EXTERNAL SCIENTIFIC REPORT

Literature review on in vitro and alternative Developmental Neurotoxicity (DNT) testing methods[1]

[2]Ellen Fritsche, [2]Henrik Alm, [2]Jenny Baumann [3]Lieve Geerts, [4]Helen Håkansson, [2]Stefan Masjosthusmann, [3]Hilda Witters

[2]Leibniz Research Institute for Environmental Medicine (IUF), Group of Sphere Models and Risk Assessment, Auf'm Hennekamp 50, 40225 Düsseldorf, Germany

[3]Flemish Institute for Technological Research (VITO), Environmental Risk & Health, Boeretang 200, B-2400 Mol, Belgium

[4]Karolinska Institute (KI), Institute of Environmental Medicine (IMM), Unit of Environmental Health Risk Assessment, SE-171 77 Stockholm, Sweden

ABSTRACT

The goal of this systematic review performed under a contract with EFSA was the evaluation of information on assessment methods in the field of developmental neurotoxicity (DNT). Therefore, a systematic and comprehensive literature search and collection of scientific literature and all other relevant grey literature and website information (in English) from past 20 years until mid of April 2014 on the state of the art of in vivo DNT testing methods including novel and alternative non-mammalian models, in vitro test methods, in silico methods, read across and combination of testing methods in test batteries was performed. This systematic review identified a variety of methods covering early and later stages of neurodevelopment that have the ability to predict DNT of chemicals. Few in vivo models alternative to OECD TG 426 were identified. In general, the available published in vivo data is scattered and heterogeneous, making comparisons across exposure paradigms and compounds very difficult. Thus, to make more useful evaluations, publications with more consistent experimental designs are needed, and more focused in vivo-in vitro comparisons are needed to establish useful effect biomarkers and testing models. From the alternative methods, a testing strategy covering early and later neurodevelopmental stages (from stem cell to zebrafish larvae motor behaviour) can be assembled. For gaining regulatory acceptance, definition of biological application domains of alternative methods by performing e.g. in vitro - in vivo validation is needed. Moreover, protocols for cell-based and zebrafish assays need international standardization. With such standardized protocols, the test battery needs to be tested for its sensitivity and specificity by testing concentration-responses of known DNT positive and negative compounds across the different assays. Such data might then be used either for regulatory decision-making or compound prioritization in favour of a reduction of rodent guideline in vivo testing.

[1] Question No EFSA-Q-2013-00567

Any enquiries related to this output should be addressed to pesticides.ppr@efsa.europa.eu

Suggested citation: Fritsche E, Alm H, Baumann J, Geerts L, Håkansson H, Masjosthusmann S, Witters H, 2015. Literature review on in vitro and alternative Developmental Neurotoxicity (DNT) testing methods. EFSA supporting publication 2015:EN-778, 186 pp.

Available online: www.efsa.europa.eu/publications

KEY WORDS

literature review, DNT, developmental neurotoxicity, in vivo, in vitro, in silico, alternative organism

DISCLAIMER

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

Human health evaluations of active substances in pesticide formulations under Regulation (EC) No 1107/2009[2] require a series of animal testing as specified in Regulation (EU) No 283/2013[3]. Recognised testing methods include the internationally harmonised OECD Test guidelines, as well as methods listed in the Commission Communication No 2013/C 95/01[4]. To date, there is only a small amount of data available for developmental neurotoxicity (DNT) in support of the authorisation of pesticides. This is mainly due to the fact that there is no *a priori* requirement for pesticides or other chemicals to be tested for DNT effects prior to their registration and use under the present regulation. While developmental toxicity testing is mandatory, DNT studies are only to be carried out when relevant observations are made in other studies or when suggested appropriate on the basis of the mode of action of the active substance.

Although systematic DNT testing is lacking for pesticides, many chemicals have been shown to be neurotoxic in *in vitro* and *in vivo* assays and this, according to some authors[5], raises particular concern for the developing brain, inherently much more susceptible to toxic agents than adult brains. A recent review[6] listed about 200 chemicals known to be neurotoxic in humans, and just 5 of these substances have been firmly documented as causes of developmental neurotoxicity. Given the potential vulnerability of the developing brain, it is likely that many of these substances are capable of causing also developmental neurotoxicity but that such effects have not been captured because of insufficient testing.

Developmental neurotoxicity represents therefore an emerging issue and EFSA has been recently asked to provide an opinion on the developmental neurotoxicity of the pesticide active substances acetamiprid and imidacloprid[7]. In addition, DNT is considered as a 'critical effect of particular significance', as highlighted in the Annex II of the Regulation No 1107/2009.

At the regulatory level, the current OECD guideline[8] for DNT refers only to *in vivo* studies mainly performed in the rat: however, the experimental conditions of these tests are difficult to apply in a standardised manner, thus resulting in limited reproducibility of results. In addition, *in vivo* studies are unsuitable for screening large numbers of chemicals, due to the use of large number of animals and long duration of tests. Despite this, the use of *in vivo* studies based on animal models remains an important factor when data are to be extrapolated to the human case.

Reliable, fast and efficient screening and assessment tools are needed to improve the identification, and evaluation of chemicals with the potential to induce DNT. Ongoing international trends in DNT testing have nowadays shifted from the use of rodent studies to alternative methods with the aim to reduce the DNT *in vivo* testing (taking into account the 3Rs concept aiming at refining, reducing and replacing tests

[2] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC

[3] Commission Regulation (EU) No 283/2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market.

[4] Commission Communication in the framework of the implementation of Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market.

[5] Grandjean, P, Landrigan, PJ. Developmental neurotoxicity of industrial chemicals. Lancet 2006, 368:2167-2178.

[6] Bjorling-Poulsen, M, Raun Andersen, H, Grandjean, P. Potential developmental neurotoxicity of pesticides used in Europe. Environ Health 2008, 7(50):1-22

[7] M-2012-0336. Draft scientific opinion on the potential developmental neurotoxicity of acetamiprid and imidacloprid (EFSA-Q-2012-00958) unpublished

[8] OECD/OCDE Guideline 426 for the testing of chemicals-Developmental Neurotoxicity study, October 2007.

with vertebrates)[9] with integrated testing strategies which combine *in vivo* data sets with *in vitro* approaches.

TERMS OF REFERECE AS PROVIDED BY EFSA

The overall objective of the contract resulting from the present procurement procedure is to perform a literature search and analysis on the state of the art of DNT testing methods currently available or under development in order to support the peer review of active substances under Reg. 1107/2009. In particular, *in vivo* DNT testing methods including novel and alternative models (e.g. non-mammalian animal models), *in vitro* tests, *in silico* and read across methods, combination of test methods in test batteries and alternative tests that incorporate different DNT-relevant endpoints in high-throughput systems (thus testing a wide range of concentrations of test chemicals), need to be considered for this review and analysis.

This contract/grant was awarded by EFSA to: IUF – Leibniz Research Institute for Environmental Medicine

Contractor/Beneficiary: IUF – Leibniz Research Institute for Environmental Medicine, with subcontracting VITO – Flemish Institute for Technological Research and KI – Karolinska Institute.

Contract/grant title: Literature review on *in vitro* and alternative Developmental Neurotoxicity (DNT) testing methods

Contract/grant number: OC/EFSA/PRAS/2013/03

[10] Existing Approaches incorporating replacement, reduction and refinement of animal testing: applicability in food and feed risk assessment. EFSA Journal 2009; 1052:1-77.

GLOSSARY

Citation	refers to part of a publication, dedicated to the study of one compound for one endpoint in one test system, which is entered as a separate row in the data collection sheets
Publication	bibliographic reference
Study	refers to part of a publication dedicated to the evaluation of a specific model system, covering one or more endpoints for one or more chemicals
Hpf	hours post fertilisation
Validation	see Appendix L

1. Project Summary

1.1. Background

The intelligence of a population is its largest socio-economic potential. Therefore, it is of paramount importance to ensure individual evolvement of maximum intellectual potential. Poisoning disasters with e.g. polychlorinated biphenyls or mercury demonstrate that chemicals can interfere with developmental processes of the human brain (reviewed by Grandjean and Landrigan, 2006b). Thereby, a reduction in mean intelligence quotient (IQ) of a population by only 5 points leads to a decrease in the number of gifted and simultaneously an increase of less gifted by 57% each (Figure 1).

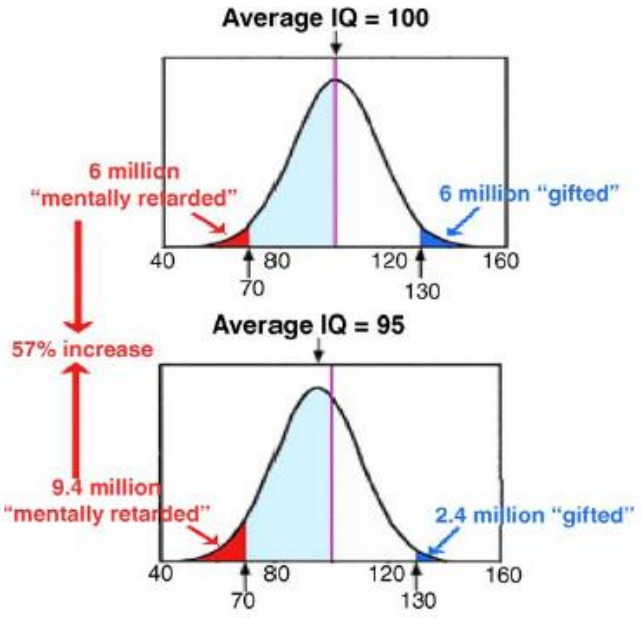


Figure 1: **Socio-economic impact of a reduction in the average population IQ by 5 points.** A reduction in the average population IQ by only 5 points means that the proportion of mentally retarded in a population increases by 57%, while the proportion of gifted decreases by 57% (source: Mt Sinai Children's Environmental Health Centre).

Animal experiments with rats are currently the gold standard in developmental neurotoxicity (DNT) testing. These are specified in the OECD test guideline 426 (OECD/OCDE, 2007). The OECD TG 426 (OECD, 2007) for 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). Other national or regional guidelines for DNT-testing have not been identified. Animal tests following OECD 426 are currently required as so-called Tier II studies for pesticides when there is an indication of toxicity from neurotoxicological or reproductive toxicity studies or tests for endocrine

disruption. European regulators are currently considering DNT studies as Tier II regulatory requirements not only for pesticides, but also for other chemicals and drugs. Also within the European REACH (Regulation on Registration, Evaluation, Authorization and Restriction of Chemicals) legislature, which requires testing of more than 30,000 substances with respect to their toxic potential, DNT tests are recommended if evidence from other *in vivo* studies exist (ECHA, 2008a, Rovida et al., 2010). These activities will lead to a significant increase in the world's DNT tests and thus to a large increase in the number of animals in DNT studies.

However, presently practiced animal tests for DNT are problematic in several ways: (i) they are ethically questionable, since for testing one substance about 140 dams and 1,000 juveniles are required; (ii) they last up to 12 months for the testing of one compound; (iii) they are very expensive (up to € 1,000,000/substance) and (iv) their predictivity for protection of the human brain is questionable. This uncertain transferability of animal experiments to humans (iv) is in part due to a lack of information on pharmaco-/toxicodynamics of the developing brain of rodents compared to humans (Dorman et al., 2001, Kaufmann, 2003). It is therefore essential to utilize alternative test methods in regulatory toxicology, which are able to predict DNT of compounds in a shorter time, less expensive and based on human-specific toxicity pathways (Crofton et al., 2011). However, so far there are no validated alternative methods available, which are able to predict DNT capacities of compounds for the human brain. Nevertheless, there has been a vast amount of method development possibly predictive for human DNT spanning from stem/progenitor cell methods to alternative organisms (reviewed in Lein et al., 2005, Coecke et al., 2007, Bal-Price et al., 2012, de Esch et al., 2012). Evaluation of such existing methods in a systematic way for identifying methods and/or tiered testing strategies for alternative DNT testing is thus necessary to improve the currently unsatisfactory DNT regulatory situation.

1.2. Review Question and Objectives

The **review question** is summarized as **"Which test methods or approaches are available to evaluate developmental neurotoxic effects of chemical exposure?"**

With this question the main goal of the project is a literature search and analysis on state of the art of DNT testing methods currently available or under development that should support the EFSA Pesticides unit with respect to the peer review of active substances under Reg. 1107/2009.

In particular, the 1st objective is to perform a systematic and comprehensive literature search and collection of all relevant information in English, German, Dutch and Swedish for the period 1990- April 2014 on the state of the art of 1) *in vivo* DNT testing methods, including novel and alternative models (e.g. non-mammalian animal models), 2) *in vitro* tests (cell-free and cell based) that allow testing of large number of chemicals in a high-throughput set-up, 3) *in silico*, 4) read across and 5) combination of testing methods in test batteries that include different DNT relevant endpoints. Next to validated methods, also methods currently under validation and methods at research stage should be considered (For definition of 'validation' see Appendix L). The search has to include publicly available peer reviewed research publications and publicly available 'grey literature'.

These results will feed into an overall analysis of the suitability of selected methods to support international regulatory assessments for DNT. Therefore, current international regulatory requirements for DNT testing are reviewed and critically analysed with regard to applicability, appropriateness, feasibility and predictivity for human risk assessment. This analysis will then allow identifying gaps and experimental needs with regard to DNT testing.

This comprehensive literature search on state of the art of DNT test methodologies and critical analysis for suitability to support regulatory assessments will be made from a Plant Protection Products regulatory point of view, with the focus to the needs and possible strategies for future research and risk assessment.

2. Methodology

2.1. Search strategy

The principle of the initial information gathering is illustrated in Figure 2 and Figure 3 and will cover a **time span of the last 24 years** (01st January 1990 to 15th April 2014). For the first search on DNT testing and non-testing methods six search strings were designed (Table 1). These search strings were designed in order to identify all publication that use *in vivo*, including novel and alternative models (e.g. non-mammalian animal models) and *in vitro* DNT testing and non-testing methods to evaluate developmental neurotoxic effects of chemical exposure. However, publications on models that predict chemical parameters relevant for DNT may be overlooked by the above mentioned search strings. Substances not reaching the bloodstream are considered to be of less toxicity concern. Hence, the blood-brain and placental barrier permeability may be an important parameter in the strategy for DNT screening. A complementary search for publications on *in silico* methods predicting parameters of biological barriers was performed (Table 1 'BB search string').

Table 1 Search strings for 'DNT' and 'BB' database search: Combination of the different search terms to make up each search string. For the complete search string these 6 strings are combined with the Boolean operator 'OR'. * is used as wildcard in the search.

String	Combined by 'OR'	AND	Combined by 'OR'	AND	Combined by 'OR'
1	DNT, Developmental neurotoxi*, Neurodevelopmental toxi*				
2	Neonatal, immature, prenatal, postnatal, perinatal, in utero, fetal	AND	neurotoxi*, toxi*	AND	development*'
3	Brain, CNS, nervous system	AND	developmental toxi*		
4	Developing brain, fetal Brain, young brain	AND	toxi* OR neurotoxi*		
5	Neonatal/prenatal/postnatal/perinatal/in utero/fetal exposure	AND	behaviour, behavior, neurobehaviour*, neurobehavior*		
6	Embryo*	AND	motor activity, locomotor activity, neurotoxi*		
BB search string	Blood brain barrier*, Brain blood barrier*, Placental barrier*	AND	*in silico*, computational, prediction*, (Q)SAR (Q)SARs, QSAR, QSARs		

Articles in English, German, Dutch and Swedish language will be taken into consideration. The search results of the DNT search are expected to contain every method dealing with developmental neurotoxicity. The search results of the biological barrier search are expected to contain every method dealing with the modelling of biological barriers. The first bibliographic database searched was 'Web of

Science®' (http://apps.webofknowledge.com), as it seems to provide the most comprehensive collection of articles (15710 with the DNT search terms). Next, the information source PubMed (http://www.ncbi.nlm.nih.gov/pubmed) was searched. As 'Science Direct' publications are embedded into PubMed, no extra search needs to be performed in this database.

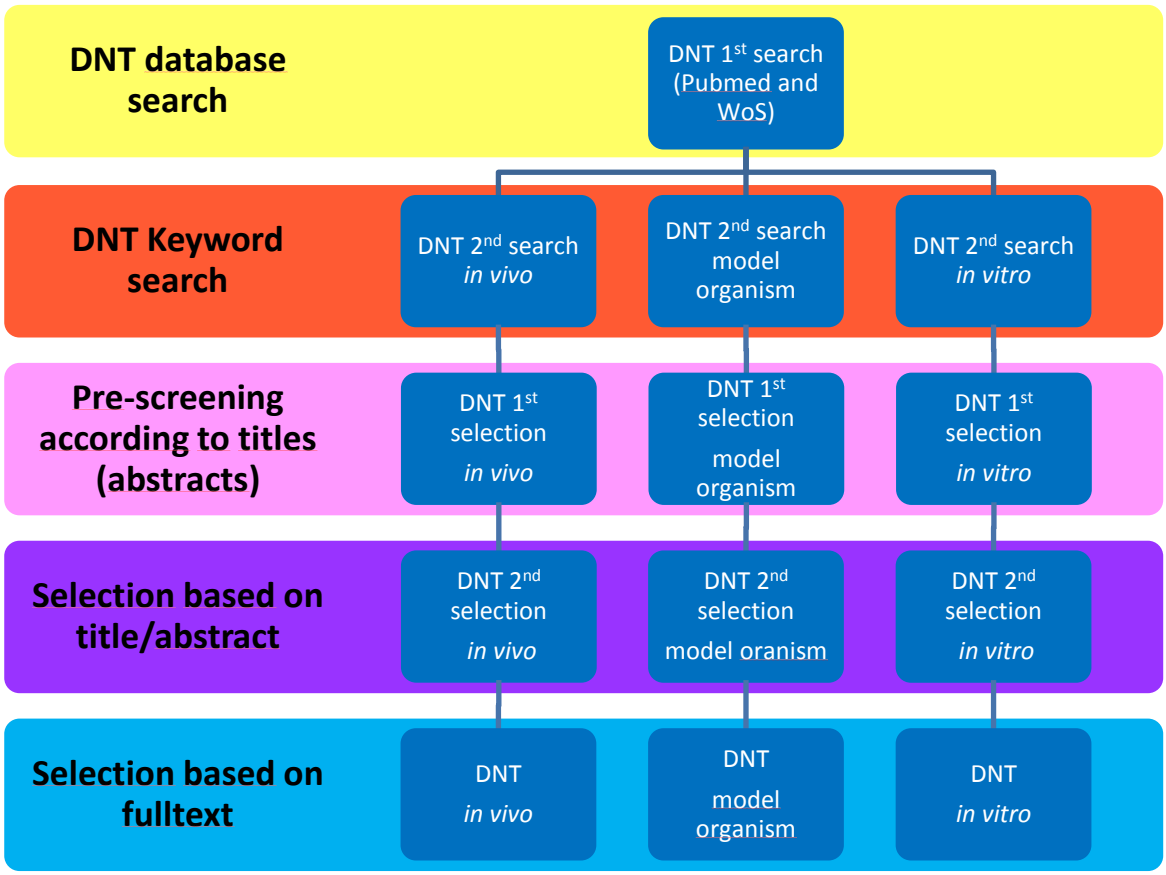


Figure 2: **Graphical abstract of primary search/selection strategy for peer-reviewed literature:** The databases Pubmed and Web of Science (WoS) were searched with the **DNT search strings**. Results were imported into EndNote. Within EndNote, a second search with additional keywords was performed. Results were divided into methods for *in vivo*, alternative organisms and *in vitro*. Subsequently a first manual pre-screening was performed to eliminate publications that to not deal with DNT assessment methods. Within the selection i) based on abstract/title, ii) based on fulltext, articles were select or excluded due to pre-defined criteria.

2.1.1. Search in Web of Science

The six different search strings (Appendix A) were entered one by one into the search field 'Topic' which includes a search in title, abstract, author keywords and keywords plus® of the article. The timespan was set from 1990 to 2014. Only for search string #6 the timespan was set from 2004 to 2014. This restriction of the timespan was used to limit the search results to a feasible amount. Without this restriction, we would have received more than 20 000 publications just out of WoS. Search string 6 was designed to obtain publications on methods that use non-mammalian animal models, especially zebrafish embryos. As scientific research using zebrafish as a model organism is just been growing in the past 10 years (Kinth et al., 2013), we expect that no relevant publication will be missed due to this restriction. The search was performed on the 22nd of April, which means that all publications that had been published until this date are considered in this literature review. To combine the search terms within one search string, all search terms combined by the operator 'OR' were copied in one field, search terms combined with an 'AND' are

copied in a second or third search field. Each search was automatically saved in the search history. After all searches were conducted, they were selected in the search history and combined by the Boolean operator 'OR'. The search was opened and the document types: 'case report', 'review' and 'meeting' were excluded. Additionally the search was limited to English, German, Dutch and Swedish language. All references were imported into the EndNote library 'Systematic review on DNT_search results' and saved in the group set 'DNT 1st search' and the folder 'Web of Science'. The procedure was the same for the BB search with the exceptions that there was no need to combine different search strings and there was no exclusion of document types or restriction to language. The references were imported into the EndNote library 'Systematic review on biological barrier_search results' and saved in the group set 'BB 1st search' and the folder 'Web of Science'.

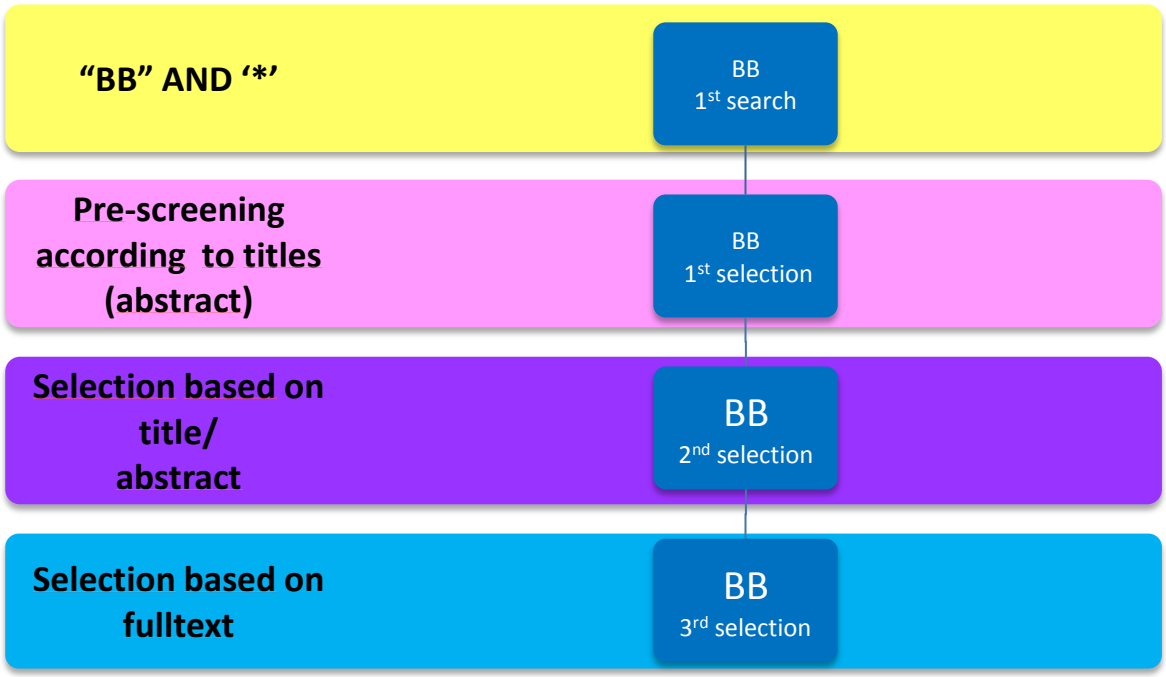


Figure 3: **Graphical abstract of primary search strategy for peer-reviewed literature concerning Biological Barriers (BB) and '*in silico*' methods**: The databases Pubmed and Web of Science were searched for the BB search terms 'Blood brain barrier' OR 'Brain blood barrier' OR 'Placental barrier' AND '*'. Thereby, '*' was '*in silico*' OR 'QSAR' OR (Q)SAR OR computational OR prediction. Results will be imported into EndNote.

2.1.2. Search in Pubmed

A search string was prepared that combined the six different search stings with the Boolean operator 'OR' (Appendix A). This final search string was copied into the PubMed advanced search builder. The search was performed in 'Title/Abstract', which includes a search in title, abstract and author keywords of the article. The timespan was set from 01.01.1990 to 30.04.2014 including all articles that were published until the 22nd of April (day the search was performed). The search was opened and all references were imported into the EndNote library 'Systematic review on DNT_search results' and saved in the group set 'DNT 1st search' and the folder 'Pubmed'. The procedure was the same for the BB search. The references were imported into the EndNote library 'Systematic review on biological barrier_search results' and saved in the group set 'BB 1st search' and the folder 'Pubmed'.

2.1.3. Searches in EndNote

To filter the DNT relevant articles from all search results and thus limit the number of not relevant articles, we searched all articles within EndNote for a set of keywords (Appendix B). These keywords are grouped in four categories: I. method related, II. target system/organ related, III. toxicological endpoint related and IV. study design related and were divided according to the individual method types : e.g. '*in vivo*' and 'mammal*' for the *in vivo* methods, '*in vivo*', 'non-animal alternative', 'non-mammal*', 'zebrafish', 'sea urchin' and 'nematode', 'QSAR', '*in silico*' for alternative organisms & *in silico* methods and '*in vitro*', 'cell based', 'cell culture' for the *in vitro* methods. Keywords from group I. and II. and keywords from group III. and IV were each combined with the operator 'OR'. The two combined groups were than combined with the operator 'AND' ((I. OR II.) AND (III. OR IV)).

