of the studies on BPA (Jones *et al*., 2011). The US EPA guidance document on neurotoxicity risk assessment briefly states that both non-monotonic (biphasic) and monotonic dose-response curves are likely in developmental neurotoxicity data, depending on the function being tested, but does not provide further guidance on how to handle such data (US EPA, 1998b). This issue is also discussed by Tyl *et al*. (2008) who state that "an orderly trend to increasing effect with increasing dose, a basic tenet in pharmacology and toxicology… lends greater credence to concluding that there is a significant biological effect caused by exposure… The biological relevance of treatment effects involving nonlinear or multimodal dose-related responses should be evaluated within the context of the other factors, including the occurrence of other significant treatment related effect(s)". The OECD guidance documents reviewed here do not discuss the shape of dose-response curves.

Question for reviewers: How common is it to observe non-monotonic (U-shaped, biphasic) dose-response curves in DNT test results? If not common, is it because such data are systematically disregarded since they are not considered plausible?

Reviewer comments:
The reviewers disagreed slightly in regard to this question, the majority claiming that non-monotonic dose-response curves are not very common and some stating that they are not uncommon. Most reviewers did not comment on whether such data is systematically disregarded. It was recognized by one reviewer that it may be difficult to understand how large differences in exposure can result in such responses. A few stated that such data (if considered "robust") should at least give rise to some uncertainty.

A couple of the reviewers commented that non-monotonic dose-response curves may be hard to detect as it requires many dose groups (more than 3, which is required in TG 426).

Question for reviewers: What (if any) functions are the most likely to show non-monotonic dose-response curves?

Reviewer comments:
Only a few reviewers answered this question. Some answers seemed based on personal experience while others may be more theoretical.

Motor activity was indicated by a few reviewers as one function which may show non-monotonic dose-response. However, a couple of the reviewers commented that the response profile would likely be different at low compared to high doses, *e.g.* at low doses the effect may appear as an increase in activity while at high doses it may appear as an alteration in the habituation profile.

That auditory startle and learning in Y-maze may show non-monotonic dose-response were also brought up by individual reviewers.

One comment was that the occurrence of non-monotonic dose-response curves is due to the range of doses included rather than the function tested.

Question for reviewers: What are important factors to consider if non-monotonic dose-response curves are observed?

Reviewer comments:
The reviewers mentioned some considerations that should be made:

- If effects are similar in both sexes
- Kinetics
- Bioavailability
- Possible saturation of receptor binding capacity
- Compensatory mechanisms
- If data is consistent with results in related endpoints and previous toxicity data
- The biological plausibility
- Reliability of results from the control group
- Reliability of dosing and analysis
- Sufficiency of the dose range tested

A couple of the reviewers implied that a monotonic/traditional dose-response curve is not crucial for drawing conclusions about the neurotoxicity of the compound; "if any of

the doses induces deleterious effects the compound should be considered neurotoxic".

Flexibility of the TG 426

The TG 426 is quite flexible, both in terms of study design, including choice of animal strain, exposure route in pre-weaning pups (if direct exposure is deemed necessary) and choices of behavioural tests, and in the design of the individual tests themselves. In the guidance documents there are many different tests listed and described, but little guidance on how to choose the most appropriate/sensitive test (if there is no specific guidance in the published literature on the particular compound).

Raffaele *et al*. (2008) discuss some examples of critical test design features which are not specified in TG 426, for example:

- the delay between trials, *e.g.* inter-trial interval in auditory startle or learning paradigms
- stimulus strength, *e.g.* shock intensity or sound amplitude
- apparatus configuration, *e.g.* shape and size of apparatus used to measure motor activity
- number of trials per day
- the length of retention intervals in learning tasks
- definitions of learning or memory criteria, *e.g.* number of consecutive correct trials and maximum latency
- standardization procedures for handling of animals, *e.g.* acclimation time and placement in the device or maze

In these areas, discretion is given to the investigator with the purpose to enable the most appropriate study design for the substance being investigated. However, this flexibility also means possibly introducing different sources of variability in test results (Raffaele *et al*., 2008) and that there may be little conformity in DNT testing between and even within different laboratories.

Question for reviewers: Does the TG 426 need to be as flexible as it is, or could it be made more specific?

Reviewer comments:

All reviewers agreed that the TG must remain flexible as it has to cover a broad spectrum of compounds and that less flexibility could mean a reduced probability of identifying some specific neurotoxic effect. Flexibility also allows the labs to use the methods where they have the most expertise and experience to get reliable and robust data. The importance of experience in setting up and conducting tests as well as evaluating and interpreting results was highlighted.

Several reviewers suggested that the refinement and further development of guidance documents on DNT testing and assessment for testing laboratories as well as for authorities may be warranted. For example, guidance on different methodologies (including the sensitivity of different tests) and on the choice of parameters and tests when designing a DNT study, as well as important factors to consider when conducting studies and interpreting results.

Question for reviewers: What aspects (handling animals, animal storage conditions, choice of tests, how to perform tests, etc) would need harmonisation to decrease variability between laboratories?

Reviewer comments:
Reviewers' comments varied between "at least all of the above" to that TG 426 has shown generally very good interlaboratory agreement of results, implying that further harmonization is not needed.

It was recognized by one reviewer that it may be difficult to have standardized optimal procedures for different tests as each testing facility must know how to perform tests properly and should have reached the best conditions possible. One issue that was brought up was that it is more important to harmonize these aspects within laboratories, *i.e.* have standardized practices at the laboratory which are well recorded and reported. Another point was that "excessive standardization across laboratories may overestimate specific toxic effects. Reproducible findings across laboratories in spite of not completely harmonized conditions have more weight."

In addition to the above mentioned aspects other aspects that were discussed as important to the reliability/reproducibility of the DNT study were:

- Reporting of circadian period
- Balancing the testing of animals over the course of the day and across test devices
- Controlling test conditions
- Reporting the age of the animals
- Harmonization of dosing schedules
- Labs should be certified
- Well trained staff at all levels of testing as well as data evaluation and interpretation
- Availability of positive/historical controls
- Labs should operate according to standard operating procedures
- Use of sufficient sample sizes
- Application of generally accepted statistical methods
- Use blinded observers to record data

Screening for neurotoxicity
By including certain tests in general toxicity studies, *e.g.* 28-day or 90-day studies, indications of neurotoxic properties of a substance can be obtained from non-neurotoxicity studies. The term "Functional Observational Battery" (FOB) seems to be generally used by the US EPA while OECD guidelines refer to "functional observations". However, the endpoints evaluated cover much the same functional domains. The functional observations included in OECD guidelines are generally less defined but should include measurements of sensory reactivity to stimuli of different types, *e.g.* auditory, visual and proprioceptive stimuli, as well as assessment of grip strength and motor activity.

Question for reviewers: Are the endpoints/tests recommended in the FOB appropriate and sensitive enough to be used as a screening tool for (developmental) neurotoxicity?

Reviewer comments:
About half of the reviewers that commented on this question stated that the FOB is appropriate as a screening tool while the rest argued that it is not enough on its own and should be complemented with other tests, *e.g.* startle testing and motor activity. One comment was that the FOB is sufficient if high enough doses are used.

A couple of the reviewers noted that a lack of effect in the FOB does not mean that there are no DNT effects, and that results may be different under different conditions and more subtle effects may become apparent in other tests.

Question for reviewers: Is motor activity a sensitive and non-specific endpoint that could be useful for screening purposes?

Reviewer comments:
Also here reviewers' opinions diverged. Some reviewers stated that it is not; as motor activity is an apical endpoint and is also very dependent on systemic toxicity it must be evaluated in relation to other endpoints. Conversely, some reviewers argued that it would be a good screening test *because* of its apical nature.

One reviewer commented that motor activity can be a screening tool if conducted correctly, *i.e.* you need to control for stress and anxiety by conducting the test in a novel home environment and not in the Open Field.

It was also noted that "according to the experience of the reviewer, frequently motor activity appears to be the only and most frequently observed altered end-point, despite the fact that it has a significant inter and intra-individual variability."

Positive and historical controls
Both positive and historical controls are important features in the evaluation of DNT data. Historical control data are important in evaluating the performance and variance in the concurrent negative controls in the DNT study as well as in interpreting statistically significant effects in treated animals (Tyl *et al*., 2008). For example, in some of the studies of the model compounds in this report, statistically significant effects were dismissed as not being treatment related because the effects observed were within the normal range of historical controls (*e.g.* Butenhoff *et al*., 2009; Stump *et al*., 2010; Beck 2009/Biesemeier *et al*., 2011). Positive control data are argued as being crucial in determining the proficiency of a laboratory in detecting chemically-induced changes in measured endpoints and are also valuable when identifying and interpreting the significance of effects observed in DNT tests (Crofton *et al*, 2008).

The US EPA DNT guideline states that "positive control data from the laboratory performing the test that demonstrate the sensitivity of the procedures being used" should be provided, with no mention of historical control data (US EPA, 1998a), while the OECD TG 426 states that "available positive and historical control data should be discussed, especially when there are no treatment-related effects" (OECD, 2007).

It should be noted that it is not possible to identify a common positive control chemical that demonstrates effects in all endpoints included in the DNT study. In the guidance documents from US EPA (1998b) and OECD (2004) tables of positive controls for different functions tested are included. In addition, a thorough review of critical aspects in conducting positive control studies and which provides "a context for the use, interpretation and reporting of positive control studies for regulatory developmental neurotoxicity testing" is available (Crofton *et al*., 2008).

A few of the studies reviewed in this report have used or referred to studies on a "positive control" with the purpose of justifying the sensitivity of the experimental design and the laboratory. 17-beta-Ethinyl estradiol has been used as a positive control for BPA and chlorpyrifos has been used to justify the design and laboratory of the Beck study (on deltamethrin).

Question for reviewers: Are there any neurotoxic substances that could be used as positive controls particularly suitable for the identification of developmental neurotoxicity (for certain groups of compounds or more generally)?

Reviewer comments:
Reviewers generally confirmed that there is no single compound that should be included as a concurrent positive control, but that there are specific chemicals that could be used as positive controls for specific endpoints.

It was suggested that a concurrent positive control could potentially be included in a study for the testing of certain endpoints but that this is hardly necessary. Two different reasons were given: 1) the comparison to the negative control should be enough or 2) concurrent positive control is not necessary as there should be historical positive control data available at the lab. Also the arguments of increased costs and animal use were given against the inclusion of positive controls.

Question for reviewers: To what extent does the absence of adequate positive and historical control data limit the reliability of studies conducted at laboratories that do not have these resources?

Reviewer comments:
Reviewers disagreed on this point. Some reviewers stated that the lack of historical and positive control data reduces the reliability of the study and increases the uncertainty of the results and that positive control data are important to show that the lab is carrying out the test correctly and can get correct results.

The importance of positive and historical control data was specifically highlighted for non-standardized (*i.e.* non-guideline) studies, stating that "Though some laboratories may not have the resources to generate positive and historical control data, all laboratories have the capability of maintaining accurate data logs and archiving raw data. Therefore, a simple solution is to require all laboratories to make their underlying raw data available to regulatory agencies for independent review and evaluation."

Other reviewers commented that positive and historical control data are not necessary for the reliability of the study, if it is properly performed. One comment was specifically that historical controls are not useful since results must be compared to controls tested under the exact same conditions and these vary over time. However, it was acknowledged that historical controls are useful to detect if the study "has had some unexpected mistake or some problem with the animals" and when concurrent controls are significantly different from historical controls.

A few comment that even though the lack of such control data may limit the reliability of the study it does not invalidate it if it is otherwise well performed.

Question for reviewers: Is it always relevant to disregard observed effects based on the argument that they are within the normal range of historical controls?

Reviewer comments:
The reviewers were in general agreement that this argument is not enough to disregard observed effects but rather that all data has to be evaluated carefully and effects within the range of historical controls should be used as "a piece of the puzzle", *i.e.* the investigator must also consider data of other endpoints, the "constellation of differences" as well as dose-response relationships and standard deviations. One comment was also that interpretation should depend on the "quality" of the historical controls, *e.g.* "consistent results over time strengthens an argument that results for treated animals within the range of historical controls are not due to treatment".

Another comment was that the concurrent control should be the most important as those animals have been tested under the same environmental conditions.

Human relevance
At a workshop held by the US EPA in 1989 the qualitative and quantitative comparability of animal and human DNT data was evaluated by experts from government, industry, academia and public interest groups (Francis *et al*., 1990). A number of known developmental neurotoxicants were used as model compounds. The conclusion from this workshop was that the then available protocol for evaluating DNT would have identified each of the model compounds as potential developmental neurotoxicants but would probably have underestimated human risk (Stanton and Spear, 1990).

Based on the conclusions from the Williamsburg workshop the US EPA Guidelines for Neurotoxicity Risk Assessment (US EPA, 1998) state that even though direct extrapolation of developmental neurotoxicity in animals to humans is limited many similarities in effects have been shown when comparing animal and human developmental neurotoxicity data. Thus, human relevance of developmental neurotoxicity observed in animal studies can be assumed.

Research, development and training needs

Question for reviewers: Which are the major research and development needs for improving the regulatory tools for developmental neurotoxicity testing?

Reviewer comments:
Most reviewers brought up several issues that, in their views, could improve DNT testing and evaluation, while some felt that there is no need to further improve the DNT testing beyond the TG 426 at this time.