For the biological barrier search, no additional keyword search was necessary, as relevant keywords were already included in the database search.

Before starting with the keyword search in EndNote all articles retrieved from WoS and PubMed were copied into the subfolder 'complete'. In this folder all duplicate articles (based on same title and year) were identified by EndNote and deleted. To restrict the search to articles dealing with developmental neurotoxicity, all studies that deal with Dinitrotoluene, which is also abbreviated with DNT had to be excluded. Therefore, we ran the following search in the folder 'complete' (Figure 4).

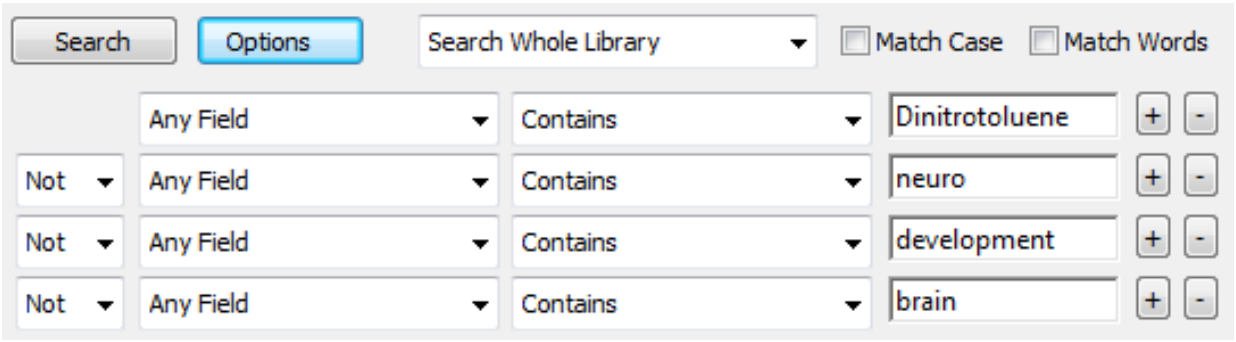


Figure 4 Search in EndNote to exclude all studies that deal with Dinitrotoluene and not Developmental neurotoxicity or brain.

All hits were deleted from this folder.

To exclude review articles from Pubmed (for WoS they were excluded before importing to EndNote) we ran the following search again in the folder 'complete' in the field 'Notes' (Figure 5).

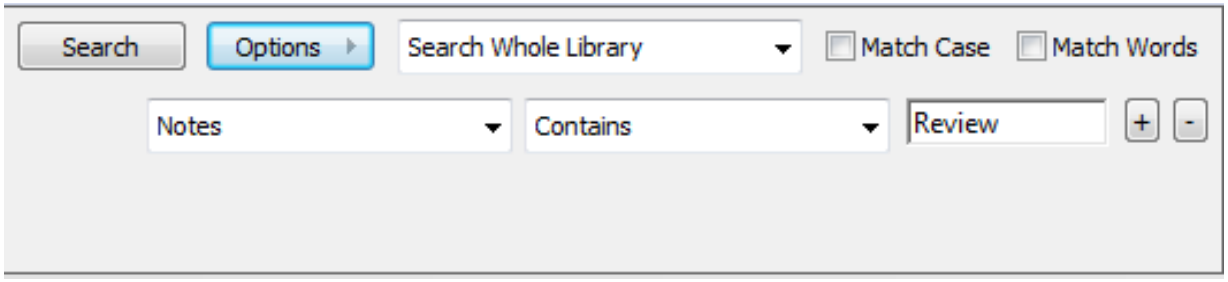


Figure 5 Search in EndNote to exclude all review articles from Pubmed.

All hits were deleted from this folder and saved in the folder 'Review articles (Pubmed)'.

After deletion of all duplicates, dinitrotoluene- and review articles we retrieved 17 272 publications.

The first keyword search was performed in the folder 'complete'. This folder was searched for all keywords from group I. and II. for '*in vitro*' (Appendix B) in all fields of the articles. Among all fields, the relevant are title, keywords and the abstract of the article. Search results were copied into the group set 'DNT 2nd search' and the folder '*in vitro* after Group I+II'. The same search was repeated with the keywords from group I. and II. for 'AO' and '*in vivo*'. The results were copied into the group set 'DNT 2nd search' and the folder 'AO after Group I+II' or '*in vivo* after Group I+II' respectively. The second search was performed in the folder '*in vitro*/AO/*in vivo* after group I+II' with the keywords from group III. and IV. for each model type (*in vitro*/AO/*in vivo*) (Appendix B) again in all fields. Search results were copied into the folder '*in vitro*/AO/*in vivo* after Group III+IV' of the respective partner. As some of the keywords are identical for the individual method types and because different keyword combinations can be found in one article it might be that one article appears in more than one folder. To avoid that one articles appears in different folders, all articles from the folders 'AO/*in vivo* after Group III+IV' were copied into the folder 'AO final (without *in vitro*)' or '*in vivo* final (without *in vitro* & AO)'. All articles that were already in the folder '*in vitro* after Group III+IV' were deleted from final 'AO' and '*in vivo*' folder additionally all articles that are in the final 'AO' folder were deleted from the final in 'vivo' folder.

Appendix C shows a screenshot of the EndNote library with the folder structure and the number of articles after different steps of the search.

Some articles in the EndNote library were imported without abstract. Most of these articles were not captured in the keyword search as the search field 'abstract' is missing. To avoid missing one of these articles, we decided to filter all studies with missing abstract from the folder 'complete' by searching for an empty abstract field (Figure 6).

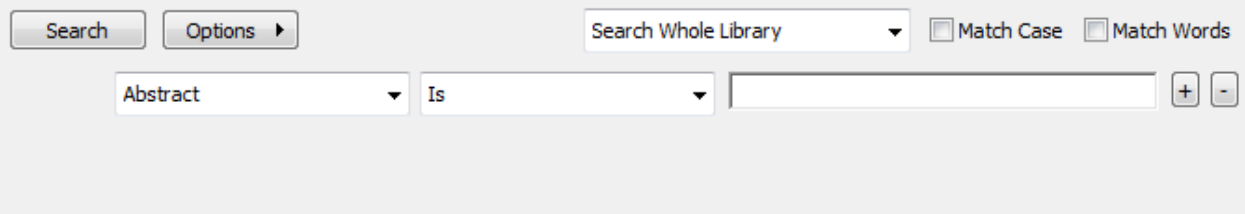


Figure 6 Search for studies with missing abstract

The results of this search were copied into the folder 'missing abstract' and screened separately. To make sure that no articles appears in one of the folders 'after keyword search' and the 'missing Abstract' folder, we deleted all articles that are already captured by the keyword search from the folder 'missing abstract'.

2.1.4. 'Grey' literature search

Data mining for 'grey literature' is performed by a different strategy than scientifically published information involving a peer-review process.

We performed a search on all websites listed in Appendix D in the 'Search' function of the respective websites with the keywords 'DNT', 'Developmental Neurotoxicity' and 'Neurodevelopmental toxicity' each. For websites related to read-across, models for ADME characteristics and QSAR models we used the addition search term 'Placental barrier'. Results of this search were screened for their relevance in the development of DNT testing methods. Relevant results were collected and saved for later selection/evaluation. In case the website search retrieved more than 100 hits for the combined search terms the search on this website was omitted. In addition to the keyword search, we designed an email (Appendix E) and contacted the contact persons of each of the above-mentioned website asking if they

can provide information regarding unpublished DNT methods. If no response was received 2 weeks after sending the emails requesting information, a reminder was sent. A contact list with specification on response success was prepared (Appendix D). Additionally we read background information on the *in silico* models for reproductive and/or developmental toxicity to see whether developmental neurotoxicity is addressed. We also contacted Dr. M. Cronin of the John Moores University of Liverpool who is corresponding author for a publication on a QSAR model developed for placental barrier diffusion, to ask for information on recent developments in that field (Hewitt et al., 2007).

Personal contacts were utilized to contact organizations/consortiums currently involved in DNT method development. Therefore, we designed another email that contained the question of unpublished methods for DNT testing, which was sent to the personal contacts by each of the partners. Specifically, we contacted:

- Coordinator of Denamic (Pim Leonards), Coordinator of SEURAT-1 (Michael Schwarz), EPA (Stefanie Padilla), Coordinator of ESNATS (Jürgen Hescheler), Coordinator of AXLR8 (Horst Spielmann), EURL-ECVAM (Anna Price), EPA (Kevin Crofton), ZEBET (Andrea Seiler), HESI, CAAT-US (Helena Hogberg), CAAT-Europe (Marcel Leist)

All the information (from website search, website contact, personal contact) was collected and sorted with regard to the specific topics: *in vitro*, *in vivo* alternative organisms, *in vivo* rodents and *in silico*. Subsequently, results were distributed to the related experts for further selection.

The language restrictions for 'grey' literature were in concordance to literature from bibliographic databases English, German, Dutch and Swedish.

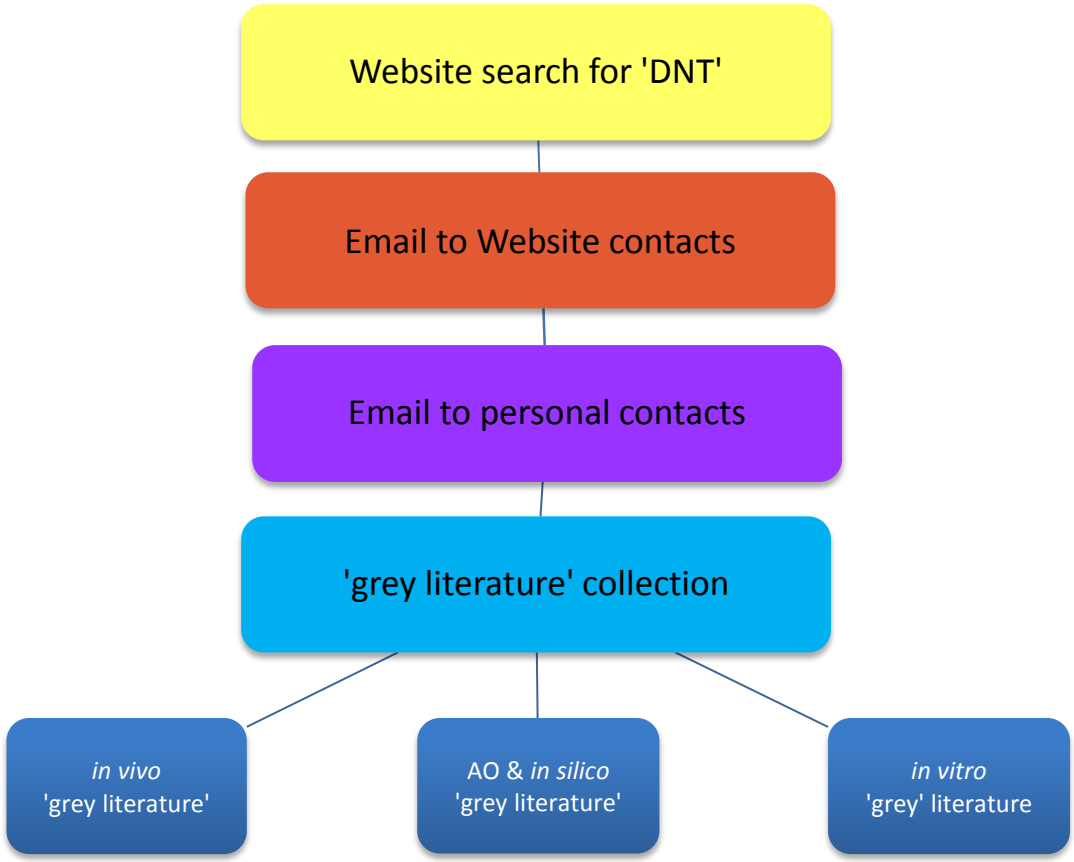


Figure 7: Graphical abstract of primary search strategy for 'grey' literature: Websites of national and international organization, international societies, EU-projects, industrial stakeholders and NGO's were searched for the terms 'DNT', 'developmental neurotoxicity' and 'neurodevelopmental toxicity'. Furthermore an email was send to these Website contacts, asking for information on DNT methods. Results were collected, screened and distributed by DNT topics: *in vitro*, *in vivo* alternative organisms, *in vivo* rodents and *in silico*.

2.2. Selection

2.2.1. Pre-screening of publications

The first selection process was a pre-screening based on titles and -if required- abstract of all research publications in the group set 'DNT 2nd search' and the respective 'final' folder of the individual method type (*in vitro, in vivo*, AO) or the group set 'BB 1st search' and the folder '*in silico*'. Within this pre-screening potentially relevant publications and publications with uncertainty with regard to their relevance were selected. Obviously non-relevant research publications, review articles, workshop reports or other secondary sources were dismissed. DNT or BB relevant secondary sources were stored in a different folder called 'DNT/BB relevant (pre-screening)' in the group set 'Secondary literature'. For all publications with missing abstract the decision for study selection was only based on the title of the study. In case of uncertainty the study was selected.

The pre-screening resulted in new group set entitled 'DNT 1st selection' with the subfolders '*in vitro*', 'AO' and '*in vivo*' or 'BB 1st selection' with the subfolder '*in silico*'. In this process 4326 publications were selected for the 'DNT' search and 221 for 'BB' search. The number of all selected publications in each of the selection steps is summarized in Figure 8.

The decision process in the study selection for pre-screening was as follows:

For *in vitro*:

- Is the publication a review article, workshop report or other secondary source?
  - If yes, does it deal with DNT?
    - If yes, publication is marked as secondary literature
    - If no, publication is excluded
  - If no, next criteria is checked
- Does the method used deal with developmental neurotoxicity (not neurotoxicity, not neurodegenerative diseases)?
  - If yes, is it a cell based method in human or rodent cells?
    - If yes, publication is selected
    - If no, is it an animal study performed in zebrafish (early life stage), *D. rerio*, sea urchin or nematode, *C. elegans* with molecular, histopathological and/or behavioural readout OR animal study performed in mice or rats with molecular, histopathological and/or behavioural readout OR any *in silico* method that models biological barriers
      - If yes, publication is selected for 'AO' OR '*in vivo*'
      - If no, publication is excluded
  - If no, leave it in the folder

For AO:

- Is the publication a review article, workshop report or other secondary source?
  - If yes, does it deal with DNT?
    - If yes, publication is marked as secondary literature
    - If no, publication is excluded
  - If no, next criteria is checked
- Does the method used, deal with developmental neurotoxicity (not neurotoxicity, not neurodegenerative diseases)?
  - If yes, is it an animal study performed in zebrafish (early life stage), *D. rerio*, sea urchin or nematode, *C. elegans* with molecular, histopathological and/or behavioural readout?
    - If yes, publication is selected
    - If no, is it a cell based method in human or rodent cells OR animal study performed in mice or rats with molecular, histopathological and/or behavioural readout
      - If yes, publication is selected for '*in vitro*' OR '*in vivo*'
      - If no, publication is excluded
  - If no, publication is excluded

For *in vivo*:

- Is the publication a review article, workshop report or other secondary source?
  - If yes, does it deal with DNT?
    - If yes, publication is marked as secondary literature

        - If no, publication is excluded
    - If no, next criteria is checked
- Does the method used, deal with developmental neurotoxicity (not neurotoxicity, not neurodegenerative diseases)?
    - If yes, is it an animal study performed in mice or rats with molecular, histopathological and/or behavioural readout?
        - If yes, publication is selected
        - If no, is it a cell based method in human or rodent cells OR an animal study performed in zebrafish (early life stage), *D. rerio*, sea urchin or nematode, *C. elegans* with molecular, histopathological and/or behavioural readout OR any *in silico* method that models biological barriers
            - If yes, publication is selected for '*in vitro*' OR 'AO'
            - If no, publication is excluded
    - If no, publication is excluded

For *in silico* (BB search):

- Is the publication a review article, workshop report or other secondary source?
    - If yes, does it deal with blood-brain barrier, placental barrier or ADME?
        - If yes, publication is marked as secondary literature
        - If no, publication is excluded
    - If no, next criteria is checked
- Does the article deal with blood-brain barrier, placental barrier or ADME on biological barriers?
    - If yes, is it an *in silico* model?
        - If yes, publication is selected
        - If no, publication is excluded
    - If no, publication is excluded

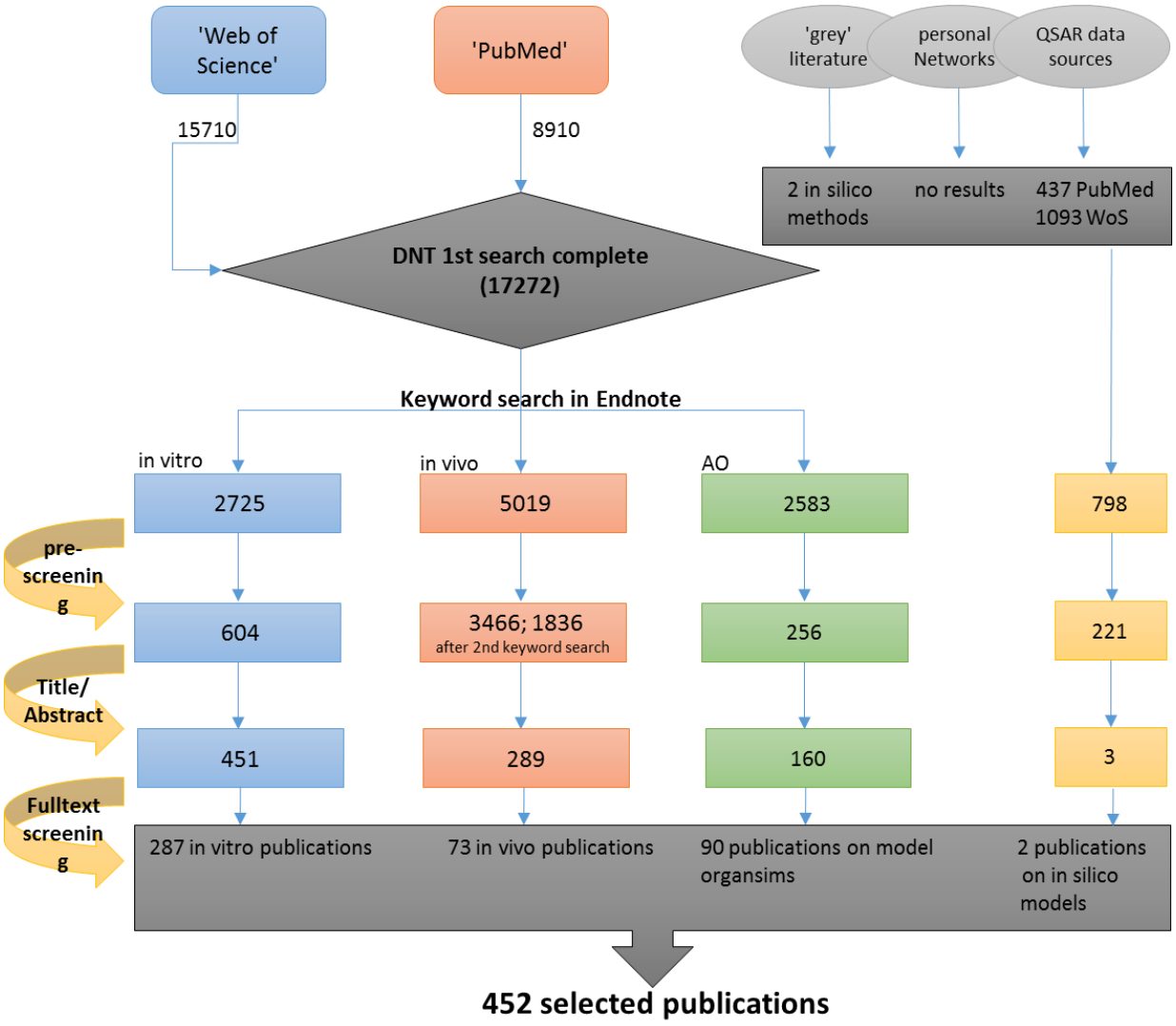


Figure 8 Graphical summary of publication selection

2.2.2. Selection based on title and abstract

The second selection process was an unmasked assessment in which titles and abstracts of the pre-selected publications were screened according to the pre-defined criteria (see below). In case of uncertainty, a second expert screened the publication and both experts made the final decision (in case of uncertainty the second expert decided for inclusion). In case of uncertainty for the '*in vivo*' the publication was not screened by a second expert, it was directly included. For documentation of this selection, the EndNote record number, the first author and the publication year of each article were transferred to an Excel file. In this file, 'yes' for included or 'no' for excluded was documented (Appendix F). In case of exclusion the argumentation for publication exclusion was also given in the excel sheet.

Publications, which were included based on the following criteria, were copied into the group set 'DNT 2nd selection' and into the respective folder (*in vitro*, *in vivo*, AO). The publications included from the

EndNote library 'systematic review on biological barrier search results' were copied into the group folder 'BB 2nd selection' with the Subfolder '*in silico*'.

Publication was excluded if one of the following criteria was met:

- Exposure is later than PND21 **(*in vivo*)**
- Exposure is other than chemical or particle (e.g exposure to mixtures or radiation)
- Endpoints used assess neurodegenerative diseases
- Transgenerational effects or multiples exposures of different substances were investigated **(*in vivo*)**
- Publication is on neurotoxicity, not DNT
- Publication shows no endpoint which is specific for brain development
- Species other than human or rodent
- Publication is not on DNT/Study does not use a DNT Method
- Publication investigates disease models
- Exposure and/or effect evaluation of an non alternative life stage **(AO)**
- Publication is secondary literature
- The test method does not have the ability to measure the intended endpoint or it is not the right application for the studied endpoint.
- Tumor cells are not differentiated for studying a DNT endpoint
- No compound is tested in the publication
- Publication is not on a model, but a consideration on a modelling approach (yound life stage not mentioned; ***in silico***)
- Publication on model evaluation for chemical properties (not the placental barrier permeation or blood-brain barrier for the young state; ***in silico***)
- Publication deals with drugs: e.g. structural and physical-chemical properties, artificial membrane permeation (***in silico***)
- Publication is not on the young life stage (***in silico***)
- Publication is on model development for drugs on diseases not relevant for the younger life stage, such as Alzheimer? **(*in silico*)**
- In the publication existing models are used; the models are not on the young life stage (***in silico***)
- Publication is not about *in silico* models (***in silico***)

For *in vivo*:

A general challenge in the selection process was the unexpected high amount of publications retrieved. This was especially the case for the *in vivo* selection with 3466 articles for title/abstract screening. In order to limit the number of articles to a feasible amount we included an additional keyword search and set up very conservative criteria for the title abstract screening. Both measures aimed at limiting the selection to publications with any type of mechanistic (molecular) readout that may be combined with a behavioral read out (methods from OECD TG 426). The rationale behind this was that we were searching for methods that give additional information to the OECD TG 426. Based on the additional search and exclusion criteria all publications that use the TG 426 or only parts of the TG 426 and no additional molecular readout were excluded (as they are no alternative approaches). Another exclusion criterion aimed at the scientific quality of the studies excluding publications with less than four treatment groups (control plus three doses). This criterion is in accordance with the OECD TG 426 and was chosen to exclude methods that cannot be validated in terms of their potential to complement or replace the guideline study.

Similar to the keyword search in 2.1.3 we searched for the keywords listed in Appendix B in the group set 'DNT 1st selection' and the folder '*in vivo*'. All results were copied into the folder '*in vivo* (after 2nd keyword search' and screened for title and abstract according to the criteria (see above). Additional exclusion criteria that were just applied to *in vivo* studies were:

- The method gives no additional mechanistic (molecular) information other than assessed by the OECD TG 426 (neuropathology, sexual maturation, clinical observations, behavioral ontogeny, sexual maturation, motor activity, motor and sensory function, learning and memory)
- Less than 4 treatment groups are used (control plus 3 doses)

This selection step led to a total of 900 publications for the DNT search (451 *in vitro*, 289 *in vivo*, 160 AO) and 3 publications for BB search, which qualified for full-text screening.

For the full-text screening, full text copies of most publications were retrieved with the automated full text search of EndNote. All full texts found in this process were automatically attached to the Reference in 'pdf' format. For all references for which EndNote was not able to retrieve the full text automatically, a manual search was conducted.

Description of search procedure: e.g.

1) via google scholar
2) via contacting the corresponding author of the publication by email.

The file is than manually attached to the respective EndNote reference.

In case the author contact was not available or the contact did not answer our request, the studies were excluded with the exclusion criteria 'no fulltext' (Appendix G).

2.2.3. Selection based on full text

This third selection process was an unmasked assessment in which the experts independently screened the full text of all publications selected based on title and abstract screening. The expert than made a decision for inclusion or exclusion based on the selection criteria listed in 2.2.2 and the additional 'fulltext' criteria below. In case of uncertainty, a second expert screened the publication and both experts agreed on a final decision. The selection process is documented in an Excel file as described in 2.2.2 (Appendix G).

Additional criteria for full text screening

Publication was excluded if one of the following criteria was met:

- The publication does not give sufficient information (information on at least endpoint, test method, compound, compound concentration, exposure time and exposure scheme should be available)
- Publication investigates a toxicity mechanism with no real compound testing (e.g. compounds are specific inhibitors)
- Publication is an operational procedure or guidline rather than a scientific publication
- Language is other than English, German, Swedish or Dutch
- Full text is not available

This selection step led to a total of 451 publications for the DNT search (287 *in vitro*, 73 *in vivo*, 91 AO) and 2 publications for BB search, which were summarized in the data collection sheets (see 2.3).

2.2.4. Selection of 'grey' literature

Within data mining of grey literature the consortium agreed that 'work in progress', e.g. granted projects on method validation, ring trials, etc. are important to recognize. However, methods, which are developed and presented in various media, but have not yet undergone scientific peer review e.g. through a standard scientific publication process should not be taken forward in the process on the same grounds as publications from the peer-reviewed literature since such data can be preliminary and open to final adjustments. Due to this fact the following selection criteria were applied for 'grey literature':

- Information on granted projects currently being performed
- Information containing preliminary results of on-going projects not yet finished

2.3. Data collection

Data collection sheets were prepared with respect to the biological method type (*in vitro*, alternative organism or *in vivo*). These sheets were designed to summarize all information from the selected scientific papers that are relevant for an evaluation and comparison of the different test systems with regard to the endpoints they are able to assess.

To fill the data collection sheets, the experts read all selected publications and extracted the following information.

For '*in vitro*' and 'AO'

Record number, first author, journal and year

- To identify each publication.

Method type

- To distinguish the three different method types, *in vivo*, *in vitro*, alternative organisms or a combination of these. In case one publication presents data on more than one method type the information is collected in the respective data collection sheet.

Test system, cell type (species, celltype, brain region; for '*in vitro*' or species, life stage for 'AO')

- To classify the test system that is used in each publication. The test system/cell type there by as any animal, cellular or subcellular system used in a study. Each test system is entered in a separate row.

Serum (excluded for 'AO' as it is not relevant)

- Serum can alter chemical properties *in vitro* and thus may change the biological response to a chemical. To identify the reason for different effects of test systems after chemical exposure it can be crucial to know if the system was treated in the presence of serum (yes/no).

Grouped endpoints (*in vitro*), Endpoints

- The biological or chemical process, response or effect assessed by a test method. A grouping of the endpoints according to cell type specific and general cellular effects was necessary to be able to deal with the heterogeneity of endpoints from the different publications. Each endpoint assessed with one test system was entered in a separate row.

Multiples endpoints

- To see if more than one Endpoint was assessed with one test system in one publication (yes/no).

Test method

- The process or procedure used to obtain information on the biological effects of a substance or agent. The test method describes the method that is used to do the endpoint measurement. In case of gene or protein expression this column gives the genes/proteins that were analysed. If more than 4 genes/proteins are analysed, test method is 'multiple' and the genes/proteins were listed in the comments.

Detection methods

- Gives the technical methods that was used to measure the endpoint.

Endpoint-specific controls

- Shows if endpoint-specific controls were used (yes/no)?

Compound

- Each compound with each endpoint in one test system was entered in a separate row. If the combination of different test systems, endpoints and compounds exceeded 20 rows for one publication, the compound was entered as 'multiple' and all compounds were listed in the comments.

CAS number

- To identify the compound

Compound classification a priori

- Gives a classification off compounds with respect to effects on the developing nervous system. See below for an argumentation of classification and Appendix H for the a priori classification of each compound.

Concentration range

- Gives the range of exposure concentration in 1x 10^3 steps.

Exposure scheme

- Shows in which developmental period/culture time the system is exposed to the compound.

Exposure duration

- Gives the duration of compound exposure.

Effect analysis

- Describes how an effect on this endpoint is analysed (e.g EC_{50}, concentration-response, induction/reduction).

Effect concentration

- Gives the concentration range in which the endpoint is affected (in 1x 10^3 steps).

Harzard

- Shows if the endpoint was affected (yes/no).

Effect on viability

- Shows how the viability is affected in comparison to an effect on the endpoint (e.g viability is more sensitive, less sensitive than endpoint)

Data analysis

- Describes how the data was analysed (quantitative, qualitative, statistical)

Statistical evaluation

- Describes the statistical method used for data evaluation.

n =

- Gives the number of biological replicates (one 'n' means one independent experiment)

Throughput

- Shows if a test system with this specific endpoint is adaptable for high/medium throughput.

Category of reliability, justification for reliability category

- Gives the reliability of the endpoint measurement in this test system and the justification for this reliability.

Comments, summary

For '*in vivo*':

Record number, first author, journal, year, method type,

- See '*in vitro*' and 'AO'

Species/Strain, sex

- To classify the test system that is used in each publication. Each test system is entered in a separate row.

number of animals per dose group, number of dose levels

Brain part evaluated

- Shows which part of the brain was evaluated.

Parameter monitored

- Describes the parameter or outcome, which was monitored in the study (comparable to endpoint for '*in vitro*'/'AO'). Each parameter assessed with one test system was entered in a separate row.

Multiple parameters monitored

- To see if more than one parameter was assessed with one test system in one publication.

Method to monitor this parameter

- Gives the method that is used to measure the parameter. In case of gene or protein expression this column gives the genes/proteins that were analysed. If more than 4 genes proteins are analysed, test method is 'multiple' and the genes/proteins were listed in the comments.

Detection methods

- See '*in vitro*' and 'AO'

Compound, CAS number, compound classification a priori, multiple test compounds, concentration/dose range, Exposure scheme, exposure duration

- See '*in vitro*' and 'AO'

Exposure scheme

- Shows in which developmental period the animal is exposed to the compound.

Exposure duration

- Information about the total length of the exposure

Expini

- Time point for initiation of exposure

Expterm

- Time point for the termination of the exposure

Administration route

- The route by which the compound entered the animal. "Oral gavage" and "oral injection" have been collected under "oral".

Vehicle, diet

- Gives information on the vehicle that was used for compound administration and the diet the animals received during the experiments.

Effect analysis, effect concentration/dose, hazard, data analysis, statistical evaluation, category of reliability, justification for reliability category, comments, summary

- See '*in vitro*' and 'AO'

For *'in silico'*:

In the data collection sheet, the models are evaluated on their accordance with the OECD validity and performance criteria for (Q)SARs for regulatory use, as described in the OECD guidance document No 69.

Defining the endpoint (OECD Principle 1)

- Gives information on the endpoint, the endpoint units and the dependent variable(s)

Defining the algorithm (OECD Principle 2)

- Gives information on the algorithm and its descriptors, on how descriptor values are generated, and on the ratio of chemicals used to build the model and the number of descriptors

Defining the applicability domain (OECD Principle 3)

- Gives information on the applicability domain and the limits of applicability of the model

Internal validation (OECD Principle 4)

- Gives information on the training set, on the presence of data for the dependent variable and the descriptors, on the statistics for goodness-of-fit, the robustness of the model, the presence and predictivity of the external validation set, and on whether the validation set is within the applicability domain of the model

Mechanistic interpretation (OECD Principle 5)

- Gives information on the mechanistic base for the model (if available) and on the a priori or a posteriori mechanistic interpretation

Comments and summary

Argumentation for Compound classification:

All test compounds that were listed in the data collection sheet were a priori classified according to their DNT properties to be known to be toxic to the developing brain (positive compounds), potentially toxic to the developing brain (possible positive), not known to be toxic to the developing brain (negative compounds) or not classified for DNT (not classified).

A positive compounds is classified as such, if it is a drugs or a chemical that is known to be developmentally neurotoxic in humans (rev. in Grandjean and Landrigan, 2014, Kadereit et al., 2012). Additionally to the compounds listed in these two publications, the following compounds were classified as positive compounds. Methylazoxymethanol, the aglcycosid of cycasin occurring in cycads, was included in the list as it used as a model compound for DNT (Kabat et al., 1985) and is known to increase the prevalence of amyotrophic lateral sclerosis and parkinson/dementia diseases in the western pacific (rev. in KISBY et al., 1999). Cyclopamine is a steroidal alkaloid causing an undivided forebrain (holoprosencephaly) by inhibiting sonic hedgehoc signalling (Incardona et al., 1998) and is therefore clearly disrupting brain development. Diphenylhydantoin is an antiepileptic drug known to cause intellectual impairment in developmentally exposed children (rev. in Adams et al., 1990). Last, nicotine as the major compound of cigarette smoke is also prominent for causing developmentally neurotoxic effects after maternal smoking (rev. in Slotkin, 2004).

A possible positive compound is classified as such, if it is a neurotoxic, embryotoxic or endocrine disrupting compounds that in general may have the potential to disrupt brain development due to the

inherent properties. Moreover, pesticides and their metabolites which were not included in the list of positive compounds were also classified as possible positives.

Negative compounds were classified if there was no indication for any developmentally neurotoxic effects in the literature. Moreover, Kadereit and Coworkers published a list of negative control compounds (Kadereit et al., 2012) which served as an orientation for classification.

Compounds that are known to have no DNT effects were sorted in the category "not classified". Besides, chemical model compounds that specifically alter an endpoint by interacting with e.g. certain receptor were categorised as "not classified" as those are usually compounds that are pharmacologically designed for a specific purpose and have no relevance for any exposure scenario.

3. Results

3.1. In vitro

All publications containing *in vitro* studies for assessment of DNT (n = 287) were grouped according to (I) species (Human, Rat, Mouse), (II) cell type category (Stem/progenitor cells, Primary cells, Tumor/immortalized cells) and (III) endpoint (for comprehensive list of endpoints see Appendix I).

After grouping according to **species**, we counted 46 studies with human, 181 with rat and 80 with mouse cells. Of these, 74, 149 and 87 belonged to the category of stem/progenitor cells, primary cells and tumor/immortalized cells, respectively (Figure 9). The total number of studies using certain cell types (human (46) & rat (181) & mouse (80) is 307) is, however, larger than the total number of publications identified (287) because some publication contain more than one cell model and here the cell model is the evaluated unit. Publications containing multiple cell models are counted numerous times.

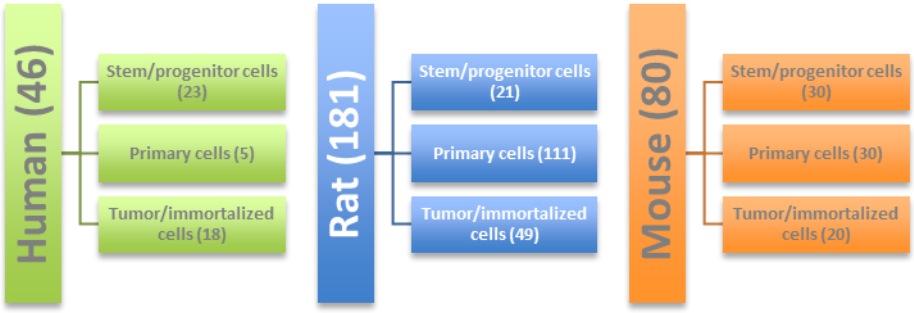


Figure 9 Within 287 original DNT (selected) *in vitro* publications human, rat and mouse, stem/progenitor, primary and tumor/immortalized cells (= 9 cell type categories) were published giving 307 studies. Thereby, some publications contain multiple cell types and are thus counted multiple times.

While due to obvious availability reasons human primary cells are hardly ever employed in *in vitro* DNT studies (5), most data is produced with rat primary cells (111) followed by rat tumor/immortalized cells (49), mouse stem/progenitor cells (30), mouse primary cells (30), stem/progenitor cells from human (23) and rat (21), pursued by mouse (20) and human (18) tumor/immortalized cells. Of all studies performed with human cells (46), 23 (50%), 5 (11%) and 18 (39%) utilized stem/progenitor, primary and tumor/immortalized cells, respectively. Within the group of rat cells (181), 21 (12%), 111 (61%) and 49 (27%) of stem/progenitor, primary and tumor/immortalized cells were employed, respectively. And

within the species mouse (80), the studies divide into 30 (37.5%), 30 (37.5%) and 20 (25%) stem/progenitor, primary and tumor/immortalized cells, respectively (Figure 9).

According to the different **cell types**, most data was generated with primary cells (146), followed by tumor/immortalized (87) and stem/progenitor cells (74; Figure 10).

Of the 74 studies performed with stem/progenitor cells, 23 (31%), 21 (28%) and 30 (41%) were derived from human, rat and mouse, respectively. Within the group of 146 primary cells, 5 (3%) originated from human, 111 (76%) from rat and 30 (21%) from mouse. And within the tumor/immortalized cell group (87), the species divide into 18 (21%) for human, 49 (56%) for rat and 20 (23%) for mouse (Figure 10). The pie graphs generated by this data clearly show, that DNT *in vitro* studies performed in primary and tumor/immortalized cells mainly utilize rat cells (76% and 56%, respectively), whereas the stem/progenitor-based studies divide almost equally across the three species with human stem/progenitor cell studies (31%) almost as prevalent as studies with mouse or rat cells (41% and 28%, respectively). However, the majority of studies originated from rat primary cells (111 of 307 total (37%)).

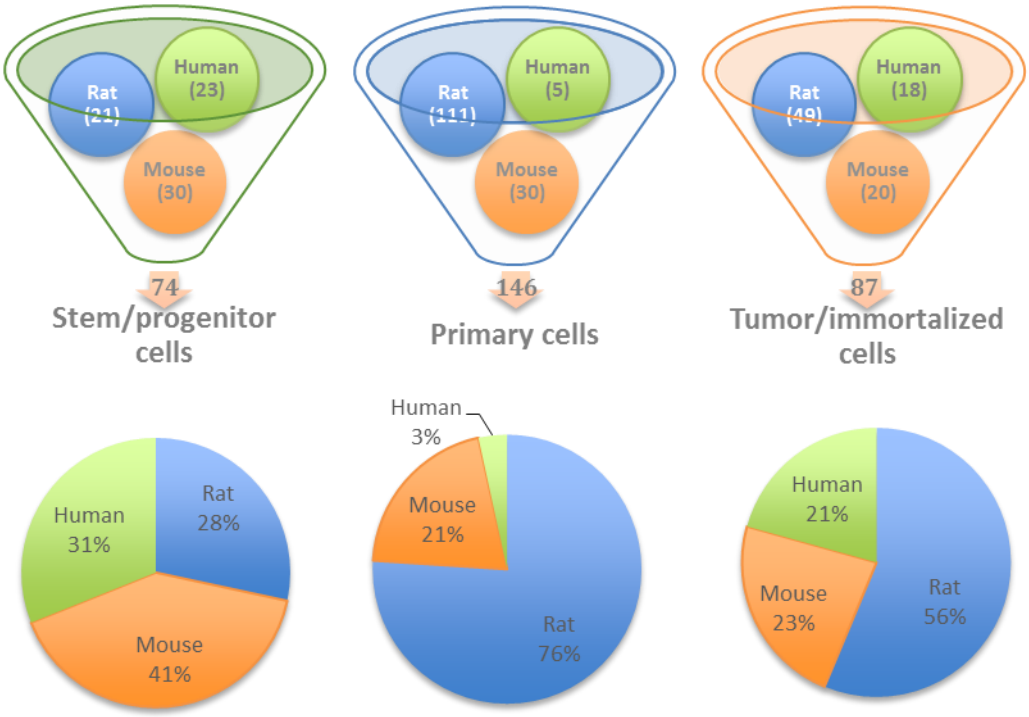


Figure 10 Number of DNT *in vitro* studies (287) employing stem/progenitor, primary and tumor/immortalized cells from human, rat and mouse. Given are the distributions across species within each cell type category. The lower pie plots show the same data in % of all the studies from the respective cell type categories.

When looking at the use of different cell models and species over time it becomes clear that rat primary cells make up the biggest part of the cell type categories because they have been used longest. From our pool of selected publications rat cells have been used from the beginning of the twentieth century, whereas there were only few publications with mouse cells before 2004 and even less publications using human cells before 2006. However, starting 2004 and 2006 the number of publications using mouse and human cells respectively increased compared to the number of publications using rat cells from 2011 on (Figure 11a). From the three different cell types categories (primary cells, tumor/immortalized cells and

stem/progenitor cells) primary cells have been used longest (first publication 1991) followed by tumor cells that have been introduced in 1994 but where only sporadically used until 2004. The first publication using stem/progenitor cells was published in 2002 with a strong increase in number of publications after 2009. This demonstrates that stem/progenitor cells which are a relatively young model are increasingly used for studying effects on the developing nervous system (Figure 11b).

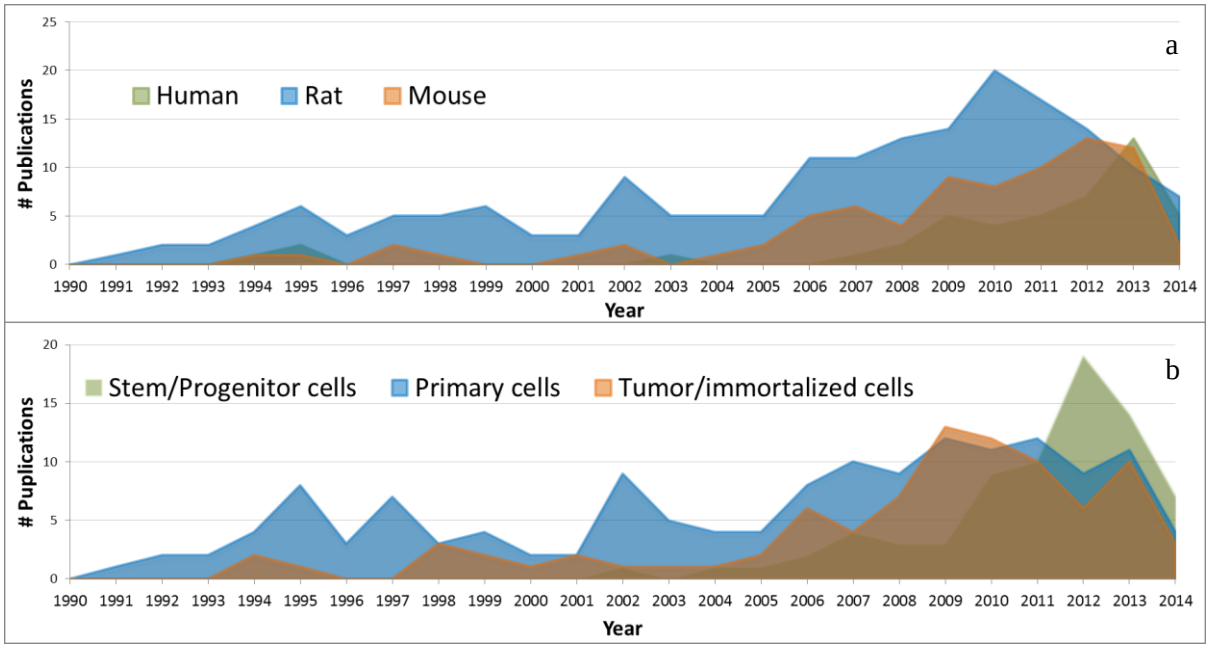


Figure 11 Use of species (a) and cell types categories (b) over time within the selected in vitro publications.

Table 2 Number of different cell models within the different cell type categories and for the different species. Some of the cell models are also prepared from different brain regions. A detailed list of all individual cell types is given in Appendix J.

Cell type category	Stem-/progenitor cells			Primary cells			Tumor/immortalized cells		
Species	Human	Rat	Mouse	Human	Rat	Mouse	Human	Rat	Mouse
Number of cell models	3	2	4	3	11	9	6	5	4
from number of brain regions	2/n.a.*	8	9	3	21	10	n.a.	n.a.	n.a.

n.a.; *only applicable for neural progenitor, not for stem cells.

The grouping of the individual cell types into cell models and than cell type categories results in 9 different categories depicted in Figure 9. These categories consist of a variety of different cell models. The numbers of cell models for each of the 9 categories are shown in Table 2 and the models are listed in

Table 3 to Table 5. These models (e.g. 'hESC-derived neural cells') are still comprised of a variety of different individual cell types because e.g. different hESC-lines (H1, H9, HUES) were employed. In addition, the primary cells (e.g. 'rat neurons') were generated from a large variety of different brain regions. A comprehensive list of all individual cell types is given in Appendix J.

The goal of this systematic review was the evaluation different test systems for their suitability to assess DNT-relevant endpoints. Because most publications contain several endpoint evaluations measured with different methods, different endpoints, different test systems (individual cell type) or different compounds, it was not effective for the evaluation to work with 'number of publications'. Therefore, we evaluated the numbers of times a test system was used for each endpoint evaluation and/or chemical during the rest of this report. E.g., when 5 compounds are evaluated in the same publication for two endpoints and one test system, this will result in 10 citations for this one test system. In the end, each of these citations is then evaluated with respect to the assay performance for each compound and each endpoint. Thus, the number of citations for each test system or cell type category is higher than the amount of studies, and both the number of citations and studies are higher than the number of publications evaluated within this review (Figure 12 and Figure 13). For each of this test system/endpoint measure/compound evaluation one line in the 'Data Collection Sheet' was filled. This resulted in 1.358 cell type citations within the 287 publications and 307 studies.

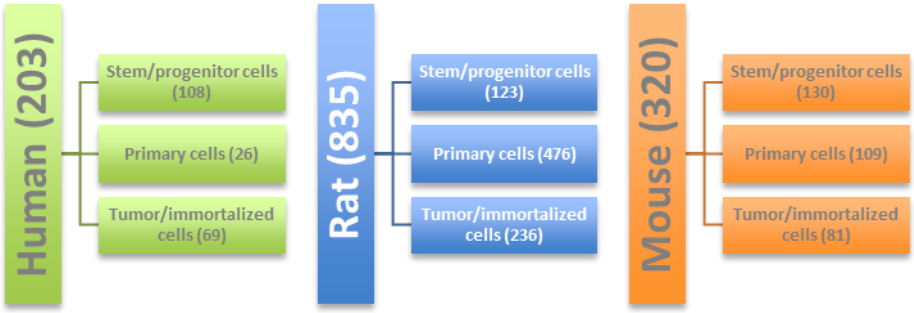


Figure 12 Within 287 original DNT (selected) *in vitro* publications human, rat and mouse, stem/progenitor, primary and tumor/immortalized cells (= 9 cell type categories) were evaluated 1,358 times, giving 1,358 citations. This number results from counting the numbers of times a cell method was used for each endpoint evaluation and/or chemical. For each of this cell method/endpoint measure/compound evaluation one row in the 'Data Collection Sheet' was filled.

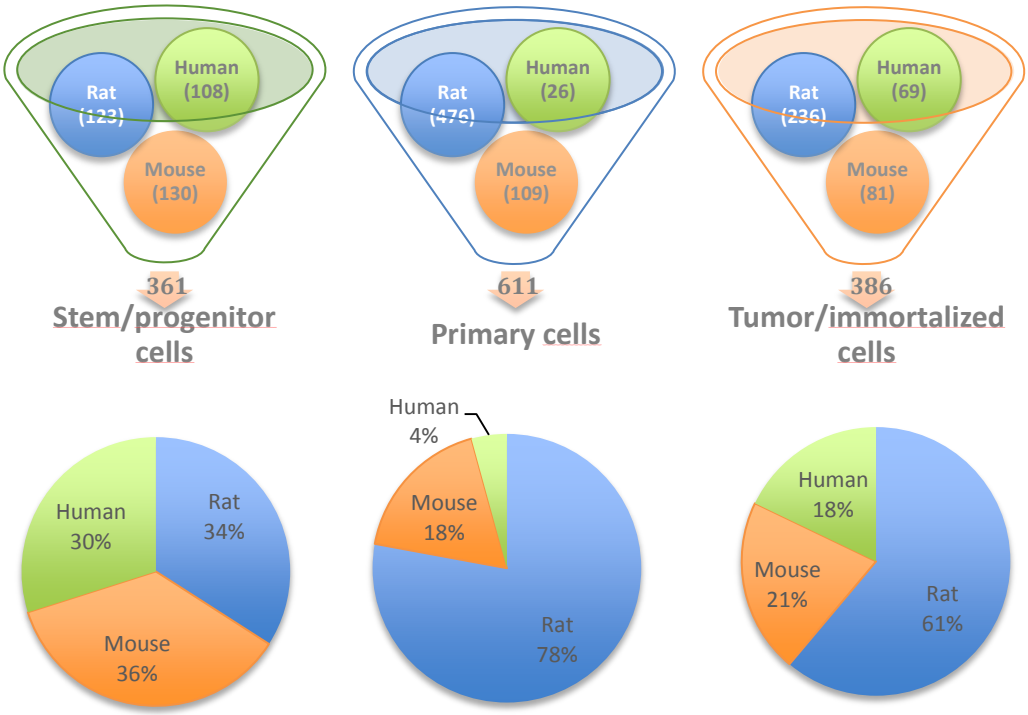


Figure 13 Numbers of times a cell method was used/cited for each endpoint evaluation and/or chemical in 287 *in vitro* publications Given are the distributions across species within each cell type category. The lower pie plots show the same data in % of all the individual cell method citations from the respective cell type categories.