Most reviewers mentioned the complexity of the topic and the need of expertise. Specific comments included eg.:

- Improving the knowledge in the area of developmental neurotoxicity in general, especially the understanding of mechanisms modulating different behavioural functions
- Introducing a more precautious approach in the interpretation of negative results
- Encouraging the practice of considering the individual data (distribution), not just mean values when evaluating data for risk assessment

Most reviewers mentioned the need to generate new data to fill knowledge gaps and better understand the significance of testing observations. Specific comments included e.g.:

- Increasing, in general, the scientific quality of the research and including more scientific knowledge in the DNT studies
- Conducting more studies of the reversibility of behavioural alterations and underlying mechanisms
- Development of automated tools for tests of *e.g.* learning and memory, motor coordination and circadian rhythms
- Further exploration of several of the questions in this review
- Development of defined protocols to be applied for specific needs, appropriate for the intended use of the data generated (modelling, risk assessment, etc)
- Conducting toxicokinetic investigations so that kinetics are understood
- Conducting studies on more test compounds

Some reviewers mentioned the need of continuous procedures for revision of DNT test guidelines, such as OECD TG 426; as well as of corresponding guidance revision documents. Specific comments included e.g:

- Making sure that if new methodologies are incorporated in TG 426 these must first be validated across laboratories
- Workshop format is useful to share knowledge and harmonize opinions between different categories of stakeholders

Taken together, the reviewer answers to all the questions of this report demonstrated a general agreement on the following issues in regard to DNT testing:

- When exposure is via the mother the statistical unit should be the litter and not the individual pup when analysing test results.
- The exposure should include both the pre- and postnatal periods if the mechanism of action of the substance under investigation and sensitive windows for the onset of effects are not known
- Investigation of kinetics is useful, but usually not feasible to include in the DNT study design

- The choice of test method for investigating a certain function can influence the study outcome, especially in learning and memory tests
- No test method for a certain endpoint is the most sensitive and appropriate for all compounds
- The functional and behavioural tests that can be included in the study according to TG 426 are appropriate for testing of young animals
- Timing of testing has a large influence on the ability to identify effects, especially in developing animals
- The TG 426 must remain flexible
- There are reports on difference in sensitivity among rodent species, strains and sexes.
- No single compound can be included as a concurrent positive control, but there are specific chemicals that can be used for specific endpoints
- A high level of expertise is necessary to design and conduct the study as well as to evaluate the analytical data and interpret the results
- Training needs are substantial within the field of developmental neurotoxicity and extends from detailed knowledge about study design and performance, to data evaluation and interpretation among several categories of professionals.
- Research needs to fill knowledge gaps both in neurodevelopment and neurotoxicology are large, and so is the need to transfer and communicate the new knowledge to the different categories of professional end-users.

On the other hand, the reviewers did not agree on some other issues, which were brought up in the questions, such as:

- If automated devices/methods or human observers are most appropriate for measuring motor activity and if motor function also can be measured in tests for motor activity.
- The value of adding requirements for additional tests in TG 426 for *e.g.* anxiety and social/sexual behaviour to increase the sensitivity of the study.
- The benefit of using a pharmacological or physiological challenge to “unmask” DNT effects that may elude the general test paradigm of TG 426.
- The importance of controlling for estrous cycle when conducting behavioural tests in adult females.
- If it is common to observe non-monotonic dose-response relationships in DNT test results.
- Whether or not increased harmonization between laboratories is needed to increase reliability of the DNT study and, if so, what aspects would require further harmonization.
- If the FOB or tests for motor activity are appropriate and sensitive enough to be used as screening tools for developmental neurotoxicity.
- The importance of historical and positive control data for the interpretation and evaluation of DNT study results.
- Whether the pup or the litter is the appropriate unit to analyse when pups are only exposed directly (and not in utero via the mother)

5. Conclusions

The areas of neurodevelopment and neurotoxicity are inherently very complex. This complicates both DNT testing and risk assessment of compounds. Also, it makes defining strict criteria for testing and risk assessment difficult.

There is considerable flexibility in TG 426 concerning the study design, such as the choice of behavioural tests included in the study, and also the design of those individual tests, *e.g.* size and shape of testing apparatus, strength of stimulus, intervals between testing trials and sessions and number of trials per day. Given the inherent complexity of DNT it can be argued that this flexibility is necessary, that strict criteria for *e.g.* study design aspects are in fact not desired, and that it must be up to expert judgment of the investigator to design the most sensitive and appropriate study relevant for the exposure and toxicity of the compound under investigation. The demands on the proper expertise of the investigator/investigating laboratory are thus very large. On the other hand, it can also be argued that the flexibility of the DNT guideline makes it possible to deliberately design a negative study. In any case, it should be noted that leaving the decision concerning important study design aspects to the discretion of the investigator introduces potential sources of variability in DNT study results (Raffaele *et al.,* 2008). Consequently, reporting of DNT studies conducted for regulatory testing, as well as basic research studies within this field, requires a high level of transparency and detail concerning the study design and methods used.

It is clear from the literature reviewed for the purposes of this investigation as well as from the comments received from reviewers that expert judgment makes up an integral part of DNT testing and risk assessment. A substantial amount of expertise within the field of developmental neurotoxicity is required when designing studies and interpreting their results, as well as when evaluating the reliability and relevance of DNT data for risk assessment.

To ensure proper use and the reliability of the DNT study it is evident that good guidance documents are essential to assist investigators both within research and regulatory testing in the designing of DNT studies as well as for both investigators and regulatory authorities in the interpretation of study results. However, the present review of the OECD guidance documents for Neurotoxicity Testing number 20 (2004) and Mammalian Reproductive Toxicity Testing and Assessment number 43 (2008) as well as the US EPA Guidelines for Neurotoxicity Risk Assessment (1998b) shows that these documents actually provide little structured guidance on how to interpret the results of the different DNT tests. Feedback from the external reviewers of this report indicates that further development and improvement of guidance documents is warranted on a periodic basis as new knowledge becomes available. A good basis to further develop and improve testing guidelines and guidance documents is the work conducted within the ILSI expert working group on neurodevelopmental endpoints (Tyl *et al*., 2008). The reviews published from those activities provide detailed discussions and guidance on important issues for the interpretation of DNT study data (Tyl *et al*., 2008), the variability in such data (Raffaele *et al*., 2008), as well as the use of positive controls (Crofton *et al*., 2008) and appropriate statistical

techniques (Holson *et al*., 2008) in DNT-testing. These publications provide concrete guidance on several aspects from designing the study to interpretation of results.

6. Recommendations for the future

It is clear from the compound evaluation work and the reviewer survey of this investigation that the areas of neurodevelopment and neurotoxicology are inherently very complex, and, in particular, there are massive gaps in knowledge about normal brain development on the functional, structural and molecular levels, which complicates both neurotoxicity testing and data interpretation. Consequently, safety and health risk assessment of neuromodulatory compounds also becomes complicated.

Decreasing the flexibility of TG 426 in order to make testing more standardized or to facilitate the evaluation of study results is not considered the right way forward. Instead, efforts to further develop and improve guidance documents provided for testing laboratories and authorities to ensure reliable and relevant testing and assessment of DNT are urgently needed in order to ensure that the health protection aims of chemicals legislation are fulfilled. It is also important to make testing and guidance documents well known among scientists in general in the fields of neurodevelopment and neurotoxicity.

Research to decrease knowledge gaps, training to assure access to the right expertise in technical as well as scientific matters, including evaluation, interpretation and judgement abilities, as well as procedural efforts to allow for knowledge exchange and periodic revision of regulatory guidelines and guidance documents as new knowledge becomes available and adopted, are all important contributors to such efforts. It is important that efforts along these lines, which will benefit from comprehensive and open-minded data-sharing, will take place on a periodic basis as new knowledge becomes available. It is also important that such efforts will involve all relevant stake-holders, as well as experts in the fields of neurodevelopment, neurotoxicity, and health risk assessment sciences. The ILSI expert working group on neurodevelopmental endpoints is one good example of such a working model in the DNT field, which could be acting on a periodic basis.

8. References

Beck MJ (2009) An oral (gavage) developmental neurotoxicity study of decabromodiphenyl oxide in rats. WIL Research Laboratories, LLC.

Biesemeier JA, Beck MJ, Silberberg H, Myers NR, Ariano JM, Radovsky A, Freshwater L,, Sved DW, Jacob SJ, Stump DG, Hardy ML, Stedeford T (2011) An oral developmental neurotoxicity study of decabromodiphenyl ether (DecaBDE) in rats. Birth Defects Res (Part B) 92:17-35.

Butenhoff JL, Ehresman DJ, Chang S-C, parker GA, Stump DG (2009) Gestational and lactational exposure to potassium perfluorooctanesulfonate (K^+PFOS) in rats; Developmental neurotoxicity. Reprod Toxicol 27: 319-330.

Crofton K, Foss JA, Hass U, Jensen KF, Levin ED, Parker SP (2008) Undertaking positive control studies as part of delvelopmental neurotoxicity testing. A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:266-287.

Crofton KM, Mundy WR, Lein PJ, Bal-Price A, Coecke S, Seiler AE, Knaut H, Buzanska L, Goldberg A (2011) Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. ALTEX 28:9-15.

Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazono A (2001) Rat two-generation reproductive toxicity study of bisphenol A. Reprod Toxicol 15:505-523.

Eriksson P, Fredriksson A (1991) Neurotoxic effects of two different pyrethroids, bioallethrin and deltamethrin, on immature and adult mice: changes in behavioral and muscarinic receptor variables. Toxicol Appl Pharmacol 108:78-85.

Fischer C, Fredriksson A, Eriksson P (2008) Neonatal co-exposure to low doses of an ortho-PCB (PCB 153) and methyl mercury exacerbate defective developmental neurobehavior in mice. Toxicology, 244:157-65.

Fitzpatrick J, Mendez E, Walls I (2008) Introduction to the ILSI Research Foundation/Risk Science Institute reports from the expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:263-265.

Gilmore RG *et al.* (2006) A developmental neurotoxicity screening study with technical grade deltamethrin in Wistar rats. Study summary, Document III-A, Bayer CropScience AG.

Grandjean P and Landrigan PJ (2006) Developmental neurotoxicity of industrial chemicals. Lancet 368:2167-2178.

Holson RR, Freshwater L, Maurissen JPJ, Moser VC, Phang W (2008) Statistical issues and techniques appropriate for developmental neurotoxicity testing. A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:326-348.

Johansson N, Viberg H, Fredriksson A, Eriksson P (2008a) Neonatal exposure to deca-brominated diphenyl ether (PBDE 209) causes dose-response changes in spontaneous behaviour and cholinergic susceptibility in adult mice. NeuroToxicology 29:911-919.

Johansson N, Fredriksson A, Eriksson P (2008b) Neonatal exposure to perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) causes neurobehavioral defects in adult mice. Neurotoxicol 29:160-169.

Jones BA, Shimell JJ, Watson NV (2011) Pre- and postnatal Bisphenol A treatment results in persistent deficits in the sexual behavior of male rats, but not female rats, in adulthood. Horm Behav 59:246-51.

Makris SL, Raffaele K, Allen K, Bowers WJ, Hass U, Alleva E, Calamandrei G, Sheets L, Amcoff P, Delrue N, Crofton KM (2009) A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. Environ Health Perspect 117:17-25.

Negishi T, Kawasaki K, Takatori A, Ishii Y, Kyuwa S, Kuroda Y, Yoshikawa Y (2003). Effects of perinatal exposure to bisphenol A on the behavior of offspring in F344 rats. Environ Toxicol Pharmacol 14:99-108.

OECD Guideline for the testing of chemicals (1997) Neurotoxicity study in rodents, TG 424.

OECD Guideline for the testing of chemicals (2007) Developmental Neurotoxicity study, TG 426.

OECD Series on testing and assessment, Number 20 (2004) Guidance document for neurotoxicity testing (ENV/JM/MONO(2004)25).

OECD Series on testing and assessment, Number 43 (2008) Guidance document on mammalian reproductive toxicity testing and assessment (ENV/JM/MONO(2008)16).

Onishchenko N, Fischer C, Wan Ibrahim WN, Negri S, Spulber S, Cottica D, Ceccatelli S (2011) Prenatal exposure to PFOS or PFOA alters motor function in mice in a sex-related manner. Neurotox Res. 19:452-461.

Piedrafita B, Erceg S, Cauli O, Monfort P, Felipo V (2008) Developmental exposure to polychlorinated biphenyls PCB153 or PCB126 impairs learning ability in young but not in adult rats. Eur J Neurosci. 27:177-82.

Raffaele K, Fisher Jr JE, Hancock S, Hazelden K, Sobrian SK (2008) Determining normal variability in a developmental neurotoxicity test. A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:288-325.

Raffaele K, Rowland J, May B, Makris S, Schumacher K, Scarano LJ (2010) The use of developmental neurotoxicity data in pesticide risk assessments. Neurotoxicol Teratol. 32:563-572.

Rice DC, Reeve EA, Herlihy A, Zoeller RT, Thompson WD, Markowski VP (2007) Developmental delays and locomotor activity in the C57BL6/J mouse following neonatal exposure to the fully-brominated PBDE, decabromodiphenyl ether. Neurotoxicol Teratol. 29:511-20.

Rice DC, Thompson WD, Reeve EA, Onos KD, Assadollahzadeh M, Markowski VP (2009) Behavioral changes in aging but not young mice after neonatal exposure to the polybrominated flame retardant decaBDE. Environ Health Perspect. 117:1903-11.

Schantz SL, Moshtaghian J, Ness DK. (1995) Spatial learning deficits in adult rats exposed to ortho-substituted PCB congeners during gestation and lactation. Fundam Appl Toxicol 26: 117-126.

Stump DG, Beck MJ, Radovsky A, Garman RH, Freshwater LL, Sheets LP, Marty MS, Waechter, Jr. JM, Dimond SS, Van Miller JP, Shiotsuka RN, Beyer D, Chappelle AH, Hentges SG (2010) Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. Toxicol Sci 115:167-182.

Tyl RW, Crofton K, Moretto A, Moser V, Sheets LP, Sobotka TJ (2008) Identification and interpretation of developmental neurotoxicity effects. A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:349-381.

US EPA (1998a) US Environmental Protection Agency Health Effects Test Guidelines. OPPTS 870.6300. Developmental Neurotoxicity Study.