Table 3 Cell models from the category 'stem/progenitor cells' utilized for DNT *in vitro* publications. The numbers in brackets indicate the times the model was studies/cited, the latter referring to distinct endpoints and chemicals

Cell type category	**Stem-/progenitor cells**		
Species	**Human**	**Rat**	**Mouse**
	hESC-derived neural cells (14/58)	rNSC (14/82)	Neural Crest-derived cells (1/1)
	Umbilical cord-derived hNSC (4/15)	rNPC (7/41)	mESC (20/94)
	hNPC (5/35)		mNSC (1/2)
			mNPC (7/30)

ESC = embryonic stem cell; NPC = neural progenitor cell; NSC = neural stem cell; h = human; r = rat; m = mouse

Table 4 Cell models from the category 'primary cells' utilized for DNT *in vitro* publications. The numbers in brackets indicate the times the model was published/cited, the latter referring to distinct endpoints and chemicals

Cell type category	**Primary cells**		
Species	**Human**	**Rat**	**Mouse**
	Aggregated brain cells (1/1)	Re-aggregate cultures (9/62)	Whole embryo culture (1/1)
	Neurons (2/14)	Slice cultures (9/23)	Re-aggregate cultures (1/2)
	Astrocytes (3/11)	Whole brain homogenates (1/4)	Slice cultures (3/5)
		Neurons (49/218)	Neurons (13/60)
		Astrocytes (10/38)	Astrocytes (5/5)
		Oligodendrocytes (2/4)	CGC (6/18)
		CGC (14/58)	Mixed cell preparations (3/8)
		Mixed cell preparations (15/48)	Neuronal & glia co-cultures (1/6)
		Neuronal & glia co-cultures (2/4)	Microwell cultures (1/4)
		Microwell cultures (1/8)	

CGC = cerebellar granule cells

Table 5 Cell models from the category 'tumor/immortalized cells' utilized for DNT *in vitro* published. The numbers in brackets indicate the times the model was published/cited, the latter referring to distinct endpoints and chemicals

Cell type category	**Tumor/immortalized cells**		
Species	**Human**	**Rat**	**Mouse**
	NT-2 (6/35)	PC12 (44/203)	N2a (9/31)
	SH-SY5Y (4/8)	C6 (2/11)	C17.2 (7/36)
	LUHMES (3/3)	PC12-NS1 (2/17)	P19 (4/14)
	hN2 (2/17)	CG-4 (1/5)	
	ReN CX cells (1/1)		
	NB69 (1/3)		
	AScH-7 (1/2)		

The individual listing of **cell types** across the 9 categories of cell type groups and species (Appendix J) demonstrates the heterogeneity of all systems used for DNT *in vitro* studies. When this data is analyzed for the number of times each model is employed within the different publications, it becomes obvious that most individual cell types are used only one to ten times (e.g. >90 individual cell types were used in less than ten citations; Figure 14). This instant makes the evaluation and a standardized comparison of *in vitro* test methods for DNT difficult and should be kept in mind when further evaluation of data is performed with grouped data.

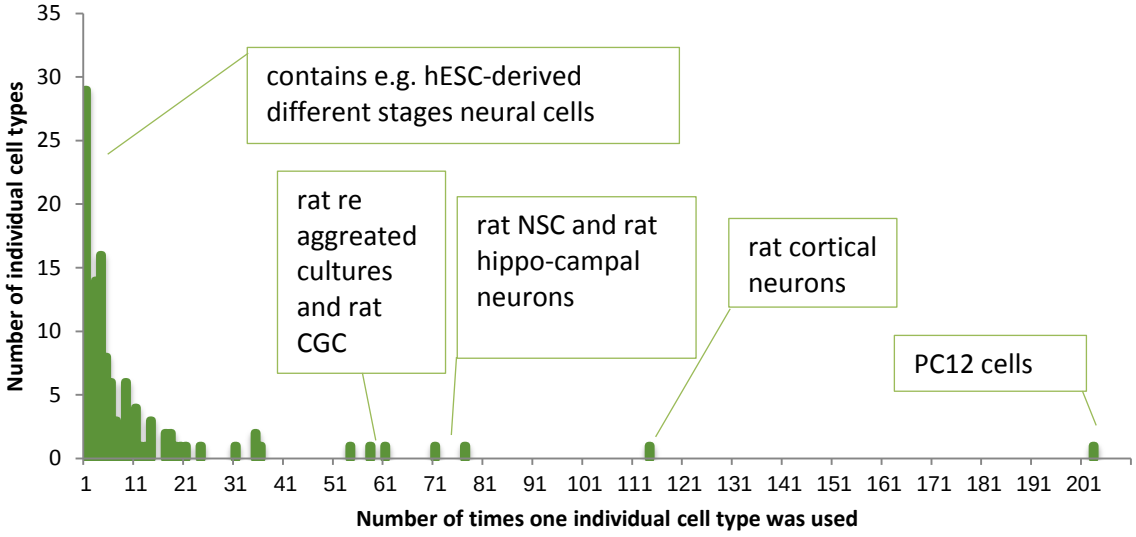


Figure 14 Frequency of usage of individual cell types for DNT *in vitro* publications.

Evaluation of the data not according to individual cell types, but with the grouped cell types within cell models (e.g. mouse neurons, hNPC, etc.) still looks similar: most cell models were cited only a few times while a few cell models were cited more frequently. E.g., Rat tumor cells were used in 49 publications and cited 236 times (most of them being PC12 cells;

Figure 15) and rat neurons in 49 publications, cited 216 times. These are the most data-rich cell models (

Figure 15). In the middle with 20-100 citations were primary mouse neurons, hESC-derived different stages neural cells, rat CGC, ratNPC, rat NSC, rat cell preparations from different brain regions, rat astrocytes and mESC. Only very few citations were found for human neural crest-derived cells, rat and mouse whole embryo cultures, mouse and human re-aggregate cultures, human cortical neurons, rat P19 and CG-4 cells, mouse N1-E115 cells, human ReN CX, NB69 and AScH-7 cells (

Figure 15). All the other cell models were found in a frequency between 2 and 9 publication or 1 to 20 citations/cell model. Although there is a relatively high publication frequency for e.g. hESC-derived different stages of neural cells (58 citations;

Figure 15) one has to be aware that within this cell model, each individual cell type is cited only 1-10 times each (Figure 14).

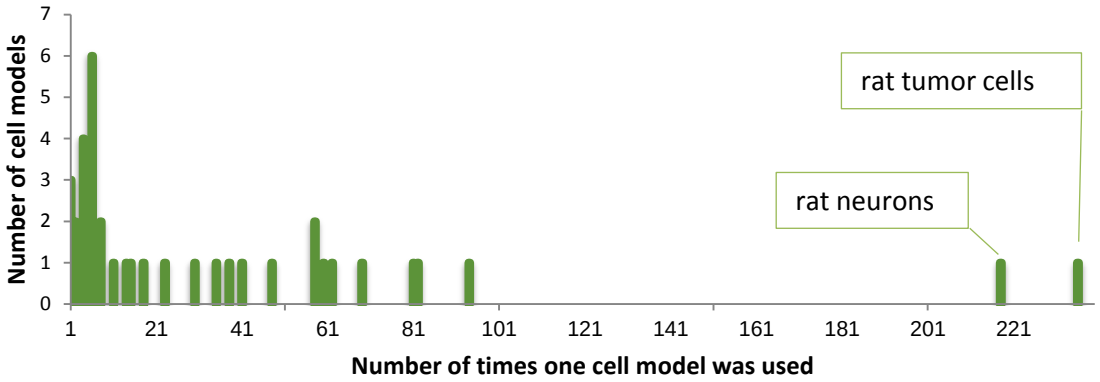


Figure 15 Frequency of usage of cell models for DNT *in vitro* publications.

With these published individual cell types for DNT *in vitro* publications, a large variety of **endpoints** were analyzed. In total, we assessed 149 different endpoints in a DNT context/cell model relevant for DNT (Data Collection Sheet, column 'Endpoint'). Of these, 48 and 101 were grouped into 'non-DNT-specific' and 'DNT-specific' endpoints, respectively. These individual endpoints were further sorted into 'endpoint groups' as depicted in Table 6. The 'non-DNT-specific' endpoints contain the 5 endpoint groups: general cellular function, epigenetics, apoptosis/cell death, energy metabolism and oxidative stress. These are not specific for the developing nervous system, yet represent common mechanisms for toxicant-related disturbances of neurodevelopment (Abel and Zukin, 2008, Farina et al., 2011, Ornoy, 2009). For 'DNT-specific' endpoints, 20 endpoint groups are listed which contain different NS/PC-, neuron- and glia-specific endpoints exclusive for brain cells (Table 6). Thereby the endpoint groups are created in a way that a certain cell model is potentially able to measure all endpoints belonging to one endpoint group. As an example, the endpoint group 'neurite endpoints' contains the individual endpoints neurite initiation, neurite length, neurite morphology, neurite outgrowth, neuronal process formation, number of filapodia, number of neurite branches, number of branch points, number of neurites, neurite complexity and process formation, which were collected from the original publications. When 'neurite length' can be assessed in one cell model, also 'number of neurites' can potentially be measured in the same model – although it was not measured in the original publication. The list of individual endpoints and the respective endpoint groups can be found in the Appendix I.

These 'endpoint groups' are necessary for the evaluation of DNT *in vitro* methods, because there are too few individual endpoints measured by different cell types to evaluate them individually (Figure 16). The majority of endpoints (n = 78) only occurred once or twice in the literature. Only 22 endpoints were published more than 10-times. Of these, only 5 occurred more than 60-times in the literature (2 non-DNT-specific: apoptosis and protein expression-/phosphorylation and 3 DNT-specific: neuronal differentiation, neurite outgrowth and proliferation), 5 were published more than 30-times (3 non-DNT-specific: oxidative stress, cell growth and gene expression and 1 DNT-specific: astrocyte differentiation) and 11 were published more than 10-times (2 non-DNT-specific: lipid peroxidation and calcium (intracellular) and 7 DNT-specific: tyrosine hydroxylase, neural differentiation, oligodendrocyte differentiation, neuronal death, nerite length, migration, axon length, AChE activity and choline acetyltransferase; Figure 16) leading to the above-mentioned total cell method/endpoint/compound citation number of 1.359. Thereby, one citation time is defined as one individual endpoint measured after exposure to one compound in one individual cell method. All individual endpoints with their endpoint groups are listed in Appendix I.

Table 6 Endpoint groups for non-DNT-specific and DNT-specific endpoints. Number citations within endpoint groups is given in brackets.

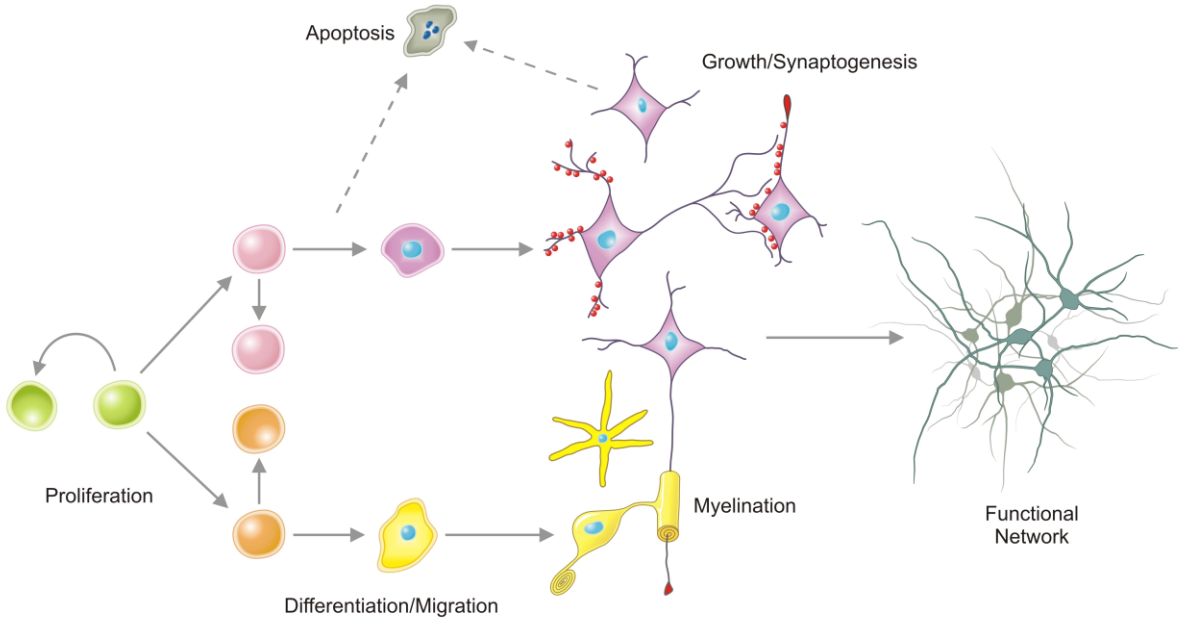
Non-DNT-specific endpoint groups	DNT-specific endpoint groups
General cellular function (243)	General neuronal endpoints (190)
Epigenetics (7)	Axonal endpoints (37)
Apoptosis/cell death (162)	Dendritic endpoints (16)
Energy metabolism (24)	Gabaergic function (9)
Oxidative stress (71)	Neurite endpoints (143)
	Synapse endpoints (10)
	Cholinergic function (60)
	Glutamatergic function (15)
	Catecholaminergic function (4)
	Chlorine signaling (2)
	Dopaminergic function (30)
	Neuronal functioning/electrophysiology (8)
	Neuronal trophic factors (4)
	Neural proliferation (98)
	Migration (25)
	NS/PC-related endpoints (50)
	Astrocyte endpoints (56)
	Oligodendrocyte endpoints (24)
	Cytoskeleton (2)
	Neuroinflammation (5)
Omics (63)	

One endpoint group that does not belong to either of the categories is 'omics'. This is due to the current difficulty to handle 'omics' studies in a regulatory context. While such 'big data' is certainly useful for potential DNT biomarker identification, more research is clearly needed to confirm the validity of such biomarkers. The 'omics' publications represented in this systematic review do not provide such biomarkers but mainly focus on gene ontologies.

Figure 16 Individual endpoints (n = 149) published in the DNT *in vitro* literature (n = 661). Endpoints available more than 60 times in the literature are typed in 12 p, the ones published more than 30-times are written in 10 p and more than 10-times occurring subjects are typed in 8 p. Endpoints occurring less than 10-times are not highlighted.

Successful formation of the brain is guaranteed by a spatiotemporal orchestration of developmental processes involving NS/PC-, neuronal- and glia cell function (Cowan et al., 1997, Jessell, 2000). Effects of toxicants on molecular and cellular functions of such developmental processes were identified within this systematic review and summarized in Table 6 as the here identified DNT-specific endpoint groups. With regards to cellular processes involved in neurodevelopment (Figure 17; (Fritsche, 2014)) the endpoint groups extracted from the DNT *in vitro* publications cover these basic neurodevelopmental processes. Moreover, endpoint groups for neuronal subtype specific toxicities (e.g. cholinergic, dopaminergic, glutamatergic functions) were identified, which complete the endpoint portfolio. The only process missing in the endpoint group lists is 'formation of the blood-brain-barrier (BBB)'. This is due to the fact that currently there is no *in vitro* model available, which mimics BBB development.

It was postulated earlier that disturbance of any of these neurodevelopmental key events listed in Table 6 under 'DNT-specific' will most likely cause an adverse outcome on the organ level (Lein et al., 2005). In addition, toxicants interfering with general cellular function, epigenetics, apoptosis/cell death, energy metabolism and oxidative stress (Table 6, 'non-DNT-specific') can also cause neurodevelopmental adversity when cellular adaptive responses are exhausted. Such in the first place 'non-DNT-specific' endpoints might in the end be early key events and cause alterations of 'DNT-specific' endpoints as later key events. One example for such a chain of events is interference with calcium signaling by PCB95 leading to induced dendritic growth (Wayman et al., 2012). Although calcium signaling belongs to the endpoint group 'general cellular function' and thus is in the first place a non-specific cellular event, it mediates the PCB95-dependent induction of dendritic growth in primary hippocampal neurons, which belongs to the endpoint group 'dendritic endpoints' and is 'DNT-specific'. In this specific example, PCB95 concentrations as low as 200 nM alter calcium signaling and induce the adverse dendritic outcome. In other cases these non-DNT-specific endpoints (early key events) might be more sensitive towards disturbance by toxicants than later events because adaptive processes might be activated that prohibit adversity going from the earlier (e.g. ROS production) to the later key events (e.g. ROS-induced neuronal apoptosis). Due to this reason, we suggest that DNT *in vitro* testing is probably more robust when later, 'DNT-specific' key events are assessed.



.

Figure 17 Basic processes of brain development necessary for proper organ function. Neural Progenitor Cells (NPCs, *green)* proliferate to provide an access amount of cells, which then migrate and differentiate into neurons (purple) and glia (yellow). These form synapses (red) and surplus cells undergo apoptosis (grey). When these processes happen in the right and coordinated way, functional neuronal networks form (olive). With courtesy from William Mundy, US Environmental Protection Agency and John Havel, SRA International, Inc.

How are the **endpoint groups** represented in the current 'DNT *in vitro*' literature? Figure 18 summarizes the publication frequencies of the 5 'non-DNT-specific', the 20 'DNT-specific' as well as the 'omics' endpoint groups. Similar to the citation frequencies of the individual, non-grouped endpoints (Figure 16), only a few endpoint groups (General Cellular Function, Apoptosis/Cell Death, Oxidative Stress (= 3 non-DNT-specific), General Neuronal Endpoints, Neurite Endpoints, Cholinergic Endpoints, Neural Proliferation, Astrocyte Endpoints (= 5 DNT-specific)) are published at a relatively high frequency (>20 publications/endpoint group). For all the other endpoint groups there is a clear lack of data.

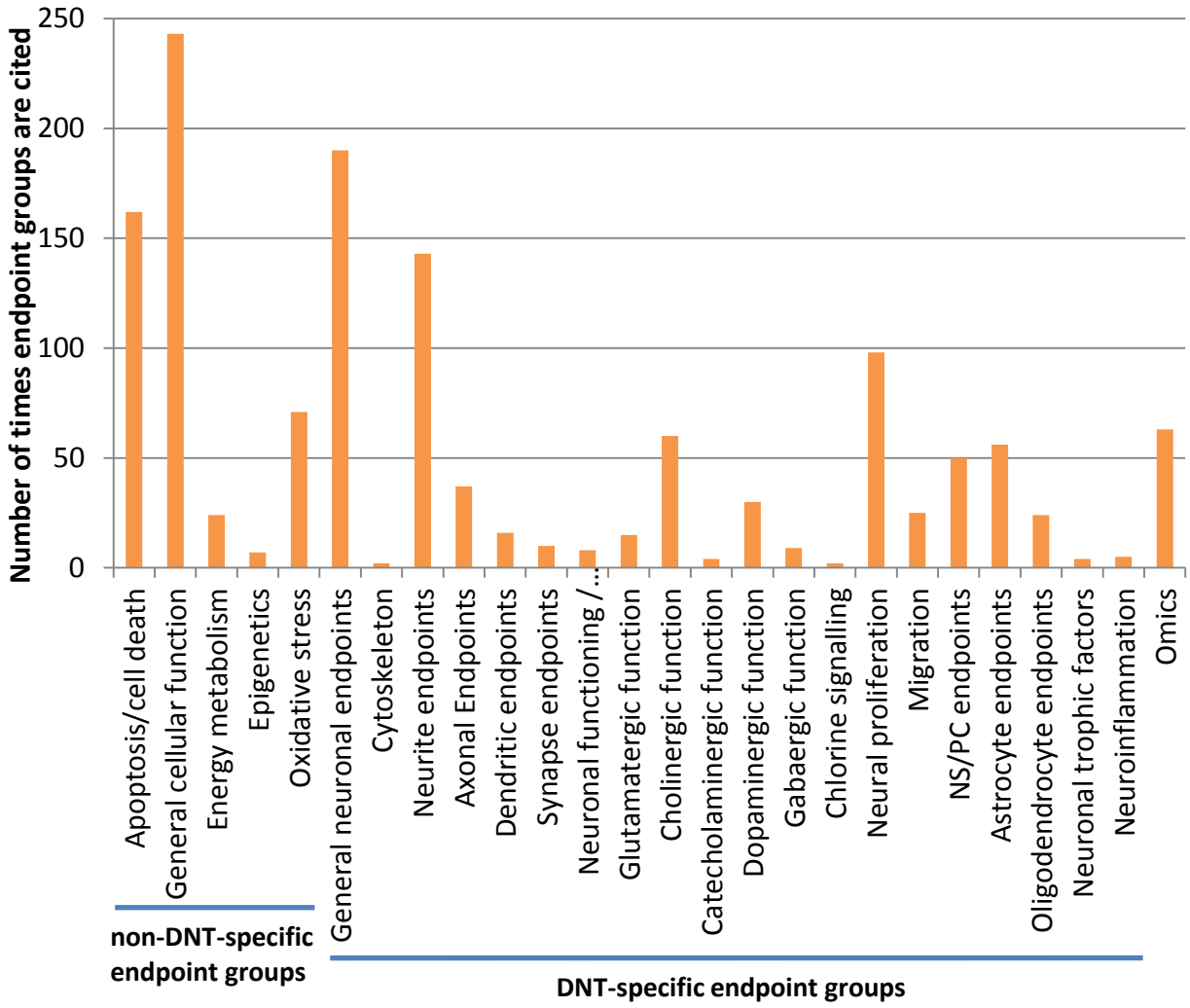


Figure 18 Citation frequency of non-DNT-specific and DNT-specific endpoint groups published in the DNT *in vitro* literature.

This systematic review is supposed to identify methods, which can reliably predict DNT with alternative approaches. One of the *in vitro* criteria for a reliable DNT method is to distinguish general cytotoxicity from specific endpoint alteration (Crofton et al., 2011). One example of two compounds, which exert such non-DNT-specific and DNT-specific toxicities for the endpoint neurite outgrowth, cadmium and methylmercury, respectively, is given in Figure 19 (taken from Crofton et al., 2011). While cadmium-induced inhibition of neurite outgrowth is paralleled by a loss in cell viability, methyl mercury-dependent shortening of neurites happens at concentrations not affecting viability. This observation demonstrates the necessity to compare data on endpoints (summarized in Figure 18) to data on cell viability (for one compound at one time/concentration and within one study). These analyses will enable the identification of cell-based methods, which are able to predict specific DNT distinguishable from overt cytotoxicity.

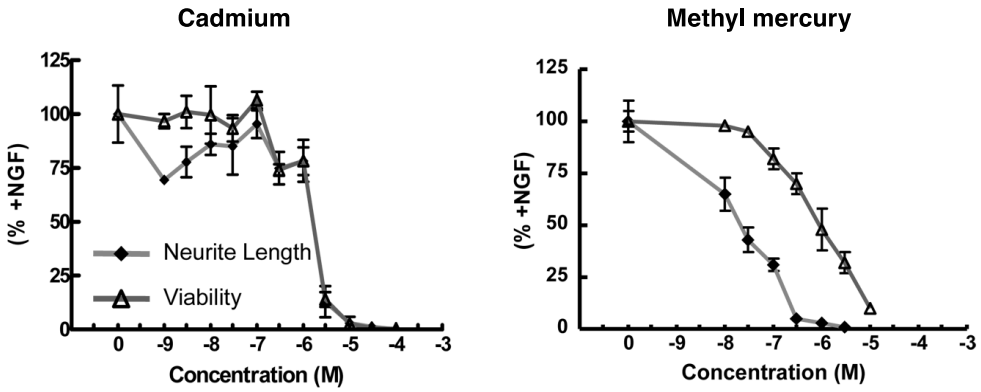


Figure 19 Toxicological specificity of neurite outgrowth. Methyl mercury decreased neurite outgrowth specifically, relative to cell viability. Cadmium chloride decreased neurite outgrowth in a non- specific manner, i.e., only at concentrations that also induced cell death (redrawn from Parran et al., 2001, Radio et al., 2008) (taken from Crofton et al., 2011).

For analyzing the total collected data of the systematic review search all endpoint groups containing data from the 9 cell type categories (stem/progenitor cells, primary cells and tumor/immortalized cells of human, rat and mouse each, Figure 9), were analyzed for chemical effects on viability versus the (non)-DNT-specific endpoints. Thereby, the categories a) 'Specific endpoint is more sensitive towards compound exposure compared to the effect on cell viability' (green), b) 'Specific endpoint is equally or less sensitive towards compound exposure compared to the effect on cell viability' (red), c) 'Viability was not assessed' (purple), d) 'not clear' (blue) or f) 'Effect on endpoint and viability are compound-dependent' (turquoise; only studies with many compounds (leading to more than 20 citations) were distinguished). Considering all measured endpoints (in n = 1358 citation), 43.3% were more and 24.3% less or equally sensitive than viability. For 30% the viability was not assessed (Figure 20).