US EPA (1998b) Guidelines for Neurotoxicity Risk Assessment (EPA/630/R-95/001F).

Viberg H, Fredriksson A, Jakobsson E, Örn U, Eriksson P (2003) Neurobehavioral derangements in adult mice receiving decabrominated diphenyl ether (PBDE 209) during a defined period of neonatal brain development. Toxicol Sci 76:112-120.

Viberg H, Fredriksson A, Eriksson P (2004) Investigations of strain and/or gender differences in developmental neurotoxic effects of polybrominated diphenyl ethers in mice. Toxicol Sci 81:344-353.

Viberg H, Fredriksson A, Eriksson P (2007) Changes in spontaneous behaviour and altered response to nicotine in the adult rat after neonatal exposure to the brominated flame retardant, decabrominated diphenyl ether (PBDE 209). NeuroToxicology 28:136-142.

9. Tables

Table 1. A summary of the requirements and recommendations stated in OECD TG 426 compared to the functional observations generally included in other OECD test guidelines.

	OECD 426	**General functional observations**
AIM	Designed to analyse the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	included to detect any indications of neurotoxicity
Recommendations for use	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types	-
Experimental design		
A. Test animals		
	Rat (commonly used strains)	-
	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	-
	Supplier of test animals to be provided	-
	Number, age at start and sex of animals	-
	Housing conditions, acclimatization etc.	-
	Unique identification for each animal and litters	-
	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	-
	Live pups to be counted and sexed	-

	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	-
	Pup identification is required	-
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	-
B. Test conditions		
1. Administration of chemical/dosing	Most relevant to potential human exposure	-
	Oral by gavage, in diet, drinking water or capsules	-
	Other forms of admin (inhalation or dermal) requires modification of procedures	-
	At least 3 dose groups and a concurrent control	-
	Repeated exposure	-
	Mated females, starting on GD 6	-
	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	-
	High dose level should induce some maternal toxicity (*e.g.* weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	-
	Dose levels should be selected to allow for illustration of dose-response	-
	Positive controls not mentioned	-
2. Duration of exposure *(Add information if any comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure via milk)*	GD6 to PND21	-
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	-

3. If other routes of administration than oral	Use OECD Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment #43 to assist in the design of the studies	-
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at study termination.	-
5. Food and water consumption	Food consumption should be measured weekly at a minimum during gestation and lactation. Water consumption should be measured at least weekly if exposure is via the water supply.	-
6. Control for estrous cyclicity in females	*Not specified if required*	**-**
C. Endpoints:		
1. Physical and developmental landmarks		
a. Body weight and clinical observations	Required	-
	Weekly during pre-weaning and at least every two weeks at adolescence and adulthood. Detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	-
	Where possible, observations to be made by the same technician	-
	Observations to be performed outside of home cage	-
	"Normal" range needs to be documented and a well-defined clinical observation criteria should be followed (some examples are given).	-
	"Unusual" responses with respect to *e.g.* activity level, need to be documented	-
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development. Body weight should be measured weekly during pre-weaning and at least every two weeks at adolescence and adulthood.	-

b. Brain weight and Neuropathology	Required, on PND 22 (or earlier between PND 11 and 22) and at termination	-
	Gross changes and lesions identified and scored	-
	Fixation; Immersion <PND21, Perfusion >PND21, Preparation is critical. Require representative sections of the central and peripheral nervous system. Areas to be examined listed.	-
	Tissue selection need both CNS and PNS	-
	General staining (H&E) should be carried out for animals at PND 22 or earlier, otherwise myelin and silver stains are recommended for CNS and PNS sections	-
	Evaluation to be conducted by trained pathologist. Stepwise procedure is recommended for qualitative and quantitative neuropathological analyses.	-
	Morphometric evaluation of tissue collected on PND 21 and at end of study	-
	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	-
c. Sexual maturation	At adolescence (as appropriate)	-
d. Other physical landmarks for pup development (eg eye opening)	only required if these will provide additional information	-
2. Functional/behavioural endpoints		
a. Behavioural ontogeny	Required, at least two measures pre-weaning using the same pup (1 pup/sex/litter), *e.g.* righting reflex, negative geotaxis, motor activity (strongly recommended)	-
b. Motor activity	Required, including habituation, at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Included
	Motor activity should be measured with an automated device that can measure both decreases and increases in activity	-

c. Motor and sensory function	Required, at adolescence (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider: Grip Strength, rotating rod, hind limb foot splay, nociception, sensory irritation, somatosensory operant discrimination task, acoustic startle, auditory discrimination, visual discrimination	Included, *e.g.* auditory, visual and proprioceptive stimuli
d. Grip strength	Not specifically required other than as a potential test for motor function	Included
e. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Required, post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider: Conditioned taste aversion, active or passive avoidance, spatial mazes, conditional discrimination, delayed discrimination, eye-blink conditioning, schedule-controlled operant behaviour	-
	<u>Two criteria</u> need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	-
3. Other endpoints		
a. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not required*	-
b. Neurochemical	Optional	-
c. Electrophysiological	Optional	-

Table 2. Summary of US EPA guidance on animal neurotoxicological studies, their endpoints and interpretation (US 1998b)

Endpoint category	Type of effect	Tests/Methods	Guidance for interpreting data
Structural/ neuropathological	Gross changes in morphology, *e.g.* lesions and changes in brain weight Histologic changes in neurons or glia	Gross observations and light microscopy.	Changes in brain weight are a more reliable indicator of alteration in brain structure than are measurements of length or width in fresh brain. It is inappropriate to express brain weight changes as a ratio to body weight. Alterations in the structure of the nervous system are regarded as evidence of a neurotoxic effect. In many cases, pathological changes require time for the perturbation to become observable, especially with evaluation at the light microscopic level. Neuropathological studies should control for potential differences in the area(s) and section(s) of the nervous system sampled; in the age, sex, and body weight of the subject; and in fixation artifacts. Various histological changes can result after exposure to neurotoxicants. Specific changes in nerve cell bodies include chromatolysis, vacuolization, and cell death. Axons can undergo swelling, degeneration, and atrophy, while myelin sheath changes include folding, edematous splitting, and demyelination.
Behavioural/ neurological	Increases or decreases in motor activity	Frequency of movement over a period of time.	Changes expressed as absolute activity counts or as percent of control values. The frequency of motor activity within a session usually decreases and is reported as the average number of counts occurring in each successive block of time. The EPA's Office of Prevention, Pesticides and Toxic Substances guidelines (U.S. EPA, 1991), for example, call for test sessions of sufficient duration to allow motor activity to approach steady-state levels during the last 20 percent of the session for control animals. Both increases and decreases in activity are possible. Motor activity may also be altered by experimental factors other than chemical exposure (no examples given). Both transient and persistent increases in motor activity are possible. Changes in motor activity associated with other overt signs of toxicity or occurring in non-dose-related fashion are of less concern than changes that are dose dependent, are related to structural or other functional changes in the nervous system, or occur in the absence of life-threatening toxicity.
	Changes in touch, sight, sound, taste, or smell	Tests for sensory function, *e.g.* hot plate test, tail flick test, auditory reflex	"Gross perturbations of sensory function can be observed in simple neurological assessments such as the hot plate or tail flick test. However, these tests may not be sufficiently sensitive to detect subtle sensory changes.

	sensations Absence or decreased occurrence, magnitude, or latency of sensorimotor reflex	modification or auditory, visual, somatosensory, pain sensitivity and olfactory discrimination conditioning.	Psychophysical procedures that study the relationship between a physical dimension (*e.g.*, intensity, frequency) of a stimulus and behavior may be necessary to quantify agent-induced alterations in sensory function. Examples of psychophysical procedures include discriminated conditioning and startle reflex modification."
	Changes in motor coordination, weakness, paralysis, abnormal movement or posture, tremor, ongoing performance Altered magnitude of neurological measurement, including grip strength, hind limb splay	Tests for motor function, *e.g.* Weakness: grip strength, swimming endurance, suspension rod, discriminative motor function Incoordination: rotorod, gait assessment, righting reflex Tremor: rating scale, spectral analysis Myoclonic spasms: rating scale	*No guidance on the interpretation of test results provided.*
	Seizures	Observations	"Observable convulsions in animals are indicative of an adverse effect. These events can reflect central nervous system activity comparable to that of epilepsy in humans and could be defined as neurotoxicity. Occasionally, other toxic actions of compounds, such as direct effects on muscle, might mimic some convulsionlike behaviors. In some cases, convulsions or convulsionlike behaviors may be observed in animals that are otherwise severely compromised, moribund, or near death. In such cases, convulsions might reflect an indirect effect of systemic toxicity and are less clearly indicative of neurotoxicity. As discussed in the section on neurophysiological measures, electrical recordings of brain activity could be used to determine specificity of effects on the nervous system."
	Changes in rate or temporal patterning of schedule-controlled operant behaviour (SCOB)	Lever-press or key-peck response	SCOB provides a measure of performance of a learned behaviour and involves training and motivational variables that should be considered in evaluating the data. The primary SCOB endpoints for evaluation are response rate and the temporal pattern of responding. Changes in SCOB may be due to effects on sensory processing, motor output, motivational, training history, and baseline characteristics. Substantial qualitative changes

			in operant performance, such as elimination of characteristic response patterns, can be evidence of an adverse effect. Small quantitative changes are not adverse. Assessing the toxicological importance of these effects requires considerable professional judgment and evaluation of converging evidence from other types of toxicological endpoints. Some agents may increase response efficiency on schedules requiring high response rates because of a stimulant effect or an increase in central nervous system excitability. Agent-induced changes in response rate or temporal patterning associated with other overt signs of toxicity are of less concern than changes that are dose dependent, related to structural or other functional changes in the nervous system, or occur in the absence of life-threatening toxicity.
	Changes in learning, memory, and attention	Tests for cognitive function, *e.g.* *Habituation*: startle reflex *Classical conditioning*: nictitating membrane, conditioned flavor aversion, passive avoidance, olfactory conditioning *Instrumental conditioning*: one-way avoidance, two-way avoidance, Y-maze avoidance, Biel water maze, Morris water maze, radial arm maze, delayed matching to sample, repeated acquisition	Alterations in learning and memory should be compared with that seen prior to exposure or with a non-exposed control group. Learning is defined as a relatively lasting change in behaviour due to experience, and memory is defined as the persistence of a learned behaviour over time. Measurement of changes in learning and memory should be separated from other changes in behaviour that do not involve cognitive or associative processes. Any apparent toxicant-induced change in learning or memory should ideally be demonstrated over a range of stimulus and response conditions and testing conditions. Older animals frequently perform poorly on some types of tests, and it should be demonstrated that control animals in this population are capable of performing the procedure. Apparent improvement in performance is not either adverse or beneficial until demonstrated to be so by converging evidence with a variety of experimental methods.
Developmental neurotoxicity	Chemically induced changes in the time of appearance of behaviours during development Chemically induced changes in the growth or organization of structural or neurochemical	*The various tests mentioned for the other endpoint categories apply*	There are particular issues of importance in the evaluation of developmental neurotoxicity studies: Many known developmental neurotoxicants cause functional deficits at dose levels below those that are toxic in adults. Such effects may be transient, but generally are considered adverse. Developmental exposure to a chemical could result in transient or reversible effects observed during early development that could reemerge as the individual ages. Important study design issues include having enough litters for adequate

	elements		statistical power, randomization of animals to dose groups and test groups and using the litter as the statistical unit. A pharmacological or physiological challenge may be valuable in evaluating neurological function and “unmasking” effects not otherwise detectable. A battery of functional tests, in contrast to a single test, is usually needed to evaluate the full complement of nervous system functions in an animal. There are critical developmental periods for the disruption of functional competence and the effect of a toxicant is likely to vary depending on the time and degree of exposure. It is also important to consider the data from studies in which postnatal exposure is included, as there may be an interaction of the agent with maternal behaviour, milk composition, or pup suckling behaviour, as well as possible direct exposure of pups via dosed food or water. Agents that produce developmental neurotoxicity at a dose that is not toxic to the maternal animal are of special concern. At doses causing moderate maternal toxicity (*i.e.*, 20% or more reduction in weight gain during gestation and lactation), interpretation of developmental effects may be confounded. Whether developmental effects are secondary to maternal toxicity or not, the maternal effects may be reversible while the effects on the offspring may be permanent. Functional effects should be evaluated in light of other toxicity data, including other forms of developmental toxicity. In the case of developmental neurotoxicity both monotonic and biphasic dose-response curves are likely, depending on the function being tested.

Table 3. Summary of the aims and designs of the selected developmental neurotoxicity studies of BPA.