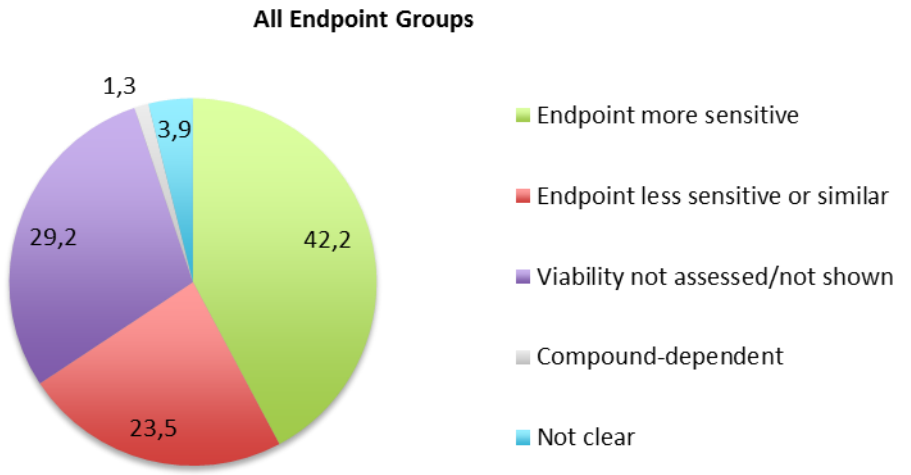
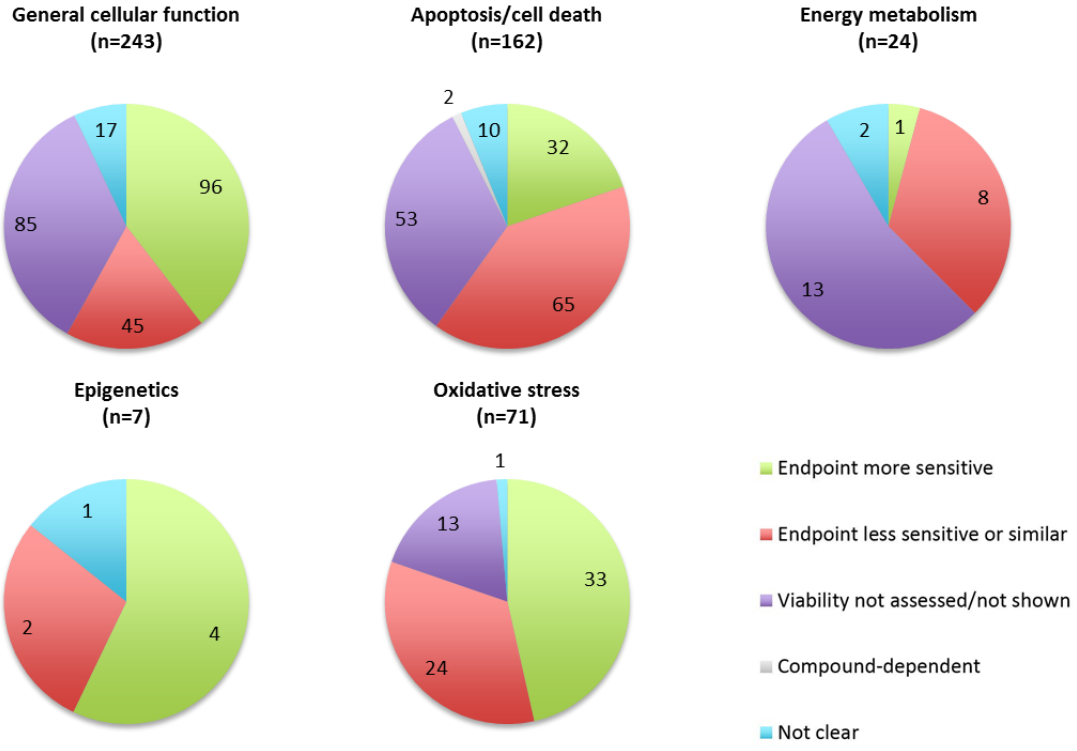
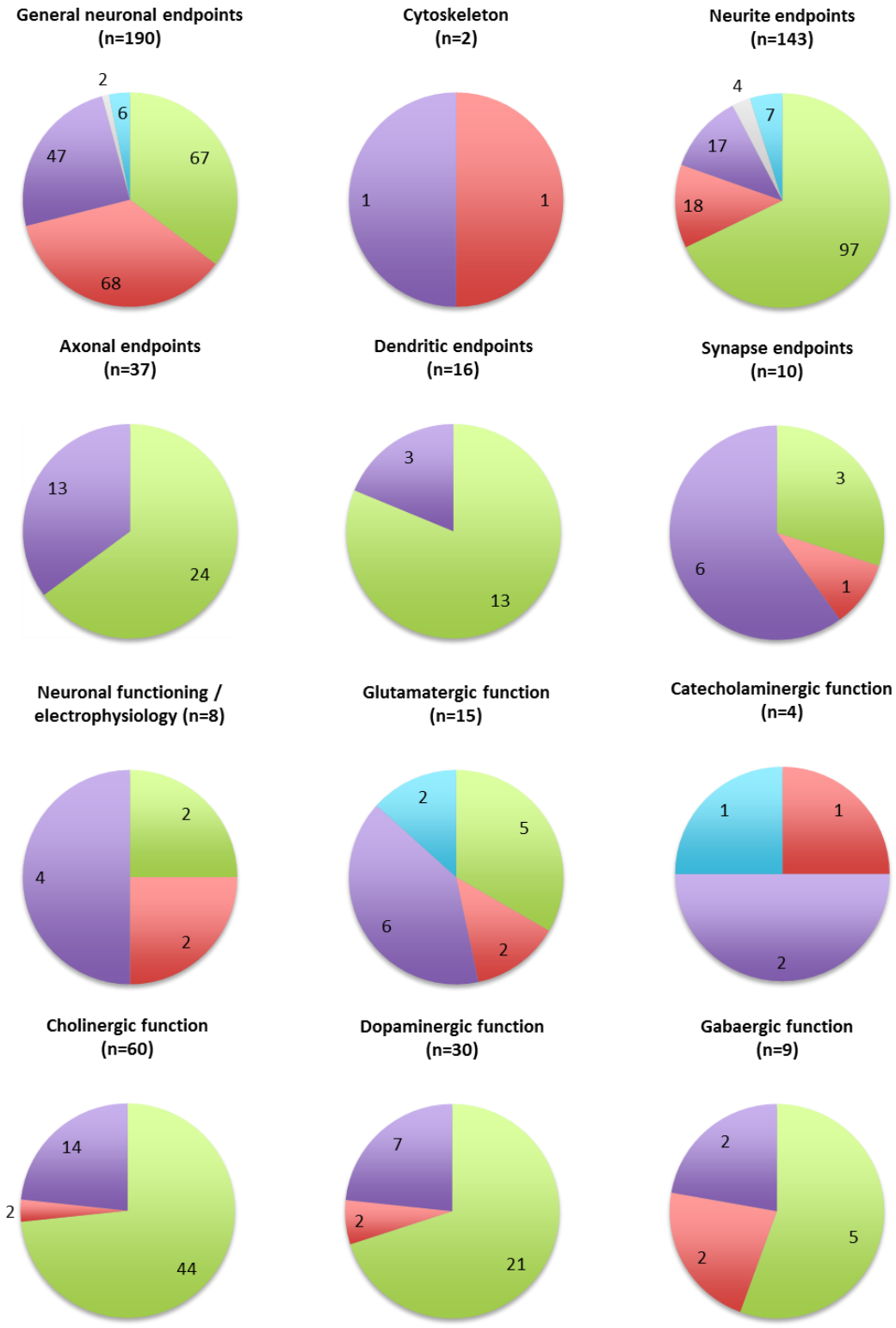


Figure 20 Distribution of specific and non-specific toxicant-induced endpoint group changes across (non)-DNT-specific endpoint groups. Thereby, a specific change is considered as an endpoint change occurring without cytotoxicity (green). When cytotoxicity is measured at the same or lower concentrations (red) than the specific endpoint, the change is non-specific. Numbers in the pie chart give the percent of citations for each classification based on the total of 1358 citations.





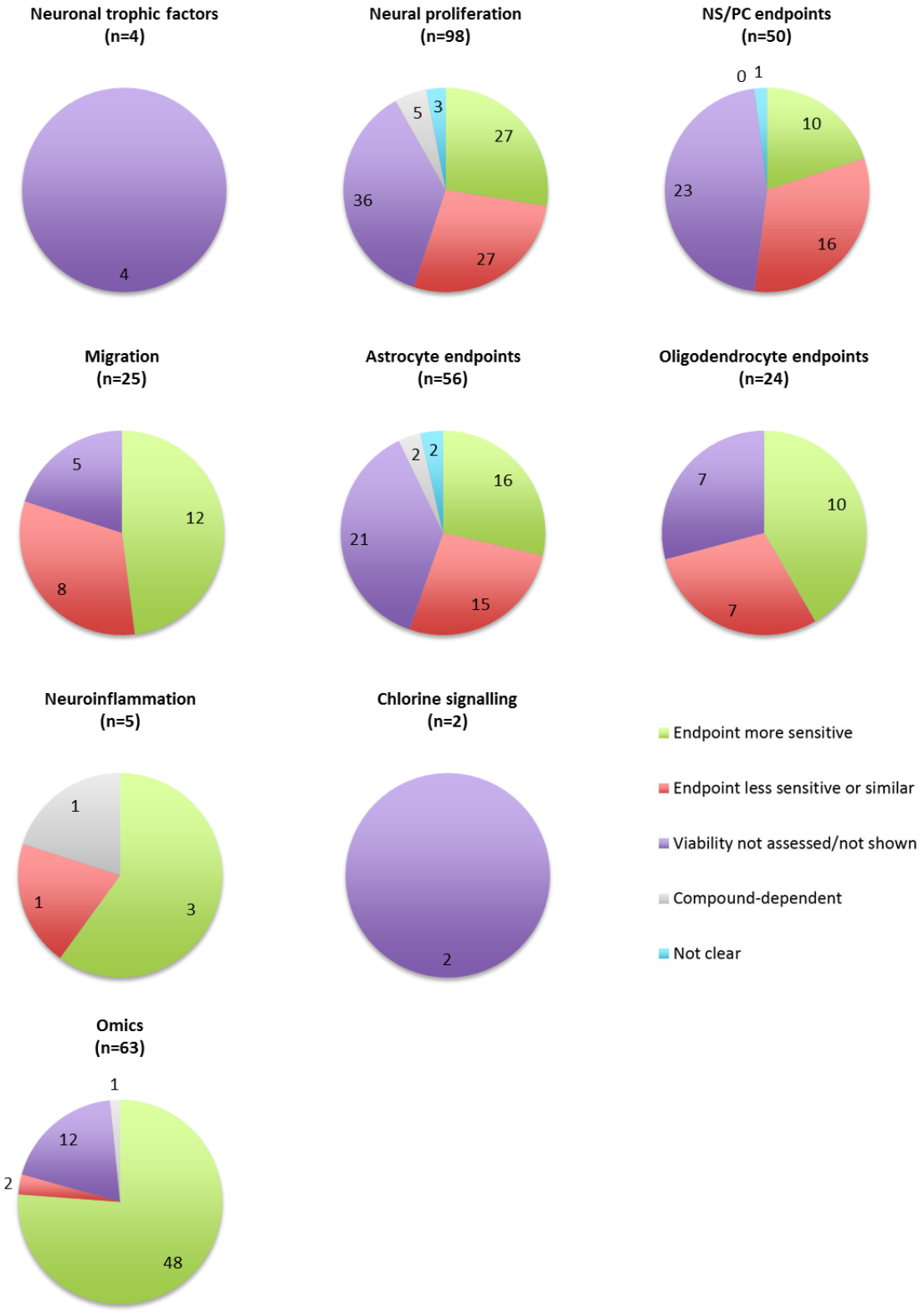


Figure 21 Distribution of specific and non-specific toxicant-induced endpoint group changes across (non)-DNT-specific endpoint groups. Thereby, a specific change is considered as an endpoint change occurring without cytotoxicity (green). When cytotoxicity is measured at the same or lower concentrations (red) than the specific endpoint, the change is non-specific. In cases viability was not assessed, the pie piece is purple. Each pie diagram represents one endpoint group published in the DNT *in vitro*

literature as listed in Table 6. The 'n' gives the total number of citations and numbers in the pie charts give the number of citations for each classification.

Splitting this information into the 26 pre-defined endpoint groups (Table 6) reveals that information on viability strongly differs across the groups. E.g., for the endpoint groups Chlorine signalling and Neuronal trophic factors there is no study available that simultaneously measured the specific endpoint and cell viability. In contrast, the endpoint groups Neurite endpoints and Dendritic endpoints are almost all accompanied by viability measures (Figure 21). These information illustrate clear data gaps on certain endpoints, where no or only very sparse viability data is available (endpoint groups Energy Metabolism, Cytoskeleton, Synapse Endpoints, Neuronal Functioning/Electrophysiology, Glutamatergic Function, Catecholaminergic Function, Neuronal Trophic Factors, NS/PC Endpoints). These endpoints cannot be further evaluated to assess the performance of different cell types because it is not known if published endpoint alterations are specific for DNT or due to general cytotoxicity. As alterations in these processes cause disturbed neurodevelopment it is important to generate tests that reliably assess compound effects on these cellular functions, especially the DNT-specific ones.

The endpoint evaluations shown in Figure 21 were performed with all compounds tested in the 287 DNT *in vitro* publications. Of these, some are true (known) DNT positive compounds (Grandjean and Landrigan, 2006b), some are possible positives, some are true negatives and some are not classified (see Appendix H for list of all compounds and the respective classification). Therefore, we next evaluated how many of the true positives resulted in a specific DNT effect and how many of the true negatives caused either cytotoxicity at the same concentrations as the specific endpoint or had no effect. We grouped all the data of the individual endpoint groups according to 'Hazard' yes/no, 'Compound Classification a priori' and 'Effects on Viability' into the categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. The category 'False Negatives' is not applicable here, because a negative endpoint evaluation of an a priori positive compound might be negative because 1) the DNT mechanism of action of this compound is not by altering this endpoint or 2) the cell system is not affected although this compound is acting through this mode of action. The latter would be a true 'False Negative', whereas the first case would be correctly identified. However, we cannot distinguish between these two because very often the precise mode of action of a DNT compound is unknown. Therefore, we name this category not 'False Negative', but 'Positives, Negative for this Endpoint'. The different groups are given in detail within Table 7A. For all studies with multiple compounds and viability data we analyzed the effect of the individual compounds to be able to also make the grouping. This evaluation will provide information that in the next step can be used for evaluating the specificities of certain cell models in assessment of DNT endpoints. In Table 7B, compounds with a priori unclear classifications are listed. Because they do not represent clear DNT positives or negatives, one cannot judge performance of a cell system according to their responses on these compounds. However, as approximately half of the studies were performed with so-called 'possible positives' or 'a priori not classified' compounds, they will help support data coming out of the known DNT positives/negatives evaluation.

Table 7 Grouping of citations **with** available viability data (n = 1155 including all studies with multiple compounds and viability data, which were analyzed with regard to the individual compound) according to 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint' (A) as well as 'Possible Positives identified as positive', 'Possible Positives identified as negative for this endpoint', 'A priori not classified, but identified as positive', A priori not classified, but identified as negative for this endpoint' and studies testing 'Multiple compounds from different categories'. The number of total citations in each group across all endpoints is given.

A	Hazard	Compound a priori classification	Viability	Number of citations
True Positives	'yes'	'Positive'	'Not affected' or 'Endpoint more sensitive than viability'	**241**
True Negatives	'no'	'Negative'	'Not affected' or 'Endpoint less sensitive than viability'	**50**
	'yes'	'Negative'	'Endpoint equally or less sensitive than viability'	
False Positives	'yes'	'Negative'	'Not affected' or 'Endpoint more sensitive than viability'	**0**
Positives, Negative for this Endpoint	'no'	'Positive'	'Not affected' or 'Endpoint less sensitive than viability'	**262**
	'yes'	'Positive'	'Endpoint equally or less sensitive than viability'	

B	Hazard	Compound a priori classification	Viability	Number of citations
Possible Positives identified as positive	'yes'	'Possible Positive'	'Not affected' or 'Endpoint more sensitive than viability'	**259**
Possible Positives identified as negative for this endpoint	'no'	'Possible Negative'	'Not affected' or 'Endpoint less sensitive than viability'	**213**
A priori not classified, but identified as positive	'yes'	'Not classified'	'Not affected' or 'Endpoint more sensitive than viability'	**72**
A priori not classified, but identified as negative for this endpoint	'no'	'Not classified'	'Not affected' or 'Endpoint less sensitive than viability'	**57**

These data shown in Table 7A is now evaluated for cell systems having the ability to make correct predictions for DNT endpoints. Therefore, for each endpoint, cell systems are analyzed for the four

categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. The category 'Positives, Negative for this Endpoint' is frequently occurring in the evaluation. Therefore, we did not exclude this category. In addition, according to the definition of DNT compounds, only ONE of the multiple processes needed for brain development has to be disturbed to cause an adverse outcome. Thus, it is highly probable that some compounds affect only a small number of endpoints actually pointing to the ability of the cell method to specifically identify key events involved in particular modes of action of compounds. E.g. if a DNT compound acts through inhibition of cell migration, we do not expect to see an effect on e.g. oligodendrocyte differentiation at the same time. The latter would then occur in the category 'Positives, Negative for this Endpoint' and support the specificity of the cell method. By this categorization, we will retrieve information on 1) cell type categories, which are able to evaluate individual endpoints correctly, 2) cell methods, which have the ability to assess a variety of DNT endpoints in a correct way and 3) data gaps and needs with regard to cell methods and endpoints. First, point number 1) will be assessed by walking through the endpoints individually identifying 'True Positives' and 'True Negatives' across individual DNT compounds. 'Positives, Negative for this Endpoint' will be included here. From this data 2) cell type categories will be associated with endpoint evaluations. 3) We will identify data gaps by evaluating the endpoint groups and cell type categories with regards to a DNT alternative testing strategy by just including 'True Positives' and 'True Negatives'. This will be repeated for the compounds 'Not Classified' and 'Possible Positives' and compared to the results for the human DNT compound data.

General aspects on cell methods due to the current 'State of the Art'

Before evaluating the different cell methods for their abilities to assess effects of chemicals on neurodevelopmental processes, the general current 'State of the Art' concerning *in vitro* testing with cell-based methods will be summarized briefly.

Within the last years, the topic of species-specificities in cell physiology has been emphasized. Pharmacokinetic specificities between species have been long recognized and have thus been implemented into risk assessment by using species-specific uncertainty factors that compensate for the kinetic differences between animals and humans (Falk-Filipsson et al., 2007). Besides pharmacokinetics, evidence is evolving that also profound pharmacodynamics differences exist between species (e.g. Abbott et al., 1999, Gassmann et al., 2010, Hansen et al., 1999, Jenkins et al., 2012, Stebbings et al., 2007, Strasser et al., 2013, Seok et al., 2013, Perreault et al., 2013). Such interspecies variations in pharmacokinetics and –dynamics might cause drug effects in humans that are not observed in animals or vice versa (rev. in Hanke, 2006, Leist and Hartung, 2013). These lessons have mainly been learned from pharmacology but certainly also hold true for toxicology (e.g. Abbott et al., 1999, Gassmann et al., 2010, Harrill et al., 2011). Therefore, if possible, toxicological testing is thought to produce more predictive results for human hazard assessment when cells of human origin are used (Krewski et al., 2010b, Seidle and Stephens, 2009a). This is supported by the observations that rodent brain cells taken out of the *in vivo* situation placed into a tissue culture dish maintain their signaling functions or their responses towards xenobiotics (Burke et al., 2006, Foti et al., 2013, Go et al., 2012, L'Episcopo et al., 2013, Simpson et al., 2011) and thus primary human cells are thought to behave in a similar way when going from *in vivo* to *in vitro*. The easiest available sources of human cells with no ethical concerns or material restriction are tumor or immortalized cells. Yet, tumor cells have a rather different cellular signaling program than normal cells, which is geared towards unrestricted growth instead of specific organ functions (Drobic et al., 2006). Some tumor/immortalized cell lines have been widely used for restricted cellular applications. One very prominent example in the neurodevelopmental field is the usage of PC12 cells, a rat pheochromocytoma cell line. Currently (16. November 2014), there are 1925 articles published in PubMed when searching for the keywords 'PC12' AND 'neurite outgrowth'. However, a comparative study on chemical effects on neurite outgrowth in PC12 and primary rat cells (CGC) shows clear differences in compound effects on viability and neurite outgrowth between the tumor/immortalized and the primary cell system (Radio et al., 2008). Such a systematic testing approach lightens up difficulties

with these type of cell lines when it comes to compound testing on neurite outgrowth, while they might be very well suited for studying other basic research questions (e.g. Gerdin and Eiden, 2007)). Due to differences in cell cycle from normal cells, tumor/immortalized cells were excluded for studying effects on neural proliferation per se in this systematic review.

Besides tumor cells, restricted commercial sources are available for purchasing primary human brain cells. For one, primary human neural progenitor cells (hNPCs) are available from Lonza (Verviers, Belgium) and have been used to set up a testing strategy covering multiple neurodevelopmental endpoints (Moors et al., 2009; record nr. 2682, Baumann et al., 2014) up-scalable for medium-throughput testing (Gassmann et al., 2012) Immortalized hNPC are also commercially available (Millipore; ReNcell), yet they have the drawback of myc-immortalization. Occasionally human tissue for preparation of primary cells is available (e.g. record number 675), but using such tissue is not practicable for a higher throughput screening procedure.

Next to primary cells, the group of stem cell-based methods is currently rapidly increasing. For one, there is the large group of different – and still to some extent inhomogenous – embryonic stem cells (Allegrucci and Young, 2007). These cells have been very nicely characterized for their properties to mimic early neurodevelopment *in vitro* and *in vivo*. However, due to ethical concerns, the development of human induced pluripotent stem cells (hiPSC; Yamanaka, 2012) has quickly overtaken the stem cell field because iPSCs can be generated without any ethical concerns, they behave similarly to their hESCs siblings, are available without limitations and are thought to be very similar to hESC (Broccoli et al., 2014). Due to these reasons, iPSCs are currently revolutionizing the possibilities of chemical testing in human-based primary cells. However, actual chemical testing has so far not been performed with hiPSCs-derived neural cells, yet with different stem-/progenitor cell methods (rev in Fritsche, 2014). Because 'lack of a chemical' belonged to the exclusion criteria of this systematic review, there is no publication with hiPSC-derived neural cells present. One issue with human stem cell-based methods is, that so far it has not been evaluated to what extent hESC/hiPSCs mimic neurodevelopmental processes in humans *in vivo*. Therefore, although very promising methods, an *in vitro/in vivo* validation is urgently **needed**.

Taken together, when evaluating the studies and cell methods for their suitability for DNT testing it should be kept in mind that according to the current state of the art (i) primary cells are reflecting cell physiology of the respective species very well. Due to this reason (ii) human non-tumor/immortalized cell-based systems are wanted for chemical DNT testing and (iii) stem-/progenitor cell approaches are promising in this regard, but need *in vitro/in vivo* validation and a harmonization of differentiation protocols for diminishing assay variability due to different culture conditions (Allegrucci and Young, 2007).

Non-DNT-specific Endpoints:

1. General Cellular Function

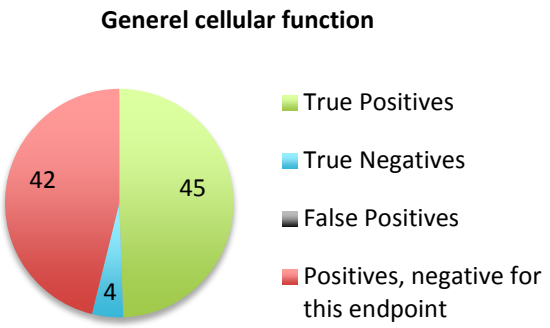


Figure 22 Distribution of citations for the endpoint 'General Cellular Function' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

From all endpoint citations 'General Cellular Function' with viability data of the categories depicted in Table 7A is shown in Figure 22. From the total of 91 citations for the endpoint 'General Cellular Function' only with DNT positive and negative compounds, 45 are identified as 'True Positives', 4 as 'True Negatives' and 42 as 'Positives, Negative for this Endpoint'. These data were obtained with diverse cell methods (see Table 8). The majority of data was obtained with primary cells (42; manly rodent), stem/progenitor cells (26) followed by tumor/immortalized cells (23). The majority of these 'General Cellular Functions' are gene/protein expression or protein phosphorylation concerning molecules not specific for neurodevelopment. E.g. GFAP, Neurofilament, NMDA-R gene or protein expression are not grouped to 'General Cellular Function', but to 'Astrocyte Endpoints' and 'Neuronal Endpoints', respectively. However, as these non-neural-specific gene/protein expressions/phosphorylations are rather unspecific, they will certainly not be first choice in biomarker identification for DNT. True positives were identified with stem/progenitor cells as well as with tumor/immortalized cells.