Study	**Stump *et al*., 2010**	**Jones *et al*., 2011**	**Negishi *et al*., 2003**	**Ema *et al*., 2001**
Full reference	Stump DG, Beck MJ, Radovsky A, Garman RH, Freshwater LL, Sheets LP, *et al*. (2010) Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. Toxicol Sci 115(1): 167-182.	Jones BA, Shimell JJ, Watson NV. (2011) Pre- and postnatal Bisphenol A treatment results in persistent deficits in the sexual behavior of male rats, but not female rats, in adulthood. Horm Behav 59: 246-251.	Negishi T, Kawasaki K, Takatori A, Ishii Y, Kyuwa S, Kuroda Y, *et al*. (2003) Effects of perinatal exposure to bisphenol A on the behavior of offspring in F344 rats. Environ Toxicol Pharmacol 14: 99-108.	Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazono A. (2001) Rat two-generation reproductive toxicity study of bisphenol A. Reprod Toxicol 15(5): 505-523.
Aim	To determine the potential of BPA, administered in feed to Sprague-Dawley rats, to induce functional and/or morphological effects in the nervous system that may arise in the offspring from exposure to the mother during pregnancy and lactation	To address three questions: Does chronic BPA treatment during the perinatal period alter sexual behaviour in adulthood? Is this effect dose-dependent? Does sexual experience mitigate any initial deficits in sexual performance?	To determine whether perinatal maternal exposure to BPA affects the behaviour of offspring	To determine the low-dose effects of bisphenol A in a rat two-generation reproduction study.
Conducted in accordance with guidelines?	Yes OECD TG 426 US EPA OPPTS 870.6300	No	No	Yes OECD TG 416 Ministry of Agriculture, Forestry and Fisheries of Japan Guidance for Application of Agriculture Chemical Registration ICH Guideline for Detection of Toxicity to Reproduction for Medicinal Products
Animals	Rat (Sprague Dawley)	Rat (Long-Evans)	Rat (F344)	Rat (Crj: CD (SD) IGS)
Exposure route	In diet to mated females	In corn oil to mated females, allowed to drink spontaneously	In olive oil to mated females, not stated if by gavage (but assumed)	In distilled water by gastric intubation to males and females starting before mating and continued

				through gestation and lactation
Doses (mg/kg bw/day)	*During gestation:* 0, 0.01, 0.12, 5.85, 56.4, 164 *During lactation:* 0, 0.03, 0.25, 13.1, 129, 410	0, 0.005, 0.05, 0.5, 5	0, 4, 40, 400	0, 0.0002,0.002, 0.02, 0.2
Exposure duration	GD 0 to PND 21	GD 7 to PND 14	GD 10 to PND 20	F0 from before mating through gestation and lactation. Offspring via gastric intubation after weaning.
Functional/behavioural endpoints				
Behavioural ontogeny	**X**	--	**--**	**X**
Motor activity	**X**	--	**X**	**X**
Motor and Sensory function	**X**	--	--	--
Learning and memory	**X**	--	**X**	**X**
Other neurobehavioural	--	**X** (sexual behaviour)	**X** (anxiety)	**X** (anxiety)
Evidence of developmental neurotoxicity?	No	Yes	Yes	No

X endpoint evaluated
-- endpoint not included

Table 4. Comparison of experimental design and endpoints in the OECD TG 426, the BPA study according to TG 426 (Stump *et al.,* 2010) with some selected "low-dose" developmental neurotoxicity studies of BPA.

Reference	**OECD 426**	**Stump *et al*., 2010**	**Jones *et al*., 2011**	**Negishi *et al*., 2003**	**Ema *et al*., 2001**
		Stump DG, Beck MJ, Radovsky A, Garman RH, Freshwater LL, Sheets LP, *et al*. (2010) Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. Toxicol Sci 115(1): 167-182.	Jones BA, Shimell JJ, Watson NV. (2011) Pre- and postnatal Bisphenol A treatment results in persistent deficits in the sexual behavior of male rats, but not female rats, in adulthood. Horm Behav 59: 246-251.	Negishi T, Kawasaki K, Takatori A, Ishii Y, Kyuwa S, Kuroda Y, *et al*. (2003) Effects of perinatal exposure to bisphenol A on the behavior of offspring in F344 rats. Environ Toxicol Pharmacol 14: 99-108.	Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazono A. (2001) Rat two-generation reproductive toxicity study of bisphenol A. Reprod Toxicol 15(5): 505-523.
AIM	Designed to analyze the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	To determine the potential of BPA, administered in feed to Sprague-Dawley rats, to induce functional and/or morphological effects in the nervous system that may arise in the offspring from exposure to the mother during pregnancy and lactation	To address three questions: Does chronic BPA treatment during the perinatal period alter sexual behaviour in adulthood? Is this effect dose-dependent? Does sexual experience mitigate any initial deficits in sexual performance?	To determine whether perinatal maternal exposure to BPA effects the behaviour of offspring using F344 rats.	To determine the low-dose effects of bisphenol A in a rat two-generation reproduction study.
	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types	Performed as a separate study	Performed as a separate study	Also evaluates offspring development and organ weights	Behavioural tests incorporated in this repro study
Stated that performed according to OECD TG 426 or other guidelines (specify)?		Yes OECD TG 426	No	No	OECD 416; MAFF No 4200 (Japan, for agricultural chemicals); ICH (guideline for reprotox)
Experimental design					
A. Test animals					

	Rat (commonly used strains)	Rat (Sprague Dawley)	Rat (Long-Evans)	Rat (F344)	Rat (Crj: CD (SD) IGS)
	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	N/A	N/A	N/A	N/A
	Supplier of test animals to be provided	Charles River Laboratories, Inc	Charles River Laboratories, Inc	SLC, Shizuoka, Japan	Charles River Japan
	Number, age at start and sex of animals	Clearly stated	Clearly stated	age not stated	clearly stated
	Housing conditions, acclimatization etc	Clearly stated, apart from acclimatization, as per OECD 426	Not stated	clearly stated	clearly stated, comparable to TG 426
	Unique identification for each animal and litters	Clearly stated, metal ear tags	Not stated	Not stated	Not stated
	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	Clearly stated, 24 mated females per dose group (no report on final # of litters)	3 dams per group (only 2 in the next to highest dose group)	8-9 dams in each group (not stated more specifically)	25 dams per dose group
	Live pups to be counted and sexed	Done	done (not stated when)	done on PND 0	done on PND 0
	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	Clearly stated, litters adjusted to 8 pups of equal sex distribution on PND 4 (litters where 4 pups/sex could not be achieved were not included in statistics)	Litters culled to 4 males and 4 females (not stated when) except for next to highest dose group, which was culled to 6 males and 6 females because only 2 litters instead of 3.	Clearly stated, litters adjusted to 8 pups on PND 0, of equal sex distribution when possible	litters adjusted to 4 males and 4 females on PND 4

	Pup identification is required	Pups selected to continue on study identified by foot markings. Metal ear tags after weaning.	not stated	"Pups identified individually on PND 7" not stated how	Not stated
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	Clearly stated, as per TG 426	Not stated. All pups subjected to testing?	Not stated	25 pups/sex/group selected at weaning to continue as parental animals. Not clearly stated if siblings but assumed 1/sex/litter. Only F1 subjected to behavioural tests.
Statistical unit	The statistical unit of measure should be the litter (or dam) and not the pup.	litter	pup	pup (assumed)	litter
B. Test conditions					
1. Administration of chemical/dosing *(Include information on vehicle as well as route of exposure)*	Most relevant to potential human exposure	Yes, oral	Yes, oral	Yes, oral	Yes, oral
	Oral by gavage, in diet, drinking water or capsules	In diet to mated females	In corn oil, allowed to drink spontaneously	In olive oil, not stated if by gavage (but assumed)	in distilled water by gastric intubation
	Other forms of admin (inhalation or dermal) requires modification of procedures	N/A	N/A	N/A	N/A
	At least 3 dose groups and a concurrent control	5 dose groups + neg control; 0, 0.15, 1.5, 75, 750, 2250 ppm	4 dose groups + neg control; 0, 5, 50, 500, 5000 µg/kg bw/day	3 doses + neg ctrl; 0, 4, 40, 400 mg/kg bw/day	4 doses + neg control; 0, 0.2, 2, 20, 200 µg/kg bw/day
	Repeated exposure	Continuous in feed, from GD 0 to lactation day 21	Repeated	Repeated	repeated
	Mated females, starting on GD 6	Mated females from GD 0	Mated females from GD 7	Mated females from GD 10	to males and females starting before mating and continued through gestation and lactation
	Motivation if starting exposure earlier	None given	N/A	N/A	*done according to reprotox TGs*

	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	Dose selection based on published studies and previous testing of BPA	Rationale for dose selection not stated	not stated, stated that the lowest dose is well below the LOAEL	dose levels determined based on previous published studies in which effects on repro were seen at the 2 and 20 µg/kg bw/day
	High dose level should induce some maternal toxicity (eg weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	Doses selected to include high dose expected to result in systemic toxicity in F0 dams and to cover low doses reported in some published studies to cause dev neurotox. Low dose not clearly stated to be without effect.	Not discussed	Not discussed	not stated
	Dose levels should be selected to allow for illustration of dose-response	includes high and low doses with appropriately spaced dose groups	includes high and low doses with appropriately spaced dose groups	done	done
	Positive controls not mentioned	No pos control included as the reliability and sensitivity of the neurotox methods in SD rats have been established using other chemicals	not included	Not included	Not included
2. Duration of exposure *(Add information if any comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure via milk)*	GD6 to PND21	GD 0 to PND 21	GD 7 to PND 14	GD 10 - PND 20	F0 from before mating through gestation and lactation. F1 exposure via gastric intubation from PND 23 through mating, gestation and lactation. F2 males and females exposure via gastric intubation from PND 22 for 4 and 11 weeks, respectively.

	Motivation for using other exposure duration if not GD 6 to PND 21	None given	None given	None given	*done according to reprotox TGs*
	Exposure via lactation generally assumed but evidence of continuous exposure [during lactation] can be retrieved from *e.g.*, pharmacokinetic information, offspring toxicity or changes in bio-markers	not included, refers to literature	not included, exposure via lactation assumed without reference to the literature	not included	not included
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	Pups were exposed via direct consumption of feed during the third week of the lactation period.	No direct dosing of pups other than from mother's milk	No direct dosing of pups other than from mother's milk	pups exposed via gastric intubation from PND 23 (F1) or 22 (F2)
3. If other routes of administration than oral	Use OECD Guidance Document on Neurotoxicity Testing Strategies and Methods #43 to assist in the design of the studies	N/A	N/A	N/A	N/A
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at study termination.	Pups sacrificed on PNDs 21 and 72, not stated for dams	not stated	dams sacrificed at weaning, 8 randomly selected offspring per sex and treatment sacrificed on PND 62	dams sacrificed after weaning, offspring not selected to become parents terminated after weaning, parental males terminated after mating, parental females terminated after weaning of offspring
5. Control for estrous cyclicity of female offspring	*Not specified if required*	not reported	yes	not reported	cyclicity included as separate endpoint but seems not to have been controlled for in behavioural tests
C. Endpoints*:					
1. Physical and developmental landmarks					

a. Body weight and clinical observations, weekly during pre-weaning and at least every two weeks at adolescence and adulthood	Required	Provided	not provided	Provided. "Dams were examined for clinical signs of toxicity and weighed daily before dosing"	Provided (daily) for adult animals.
	During treatment and observation period, detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	DCO performed on all females on GD 10 and 15 and LD 10 and 21, and on 1 pup/sex/litter on PND 4, 11, 21, 35, 45, 60. Performed by trained observers w/o knowledge of group assignment	-	not stated	Not stated
	Where possible, observations to be made by the same technician	Not specified	-	not stated	Not stated
	Observations to be performed outside of home cage	Outside of home cage in open field for 2 min	-	not stated	Not stated
	“Unusual” responses with respect to eg activity level, need to be documented	Provided (No treatment-related clinical findings were noted in the dams or pups.)	-	not provided	not provided
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development	Dams and pups weighed every 3 or 4 days until PND 21. Pups weighed every week thereafter until euthanasia	Not stated	dams weighed daily during dosing. Pups weighed on PNDs 0, 7, 14, 28, 56 and 84	dams weighed on GDs 0, 7, 14 and 20. Pups weighed on PNDs 0, 4, 7, 14 and 21
b. Brain weight and Neuropathology, on PND 22 (or earlier between PND 11 and 22) and at termination	Fixation; Immersion <PND21, Perfusion >PND21	PND 21 and 72 pups perfused in situ, brains processed. On PND 72 also additional CNS and PNS	-	brains collected from randomly selected pups on PND 62 but not stated how	-
	Tissue selection need both CNS and PNS	Brains from pups on PND 21 and 72. On PND 72 also additional CNS and PNS.	-	brains only (and other organs but not PNS)	-

	Morphometric evaluation of tissue collected on PND 21 and at end of study	Done (PND 21 and 72)	-	brains only on PND 62 but methods or results not provided	-
	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	Not provided (no treatment related lesions noted)	-	not provided	-
c. Sexual maturation	At adolescence (as appropriate)	Females (1/litter): daily for vaginal patency from PND 25. Males (1/litter): daily for preputial separation from PND 35	-	-	testes descent, preputial separation, vaginal opening
d. Other physical landmaks for pup development (*e.g.* eye opening)	only required if these will provide additional information	-	-	-	AGD, pinna detachment, incisor eruption, eye opening
2. Functional/behavioural endpoints					
a. Behavioural ontogeny, at least two measures pre-weaning using the same pup (1 pup/sex/litter). Some tests to consider:	righting reflex	all pups, daily beginning on PND 2 until attainment.	-	-	surface righting reflex daily from PND 6, mid-air righting reflex daily from PND 13
	negative geotaxis	-	-	-	daily from PND 7
	motor activity (strongly recommended!)	Motor activity on PND 13, 17, 21 by Kinder Scientific Motor Monitor System (1 pup/sex/litter)	-	-	-
b. Motor activity (including habituation) at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Motor activity should be monitored using an automated activity recording apparatus	Motor activity on PND 13, 17, 21, 61 by Kinder Scientific Motor Monitor System	-	Spontaneous motor activity using Supermex at 4 weeks of age (PND 28-34), locomotion in the open field at 8 weeks (PND 56-62)	Open field test on F1 rats at 5-6 weeks of age
	Habituation	done	-	not reported	not reported
c. Motor and sensory	Grip Strength	-	-	-	-

function at at adolesence (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Rotating Rod	-	-	-	-
	Hind limb foot splay, landing foot spread	-	-	-	-
	Nociception (hotplate, tail flick)	-	-	-	-
	Sensory irritation	-	-	-	-
	Somatosensory operant discrimination task	-	-	-	-
	Acoustic startle response and prepulse inhibition	On PND 20 and 60 by Kinder Scientific Startle Monitor System	-	-	-
	Auditory discrimination procedure	-	-	-	-
	Visual discrimination task	-	-	-	-
d. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70), for example:	Conditioned taste aversion	-	-	-	-
	Active avoidance	-	-	In a two-way shuttle-box at 4 or 8 weeks of age (PND 28-34 or 56-62)	-
	Passive avoidance	-	-	-	-
	Spatial Mazes, *e.g.* Morris water maze, Biel water maze, T-maze	Biel water maze on PND 22 and 62, 1 pup/sex/litter (different pups at both times)	-	-	T-maze on F1 rats (only 6/sex/group, selection process not stated) at 6-7 weeks
	Conditional discrimination, *e.g.* simple discrimination, matching to sample	-	-	-	-
	Delayed discrimination, *e.g.* delayed matching to sample, delayed alternation or repeated acquisition	-	-	-	-
	Eye-blink conditioning	-	-	-	-