Table 8 For the endpoint 'General Cellular Function' grouping of Cell Type Category and Species across the 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	5	4	2
Rat	2	10	16
Mouse	5	-	1
True Negatives			
Human	-	-	-
Rat	-	4	-
Mouse	-	-	-
Positives, Negative for this Endpoint			

Human	6	-	2
Rat	4	24	1
Mouse	4	-	1
Sum	26	42	23

2. Apoptosis

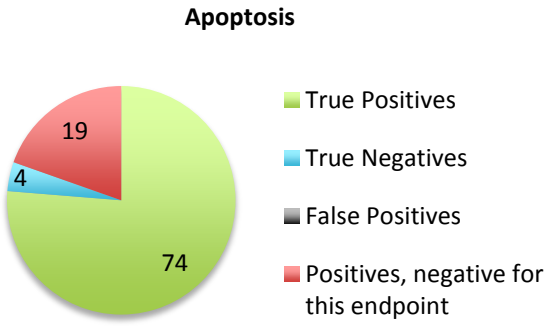


Figure 23 Distribution of citations for the endpoint 'Apoptosis' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

For the endpoint citations 'Apoptosis' it does not make sense to call compounds, which produce cell death at similar concentrations than inducing parts of the apoptosis machinery (e.g. activation of caspases) 'Positives, Negative for this Endpoint' because Apoptosis induces cell death. Therefore, the initial events inducing apoptosis and the result of apoptosis – cell death are a function of time in any *in vitro* method. Therefore, we classified this endpoint by analyzing 'Hazard yes/no' of DNT positive and negative compounds within the data collection sheet regardless of effects on viability. From the total of 97 citations for the endpoint 'Apoptosis' only with DNT positive and negative Compounds, 71 are identified as 'True Positives', 4 as 'True Negatives' and 19 as 'Positives, Negative for this Endpoint' (Figure 23). These data were obtained with diverse cell methods (see Table 9). The majority of data was obtained with stem/progenitor cells (31), followed by rodent primary cells (29) and tumor/immortalized cells (14). The only data available for 'True Negatives' was performed with a human immortalized cell line and a mouse stem cell model (record number 1138). This clearly shows that there is the **need** to also evaluate DNT negative compounds for the endpoint 'Apoptosis' to be able to evaluate assay performance. One compound intensively investigated for induction of apoptosis during brain development is ethanol. It is a documented phenomenon that developmental stages or brain regions confer to either vulnerability or resistance to ethanol-induced apoptosis (discussed within record number 13548 or Thomas et al., 1998, Heaton et al., 2003). Thus, it is not surprising that amongst all cell methods reacting towards ethanol with apoptosis only the stem-/progenitor cell-based methods in some instances do not respond towards ethanol with induction of apoptosis and are therefore grouped into 'Positives, negative for this endpoint' (record numbers 10593, 11819, 17312). Due to this instance there is a strong **need** to establish stem-/progenitor cell-based methods, which clearly reflect defined stages of development. Although primary rodent cells prepared at different times during development and from different brain regions should as well be able to reflect these stage-specific properties, they have the limitation that they do not necessarily reflect human physiology. However, brain regions can so far not be pictured by human ES/PC systems. With increasing knowledge in basic neuroscience, i.e. learning on how soluble factors drive brain region specificities during normal development (Irmady et al., 2011), this might eventually be possible to implement into *in*

vitro testing. Ethanol is taken here as an example for also other compounds acting on the developing nervous system in a developing time-dependent manner like valproate (Go et al., 2012) underlining the necessity of the reflection of developmental timing into any *in vitro* testing effort for DNT.

Table 9 For the endpoint 'Apoptosis' grouping of Cell Type Category and Species across the 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	6	-	-
Rat	11	20	3
Mouse	14	9	11
True Negatives			
Human	-	-	2
Rat	-	-	
Mouse	2	-	-
Positives, Negative for this Endpoint			
Human	2	0	5
Rat	0	6	0
Mouse	5	0	1
Sum	**40**	**35**	**22**

3. Energy Metabolism

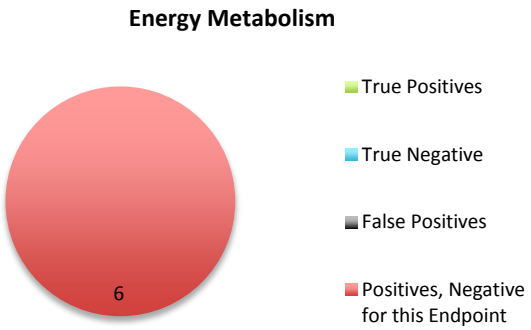


Figure 24 Distribution of citations for the endpoint 'Energy Metabolism' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

From all endpoint citations 'Energy Metabolism' with viability data of the categories depicted in Table 7A are shown in Figure 24. All 6 citations for this endpoint belong to the category 'Positives, Negative for this Endpoint' indicating that either decreased viability is accompanying alterations in energy metabolism or the endpoint was not affected. Again, this is a very general endpoint, which we do not expect to be specifically relevant for DNT.

Table 10 For the endpoint 'Energy metabolism' grouping of Cell Type Category and Species for the only category available 'Positives, Negative for this Endpoint'.

Positives, Negative for this Endpoint	Primary Cells	Tumor/immortalized Cells
Human	-	-
Rat	1	2
Mouse	1	2
Sum	**2**	**4**

4. Epigenetics

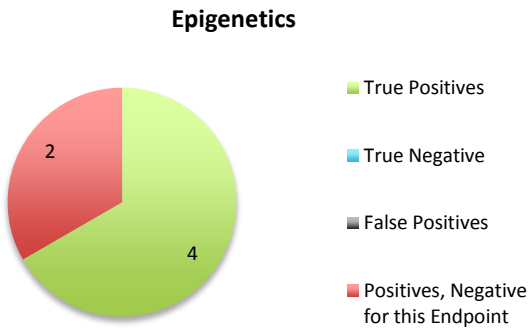


Figure 25 Distribution of citations for the endpoint 'Epigenetics' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

The distribution of all endpoint citations 'Epigenetics' with available viability data across the 4 categories is shown in Figure 25. Of the 6 citations for this endpoint 4 belong to the category 'True Positives' and 2 to the category 'Positives, Negative for this Endpoint'. Data for all of these 4 'True Positives', were generated in Stem/progenitor cells, while data for the 2 'Positives, Negative for this Endpoint' were assessed in Primary cells and Tumor/immortalized cells (Table 11).

Table 11 For the endpoint 'Epigenetics' grouping of Cell Type Category and Species for the categories 'True Positives' and 'Positives, Negative for this Endpoint'.

True Positives	Stem/progenitor cells	Tumor/immortalized Cells
Human	2	-
Rat	2	-
Mouse	-	-
Positives, Negative for this Endpoint		
Human	-	-
Rat	-	1
Mouse	-	1
Sum	**4**	**2**

5. Oxidative Stress

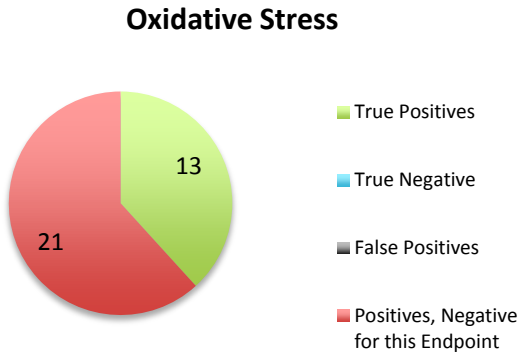


Figure 26 Distribution of citations for the endpoint 'Oxidative Stress' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation

From all endpoint citations 'Oxidative Stress' with viability data of the 4 categories results are shown in Figure 26. From the total of 34 citations for the endpoint 'Oxidative Stress' only including DNT positive and negative compounds, 13 are identified as 'True Positives' and 21 as 'Positives, Negative for this Endpoint'. 'True Negatives' were not present for this endpoint group. These data were obtained with diverse cell methods (Table 12). 'True Positives' were found with (mainly rat) tumor/immortalized cells (8), followed by primary cells (4) and stem/progenitor cells (1). These data are obviously too sparse to evaluate cell systems appropriately for their function with regards to this endpoint. However, as in the series of key events leading to an adverse outcome, oxidative stress is again a very early key event, which is probably not as predictive as later key events for organ and organism malfunction.

Table 12 For the endpoint 'Oxidative Stress' grouping of Cell Type Category and Species for the categories 'True Positives' and 'Positives, Negative for this Endpoint'. 'True Negatives' were not present.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	-	1	-
Rat	-	2	7
Mouse	1	1	1
Positives, Negative for this Endpoint			
Human	1	-	2
Rat	1	8	2
Mouse	2	4	1
Sum	**5**	**16**	**13**

DNT-specific Endpoints:

6. General Neuronal Function

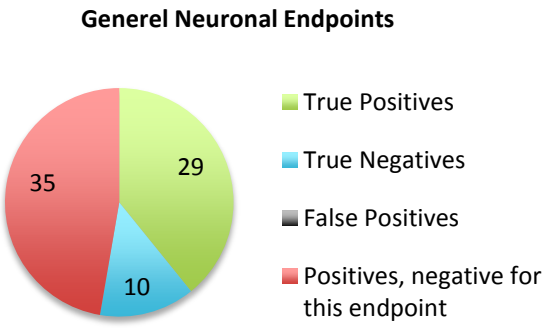


Figure 27 Distribution of citations for the endpoint 'General Neuronal Endpoints' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation

Table 13 For the endpoint 'General Neuronal Endpoints' grouping of Cell Type Category and Species for the categories 'True Positives' and 'Positives, negative for this Endpoint'. 'True Negatives' were not present.

True Positives	Stem-/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	5	-	-
Rat	5	5	4
Mouse	9	-	1
True Negatives			
Human	-	-	2
Rat	-	3	-
Mouse	5	-	-
Positives, Negative for this Endpoint			
Human	8	-	5
Rat	1	9	2
Mouse	10	-	-
Sum	**43**	**17**	**14**

Within the 'General Neuronal Function' endpoint group (n = 74), we identified 29 'True Positives' and 10 'True Negatives'. The 'True Positives' were mainly detected with stem-/progenitor cells (19) and some primary (5) and tumor/immortalized cells (5) also classified the compounds correctly. It is remarkable that for the 18 citations with compounds, which were a priori classified as 'Positive Compounds' and investigated in primary cells, only 5 came out as 'True Positives' (record numbers 2517, 2685, 7058 and 22537). Similarly, for the 21 citations with compounds, which were a priori classified as 'Positive Compounds' and investigated in tumor/immortalized cells, only 4 came out as 'True Positives' (record numbers 683, 3803, 4695 and 12741). In the stem-/progenitor cell group, 19 out of 38 citations with compounds a priori classified as 'Positive Compounds' were 'True Positives'. The majority of 'True

Negatives' were detected with stem-/progenitor cells (5, primary (3) and tumor/immortalized cells (2)). For the 35 'Positives, Negative for this Endpoint' again stem-/progenitor cells provided the majority of these data (19) and some primary (9) and tumor/immortalized cells (7) also contributed to this group. Within the 'True Positives' of the endpoint group 'General Neuronal Endpoints', the main endpoint studied was **neuronal differentiation** (24/29). This was also the case for the 'True Negatives' (9/10) and the 'Positives, Negative for this Endpoint' (29/35). Because neuronal differentiation is a crucial endpoint for brain development and thus should be an integral part of a DNT alternative testing strategy, we evaluated the cell methods detecting this endpoint in more detail taking e.g. compounds into consideration.

Table 14 For the positive compound methylmercury (MeHgCl) cell methods were compared having specific ('True Positives') or non-specific ('Positive, negative for this endpoint') effects on the DNT endpoint 'Neuronal Differentiation'. Publications with 'Reliability Category 3' are printed in red.

		Record Number	Positive, negative for this endpoint		Record Number
Primary cells	Rat midbrain cells	7058	Primary cells	Rat re-aggregate cultures	7273
Stem/Progenitor cells	hESC (H1 cell line)-derived hNSC	2963	Stem/Progenitor cells	mESC, D3	1069
Stem/Progenitor cells	hNPC	2682	Stem/Progenitor cells	mESC, D3	2286
Stem/Progenitor cells	mESC, CGR8	13223	Stem/Progenitor cells	hUCB-NSC	2712
Stem/Progenitor cells	mESC, CGR8	780	Stem/Progenitor cells	hUCB-NSC	10075
Stem/Progenitor cells	mESC, CGR8	13223	Stem/Progenitor cells	hESC (H1 cell line)-derived hNSC	2963
Stem/Progenitor cells	Rat NSC	1546	Tumor/Immortalized cell lines	NT-2	432

HUCB-NSC = human umbilical cord blood stem cells-derived neuronal cells; mESC = mouse embryonic stem cells; hESC = human embryonic stem cells; hNPC = human neural progenitor cells; NSC = neural stem cells

MeHg is a well-studied DNT compound for humans (rev in Grandjean and Landrigan, 2006b). Mechanistic information from *in vivo* data reveals that MeHgCl affects NPC proliferation, migration and differentiation processes (rev. in Johansson et al., 2007). Therefore, this compound is the most popular model compound also in this systematic review (147 total hits with MeHgCl for *in vitro*). For the endpoint 'neuronal differentiation', 7 (out of the 29 'True Positives') were performed with MeHgCl. Of these 7, 6 'True Positives for Neuronal Differentiation' were assessed with stem-/progenitor cells (1 with rat midbrain primary cells) indicating that this cell category has a good ability to assess this important endpoint for DNT. However, analyzing the cell methods that did not show specific MeHgCl effects on neuronal differentiation ('Positives, Negative for this Endpoint'), we found 8 hits. Of these, 6 were also assessed with stem-/progenitor cells, 1 with tumor/immortalized and 1 with primary cells. From this assessment (Table 14) the question arose what the difference between these stem/progenitor cell publications were. For one, 4 of the 7 'Positives, negative for this Endpoint' (record numbers 1069, 2286, 2712 and 432) were categorized into the 'Reliability Category 3' due to either missing statistical evaluations (e.g. because an n<3 was studied) or insufficient study documentation. For 2 of the remaining 3 publications exposure time was rather long (10 and 12 days for record number 7273 and 2963,

respectively), which might have caused the similar effect of MeHgCl on viability and neuronal differentiation as a function of time on the rat re-aggregate cultures (record number 7273). For record number 2963, the hESC (H1 cell line)-derived hNSC appears in both categories, 'True Positive' and 'Positives, negative for this endpoint'. Here, the exposure scheme makes the difference: while MeHgCl treatment for 12 days during neural induction of ESCs causes a specific DNT effect, exposure during neuronal differentiation over these 12 days however, produces non-specific toxicity. Again, the lack in DNT specificity of exposure during differentiation could be a function of time as treatment of primary hNPCs directly for 2 days produced a specific DNT response of MeHgCl on neuronal differentiation.

This evaluation points to stem-/progenitor cell models as most appropriate for studying neuronal differentiation. With regards to species, it is now commonly accepted that best *in vitro* test results are expected when human-based cell models are used (rev. in Andersen and Krewski, 2009, Krewski et al., 2010b, Leist and Hartung, 2013). Taken together, there is a strong **need** for compound screening in human neural stem-/progenitor cell methods for evaluation of neuronal differentiation to assess test performance (sensitivity/specificity). Hereby, a common substance-testing paradigm for specific stem-/progenitor cell types and differentiation protocols should be elaborated (e.g. specific treatment durations for hiPSC-, hESC-or hNPC-derived neurons), which perform appropriately for model compounds altering neuronal differentiation.

7. Cytoskeleton

Within the 'Cytoskeleton' endpoint group (n = 1), no 'True Positives' and 1 'Positive, negative for this Endpoint' was identified (record number 3010). This was a study investigating the effect of rat *in utero* exposure towards MeHgCl on microtubule disruption in primary rat cortical neurons of the offspring after 8 days *in vitro*. However, this study received 'Reliability Category 3' due to relevant methodological deficiencies. Because cytoskeleton alterations are an important 'common key event' (Bal-Price, 2014), there is a clear data gap for assessing this endpoint in the current literature. With regard to the need identification for the endpoint 'Neuronal Differentiation', it would be most appropriate to evaluate the endpoint 'Cytoskeleton' with model compounds also in stem-/progenitor cell types (e.g. hiPSC-, hESC-or hNPC-derived cells). There, predictivity of the methods needs to be shown. This recommendation is based 1) on the fact that human-based cell models are thought to be most predictive for humans (rev. in Andersen and Krewski, 2009, Krewski et al., 2010b, Leist and Hartung, 2013, Seidle and Stephens, 2009b) and 2) on the pragmatic reason that a DNT alternative testing battery already contains a variety of endpoints and thus the number of cell methods should be kept as predictive and as small as possible.

8. Neurite Endpoints

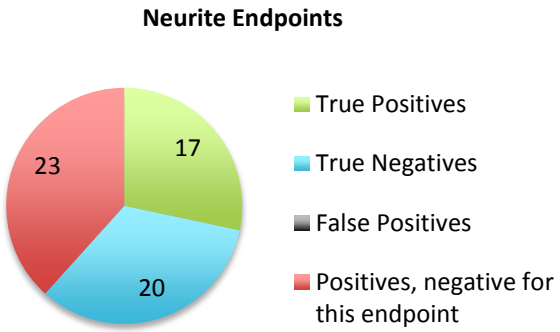


Figure 28 Distribution of citations for the endpoint 'Neurite Endpoints' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

Within the 'Neurite Endpoints' group (n = 60), we identified 17 'True Positives', 20 'True Negatives' and 23 'Positives, Negative for this Endpoint'. The 'True Positives' were mainly detected with tumor/immortalized (9) and primary cells (7), while there is only one study available with stem-/progenitor cells. Concerning this latter cell type category, there were only 4 citations in the whole database assessing neurite endpoints, the other three citations dealing with 'Possible Positives'. Two of those (record number 1452, mESC – D3) showed non-specific toxicity accompanying the 'Neurite Endpoint' and only 1 (record number 10818, rat NSC) measured specific DNT with the 'Neurite Endpoint'. Thus, there is a definite **need** for compound screening in human neural stem-/progenitor cell methods on this endpoint. The major reason for this need is the best ability of neural stem-/progenitor cell models to assess neuronal differentiation. As differentiation into neurons and their neurite endpoints are connected, it again makes sense to also measure neurite endpoints in a cell model best suited for neuronal differentiation. The majority of 'True Negatives' were detected with tumor/immortalized cells (13) and primary cells (7). There is no data in the data collection sheet for assessment of 'Neurite Endpoints' with stem-/progenitor cells. The 23 'Positives, Negative for this Endpoint' were again measured with tumor/immortalized cells (15) and primary cells (9), while there is not data on stem-/progenitor cells available.

Table 15 For the endpoint 'Neurite Endpoints' grouping of Cell Type Category and Species for the categories 'True Positives', 'True Negatives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	-	-	4
Rat	-	5	3
Mouse	1	2	2
True Negatives			
Human	-	-	4
Rat	-	7	9
Mouse	-	-	-
Positives, Negative for			

this Endpoint			
Human	-	-	3
Rat	-	7	11
Mouse	-	2	-
Sum	**1**	**23**	**36**

9. Axonal Endpoints

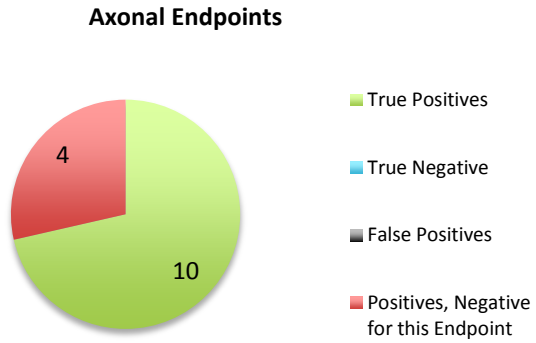


Figure 29 Distribution of citations for the endpoint 'Axonal Endpoints'.

Table 16 For the endpoint 'Axonal Endpoints' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Primary Cells
Human	-
Rat	10
Mouse	-
Positives, negative for this Endpoint	
Human	-
Rat	4
Mouse	-

Within the 'Axonal Endpoints' group (n = 14), we identified 10 'True Positives' and 4 'Positives, negative for this Endpoint', all detected with primary rat cells. These data shows that primary cells are very well suited for assessing effects of chemicals on axons. Looking at the compounds causing adverse or no effects on axons, there are two publications testing the effects of chlorpyrifos and chorpyrifos-oxon on axon length. In the first publications, both compounds produced 'True Positives' (record number 16446) and in the second publication (record number 4451) chlorpyrifos was classified as 'Positive, negative for this Endpoint' while its oxon was found as a 'True Positive'. Both publications use the same study design (chemical exposure 1 h after plating for 24 hrs) and the same cell type (primary rat cells). The only obvious difference between the publications is the location of cell preparation and thus neuronal subtype: the first publication uses rat dorsal root ganglia neurons, which are sensory neurons, while the second one assessed the effects of pesticides in rat superior cervical ganglia neurons being sympathetic neurons. This difference in neuronal cell types and their different biochemical properties might be the cause of distinct effects of chlorpyrifos. Because there are no primary human neurons easily available and chemical testing in general is preferred with human cells (rev. in Andersen and Krewski, 2009, Krewski et al., 2010b, Leist and Hartung, 2013, Seidle and Stephens, 2009b), it has to be examined if 'Axonal Endpoints' can also be studied with stem/progenitor cells, which are available from human origin. Thus, similar to 'General Neuronal Endpoints', 'Neurite Endpoints' and 'Cytoskeleton', there is a **need** for compound testing in human neural stem-/progenitor cell methods on 'Axonal Endpoints' because all neuronal endpoints – if possible – should be assessed in the same cell system. However, the ability of stem-/progenitor cells to picture neuronal cell type specificity and thus different chemical susceptibilities should also be taken into consideration. Moreover, maturation of cells is an issue when comparing primary rodent neuronal cell preparations with human stem-/progenitor cell systems because rodent cells usually show a higher degree of maturation corresponding to a different developmental stage. Being able to use as few cell systems as possible for testing as many endpoints as necessary for an alternative DNT testing strategy will in the end save a lot of resources. But it has to be ensured that these rather 'novel' systems have similar abilities in prediction of toxicity on e.g. 'Axonal Endpoints' than the long-established rodent cultures, in which these endpoints are well measurable.

10. Dendritic Endpoints

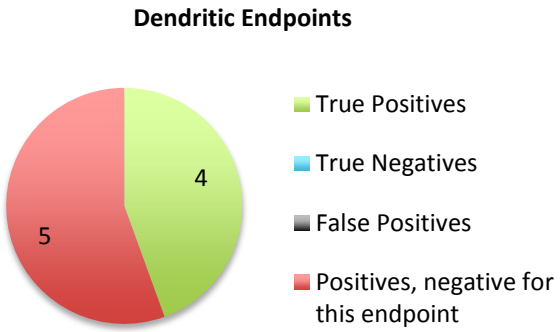


Figure 30 Distribution of citations for the endpoint 'Dendritic Endpoints'.

Table 17 For the endpoint 'Dendritic Endpoints' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Primary Cells
Human	-
Rat	4
Mouse	-
Positives, negative for this Endpoint	
Human	-
Rat	5
Mouse	-

Within the 'Dendritic Endpoints' group (n = 9), we identified 4 'True Positives' and 5 'Positives, negative for this Endpoint', all detected with primary rat cells. Similar, to the 'Axonal Endpoints', these data show that primary cells are very well suited for assessing effects of chemicals on dendrites. Because there are no primary human neurons easily available and chemical testing in general is preferred with human cells (rev. in Andersen and Krewski, 2009, Krewski et al., 2010a, Leist and Hartung, 2013), it has to be examined if 'Dendritic Endpoints' can also be studied with stem/progenitor cells, which are available from human origin. Thus, similar to 'General Neuronal Endpoints', 'Neurite Endpoints' , 'Axonal Endpoints' and 'Cytoskeleton', there is a **need** for compound testing in human neural stem-/progenitor cell methods on 'Dendritic Endpoints' because all neuronal endpoints – if possible – should be assessed in the same cell system. Again, being able to use as few cell systems as possible for testing as many endpoints as necessary for an alternative DNT testing strategy will in the end save a lot of resources. The same precautions as for the 'Axonal Endpoints' are valid for 'Dendritic Endpoints'. Neuronal subtype and developmental timing influence susceptibility to also effects of chemicals on dendrite formation. Thus, when evaluating if stem-/progenitor cells can assess this endpoint, these issues have to be critically evaluated.

11. Synapse Endpoints

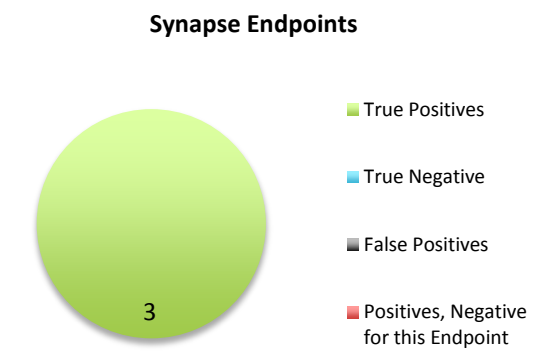


Table 18 For the endpoint 'Synapse Endpoints' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Primary Cells	Stem-/progenitor Cells
Human	-	-
Rat	2	-
Mouse	-	1

Figure 31 Distribution of citations for the endpoint 'Synapse Endpoints'.

Synaptogenesis is crucial for neuronal network formation. Thus, they are elemental in correct wiring of the developmental brain. Therefore, the identified 3 citations (within record number 3037 and 13223) with 'True Positives' and the lack of any other publication on synaptogenesis within this systematic review present a clear **data gap**. In these two publications the effects of toluene and MeHgCl, respectively, on synapse formation were measured. Alterations were assessed after 3 days in rat hippocampal neurons (for toluene) and after 6 days in mESC (CGR8; for MeHgCl). These data suggest that primary cells as well as stem-/progenitor cells are suited for assessing effects of chemicals on synaptogenesis, although the number of citations to make a sound assay performance evaluation is far too low here. Similar to 'General Neuronal Endpoints', 'Neurite Endpoints', 'Dendritic and Axonal Endpoints' and 'Cytoskeleton', there is a **need** for compound testing in human neural stem-/progenitor cell methods on 'Synaptic Endpoints' because all neuronal endpoints – if possible – should be assessed in the same cell system. Because mESC (CGR8) are able to assess adverse effects on synaptogenesis, using human stem-/progenitor cell cultures might be promising, but this has to be experimentally tested.

12. Neuronal functioning/electrophysiology

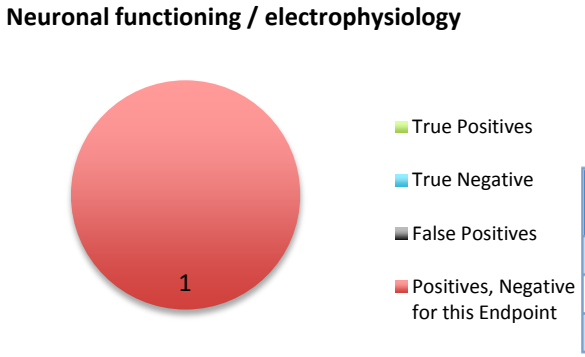


Table 19 For the endpoint 'Neuronal functioning/electrophysiology' grouping of Cell Type Category and Species for the category 'Positives, negative for this endpoint', there was no data for any other category.

Positives, negative for this endpoint	Primary Cells
Human	-
Rat	1
Mouse	-

Figure 32 Distribution of citations for the endpoint 'Neuronal functioning /electrophysiology'.

Of all citations found for the endpoint 'Neuronal functioning/electrophysiology' with DNT positive or negative compounds and viability assessment, only 1 (record number 20775; rat hippocampal neurons) could be grouped into the category 'Positives, negatives for this Endpoint' and none was found for 'True Positives' or 'True Negatives'. Exposure in this publication, however, was not during network formation but during measurement and the publication received 'Reliability 3' due to insufficient documentation. Thus, there is the urgent need for assessment of neuronal network formation with any cell system, preferably like for the other neuronal endpoints discussed above with human-based stem-/progenitor cell methods. Effectively, neuronal network activity should be altered if a compound disrupts neuronal differentiation, axon, dendrite formation or synaptogenesis as well as neuronal subtypes (see below). Thus, the hypothesis might be tested if one could either restrict alternative DNT testing to effects on neuronal network formation or if assessment of the mentioned single neuronal endpoints in high content analyses could predict network formation. However, so far there is far too few chemical testing data available to answer this question. One human ESC-based model has to be mentioned at this point which has been established for DNT testing purposes but has not been published with compounds yet – and thus was not recognized within this systematic review (Kapucu et al., 2012, Yla-Outinen et al., 2014). Differentiation of human ESC on multi electrode arrays (MEAs) was established within this work and neuronal network activities recorded. From this biotechnological 'state-of-the-art', one could launch a screening platform for neurodevelopmental toxins.

13. Glutamatergic function

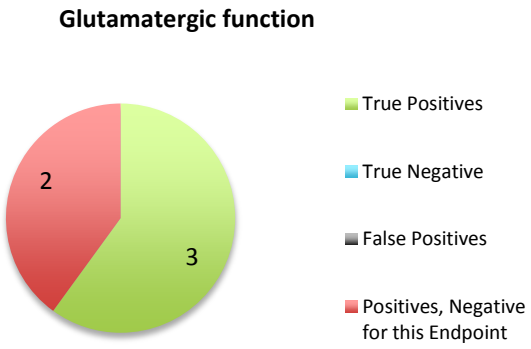


Figure 33 Distribution of citations for the endpoint 'Glutamatergic Function'.

Table 20 For the endpoint 'Glutamatergic functioning' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Primary Cells	Stem-/progenitor Cells
Human	-	-
Rat	-	-
Mouse	3	-
Positives, negative for this endpoint		
Human	-	-
Rat	-	-
Mouse	-	2

All endpoint citations 'Glutamatergic Function' with viability data of the categories depicted in Table 7A are shown in Figure 33. From the total of 5 citations for this endpoint with DNT positive compounds, 3 are identified as 'True Positives' and 2 as 'Positives, Negative for this Endpoint'. These data were obtained with mouse cortical neurons (see Table 20) and are derived from only one publications (record number 541). Investigated is one single compound (sodium arsenite) for exactly the same exposure conditions (48 hrs) but the endpoint was determined in 5 different ways. Because arsenic was correctly identified as 'True Positive' only with protein expression analyses methods of GluA1, which was the method of choice in this cell system, while GluA2/3/4 expression as well as extracellular glutamate (ELISA) did not reveal DNT-specific toxicity of arsenic. However, many more studies are missing to evaluate the best detection method for this endpoint. At this point, I would like to refer to the discussions of the previous neuronal endpoints which I do not want to repeat here but stays also true for this endpoint of 'Glutamatergic Function'.

14. Cholinergic Function

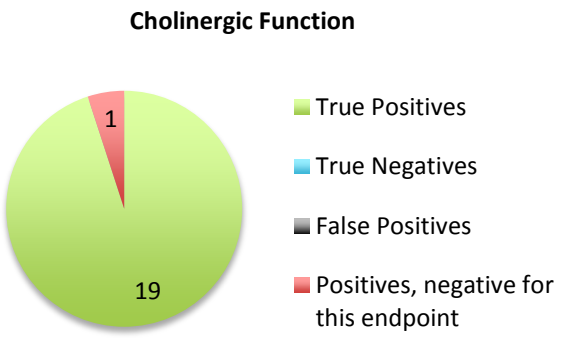


Figure 34 Distribution of citations for the endpoint 'Cholinergic Function'.

Table 21 For the endpoint 'Cholinergic Function' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells

Human	-	-	2
Rat	-	9	4
Mouse	3	-	1
Positives, negative for this Endpoint			
Human	-	-	-
Rat	-	1	-
Mouse	-	-	-

The endpoint 'Cholinergic Function' has mainly been assessed in 'Primary Cells' (9) and 'Tumor/immortalized Cells' (7), but is also assessable with 'Stem-/progenitor Cells'. This seems to be a very reliable endpoint either measuring Acetylcholin Esterase (AChE) activity as the specific enzymatic cholinergic function or Choline acetyltransferase activity as a marker for cholinergic neurons. The main compounds investigated in these cell assays were Clorpyrifos and Chlorpyrifos Oxon. Because Chlorpyrifos Oxon is the active metabolite of Chlorpyrifos mediating the anticholinergic function of this pesticide, we expect it to show its anticholinergic activity at lower concentrations than the parent compound Chlorpyrifos. For almost all cell methods collected in this systematic review and correctly identifying 'True Positives', this was true. Moreover, active concentration ranges for Chlorpyrifos were positioned in the nM-µM range, whereas the corresponding active Chlorpyrifos Oxon concentrations were positioned in the pM-nM range and differed in potency for these two compounds within one cell method. There are two exceptions. One publication (record number 11066) utilized mESC-D3, where Chlorpyrifos Oxon affected AChE activity still at lower concentrations than Chlorpyrifos, but both in the µM range, which is significantly higher than in other cell systems. The second (record number 12741) used the mouse tumor/immortalized cell line N2a, where only the Oxon of Chlorpyrifos was assessed and inhibited AChE also in the µM range, which is far above the concentrations affecting this enzyme in other cell systems. From these data it can be concluded that besides these two cell methods, the remaining tested methods are suitable for assessing the endpoint 'Cholinergic Function'. If human stem-/progenitor cells are also able to assess such cholinergic endpoints **needs** further evaluation.

15. Dopaminergic Function

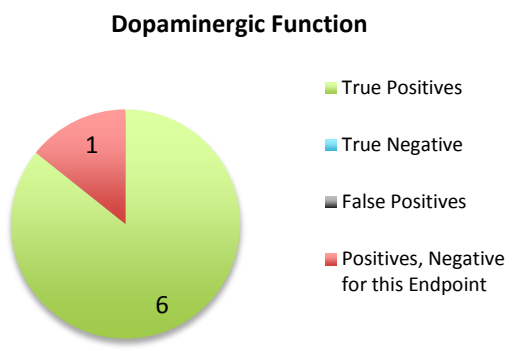


Figure 35 Distribution of citations for the endpoint 'Dopaminergic Function'.

Table 22 For the endpoint 'Dopaminergic Function' grouping of Cell Type Category and Species for the categories 'True Positives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progeni-tor cells	Tumor/immor-talized Cells
Human	-	-
Rat	-	2
Mouse	4	-
Positives, negative for this Endpoint		
Human	1	-
Rat	-	-
Mouse	-	-

The endpoint group 'Dopaminergic Function' has successfully been assessed ('True Positives') with mESC (CGR8; 4) and the rat tumor/immortalized cell line PC-12 (2), whereas the human ESC cell line (H1)-derived hNSC were not able to assess dopaminergic toxicity. In 5 of the 7 studies MeHgCl was used as the DNT compound, which might, among its other prominent cellular effects, also have an impact on the dopaminergic system during brain development (Dare et al., 2003, Gimenez-Llort et al., 2001). The 4 hits identifying MeHgCl as a true positive compound with regards to the endpoint "dopaminergic function" were all from one publication (record number 13223) which used mESC. In this publication, MeHgCl altered dopamine receptor activity and expression in the pM-low nM range after 6 days of exposure. In contrast, hESC(H1)-derived NSC were not able to detect MeHgCl as a DNT compound altering dopaminergic function (record number 2963), which might either be due to the long exposure time (12 days for hESC-NSC vs. 6 days for mESC), here cytotoxicity masking the specific DNT effects on dopaminergic endpoints as a function of time; or due to a species-specific effects of MeHgCl - possibly affecting the dopaminergic system in rodents but not in humans. This cannot be excluded from this data. Hence, stem/progenitor cells turned out to be a suitable model for testing dopaminergic function, but as this is so far based on a single publication, further evaluation on the ability of human stem/progenitor cells to assess this endpoint is **needed**.

16. Neural Proliferation

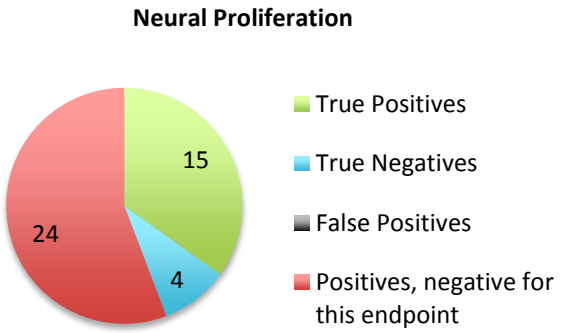


Figure 36 Distribution of citations for the endpoint 'Neural Proliferation' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation is thereby one citation of a cell model with one compound and not the number of actual papers. Thus, papers with multiple compounds and/or multiple cell methods will cause several hits in this evaluation.

Within the 'Neural Proliferation' Endpoint group (n = 43), we identified 15 'True Positives', 4 'True Negatives' and 24 'Positives, negative for this Endpoint'. The 'True Positives' were detected with primary cells (5) and stem-/progenitor cells (10), the 'True Negatives' identified with mouse stem-/progenitor cells (4) and the 'Positives, negative for this Endpoint' were mainly found with stem-/progenitor cells (15) and some with rat primary cells (9) (

Table 23). Tumor/immortalized cells do not occur in this evaluation because due to their altered cell cycle gaged to unrestricted growth, usage of tumor/immortalized cells for assessment of neural proliferation was an a priori exclusion criterion.

Table 23 For the endpoint 'Neural Proliferation' grouping of Cell Type Category and Species for the categories 'True Positives', 'True Negatives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells
Human	1	1
Rat	9	4
Mouse	-	-
True Negatives		
Human	-	-
Rat	-	-
Mouse	4	-
Positives, Negative for this Endpoint		
Human	4	-
Rat	1	9
Mouse	10	-
Sum	**29**	**14**

Evaluating the cell methods in more detail for their performance with regards to the endpoint group 'Neural Proliferation', specific compound effects have to be taken into consideration. The 12 citations revealing 'True Positives' for DNT utilized Lead acetate, Manganese(II)chloride, MeHgCl, PCB126 and Toluene. Of those, only MeHgCl is a true inhibitor of 'Neural Proliferation' *in vivo* (Bose et al., 2012, Burke et al., 2006), whereas the other compounds affect neurodevelopment through different target processes (e.g. Toluene affects astroglial differentiation (Burry et al., 2003) or Manganese alters the dopaminergic system (Tran et al., 2002)). In the 'Positives, negative for this Endpoint' group, 16 different compounds were negative, although a priori classified as 'DNT positive compounds'. Of those 16, 4 (Ethanol (rev. in Luo and Miller, 1998), MeHgCl (Bose et al., 2012, Burke et al., 2006), Methylazoxymethanol (De Groot et al., 2005) and Valproic Acid (Go et al., 2012)) are known to exert their DNT properties amongst others through alteration of neural proliferation. These 4 compounds should have been positive for this endpoint, but turned out to be negative. Table 24 summarized the cell methods, which correctly identified the 'True Positives' and the ones, which were 'Positives, negative for this Endpoint'.

Table 24 For positive compounds ('True Positives') or non-specific ('Positive, negative for this endpoint') effects on the DNT endpoint 'Neural Proliferation'. Publications with 'Reliability Category 3' are printed in red.

True Positive		Record Number	Positive, negative for this Endpoint		Record Numbe
Primary cells	Rat hippocampal astrocytes	3712	Primary cells	Rat cortical astrocytes	3712
Stem/Progenitor cells	Rat embryonic cortical NSC	967	Primary cells	Rat cerebellar astrocytes	3712
Stem/Progenitor cells	Rat embryonic cortical NSC	967	Primary cells	Rat brainstem astrocytes	3712
Stem/Progenitor cells	Rat NPC from E13 cortex	2112	Stem/Progenitor cells	mESC, D3	343

Stem/Progenitor cells	Rat NPC from E13 cortex	2112	Stem/Progenitor cells	mESC, D3	343
Stem/Progenitor cells	Rat NSC from E15 cortex	1546	Stem/Progenitor cells	mESC, D3	343
Stem/Progenitor cells	human umbilical cord blood stem	2712	Stem/Progenitor cells	Mouse cortical neural stem cells	1138
			Stem/Progenitor cells	human umbilical cord blood neural	10075
			Stem/Progenitor cells	hESC (H9)	13548

Almost all 'True Positives' for the endpoint 'Neural Proliferation' were assessed with 'Stem-/progenitor cells' (only one used primary rat hippocampal astrocytes; record number 3712). Thus, these seem to be the cell types of choice for studying NS/PC and astrocyte proliferation, respectively. Interestingly, effects of MeHgCl on developmental astrocyte proliferation seems to be brain region-specific: while hippocampal astrocytes categorize under 'True Positives', cortical, cerebellar and brain stem astrocytes proliferation is not specifically inhibited by MeHgCl in the same publication with the same treatment scheme (record number 3712) thus grouping into the category 'Positive, negative for this Endpoint'. This publication indicates nicely that known brain region differences in astrocyte physiology (Hewett, 2009) can cause distinct susceptibilities towards chemicals. Here, with the current knowledge, 'Primary Cells' are so far the only cell type that is able to assess those brain region-specific differences *in vitro* because cells keep their signaling specificity when taken out of the *in vivo* context into a dish (Sharif et al., 2009). Because very little is known so far, which signaling molecules trigger such brain-region specificity *in vivo*, it is with the current state of the art not possible to mimic brain region-specific cell differentiation, especially astrocyte formation, with stem cell-based methods *in vitro*. Nevertheless, one of these publications (record number 1546) investigated proliferation in differentiating NSC. Because neurons are postmitotic, but glia cells keep proliferative capacities while differentiating (Dimou and Gotz, 2014), this publication actually also investigates glia, and not NS/PC proliferation, yet in general, but not brain region-specific. With regards to the species, there is only one human publication available meeting the inclusion criteria of this systematic review (record number 2712), which received 'Reliability 3' due to insufficient methodological documentation. Thus, it is not possible to assess species-specific data on the endpoint 'Neural Proliferation' showing the clear **need** for more human studies assessing proliferation effects of DNT compounds. Moreover, there is only one study showing brain region-specific effects on astrocyte proliferation. That brain region-specificity is actually an important issue, also for other compounds than MeHgCl, **needs** to be shown by testing more chemicals acting as DNT compounds through inhibition of proliferation.

For the category 'Positive, negative for this Endpoint', the 3 citations with 'Primary Cells' (astrocytes) were already discussed above in the brain region-specific context. One other cell model (mESC-D3) in the category 'Positive, negative for this Endpoint' received three citations for the compounds MeHgCl, Methylazoxymethanol and Valproic Acid (record number 343). For all three DNT compounds effects on viability were not distinguished from neural proliferation when cells were exposed for 48 hrs. Also mouse cortical NSCs (record number 1138) when exposed only for 24 hrs to MeHgCl were not able to distinguish effects on proliferation from cytotoxicity. The same was true for human umbilical cord blood-derived NSC. After 48 hrs of MeHgCl exposure, specific effects on neural proliferation were not distinguished from cytotoxicity. For the DNT compound ethanol, hESC (H9) were not able to detect any hazard on proliferation when treating during early periods of the hESC differentiation towards neural rosettes (record number 13548). This cell differentiation stage is supposedly equivalent to developmental

processes in the early 1st trimester of human embryogenesis. Instead of ethanol-induced reduced proliferation, this publication identifies an increase in neural apoptosis as the major ethanol effect. As for humans, the precise mechanism of ethanol DNT is not known and thus induction of apoptosis might be a major contribution to the neurodevelopmental phenotype of the fetal alcohol syndrome. Ethanol also induces apoptosis in developing brains in mice *in vivo* (Young et al., 2003). Hence, evaluation of human cell methods with compounds having unclear precise modes of actions during brain development in humans is rather difficult. One way around this problem is the application of the parallelogram approach. When a certain cell system is predictive e.g. for rats, where there is sufficient data on mode of action of compounds *in vivo*, then the equivalent cell system will with a high probability be also predictive for different species including humans. This approach has been tested with primary NPs as well as primary rat cortical and hESC-derived neurons for different DNT-specific endpoints (Baumann et al., 2014, Gassmann et al., 2010, Harrill et al., 2011).

17. Neural Stem-/progenitor cell Endpoints

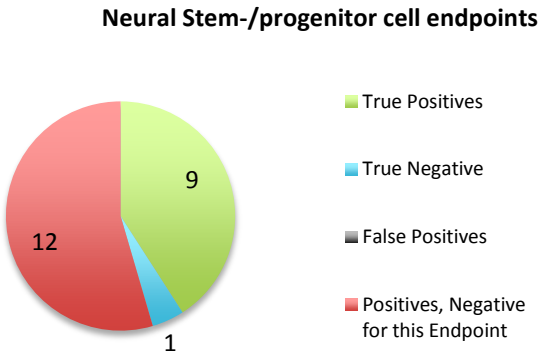


Figure 37 Distribution of citations for the endpoint group 'Neural Stem/Progenitor cell endpoints'.

The endpoint group 'NS/PC Endpoints' is mainly comprised of different methods assessing stemness- or progenitor cell markers or morphological aspects like neural rosette formation. Thereby, 9 citations with 'True Positive' compounds were identified, whereas 12 citations describe 'Positives, negative for this endpoint'. For these analyses, only stem-/progenitor cells (with the exception of one study using rat primary cells) were used (Table 25). Interestingly, the compounds producing 'True Positives' and 'Positives, negative for this Endpoint' were very similar, with both groups containing e.g. MeHgCl and Valproic acid as two human DNT compounds, which alter neural stem cell development with high probability (Theunissen et al., 2010, Foti et al., 2013, Go et al., 2012).

For valproate, all 3 citations (2 'True Positives' and 1 'Positive, negative for this Endpoint') were produced with hESC. For the first 2 (record number 1077) human ESC H9 cells (line WA09) were exposed for 6 days during neural induction towards valproate. Progenitor cell-associated gene expression changes were seen after valproate exposure in the high µM to low mM range. In contrast, for the latter the hESC HUES-1 line (record number 878) was exposed for 9 days during neural induction. Yet, the two publications employed completely different hESC-neural induction protocols. While record number 1077

uses the noggin/dorsomorphin/ROCK inhibition protocol for neural induction, record number 878 induced neural differentiation through embroid bodies and cultivation in modified DMEM/F12/ N2 medium containing FGF-2. Morphological changes in neural rosettes and gene expression alterations were observed at concentrations where cells were dying (starting from 30 µM up to 300 µM). Most gene expression data is found as supplemental data and in the data collection sheet categorized under 'Omics'. For the microarray validation, several genes were verified, but only one gene overlaps between publication 1077 and 878, which is FOXG1. This was significantly changed in the latter publication, while it was not altered in the first. However, publication 1077 investigated stem-/progenitor cell related genes like Pax-6 and Sox-2, while publication 878 did not explicitly measure those. Differences in results between the two publications could thus be attributed to i) usage of different stem cell lines, ii) different neural induction protocols, iii) different exposure times or iv) different endpoint measures. While the comparison of these two publications can be seen as a 'case study', this problem in stem cell protocols, especially with regard to method of neural induction and timing of *in vitro* exposure (during EB formation, during neural induction, during neuronal/glial differentiation, during neural cell maturation), is occurring throughout the collected literature of this scientific review. Therefore, there is an urgent **need** for harmonization of protocols employing hESC (as well as human induced pluripotent stem cells, hiPSCs) in toxicity studies.

The publications investigating the effects of MeHgCl (record numbers 1546 and 2286) identified the 5 'True Positives', while record numbers 2286 (!) and 2963 showed endpoints not distinguishable from cytotoxicity (record number 2286) or no MeHgCl-induced endpoint alteration for this endpoint (record number 2963). 'True Positives' were recognized with rat NSC (record numbers 1546) or mESC-D3 (record numbers 2286) by measuring Notch1 protein expression, Hes5 and Math1 gene expression or % outgrowth from plated embroid bodies as a morphological feature. MeHgCl-effects that were not classified as true positives were number of mESC-D3 cells expressing nestin or the pluripotency marker SSEA1 (record number 2286) or nestin gene expression of hESC (H1 line). Thereby, exposure scheme did not seem to play the major role in the classification because across these 7 citations, a variety of different exposure schemes were used. Concerning the missing effects of MeHgCl on nestin expression there is the **need** for clarification if for the stem/progenitor cell endpoint expression of Hes5 and Math1, both regulating neuron versus glia differentiation might be the better endpoint measure than nestin. Because none of the cell method assessed all markers, this question cannot be answered at this time.

Table 25 For the endpoint 'NS/PC Endpoints' grouping of Cell Type Category and Species for the categories 'True Positives' , 'True Negatives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells
Human	4	-
Rat	2	-
Mouse	3	-
Positives, negative for this Endpoint		
Human	7	-
Rat	4	-
Mouse	1	-
True Negatives		
Human	-	-
Rat	-	1
Mouse	-	-
Sum	21	1

18. Migration

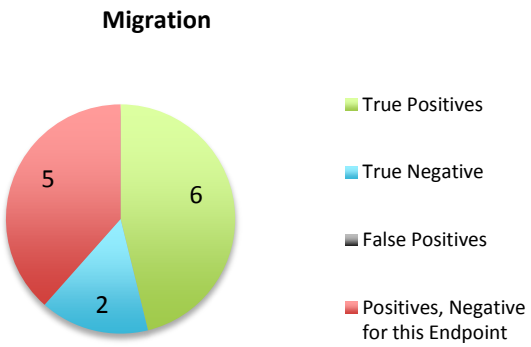


Figure 38 Distribution of citations for the endpoint group 'Migration'.

Within the endpoint group 'Migration' (n = 13, Figure 38) 'True Positives' were identified with stem-/progenitor cells (4) and primary cells (2), 'Positives, negative for this Endpoint' with tumor/immortalized cells (4) and stem-/progenitor cells (1), and the 'True Negatives' were classified tumor/immortalized cells (2; Table 26). Within the compounds tested for effects on migration, MeHgCl is the human DNT compound where migration effects during brain development were clearly described as one mode of action (Fahrion et al., 2012, Kakita et al., 2002, Mancini et al., 2009). Therefore, a migration assay used for DNT testing should be able to identify this positive compound correctly as 'True positive'. This was the case for human NPCs (record number 2682), mouse CGC (record number 3566) and rat cerebellar slice cultures (record number 15582). Because the slice cultures are more animal-, labour- and time-intensive than stem-/progenitor cell-based tissue cultures, slice cultures in general are not the first choice when it comes to routine chemical testing. Studying the cell methods showing no specific toxicity for MeHgCl on cell migration, these are the human tumor/immortalized cell line NT-2 for migration (record number 432) and the human umbilical cord blood-derived NSC line for adhesion, a prerequisite for proper migration (record number 10075). Although NT-2 cells identify 'True Negatives' correctly, the lack in specificity for effects of the model compound MeHgCl on the endpoint migration identifies this cell method as not suited for measuring this endpoint. Due to the small amount of data for this endpoint, there is a clear **need** to e.g. assess if other stem-/progenitor cell methods are also able to predict this endpoint correctly and if compounds known to affect migration *in vivo* (e.g. growth factors) also modulate migration in these assays. Looking at record number 2286 within the 'NS/PC Endpoints', migration of mESC-D3-derived embroid bodies was measured. However, in this paper the authors did not find effects of MeHgCl on migration distance, but measured % of embroid body outgrowth, which clearly has to be distinguished from NS/PC migration (covering distance/time) assuming that mESC-D3 embryoid bodies are not suited for assessing this endpoint.

Table 26 For the endpoint 'Migration' grouping of Cell Type Category and Species

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized cells
Human	4	-	
Rat	-	1	
Mouse	-	1	
Positives, negative for this Endpoint			
Human	1	-	4

Rat	-	-	-
Mouse	-	-	-
True Negatives			
Human	-	-	2
Rat	-	1	-
Mouse	-	-	-
Sum	5	3	6

19. Astrocyte Endpoints

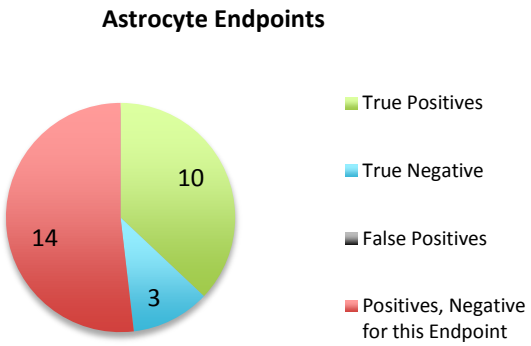


Figure 39 Distribution of citations for the endpoint group 'Astrocyte Endpoints'.