	Schedule-controlled operant behaviour	-	-	-	-
	<u>Two criteria</u> need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	fulfilled	-	not fulfilled (Shuttlebox test), acquisition tests in 50 trials per day for 3 days but no tests for retention/memory	not clear - not stated how many trials were allowed for each subject
e. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not specified if required*	-	-	Open field test at 8 weeks of age (PND 56-62)	Open field test on F1 rats ("all") at 5-6 weeks of age
f. Other neurobehavioural tests included	*Not required*	-	Male and female sexual behaviour	-	-
3. Other endpoints included					
	Not required	-	-	pup organ weights: liver, kidney, spleen, thymus, brain, testis)	repro tox parameters, hormone levels *(as per guidelines for reproductive toxicity tests)*
D. Authors' conclusions					

Evidence of developmental neurotoxicity	--	There is no evidence that BPA is a developmental neurotoxicant in rats. Some statistically significant differences were observed in the Biel maze but were not considered treatment related" because they did not occur consistently between or within testing periods, did not demonstrate any evidence of a dose-related trend, and/or were associated with atypical control responses".	"prenatal administration of low levels of BPA (50 µg/kg bw/day) impairs adult sexual performance in experienced male LE rats." High doses did not result in adverse effects. "Significant non-monotonic dose response curve". Female offspring were not affected at any dose.	"the present study demonstrated the adverse effects of perinatal exposure of BPA on the behaviour of offspring."	no effects on behaviour observed. Acknowledges that effects have been seen in other studies and discusses possible differences in study design.

* Shaded cells denote tests where a significant effect was observed in treated animals compared to controls (based on author's conclusions).

Table 5. Comparison of experimental design and endpoints in the OECD TG 426, the available guideline and non-guideline studies of developmental neurotoxicity for PBDE 209.

Reference	**OECD 426**	**Beck 2009 (Biesemeier *et al.* 2011)**	**Viberg *et al.* 2007**	**Viberg *et al.* 2003**	**Rice *et al.* 2007, 2009**
		Biesemeier JA, Beck MJ, Silberberg H, Myers NR, Ariano JM, Radovsky A, *et al*. (2011) An oral developmental neurotoxicity study of decabromodiphenyl ether (DecaBDE) in rats. Birth Defects Res B Dev Reprod Toxicol 92(1): 17-35.	Viberg H, Fredriksson A, Eriksson P (2007) Changes in spontaneous behaviour and altered response to nicotine in the adult rat after neonatal exposure to the brominated flame retardant, decabrominated diphenyl ether (PBDE 209). NeuroToxicology 28:136-142	Viberg H, Fredriksson A, Jakobsson E, Orn U, Eriksson P. (2003) Neurobehavioral derangements in adult mice receiving decabrominated diphenyl ether (PBDE 209) during a defined period of neonatal brain development. Toxicol Sci 76(1): 112-120.	Rice DC, Reeve EA, Herlihy A, Zoeller RT, Thompson WD, Markowski VP (2007) Developmental delays and locomotor activity in the C57BL6/J mouse following neonatal exposure to the fully-brominated PBDE, decabromodiphenyl ether. Neurotoxicol Teratol. 29:511-20. Rice DC, Thompson WD, Reeve EA, Onos KD, Assadollahzadeh M, Markowski VP (2009) Behavioral changes in aging but not young mice after neonatal exposure to the polybrominated flame retardant decaBDE. Environ Health Perspect. 117:1903-11.
AIM	Designed to analyze the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	To determine the potential of the test substance to induce functional and/or morphological insult to the nervous system in the offspring of dams that were administered during pregnancy and lactation via oral gavage dosage levels of 1, 10, 100 or 1000 mg/kg/d. Also to determine the concentration of the test substance in maternal and neonatal plasma and maternal milk samples.	Undertaken to elucidate whether neonatal PBDE 209 exposure can induce changes in spontaneous behaviour and cholinergic system in another species, namely the rat	Undertaken to ascertain whether PBDE 209 could be absorbed during the neonatal brain development and induce persistent neurotoxic effects on the spontaneous motor behaviour of the adult.	To examine the effects of postnatal exposure to decaBDE on basic parameters of growth and development, ontogeny of neurological function, motor activity and behaviour at two ages in male and female mice.

	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types	Performed as a separate study	Performed as a separate study	Performed as a separate study	Performed as a separate study
Stated that performed according to OECD TG 426 or other guidelines (specify)?		Yes (and GLP and US EPA guideline)	No	No	No. "Whenever possible endpoints in the USEPA Guidelines for neurotoxicity risk assessment were included"
Experimental design					
A. Test animals					
	Rat (commonly used strains)	Rat (Sprague Dawley Crl:CD)	Rat (Sprague-Dawley)	Mouse (NMRI)	Mouse (C57BL/6J)
	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	N/A	N/A	Not justified. A later publication (Viberg *et al.* 2007) showed that other mouse strain, as well as rats, are similarly sensitive.	Not justified.
	Supplier of test animals to be provided	Charles River Lab., Raleigh, NC	B&K, Sollentuna, Sweden	B&K, Sollentuna, Sweden	Jackson Lab, Bar Harbor, Maine
	Number, age at start and sex of animals	Clearly stated	Stated	Stated	Stated
	Housing conditions, acclimatization etc	Clearly stated	Clearly stated	Clearly stated	Clearly stated
	Unique identification for each animal and litters	Clearly stated, metal ear tag	Not stated	Not stated	Not stated

	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	Clearly stated, 35 mated females per dose group	Exposed postnatally. Offspring from 3-5 litters/group	Exposed postnatally. Offspring from 3-5 litters/group	Exposed postnatally. Offspring from 11-13 litters/group
	Live pups to be counted and sexed	Done	N/A	N/A	N/A
	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	Clearly stated, litters adjusted to 8 pups of equal sex distribution (if possible) on PND 4	Adjusted to 8-12 mice/litter within the first 48 hours.	Adjusted to 10-12 mice/litter within the first 48 hours.	Clearly stated, litters adjusted to 6 pups of equal sex distribution on PND 2
	Pup identification is required	Done, toe-tattoo and ear tag	Not stated	Not stated	Not stated
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	Clearly stated	N/A. Only males studied	N/A. Only males studied	Stated
Statistical Unit	The statistical unit of measure should be the litter (or dam) and not the pup.	litter	pup	pup	litter
B. Test conditions					
1. Administration of chemical/dosing *(Include information on vehicle as well as route of exposure*)	Most relevant to potential human exposure	Yes, oral (maternal exposure)	Yes, oral (directly to pups)	Yes, oral (directly to pups)	Yes, oral (directly to pups)
	Oral by gavage, in diet, drinking water or capsules	Gavage, vehicle corn oil	Gavage, vehicle 20% fat emulsion	Gavage, vehicle 20% fat emulsion	In the mouth using a micropipette, vehicle 20% emultion (egg lecithin:peanut oil)

	Other forms of admin (inhalation or demal) requires modification of procedures	N/A	N/A	N/A	N/A
	At least 3 dose groups and a concurrent control	4 dose groups + negative control; 0, 1, 10, 100, 1000 mg/kg/d	2 dose levels (6.7 and 20.1 mg/kg) + negative control, single dose on PND3.	2 dose levels (2.22 and 20.1 mg/kg) + negative control, single dose on PND3. Other groups received the dose on PND10 or 19.	2 dose groups + negative control; 0, 6, 20 mg/kg/d
	Repeated exposure	Once daily (GD 6 to PND21)	Single dose (persistent compound)	Single dose (persistent compound)	Once daily (PND 2-15)
	Mated females, starting on GD 6	Yes	N/A	N/A	N/A
	Motivation if starting exposure earlier	N/A	N/A	N/A	N/A
	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	Dose-range dinding developmental neurotoxicity study performed	Similar to previous study in mice.	Similar to previous studies of other PBDEs.	Not stated
	High dose level should induce some maternal toxicity (eg weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	Highest dose 1000 mg/kg/d is the limit level in TG 426	N/A	N/A	N/A
	Dose levels should be selected to allow for illustration of dose-response	Done	Only 2 dose levels	Only 2 dose levels	Only 2 dose levels
	Positive controls not mentioned	Not included.	Not included	Not included	Not included.
2. Duration of exposure *(Add information if any*	GD6 to PND21	GD6 to PND21	PND 3 (and longer due to long half-life)	PND 3 (and longer due to long half-life)	PND 2-15

comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure via milk)	Motivation for using other exposure duration if not GD 6 to PND 21	N/A	Refer to "brain growth spurt" during this period.	Refer to "brain growth spurt" during this period. Also studied dosing on PND10 and 19 (no effects).	No
	Exposure via lactation generally assumed but evidence of continuous exposure [during lactation] can be retrieved from *e.g.*, pharmacokinetic information, offspring toxicity or changes in bio-markers	Levels of PBDE 209 studied in maternal and offspring plasma and maternal milk	N/A	N/A	N/A
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	No direct dosing of pups	Direct dosing of pups.	Direct dosing of pups.	Direct dosing of pups.
3. If other routes of administration than oral	Use OECD Guidance Document on Neurotoxicity Testing Strategies and Methods #43 to assist in the design of the studies	N/A	N/A	N/A	N/A
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at study termination.	As soon as they were no longer needed, *i.e.* on PND 21, 22, 72, 120 and 180, respectively.	Not stated	Not stated	Not stated
5. Control for estrous cyclicity of female offspring	*Not specified if required*	Not reported	N/A	N/A	Not reported
C. Endpoints*:					
1. Physical and developmental landmarks					

a. Body weight and clinical observations, weekly during pre-weaning and at least every two weeks at adolescence and adulthood	Required	Provided	"No clinical signs of toxicity… nor any significant difference in body weight gain or adult weight"	"No clinical signs of toxicity… nor any significant difference in body weight gain or adult weight"	Not reported
	During treatment and observation period, detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	DCO performed on all females on GD 10 and 15 and LD 10 and 21, and on 30 pups/sex/group (1 per litter) on PND 4, 11, 21, 35, 45, 60.	Not stated	Not stated	FOB performed every day from PND 2-20 on 11-13 pups/sex/group (1 per litter).
	Where possible, observations to be made by the same technician	Not stated, but technicians blind to treatment	Not stated	Not stated	”Two trained observers, blind to the exposure”
	Observations to be performed outside of home cage	Out of the home cage	Not stated	Not stated	Both in and out of the home cage
	“Unusual” responses with respect to eg activity level, need to be documented	Provided	Not stated	Not stated	Not stated
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development	Recorded at "appropriate intervals"	Performed (no details given)	Performed (no details given)	Recorded every day during PND2-21
b. Brain weight and Neuropathology, on PND 22 (or earlier between PND 11 and 22) and at termination	Fixation; Immersion <PND21, Perfusion >PND21	Neuropathology at PND 21, 72	-	-	-
	Tissue selection need both CNS and PNS	Brain weight at PND 21, 72	-	-	-
	Morphometric evaluation of tissue collected on PND 21 and at end of study	Morphometry at PND21, 72	-	-	-

	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	Not stated	-	-	-
c. Sexual maturation	At adolescence (as appropriate)	Balanopreputial separation and vaginal patency (50 pups/sex/dose)	-	-	Vaginal opening and descent of the testes
d. Other physical landmaks for pup development (eg eye opening)	only required if these will provide additional information	-	-	-	Age at pinnae detach, incisors erupt and eyes open
2. Functional/behavioural endpoints					
a. Behavioural ontogeny, at least two measures pre-weaning using the same pup (1 pup/sex/litter). Some tests to consider:	righting reflex	-	-	-	Palpebral reflex
	negative geotaxis	-	-	-	-
	motor activity (strongly recommended!)	Locomotor activity on PND 13, 17, 21 and 61 (30 pups/sex/group, 1/litter)	-	-	Locomotor activity on PND 2-20 (11-13 pups/sex/group, 1/litter)
b. Motor activity (including habituation) at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Motor activity should be monitored using an automated activity recording apparatus	Locomotor activity on PND 13, 17, 21 and 61 (30 pups/sex/group, 1/litter) and 120 and 180. Nicotine challenge was conducted during/after PND 61, 120 and 180 assessments.	Motor activity (locomotion, rearing and total activity) on PNM 2 by Rat-O-Matic, ADEA Elektronik AB, Uppsala, Sweden.	Motor activity (locomotion, rearing and total activity) on PNM 2, 4 and 6 by Rat-O-Matic, ADEA Elektronik AB, Uppsala, Sweden.	Locomotor activity on PND 2-20 and 70 (11-13 pups/sex/group, 1/litter), and one year (9-12 pups/sex/group)
	Habituation	done	done	done	done
c. Motor and sensory function at at adolesence (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Grip Strength	Performed at PND 21, 35, 45 and 60 (30 pups/sex/group, 1/litter)	-	-	Performed at PND 2-20
	Rotating Rod	-	-	-	-
	Hindlimb foot splay, landing foot spread	-	-	-	-
	Noiciception (hotplate, tail flick)	-	-	-	-
	Sensory irritation	-	-	-	-

	Somatosensory operant discrimination task	-	-	-	-
	Acoustic startle response and prepulse inhibition	Auditory startle response on PND 20 and 60 (30 pups/sex/group, 1/litter)	-	-	On PND 2-20 (11-13 pups/sex/group, 1/litter)
	Auditory discrimination procedure	-	-	-	-
	Visual discrimination task	-	-	-	-
d. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Conditioned taste aversion	-	-	-	-
	Active avoidance	-	-	-	-
	Passive avoidance	-	-	-	-
	Spatial Mazes, *e.g.* Morris water maze, Biel water maze, T-maze	Learning and memory (Biel Maze swimming trial) performed at PND 22 and 62 (20 and 30 pups/sex/group, respectively, different pups at both times)	-	-	-
	Conditional discrimination, *e.g.* simple discrimination, matching to sample	-	-	-	-
	Delayed discrimination, *e.g.* delayed matching to sample, delayed alternation or repeated acquisition	-	-	-	-
	Eye-blink conditioning	-	-	-	-
	Schedule-controlled operant behaviour	-	-	-	Tested for operant procedures including a fixed-ratio schedule of reinforcement, fixed-interval schedule and light-dark visual discrimination at 3 and 16 months of age (11-13 and 9-12 mice/sex/dose, respectively)

	<u>Two criteria</u> need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	fulfilled	-	-	Not fulfilled
e. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not specified if required*	-	-	Done in a similar study (Johansson *et al.* 2008)	-
f. Other neurobehavioural tests included	*Not required*	-	-	-	-
5. Other endpoints included					
	Not required	-	-	-	-
Author's conclusions					
Evidence of developmental neurotoxicity		No evidence of developmental neurotoxicity.	Effects on spontaneous behaviour after exposure at PND3 observed at 2 months of age	Effects on spontaneous behaviour after exposure at PND3 persists and worsen with age (2, 4 and 6 months)	Suggests that decaBDE is a developmental neurotoxicant that can produce long-term behavioural changes

* Shaded cells denote tests where a significant effect was observed in treated animals compared to controls (based on author's conclusions).

Table 6. Comparison of experimental design and endpoints in the OECD TG 426, the available guideline and non-guideline studies of developmental neurotoxicity for deltamethrin.

Reference	**OECD 426**	**Gilmore *et al.* 2006** (the summary only was evaluated) Gilmore RG *et al.* (2006) A developmental neurotoxicity screening study with technical grade deltamethrin in Wistar rats. Study summary, Document III-A, Bayer CropScience AG. *(full report at KemI).*	**Eriksson and Fredriksson 1991** Eriksson P, Fredriksson A (1991) Neurotoxic effects of two different pyrethroids, bioallethrin and deltamethrin, on immature and adult mice: changes in behavioral and muscarinic receptor variables. Toxicol Appl Pharmacol 108:78-85.
AIM	Designed to analyze the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	Not stated in summary.	To investigate whether two pyrethroids, bioallethrin (type I) and deltamethrin (Type II), will affect the MAChR in the adult mouse brain and modify the behaviour of the young and adult animal when given to neonatal mice during the brain growth spurt.
	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types		Separate study
Stated that performed according to OECD TG 426 or other guidelines (specify)?		Yes. TG 426 draft (Sept 2003). Deviations stated.	No
Experimental design			
A. Test animals			
	Rat (commonly used strains)	Rat (Wistar)	Mouse (NMRI)

	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	N/A	Not justified
	Supplier of test animals to be provided	Charles River Lab Inc	ALAB, Sweden
	Number, age at start and sex of animals	Clearly stated	Stated
	Housing conditions, acclimatization etc	Not stated in summary (but according to 426 draft).	Clearly stated
	Unique identification for each animal and litters	Stated, tattooed	Not stated
	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	30 mated females. 23 litters per dose group	Exposed postnatally. Offspring from 3 litters/dose group.
	Live pups to be counted and sexed	Done	N/A
	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	PND 4, litters with minimum 7 pups, at least 3/sex, culled to yield 4/sex/litter	Adjusted to 8-10 pups per litter during first 24 hours.
	Pup identification is required	Tattooed on PND0	Not stated
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	Clearly stated. At least one male and one female from each litter (approx. 16 offspring/sex/group)	12 males from 3 litters were tested for behaviour
Statistical Unit	The statistical unit of measure should be the litter (or dam) and not the pup.	litter	pup
B. Test conditions			

1. Administration of chemical/dosing *(Include information on vehicle as well as route of exposure)*	Most relevant to potential human exposure	Yes, oral	Yes, oral
	Oral by gavage, in diet, drinking water or capsules	In diet to mated females	Via PVC tube, vehicle 20% fat/water emulsion (egg lecithin and peanut oil)
	Other forms of admin (inhalation or demal) requires modification of procedures	N/A	N/A
	At least 3 dose groups and a concurrent control	3 dose groups + neg control, 0, 20, 80, 200 ppm, corresponding to 0, 1.64, 6.78, 16.1 mg/kg/d (adjustments during lactation)	1 dose group (0.7 mg/kg/d) + negative control
	Repeated exposure	In feed, from GD6 to LD21	PND 10-16, once daily
	Mated females, starting on GD 6	Mated females, starting on GD6	N/A
	Motivation if starting exposure earlier	N/A	N/A
	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	Not justified in summary	The lower dose from previous study was chosen (did not cause any neurotoxic symptoms)
	High dose level should induce some maternal toxicity (eg weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	High dose showed reduced body weights in both dams and pups.	N/A
	Dose levels should be selected to allow for illustration of dose-response	Fulfilled	Only one dose level
	Positive controls not mentioned	A later study on chlorpyrifos showed the ability of the test laboratory to detect effect on behaviour etc.	Not included
2. Duration of exposure *(Add information if any comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure*	GD6 to PND21	GD6 to PND21	PND 10-16
	Motivation for using other exposure duration if not GD 6 to PND 21	N/A	Refer to "brain growth spurt" during this period.

via milk)	Exposure via lactation generally assumed but evidence of continuous exposure [during lactation] can be retrieved from *e.g.*, pharmacokinetic information, offspring toxicity or changes in bio-markers	Separate study on pup brain levels of deltamethrin showing evidence of lactational exposure.	N/A
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	Measures of food consumption may have included consumption by the pups, especially during the third week of lactation.	Only direct dosing of pups
3. If other routes of administration than oral	Use OECD Guidance Document on Neurotoxicity Testing Strategies and Methods #43 to assist in the design of the studies	N/A	N/A
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at study termination.	Dams were sacrificed on day 21 of lactation. Offspring were sacrificed on PND 21 or 75.	Not stated
5. Control for estrous cyclicity of female offspring	*Not specified if required*	Not mentioned in summary	N/A
C. Endpoints*:			
1. Physical and developmental landmarks			
a. Body weight and clinical	Required	Provided	Not reported

observations, weekly during pre-weaning and at least every two weeks at adolescence and adulthood	During treatment and observation period, detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	DCO and FOB performed on all females on GD13 and GD20 and on at least 10 dams/dose on LD11 and LD21. All pups DCO once daily before weaning and once weekly thereafter. FOB on PND4, 11, 21, 35, 45, 60. Trained observers and/or blind assessment not mentioned in summary.	Spontaneous behaviour tested on PND 17 and at 4 months of age.
	Where possible, observations to be made by the same technician	Not specified in summary	Not stated
	Observations to be performed outside of home cage	Yes	Not stated
	"Unusual" responses with respect to eg activity level, need to be documented	Not stated in summary	Not stated
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development	Dams weighed once weekly. Pups weighed on PND 0, 4, 11, 17, 21, thereafter once weekly.	Final bw at 4 months
b. Brain weight and Neuropathology, on PND 22 (or earlier between PND 11 and 22) and at termination	Fixation; Immersion <PND21, Perfusion >PND21	PND21 and 75 pups perfused in situ, brains processed. On PND75 also additional CNS and PNS.	-
	Tissue selection need both CNS and PNS	PND21 and 75 pups perfused in situ, brains processed. On PND75 also additional CNS and PNS.	-
	Morphometric evaluation of tissue collected on PND 21 and at end of study	AP length of cerebrum and cerebellum at PND21 and 75	-
	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	Not provided in summary (no treatment related lesions noted)	-

c. Sexual maturation	At adolescence (as appropriate)	All female pups examined for vaginal patency daily from PND29. All male pups examined for balanopreputial separation daily from PND 38.	-
d. Other physical landmaks for pup development (eg eye opening)	only required if these will provide additional information	All pups examined on PND21 for pupil constriction.	Not stated
2. Functional/behavioural endpoints			
a. Behavioural ontogeny, at least two measures pre-weaning using the same pup (1 pup/sex/litter). Some tests to consider:	righting reflex	-	-
	negative geotaxis	-	-
	motor activity (strongly recommended!)	Motor activity on PND13, 17, 21 (approx 16/sex/dose)	measured at PND 17 only
b. Motor activity (including habituation) at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Motor activity should be monitored using an automated activity recording apparatus	Motor activity on PND13, 17, 21, 60, 120 (approx 16/sex/dose) using Figure 8 maze.	Horizontal and vertical motor activity tested on PND17 and at 4 months (Rat-O-Matic, ADEA Elektronik AB, Uppsala)
	Habituation	done	done
c. Motor and sensory function at at adolesence (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Grip Strength	-	-
	Rotating Rod	-	-
	Hindlimb foot splay, landing foot spread	-	-
	Noiciception (hotplate, tail flick)	-	-
	Sensory irritation	-	-
	Somatosensory operant discrimination task	-	-
	Acoustic startle response and prepulse inhibition	Auditory startle reflex habituation in approx 16/sex/dose on PND 22 and 60 (automated system)	-

	Auditory discrimination procedure	-	-
	Visual discrimination task	-	-
d. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Conditioned taste aversion	-	-
	Active avoidance	-	-
	Passive avoidance	Tested for acquisition on PND22 and 29.	-
	Spatial Mazes, *e.g.* Morris water maze, Biel water maze, T-maze	Water maze on PND60. Animals demonstrating acquisistion tested for retention on PND 67	-
	Conditional discrimination, *e.g.* simple discrimination, matching to sample	-	-
	Delayed discrimination, *e.g.* delayed matching to sample, delayed alternation or repeated acquisition	-	-
	Eye-blink conditioning	-	-
	Schedule-controlled operant behaviour	-	-
	Two criteria need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	Fulfilled	N/A
e. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not specified if required*	-	-

f. Other neurobehavioural tests included	*Not required*	-	-
3. Other endpoints included			
	Not required	Pups examined on PND21 for pupil constriction. Ophthalmic examination on PND50-60 on min 10/sex/dose	High- and low-affinity muscarinic receptor density in synaptosomal fraction of cerebral cortex, hippocampus and striatum at 4 months.
Author's conclusions			
Evidence of developmental neurotoxicity		No. Increased incidence of vocalizations with handling in high dose males at PND4	"No clinical signs of pyrethroid poisoning". Locomotion and total activity (habituation) affected at 4 months (not 17 days).

* Shaded cells denote tests where a significant effect was observed in treated animals compared to controls (based on author's conclusions).

Table 7. Comparison of experimental design and endpoints in the OECD TG 426, the available non-guideline studies of developmental neurotoxicity for PCB 153.

.

Reference	**OECD 426**	**Piedrafita *et al*., (2008)**	**Fischer *et al*., (2008)**	**Schantz *et al*. (1995)**
		Piedrafita B, Erceg S, Cauli O, Monfort P, Felipo V (2008) Developmental exposure to polychlorinated biphenyls PCB153 or PCB126 impairs learning ability in young but not in adult rats. Eur J Neurosci. 27:177-82.	Fischer C, Fredriksson A, Eriksson P (2008) Neonatal co-exposure to low doses of an ortho-PCB (PCB 153) and methyl mercury exacerbate defective developmental neurobehavior in mice. Toxicology, 244:157-65.	Schantz SL, Moshtaghian J, Ness DK. 1995. Spatial learning deficits in adult rats exposed to ortho-substituted PCB congeners during gestation and lactation. Fundam Appl Toxicol 26(1): 117-126.
AIM	Designed to analyse the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	To assess whether exposure of rats to PCB 126 (dioxin like) or PCB 153 (non-dioxin like) during pregnancy and lactation affects the ability to learn	Investigate if PCB153 can interact with MeHG to enhance developmental neurotoxic effects on spontaneous behaviour and habituation	To assess spatial learning and memory in adult rats following combined gestational and lactational exposure to specific *ortho*-substituted PCB congeners
	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types	Performed as a separate study	Performed as a separate study	Performed as a separate study
Stated that performed according to OECD TG 426 or other guidelines (specify)?		no	no	no
Experimental design				
A. Test animals				
	Rat (commonly used strains)	Rat (Wistar)	Mouse (NMRI;Naval Medical Research Institute)	Rat (Sprague-Dawley)
	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	N/A	Justification for mouse model is provided with comparison to brain development and growth to that in humans	N/A

	Supplier of test animals to be provided	Charles River Laboratories, Inc	B&K, Sollentuna, Sweden	Harlan Sprague-Dawley
	Number, age at start and sex of animals	Not stated	Not stated	Age not stated
	Housing conditions, acclimatization etc	Stated, apart from acclimatization	Stated, apart from acclimatization	Stated, apart from acclimatization
	Unique identification for each animal and litters	not stated	not stated	not stated
	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	Number of mated females not stated.	Number of mated females not stated.	Number of mated females not stated.
	Live pups to be counted and sexed	-	-	Done on PND 0
	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	Not stated	Litters adjusted to 8-12 pups/sex/group on PND 2. Excess pups culled. Litters contained pups of both sexes	Litters adjusted to 8 with equal sex distribution to the extent possible on PND 2.
	Pup identification is required	Not stated	not stated	Not stated
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	Not clearly stated. At PND 21, two or three pups housed in cage until experiments. Unclear of gender balance.	At PND 21, male mice weaned and raised in groups of 4-7 in a room for males only, "eight mice were randomly selected from 3–4 different litters and only tested once for each test occasion"	One male and one female from each litter selected for behavioural testing on PND 21.

Statistical Unit	The statistical unit of measure should be the litter (or dam) and not the pup.	Pup	pup, stated that "We have established that studies using mice randomly selected from at least 3 different litters yield the same statistical effect and power as using litter-based studies"	Litter
B. Test conditions				
1. Administration of chemical/dosing *(Include information on vehicle as well as route of exposure)*	Most relevant to potential human exposure	Yes, oral	Yes, oral	Yes, oral
	Oral by gavage, in diet, drinking water or capsules	Mixed in sweet jelly bit (Transgel)	Oral by via gavage on PND 10	Oral by via gavage on GD10 – 16
	Other forms of admin (inhalation or dermal) requires modification of procedures	N/A	N/A	N/A
	At least 3 dose groups and a concurrent control	One dose group + neg ctrl; 0 and 1mg/kg bw/day	One dose group + neg ctrl; 0 and 0.51 mg/kg bw	Two dose groups + neg ctrl; 0, 16, 64 mg/kg bw/day
	Repeated exposure	repeated	single dose	Repeated
	Mated females, starting on GD 6	Mated females from GD 7	Male pups only PND 10	Mated females from GD 10
	Motivation if starting exposure earlier	N/A	-	.
	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	Dose selection based on previous published data	Not stated	“The doses selected so that the high dose would be at or near the threshold for developmental toxicity…and the low dose would be well below that threshold.”
	High dose level should induce some maternal toxicity (eg weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	No high dose included	No high dose included	“The doses selected so that the high dose would be at or near the threshold for developmental toxicity…and the low dose would be well below that threshold.”

	Dose levels should be selected to allow for illustration of dose-response	not stated	not stated	Not stated
	Positive controls not mentioned	Not included	Not included	Not included
2. Duration of exposure *(Add information if any comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure via milk)*	GD6 to PND21	GD 7 to PND 21	PND 10	GD 10 – 16
	Motivation for using other exposure duration if not GD 6 to PND 21	N/A	Experiment design of neonatal exposure as laboratory has performed for several years, thereby generating historical controls as well as reproducible developmental neurotoxicological data.	Not clearly stated that exposure of dams to PCBs during gestation also results in lactational exposure, but implied.
	Exposure via lactation generally assumed but evidence of continuous exposure [during lactation] can be retrieved from *e.g.*, pharmacokinetic information, offspring toxicity or changes in bio-markers	Not included	pups only exposed postnatally	Lactational transfer of PCBs discussed.
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	No direct dosing of pups	Direct dosing of male pups	No direct dosing of pups
3. If other routes of administration than oral	Use OECD Guidance Document on Neurotoxicity Testing Strategies and Methods #43 to assist in the design of the studies	N/A	N/A	N/A
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at	not stated	not stated	Not stated for dams. Pups not assigned to behavioural testing terminated at weaning.

	study termination.			
5. Control for estrous cyclicity of female offspring	*Not specified if required*	not stated	only tested male pups	Not stated
C. Endpoints*:				
1. Physical and developmental landmarks				
a. Body weight and clinical observations, weekly during pre-weaning and at least every two weeks at adolescence and adulthood	Required	Not provided	not provided	Body wt recorded on PND 0, 7, 14 and 21. Clinical obs not stated.
	During treatment and observation period, detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	Not provided	Brief statement in *Results* that no overt signs of clinical toxicity occurred during experimental period, but what was evaluated in not indicated.	Not provided
	Where possible, observations to be made by the same technician	-	Not specified	-
	Observations to be performed outside of home cage	-	Not specified	-
	"Unusual" responses with respect to eg activity level, need to be documented	-	Not provided	-
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development	-	not provided	Dam body wt recorded daily during gestation and on PND 0, 7, 14 and 21. Pups weighed on PND 0, 7, 14 and 21 then weekly and daily during periods of food restriction
b. Brain weight and Neuropathology, on PND	Fixation; Immersion <PND21, Perfusion >PND21	-	-	-

22 (or earlier between PND 11 and 22) and at termination	Tissue selection need both CNS and PNS		-	-
	Morphometric evaluation of tissue collected on PND 21 and at end of study	-	-	-
	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	-	-	-
c. Sexual maturation	At adolescence (as appropriate)	-	-	-
d. Other physical landmarks for pup development (eg eye opening)	only required if these will provide additional information	-	-	-
2. Functional/behavioural endpoints				
a. Behavioural ontogeny, at least two measures pre-weaning using the same pup (1 pup/sex/litter). Some tests to consider:	righting reflex	-	-	-
	negative geotaxis	-	-	-
	motor activity (strongly recommended!)	-	-	-
b. Motor activity (including habituation) at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Motor activity should be monitored using an automated activity recording apparatus	-	Motor activity measured in 2 and 4 month pups in automated devise for spontaneous behaviour. Locomotion, rearing and total activity were evaluated in an automated device consisting of two cages placed within two series of infrared beams (Rat-O-Matic). Cages placed in individual soundproofed boxes with separate ventilation. A pick up registered vibrations within	-

			the test cage	
	Habituation	-	Not stated as to how evaluated	-
c. Motor and sensory function at at adolescence (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Grip Strength	-	-	-
	Rotating Rod	-	-	-
	Hindlimb foot splay, landing foot spread	-	-	-
	Noiciception (hotplate, tail flick)	-	-	-
	Sensory irritation	-	-	-
	Somatosensory operant discrimination task	-	-	-
	Acoustic startle response and prepulse inhibition	-	-	-
	Auditory discrimination procedure	-	-	-
	Visual discrimination task	-	-	-
d. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (*e.g.* PND 60-70). Some tests to consider:	Conditioned taste aversion	-	-	-
	Active avoidance	-	-	-
	Passive avoidance	-	-	-
	Spatial Mazes, *e.g.* Morris water maze, Biel water maze, T-maze	Wooden Y-shaped maze at 3 and 7 months old	-	Radial arm maze starting at 90 days and continued for 7 weeks, and T-maze starting on approx. PND 165 and continued for 3 weeks
	Conditional discrimination, *e.g.* simple discrimination, matching to sample	-	-	-

	Delayed discrimination, *e.g.* delayed matching to sample, delayed alternation or repeated acquisition	-	-	-
	Eye-blink conditioning	-	-	-
	Schedule-controlled operant behaviour	-	-	-
	<u>Two criteria</u> need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	does not seem to be fulfilled for measure of memory (animals allowed to repeat the test until the completion of a criterion of ten correct responses in ten consecutive trials or until a maximum of 250 trials)	-	Seems to be fulfilled.
e. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not specified if required*	-	-	-
f) Other neurobehavioural tests included	*Not required*	-	-	-
3. Other endpoints included				
	Not required	Microdialysis: animals have probe inserted into cerebellum and are subsequently evaluated in freely moving environment of the Bioanalytical System). cGMP pathway evaluated	Evaluated MeHg in combination with PCB153. MeHg content using flameless atomic absorption spectrophotometry.	-
Author's conclusions				

Evidence of developmental neurotoxicity		PCBs 153 impair the ability to learn the Y maze task and disrupts the function of the glutamate-NO-cGMP pathway in the cerebellum of young (3 months) but not old (7 months) offspring.	PCB 153 induces neurotoxic effects on spontaneous behaviour, habituation and cognitive function. These effects are enhanced when co-exposed with MeHg.	PCB 153 at the high dose impaired learning of T-maze DSA task in adult females but not in males. No effects observed on working or reference memory in the radial arm maze.

* Shaded cells denote tests where a significant effect was observed in treated animals compared to controls (based on author's conclusions).

Table 8. Comparison of experimental design and endpoints in the OECD TG 426, the available guideline and non-guideline studies of developmental neurotoxicity for PFOS.

Reference	**OECD 426**	**Butenhoff *et al*., (2009)**	**Johansson *et al*., (2008b)**	**Onishchenko *et al*., 2011**
		Butenhoff JL, Ehresman DJ, Chang S-C, parker GA, Stump DG (2009) Gestational and lactational exposure to potassium perfluorooctanesulfonate (K^+PFOS) in rats; Developmental neurotoxicity. Reprod Toxicol 27: 319-330.	Johansson N, Fredriksson A, Eriksson P (2008b) Neonatal exposure to perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) causes neurobehavioral defects in adult mice. Neurotoxicol 29:160-169.	Onishchenko N, Fischer C, Wan Ibrahim WN, Negri S, Spulber S, Cottica D *et al*. (2011) Prenatal exposure to PFOS or PFOA alters motor function in mice in a sex-related manner. Neurotox Res. 19:452-461.
AIM	Designed to analyze the potential functional and morphological effects on the developing nervous system of the offspring that arise from exposure during gestation and lactation	Evaluate the functional and morphological changes to the nervous system in rats having gestational and lactational expisures to PFOS	Explore the developmental neurotoxic effects in mice neonatallyy exposed to PFOS, PFOA and PFDA during a critical period of brain development	To investigate whether prenatal exposure to PFOS and PFOA may have developmental neurotoxic effects at lower doses than previously established BMDL
	This guideline can be used as a separate study or incorporated into a reproductive toxicity and or adult neurotoxicity study, but it is critical to preserve integrity of both study types	Performed as a separate study	Performed as a separate study	Performed as a separate study
Stated that performed according to OECD TG 426 or other guidelines (specify)?		Yes, (EPA OPPTS 870.6300 and OECD TG 426)	no	No
Experimental design				
A. Test animals				
	Rat (commonly used strains)	Rat (Sprague Dawley)	Mouse (NMRI;Naval Medical Research Institute)	Mouse (C57BL/6/Bkl)

	Other rodents can be used; justification required, comparable days for exposure required if a different species or unusual strain	N/A	Justification for mouse model is provided with comparison to brain development and growth to that in humans	Justification for mouse model is provided with reference to the longer elimination half-life for PFOA in female mice compared to rats, making it easier to attain stable serum levels in this species
	Supplier of test animals to be provided	Charles River Laboratories, Inc	B&K, Sollentuna, Sweden	Scanbur BK, Sweden
	Number, age at start and sex of animals	Clearly stated	Clearly stated	Age not stated
	Housing conditions, acclimatization etc	Clearly stated, apart from acclimatization	Not stated	Clearly stated, apart from acclimatization
	Unique identification for each animal and litters	not stated	not stated	Not stated
	Ensure that a sufficient number of pregnant females are exposed to test substance to ensure an adequate number of offspring are produced (20 litters are recommended at each dose level)	Clearly stated, 25 mated females per dose group (no report on final # of litters). Additional 10 mated females were assigned as satellite phase rats to each of the four groups (parameters to be in companion article thyroid and gene expression)	Mated females not stated.	6 mated females exposed to PFOS, 10 mated females in the control
	Live pups to be counted and sexed	Survival evaluated daily	not stated	Not stated
	Litter size adjusted on or before PND 4 by random selection to yield a uniform litter size for all litters with equal males and females.	Clearly stated, litters adjusted to subset group A and B. Group A consisted of 20 pups of equal sex distribution per group on PND 4, while Group B consisted of 15 pups/sex/group.	Litters adjusted to 10-12 pups within 48 hours. No sex-based separation made.	Not stated

	Pup identification is required	not stated	not stated	Offspring injected with microtransponders on PND 21 used for identification
	Assignment of animals to tests: Pups should be assigned to endpoint assessment on or after PND 4. Both sexes from each litter in each dose group should be equally represented	Clearly stated, as per TG 426	At 4-5 weeks all females were killed and the male siblings were kept in litters (treatment groups) and raised in groups of 4-7 in a room for males only. "Ten mice were picked randomly from three to five different litters in each treatment group and were only tested once on each test occasion."	One or two offspring randomly selected from each litter. Assumed on PND 21. NOTE: Unclear if 1-2 offspring/sex/litter, this seems plausible as PFOS and control groups are stated to include 8 animals/sex but there were only 6 pregnant dams in each group.
Statistical Unit	The statistical unit of measure should be the litter (or dam) and not the pup.	litter	pup, stated that "Randomly selecting animals from at least three different litters will have the same statistical effect and power compared to the use of litter based studies to evaluate developmental neurotoxicity in neonatal mice"	Pup (as it seems more than 1 pup/litter was used)
B. Test conditions				
1. Administration of chemical/dosing *(Include information on vehicle as well as route of exposure)*	Most relevant to potential human exposure	Yes, oral	Yes, oral	Yes, oral
	Oral by gavage, in diet, drinking water or capsules	By gavage to mated females at volume of 5ml/kg	Oral by via gavage at PND 10	Dissolved in ethanol and applied onto food bits in a volume adjusted according to individual bw
	Other forms of admin (inhalation or demal) requires modification of procedures	N/A	N/A	N/A
	At least 3 dose groups and a concurrent control	4 dose groups + neg control; 0, 0.1, 0.3, 1.0 mg/kg/d	2 doses + neg ctrl;0, 1.4 and 21 umol/kg bw (= 0.75 or 11.3 mg/kg bw)	1 dose + neg ctr: 0, 0.3 mg/kg bw (per day not stated but assumed)

	Repeated exposure	Repeated, daily	single dose	Repeated, daily
	Mated females, starting on GD 6	Mated females from GD 0	Male pups only PND 10	Madet females from GD 1
	Motivation if starting exposure earlier	None given	N/A	None given
	Dose levels selected on any previous observed toxicity and kinetic data available for test compound or related materials	Dose selection based on published studies and previous testing of PFOS with the objective of avoiding significant neonatal toxicity	Dose levels selected on investigator's earlier studies on PCBs and PBDEs.	Dose level selected as being below previously reported BMDL
	High dose level should induce some maternal toxicity (eg weight loss). The lowest dose should not produce any evidence of maternal or developmental toxicity including neurotoxicity.	not stated	not stated	No high dose included
	Dose levels should be selected to allow for illustration of dose-response	not stated	not stated	N/A
	Positive controls not mentioned	Not included	Not included	Not included
2. Duration of exposure *(Add information if any comment has been made in the study concerning direct dosing of pups (e.g. via feed) and/or kinetic support for exposure via milk)*	GD6 to PND21	GD 0 to PND 20	PND 10	From GD1 throughout pregnancy
	Motivation for using other exposure duration if not GD 6 to PND 21	None given	Experiment design of neonatal exposure as laboratory has performed for several years, thereby generating historical controls as well as reproducibile developmental neurotoxicological data.	None given

	Exposure via lactation generally assumed but evidence of continuous exposure [during lactation] can be retrieved from *e.g.*, pharmacokinetic information, offspring toxicity or changes in bio-markers	not included	pups only exposed postnatally	Not included but refers to other literature stating that PFOS (and PFOA) can be excreted in milk but that lactational exposure does not seem to be as critical as in utero exposure in regard to developmental toxicity
	Direct dosing of pups can be considered if there is lack of evidence of continued exposure to offspring during lactation	No direct dosing of pups	Direct dosing of male pups	No direct dosing of pups
3. If other routes of administration than oral	Use OECD Guidance Document on Neurotoxicity Testing Strategies and Methods #43 to assist in the design of the studies	N/A	N/A	N/A
4. Duration of study; time for sacrifice/necropsy of dams and offspring	Maternal animals can be euthanized after weaning. Offspring to be humanely killed at PND 22 or at an earlier time point between PND 11 and 22, as well as at study termination.	dams were euthanized and subjected to a gross examination on PND 21, pups terminated and subjected to necropsy on PND 72.	not stated	not stated
5. Control for estrous cyclicity of female offspring	*Not specified if required*	not stated	only tested male pups	not stated
C. Endpoints*:				
1. Physical and developmental landmarks				
a. Body weight and clinical	Required	Provided	not provided	not provided

observations, weekly during pre-weaning and at least every two weeks at adolescence and adulthood	During treatment and observation period, detailed clinical observations to be performed for **dams** and **pups**, by trained observers unaware of the actual treatment.	Dams observed for signs of toxicity approx 1 hr following dose administration, extensive FOB conducted in offspring on PND 4, 11, 21, 35, 45 and 60	Brief statement in *Results* that no overt signs of clinical toxicity occurred during experimental period, but what was evaluated in not indicated.	--
	Where possible, observations to be made by the same technician	Not specified	Not specified	--
	Observations to be performed outside of home cage	Not specified	-	--
	"Unusual" responses with respect to eg activity level, need to be documented	-	-	--
	Both dams and pups need to be evaluated. Body weight best indicator for dam toxicity and pup physical development	Dams and pups weighed every 3 or 4 days until PND 21. Pups weighed every week thereafter until euthanasia	Not specified, pup weights reported on PND10 and PND28	Data not provided but stated in Results that exposed dams gained weight normally and that no differences in pup bw at birth were observed
b. Brain weight and Neuropathology, on PND 22 (or earlier between PND 11 and 22) and at termination	Fixation; Immersion <PND21, Perfusion >PND21	PND 21 and 72 pups perfused in situ with PFA, brains processed. On PND 72 also additional CNS and PNS	-	Brains collected from 4 pups (1/litter) at birth. Not stated how processed. Unclear if 4 pups in total of 4 pups/treatment group.
	Tissue selection need both CNS and PNS	Provided	-	Only brains
	Morphometric evaluation of tissue collected on PND 21 and at end of study	Done (PND 21 and 72)	-	Not stated

	Any neuropathological changes should be graded (grading scale should be defined) to allow for analysis of dose-response rel.	Not provided (no treatment related lesions noted)	-	Not provided
c. Sexual maturation	At adolescence (as appropriate)	Each male and female pup was observed for balanopreputial separation beginning on PND 35 or vaginal perforation beginning on PND 25, respectively	not provided	Not provided
d. Other physical landmaks for pup development (eg eye opening)	only required if these will provide additional information	All; FOB evaluated on PND 4, 11, 21, 35, 45 and 60	Not provided	Not provided
2. Functional/behavioural endpoints				
a. Behavioural ontogeny, at least two measures pre-weaning using the same pup (1 pup/sex/litter). Some tests to consider:	righting reflex	on PND 21 as part of FOB *(Stated tha NOT conducted on PND 4 and 11 "due to stage of development")*	-	-
	negative geotaxis	**-**	-	-
	motor activity (strongly recommended!)	Locomotor activity on PND 13, 17, 21 using the SDI Photobeam Activity System	-	-
b. Motor activity (including habituation) at pre-weaning (*e.g.* PND 13, 17, 21) and at adulthood (*e.g.* PND 60-70)	Motor activity should be monitored using an automated activity recording apparatus	Locomotor activity on PND 13, 17, 21 and 61 using the SDI Photobeam Activity System	Motor activity measured in 2 and 4 month pups in automated devise for spontaneous behaviour. Locomotion, rearing and total activity were evaluated with beams to detect movement and gramaphone-like pick up to detect vibrations	Locomotor activity at between 5-8 weeks of age (PND 35-56) using an automated video tracking system
	Habituation	included	included	Not provided

c. Motor and sensory function at adolescence (recommended PND 25±2) and at adulthood (e.g. PND 60-70). Some tests to consider:	Grip Strength	-	-	Muscle strenght evaluated in the ”Hanging wire test” at 3-4 months of age
	Rotating Rod	-	-	At 3-4 months of age
	Hindlimb foot splay, landing foot spread	-	-	-
	Noiciception (hotplate, tail flick)	-	-	-
	Sensory irritation	-	-	-
	Somatosensory operant discrimination task	-	-	-
	Acoustic startle response and prepulse inhibition	On PND 20 and 60 using the SR-Lab Startle Response System (same pups)	-	-
	Auditory discrimination procedure	-	-	-
	Visual discrimination task	-	-	-
d. Learning and memory post-weaning (recommended PND 25±2) and at adulthood (e.g. PND 60-70). Some tests to consider:	Conditioned taste aversion	-	-	-
	Active avoidance	-	-	-
	Passive avoidance	-	-	-
	Spatial Mazes, *e.g.* Morris water maze, Biel water maze, T-maze	Biel water maze on PND 22 for 7 consecutive days	-	-
	Conditional discrimination, *e.g.* simple discrimination, matching to sample	-	-	-
	Delayed discrimination, *e.g.* delayed matching to sample, delayed alternation or repeated acquisition	-	-	-

	Eye-blink conditioning	-	-	-
	Schedule-controlled operant behaviour	-	-	-
	<u>Two criteria</u> need to be fulfilled in learning and memory tests: 1) **original learning (acquisition**) need to be assessed as change across several repeated learning trials or (if single trial) with reference to a condition that controls for non-associative effects of the training experience, and 2) tests should include some measure of **memory** in addition to original learning	fulfilled	-	-
e. Ethology based anxiety tests, *e.g.* elevated plus maze test, black and white box test, social interaction test	*Not specified if required*	-	Elevated plus maze test at 4 months, apparatus made of plywood and videotaped	Elevated plus maze test at between 5-8 weeks (PND 35-56), apparatus made of grey plastic
f) Other neurobehavioural tests included	*Not required*	-	Nicotine-induced behaviour test at 4 months	Circadian activity in home cage (TraffiCage) at between 5-8 weeks (PND 35-56). Forced swimming test (for depression-like behaviour) at between 5-8 weeks (PND 35-56).
3. Other endpoints included				
E.g. Opthalmological, repro, organ weights	*Not required*	extensive FOB on PND 4, 11, 21, 35, 45 and 60	-	Liver weights at birth (n=4)
Author's conclusions				

Evidence of developmental neurotoxicity		The NOAEL for offspring developmental neurotoxicity was considered to be 0.3 mg/kg-d based on increased motor activity and a failure to habituate to the test environment noted in the 1.0mg/kg-d maternal dose-group males on PND 17. Serum concentrations reported in companion paper are several hundred times higher than those reported for females in the USA population	Neonatal exposure to PFOS on PND10 can cause persistent disturbances in spontaneous behaviour of adult NMRI male mice and modify the susceptibility of the adult cholinergic system	Exposure to PFOS during pregnancy caused alterations in motor behaviour in adult mice

* Shaded cells denote tests where a significant effect was observed in treated animals compared to controls (based on author's conclusions).

Table 9. Summary table on experimental design and results of all sixteen studies evaluated. Shaded cells denote tests where effects of DNT were observed (study author's conclusions).

Study	Compound	According to guideline?	Species/strain	Postnatal exposure	Motor activity (apparatus)	Learning & memory (method)	Other neurobehavioural (method)	Positive for DNT?
Stump *et al*. (2010)	BPA	Yes (OECD 426)	Rat (Sprague Dawley)	Indirect until third week of lactation, then also direct via feed	Kinder Scientific Motor Monitor System on PND 13, 17, 21, 61	Biel water maze on PND 22 and 62	-	No
Jones *et al*. (2011)	BPA	No	Rat (Long-Evans)	Indirect	-	-	Sexual behaviour between PND 90-120	Yes
Negishi *et al*. (2003)	BPA	No	Rat (F344)	Indirect	Supermex at 4 weeks of age (PND 28-34) and in the open field at 8 weeks	Active avoidance in a two-way shuttle-box at 15 weeks of age	Anxiety (open field) at 8 weeks	Yes
Ema *et al*. (2001)	BPA	Yes (for repro tox study)	Rat (Crj: CD (SD) IGS)	Indirect until PND 23 then direct via gavage	Open field at 5-6 weeks	T-maze at 6-7 weeks	Anxiety (open field) at 5-6 weeks	No
Beck (2009) (Biesemeier *et al*., 2011)	PBDE 209	Yes (OECD 426)	Rat (SD)	Indirect	Kinder Scientific Motor Monitor System at PND 13, 17, 21, 61, 120, 180	Biel maze swimming trial at PND 22, 62	-	No
Viberg *et al*. (2003)	PBDE 209	No	Mouse (NMRI)	Direct PND 3	Rat-O-Matic at 2, 4, 6 months	-	Anxiety in similar study	Yes
Viberg *et al*. (2007)	PBDE 209	No	Rat (SD)	Direct PND 3	Rat-O-Matic at 2 months	-	-	Yes
Rice *et al*. (2007, 2009)	PBDE 209	No	Mouse (C57BL/6)	Direct PND 2-20	Standard mouse operant test cage (Coulbourn) at PND 70 and one year of age	Schedule-controlled operant behaviour at 3 and 16 months of age	-	Yes

Gilmore *et al*. (2006)	Deltamethrin	Yes (draft OECD 426)	Rat (Wistar)	Indirect	Figure 8 maze at PND 13, 17, 21, 60, 120	Water maze at PND 60, passive avoidance at PND 22, 29	-	No
Eriksson and Fredriksson (1991)	Deltamethrin	No	Mouse (NMRI)	Direct PND 10-16	Rat-O-Matic at PND 17, 4 months	-	-	Yes
Piedrafita *et al*. (2008)	PCB 153	No	Rat (Wistar)	Indirect	-	Wooden Y-shaped maze at 3 and 7 months old	-	Yes
Fischer *et al*. (2008)	PCB 153	No	Mouse (NMRI)	Direct PND 10	Rat-O-Matic at 2 and 4 months	-	-	Yes
Schantz *et al*. (1995)	PCB 153	No	Rat (Sprague Dawley)	Indirect	-	Radial arm maze starting at 90 days and T-maze starting on approx. PND 165	-	Yes
Butenhoff *et al*.. (2009)	PFOS	Yes (OECD 426)	Rat (Sprague Dawley)	Indirect	SDI Photobeam Activity System on PND 13, 17, 21 and 61	Biel water maze on PND 22	-	Yes
Johansson *et al*.. (2008)	PFOS	No	Mouse (NMRI)	Direct PND 10	Rat-O-Matic at 2 and 4 months	-	Elevated Plus Maze and Nicotine-induced behaviour test at 4 months	Yes
Onishchenko *et al*. (2011)	PFOS	No	Mouse (C57BL/6)	Indirect (gestation)	In novel cage using an automated video tracking system on PND 35-56	-	Circadian activity in home cage at 5-8 weeks Elevated Plus Maze and forced swimming test at 5-8 weeks	Yes

Appendix

External reviewers and their affiliations*

Name	Affiliation
Melissa J. Beck	WIL Research Laboratories, LLC, Ashland, USA
John A. Biesemeier	Chemtura Corporation, West Lafayette, USA
Neil Carmichael	European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC)
Sandra Ceccatelli	Karolinska Institutet, Sweden
Kevin Crofton	US Environmental Protection Agency (USEPA), USA
Per Eriksson	Uppsala University, Sweden
Vicente Felipo	Centro de Investigacion Principe Felipe, Spain
Ulla Hass	Technical University of Denmark (DTU), Denmark
Sylvia Jacobi	Albemarle Europe, Belgium
Fraser Lewis	Syngenta, UK
Hellmuth Lilienthal	Ruhr University of Bochum, Germany
Sue Marty	The Dow Chemical Company, Midland, USA
Angelo Moretto	University of Milan, Italy
Hanna Silberberg	ICL Industrial Products America Inc, USA
Roland Solecki	Federal Institute for Risk Assessment (BfR), Germany
Todd Stedeford	Albemarle Corporation, Baton Rouge, USA
Matti Viluksela	National Institute for Health and Welfare, Finland

*The external reviewers provided written answers to the questions, which are outlined in the report. In no way did they influence the interpretation of the answers to the questions or to the conclusions of the report.

Institutet för miljömedicin
Box 210
171 77 Stockholm
http://ki.se/IMM