Within the endpoint group 'Astrocyte Endpoints' (n = 27, Figure 39) mainly astrocyte differentiation by assessment of astrocyte-specific markers via different techniques was measured. 'True Positives' were identified with stem-/progenitor cells (3) and primary cells (7), 'Positives, negative for this Endpoint' with primary cells (11) and stem-/progenitor cells (3), and the 'True Negatives' were classified correctly only with primary cells (3; Table 27). Within the compounds tested for effects on astrocyte endpoints, ethanol and toluene are two DNT compounds affecting astrocyte development *in vivo* (Burry et al., 2003, Vemuri and Chetty, 2005, Valles et al., 1996). Therefore, an alternative assay used for DNT testing should be able to identify this positive compound correctly as 'True Positive'. This was the case for mouse SFME cells, which have stem-/progenitor cell properties and can be induced to differentiate into astrocytes (record number 5494). These cells identified toluene as 'True Positive' for disturbance of astrocyte differentiation measured by GFAP protein expression. In contrast, the hESC H9 line identified ethanol as a 'Positive, negative for this Endpoint' when treated during proliferation or differentiation of cells because ethanol affected cell viability of hESC-H9 cells in a similar manner than GFAP expression by RT-PCR or immunofluorescence (record number 13548). Because each compound was only tested in one cell method, a comparison of methods is impossible. Thus, there is a clear **gap** identification for assessing developmental astrocyte toxicity of compounds with more cell methods and positive compounds.

Table 27 For the endpoint 'Astrocyte Endpoints' grouping of Cell Type Category and Species for the categories 'True Positives' , 'True Negatives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells
Human	-	-
Rat	3	6

Mouse	-	1
Positives, negative for this Endpoint		
Human	3	-
Rat	-	11
Mouse	-	-
True Negatives		
Rat	-	3
Sum	6	21

20. Oligodendrocyte Endpoints

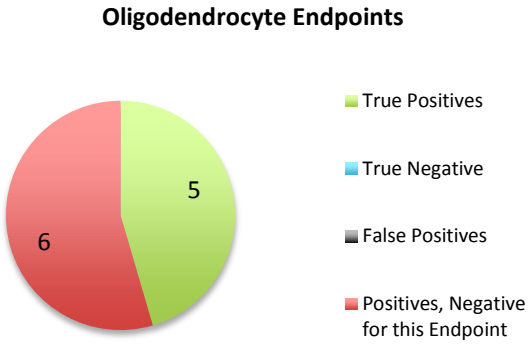


Figure 40 Distribution of citations for the endpoint group 'Oligodendrocyte Endpoints'.

For 'Oligodendrocyte Endpoints' (n = 11, Figure 40) oligodendrocyte differentiation was assessed by different methods. 'True Positives' were identified with stem-/progenitor cells (5) and 'Positives, negative for this Endpoint' with primary cells (4) and stem-/progenitor cells (2; Table 28). The oligodendrocytes endpoints are difficult to assess because 1) there are very few publications/cell models for this important endpoint available and 2) there is a lack in positive reference compounds known to affect oligodendrocyte differentiation *in vivo*. E.g. there are only 2 publications (record numbers 2313 and 4819) assessing toxicity on oligodendrogenesis categorized as 'True Positives'. One employs hNPC and the other one rat neurospheres prepared from different brain regions. Compounds used were PBDE and lead, respectively. For both compound groups, there is only indirect or *in vitro* evidence for their effects on oligodendrogenesis available: the indirect evidence would be that PBDE act on oligodendrogenesis through thyroid hormone (TH) disruption, a known feature of this class of flame retardants (Kodavanti et al., 2010, Zhou et al., 2002) because TH is indispensable for oligodendrocyte formation and maturation *in vivo* (Ibarrola and Rodriguez-Pena, 1997, Marta et al., 1998, Rodriguez-Pena et al., 1993). The *in vitro* evidence that lead affects developing oligodendrocytes was also brought earlier (Deng et al., 2001, Huang and Schneider, 2004). Due to this lack in compounds being 'True Positives' for oligodendrogenesis, the cell methods producing 'Positives, negative for this Endpoint' cannot be evaluated, especially because neither PBDE nor lead was used in these 6 publications. From these studies one can only state that rat re-aggregate cultures (record numbers 7273 and 21747) as well as hESC-H9-derived neural cells (record number 13548) are able to produce oligodendrocytes. Thus, there is a clear **need** for a further and thorough characterization of methods which have the ability to investigate oligodendrocyte differentiation *in vitro* e.g. with model compounds known to contribute to oligodendrocyte formation/maturation.

Table 28 For the endpoint 'Oligodendrocyte Endpoints' grouping of Cell Type Category and Species for the categories 'True Positives' , 'True Negatives' and 'Positives, negative for this Endpoint'.

True Positives	Stem/progenitor cells	Primary Cells
Human	2	-
Rat	3	-
Mouse	-	-
Positives, negative for this Endpoint		
Human	2	-
Rat	-	4
Mouse	-	-
Sum	7	4

21. Neuroinflammation

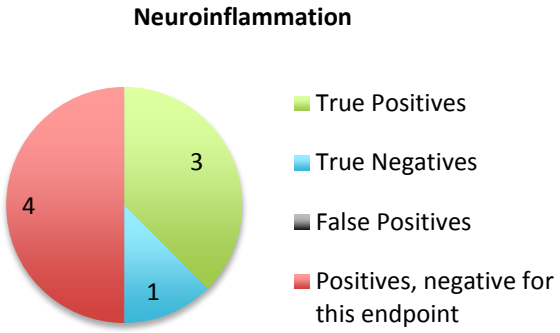


Figure 41 Distribution of citations for the endpoint group 'Neuroinflammation'

Table 29 For the endpoint 'Neuroinflammation' grouping of Cell Type Category and Species for the category 'True Positives', there was no data for any other category.

True Positives	Primary Cells
Rat	3
Positives, negative for this Endpoint	
Rat	4
True Negatives	
Rat	1

The only two published methods for assessing the effects of compounds on microglial cells are the re-aggregated rat culture and rat cerebellar granule cells. While the re-aggregated rat cells identify 3 DNT positive compounds (lead (record number 5596), mercury(II)chloride and MeHgCl (record number 7273)) as true positives by counting microglia cell number in a 3D context, CGC identify the 'True Negative' Aspirin as indeed negative, put the three potentially positives MeHgCl, lead and valproate as 'Positives, negative for this Endpoint' by counting these cells in a 2D format (record number 2685). However, amongst the possible human DNT compounds there is data from trimethyltin (TMT), which is a known stimulator for neuroinflammation (rev. in Corvino et al., 2013). CGC correctly identify TMT as an inductor of neuroinflammation by microglial proliferation (record number 2685). For this important endpoint, there is a true **need** for more data: methods and chemical testing. Inflammatory processes have recently been shown to differ tremendously between species (Seok et al., 2013). Thus, recent rodent data on assay development for assessment of neuroinflammation is very valuable and now has to be used to establish corresponding human methods. That this is technically possible in a highly sophisticated manner was just shown by combining endothelial progenitor cells, pericytes and microglia cells in hydrogels forming 3D vascular networks. In a subsequent step hiPSC-derived neurons and glia cells were added creating 'minibrains' (rev. in Hou et al., 2013). This procedure is currently being set up as a large-scale screening platform for testing compounds on their capacities to interfere with human brain development. This is one of the most promising efforts for future *in vitro* DNT chemical testing. However, the only **lack of data** here is the comparison of hiPSC-derived neural cells to primary human cells. It has to be shown that hiPSC-derived neurons and glia cells have similar physiological properties than their stem cell-derived 'siblings'. This **need** is actually there for all hESC- and hiPSC-derived methods, even if they are utilized in a simpler manner.

22. Omics

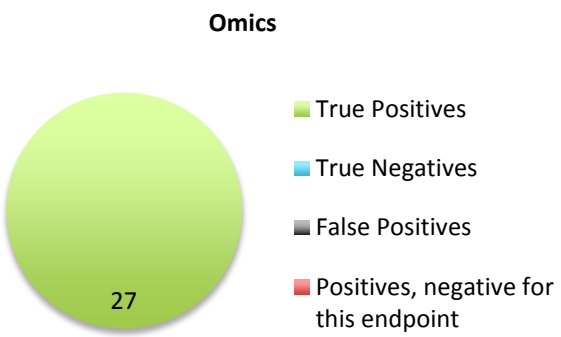


Figure 42 Distribution of citations for the endpoint group 'Omics'.

Most 'Omics' studies (17) were performed in Tumor/immortalized cells, while the number of 'Omics' studies in Primary (3) and Stem-/Progenitor cells (7) are underrepresented. For regulatory purposes, it is difficult to use studies based on expression changes because the distinction between an adaptive and an adverse change in gene expression is not easy to make, yet fundamental (Boekelheide and Andersen, 2010). A few publications have been addressing this issue in non-DNT related cell types (van Dartel et al., 2011, Andersen et al., 2010) as well as in the mESC-D3-based neuro-embryonic stem cell test (record number 1136). These publications revealed that at low compound concentrations genes regulating transcription and metabolism might indicate adaptive cell responses, while at higher concentrations transcriptome alterations for regulation of pathways involved in cell cycle, DNA damage and apoptosis as well as for specific processes like development might be indicative for adverse cell reactions. However, quantitative aspects of gene expression as well as threshold definition for adaptation versus adversity are necessary to make use of such studies in a regulatory context in the future (record number 1136).

Table 30 For the endpoint 'Omics' grouping of Cell Type Category and Species for the categories 'True Positives' and 'True Negatives'.

True Positives	Stem/progenitor cells	Primary Cells	Tumor/immortalized cells
Human	5	1	1
Rat	-	2	16
Mouse	2	-	-
Sum	7	3	17

Evaluation of cell methods with regards to correct ('True Positive', 'True Negative') endpoint evaluation of DNT compounds

From the data of DNT compounds 'True Positives' and 'True Negatives' across all endpoint groups and cell type categories we next assessed which cell type category is able to correctly identify which endpoints. Therefore, we first grouped the endpoint groups again into endpoint categories. These endpoint categories reflect important milestones of neurodevelopmental processes that have to be integrated into an alternative testing strategy. Categorization of endpoint groups is shown in Table 31.

Table 31: Categorization of 'Endpoint Groups' according to milestones of neurodevelopmental processes, which have to be integrated into an alternative testing strategy

ENDPOINT CATEGORY	ENDPOINT GROUP
NON-DNT-SPECIFIC ENDPOINTS	Apoptosis
	Energy Metabolism
	Epigenetics
	General Cellular Function
	Oxidative Stress
STEM/PROGENITOR CELL ENDPOINTS	Neural Proliferation
	NS/PC Endpoints
MIGRATION	Migration
NEURONAL ENDPOINTS	Axonal Endpoints
	Catecholaminergic Function
	Cholinergic Function
	Dendritic Endpoints
	Dopaminergic Function
	GABAergic function (+ chlorine)
	General Neuronal Endpoint
	Glutamatergic Function
	Neurite Endpoints
GLIAL ENDPOINTS	Astrocyte Endpoints
	Oligodendrocyte Endpoints
NEURONAL NETWORK ENDPOINTS	Neuronal functioning/electrophysiology
	Synapse Endpoints
NEUROINFLAMMATION	Neuroinflammation
OMICS	Omics

Evaluation of these 'Endpoint Categories' for the individual cell type categories revealed that since 1994 the majority of citations for 'True Positives' and 'True Negatives' were assessed with rat primary cells (n = 95; Table 32). This is probably due to two reasons: for one, most studies included into this systematic review were performed with rat primary cells (Figure 9 and Figure 10). Second, as stated above, primary cells taken out of the *in vivo* situation into a dish are known to reflect physiological processes *in vitro*. These two points are well reflected in this DNT compound-specific data. Because the rat has historically been the testing species of choice for DNT evaluation *in vivo* and is thus the animal used in the DNT-guidelines in Europe (OECD 426) and the US (EPA OPPTS 870.6300) there have also more data been produced with primary rat cells than with primary mouse cells (n = 17). Moreover, there is more *in vivo* data from rat than from mouse available (see *in vivo* part of this report). This might be biasing the current state of the art towards 'true' DNT effects in rats. Assuming that similar to cancer studies (Purchase, 1980) rodent species within each other do not predict each other's DNT well, this could also be a reason why relatively few citations appear in the 'True Positives' and 'True Negatives' categories for mouse primary studies. As also stated above, primary human cells are hardly ever available, there are only 7 citations.

Stem cell techniques started to evolve in toxicological sciences after the achievement was made that embryonic stem cells spontaneously differentiated into derivatives of all three primary germ layers, endoderm, ectoderm and mesoderm *in vitro* (Evans and Kaufman, 1981, Martin, 1981). The first stem cell-based test with the goal of chemical testing was the embryonic stem cell test (EST) using mESC-D3 (comprehensively reviewed by Seiler and Spielmann, 2011). A few years later, also neurodevelopmental toxicity studies were performed in stem cells. Within this systematic review, the first publication using

stem-/progenitor cell-based method was published in 2002 (record number 5494) followed by the second one in 2004 (record number 4819). Considering the late start in publications with stem-/progenitor cells, (see Figure 11b), the total number of 126 individual citations with this cell method in comparison to 119 with primary cells, which are assessed here since 1994 is impressive.

Table 32: Total number of 'True Positives' and 'True Negatives' for DNT compounds across the different cell type categories for all endpoints.

DNT compounds	Stem/progenitor cells	Primary Cells	Tumor/immortalized Cells
Human	39	7	19
Rat	47	95	61
Mouse	40	17	17
Sum	126	119	97

The abilities to correctly measure endpoints within endpoint categories (defined in Table 31) are summarized in Figure 43. The dark blue color (■) in the pie diagrams reflects the 'Non-DNT-specific Endpoints', which are not highly relevant as part of an alternative testing strategy because they represent very early non-specific key events, which most likely underlie adaptive cellular and organ reactions that can subsequently alter the adverse outcome. However, for all mouse cell type categories as well as for the few primary human studies, this endpoint category is dominating.

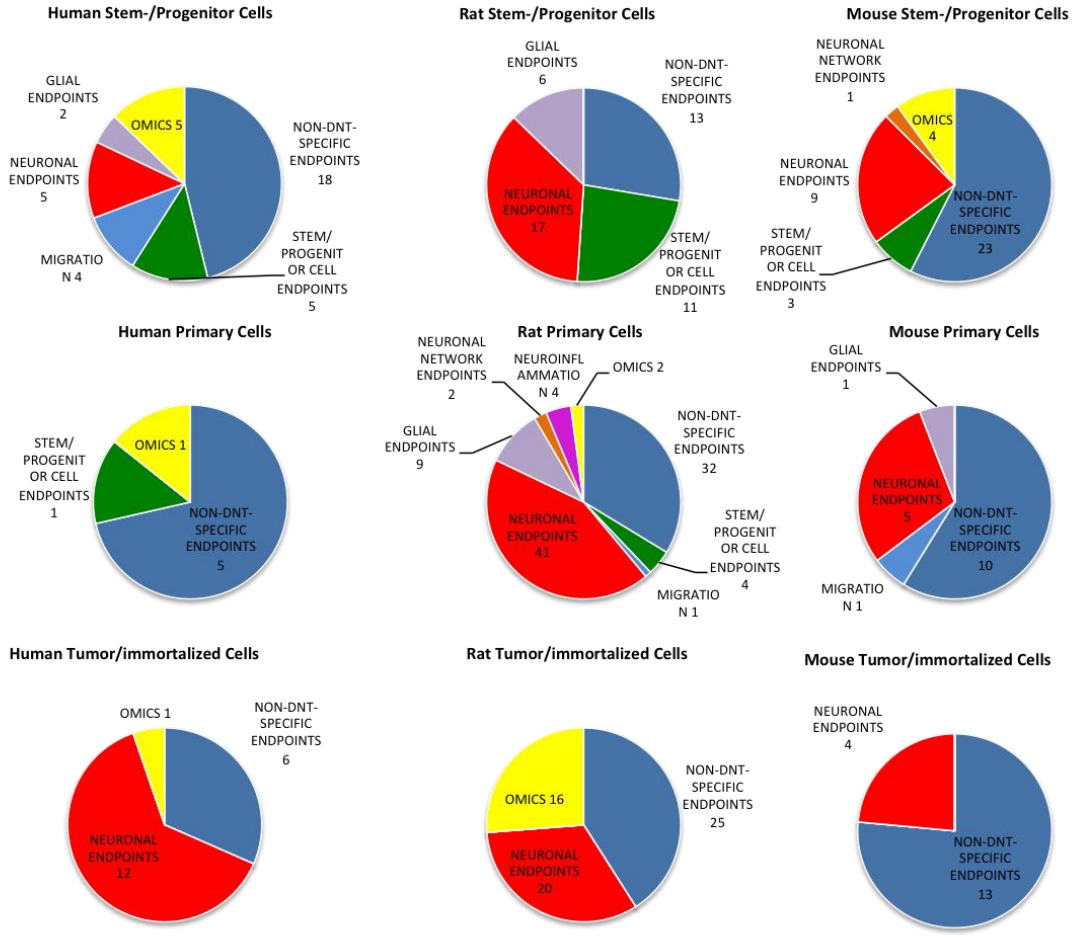


Figure 43: Abilities of cell type categories to assess the 8 endpoint categories (defined in Table 31) correctly. Numbers reflect times of citations with one compound and one specific endpoint measure, not number of publications.

The other 7 endpoint categories are successfully measurable by a number of cell methods. Rat primary cells are the only cell method with which all endpoint categories have been fruitfully assessed. In contrast, for obvious material restriction reasons, endpoint evaluation is most sparse for human primary cells. Surprisingly, besides the 'Non-DNT-specific Endpoints', Tumor/immortalized cell methods were only able to successfully measure neuronal endpoints. If the 'Omics' endpoints measured changes in gene expression in a 'correct' way, i.e. closely to human *in vivo*, has not been studied. Looking at the restricted endpoint measures in these cells though, one would expect only a restricted number of gene expression changes being predictive for human DNT. Comparing positive endpoint evaluations across the stem-/progenitor cell methods of the three species, it is obvious that the most endpoint categories have been positively assessed with human stem-/progenitor cell methods, although the absolute number of endpoint citations is similar. It needs to be pointed out that rat stem-/progenitor cell methods have the most citations (Table 32) but the least individual endpoint category measures, i.e. the endpoint categories 'Omics' and 'Migration' have not successfully been measured in these cells (Figure 43). Although so far there has been no publication measuring rat stem-/progenitor cell migration meeting our inclusion criteria for this systematic review, a migration assay for rat stem/progenitor cells is available (Chen et al., 2003, Baumann et al., 2014, Ishido and Suzuki, 2010).

Evaluation of endpoints with regards to correct ('True Positive', 'True Negative') endpoint evaluation of DNT compounds by the 9 cell type categories

From the data of DNT compounds 'True Positives' and 'True Negatives' across all endpoint categories we next assessed 1) which endpoint categories are data-rich and 2) which endpoint categories were correctly evaluated by which cell type categories. Therefore, we counted the absolute number of cited 'True Positives' and 'True Negatives' and their distribution across the different cell types and species. These absolute numbers were set in relation to circle diameters (number of cited method/15) and the circles drawn out accordingly (Figure 44). This graphical portrayal of the absolute values shows immediately that there are only 2 very data-rich endpoint categories amongst the selected publications of this systematic review: 'Non-DNT specific Endpoints' and 'Neuronal Endpoints'. Because of the secondary importance of the 'Non-DNT specific Endpoints' as stated above, this leaves us with one data-rich endpoint with importance for a potential alternative DNT testing strategy. These 'Neuronal Endpoints' were mainly examined with rat cells, there with primary cells. Because species differences e.g. in the endpoint 'Neurite Outgrowth', which is a frequent endpoint of the category 'Neuronal Endpoints' (Figure 16), are suspected (Harrill et al., 2011; record number 1855), there is the need for testing chemicals on these endpoints in human–based cell systems, preferably co-cultures. The co-culture issue, which has not been addressed within this systematic review before, is due to the fact that isolated neurons without their necessary glial partners might show different sensitivities towards DNT compounds than neurons in presence of astrocytes (Giordano et al., 2008, Giordano et al., 2009; record numbers 3205 and 2964). Targeting a higher throughput of chemicals for DNT testing, the co-culture issue is of importance because analyses of neurites are currently performed via immunocytochemical staining of neurites and 'High-Content-Image-Analyses' (HCA). HCA in mixed cultures with a relatively high cell density is a challenge because the software algorithms available cannot perform these analyses. However, as bioinformatics sciences develops there will be solutions for these challenges in the future.

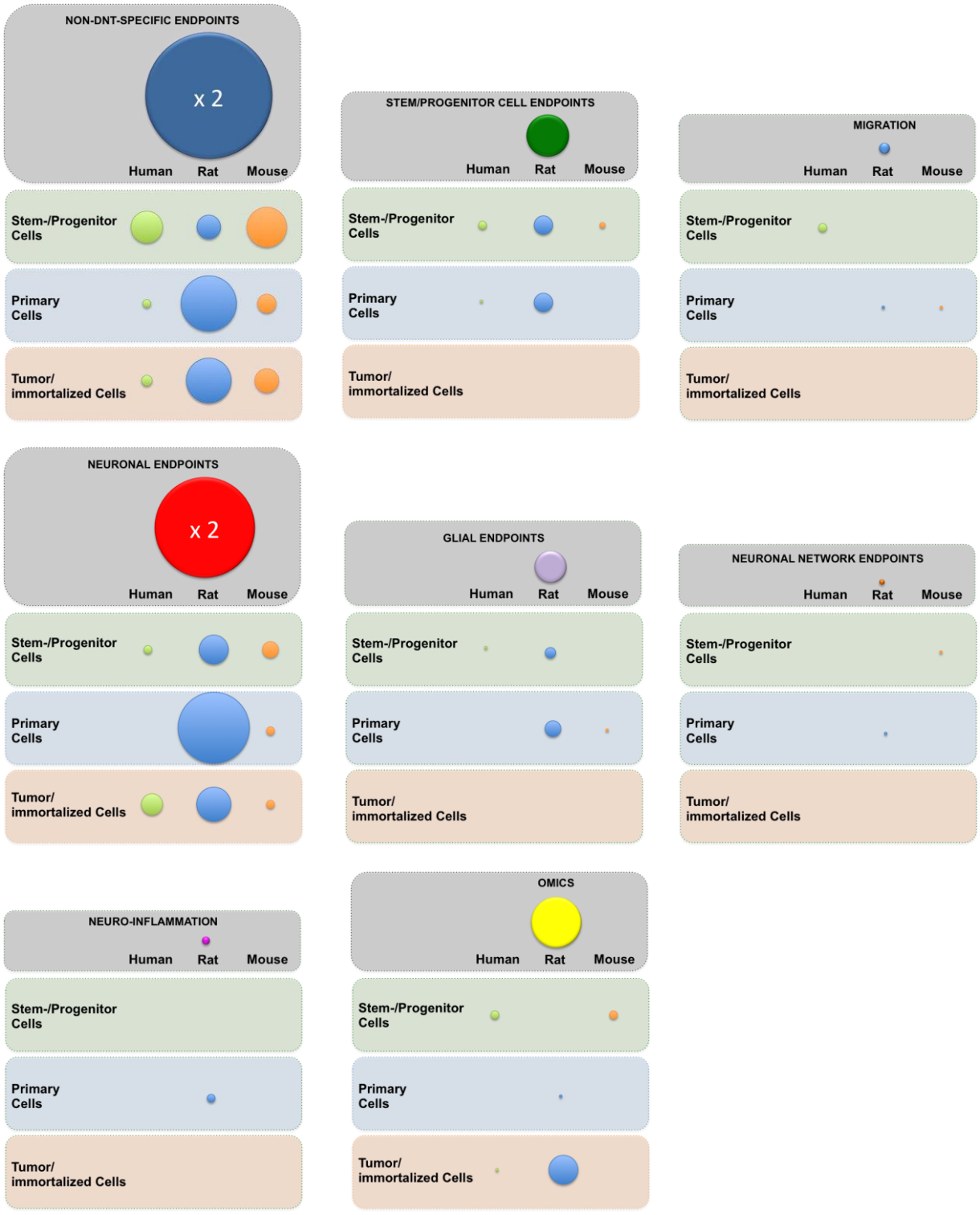


Figure 44: Graphical presentation of absolute numbers of cited cell methods (each cell method evaluated with each compound for each endpoint measure counts as 1). The endpoint colors are the same than in the pie diagrams in Figure 43. Color code for cell type and species are the same than in Figure 9. Each circle represents numbers of citations, relations of circle sizes equal relations of cited methods with the exceptions where the circle is marked with 'x 2', there the circle should be twice as big but is shown only half size due to page size issues. Number of cited methods underlying the circles in the grey field were 'Non-DNT-specific Endpoints' (145), 'Stem-/Progenitor Cell Endpoints' (24), 'Migration' (6), 'Neuronal Endpoints' (124), 'Glial Endpoints' (18), 'Neuronal Network Endpoints' (3), 'Neuroinflammation' (4) and 'Omics' (29).

For 'Stem-/progenitor Cell Endpoints' there is also some data available (n = 24), thus not as many citations as for 'Neuronal Endpoints'. Again, endpoint evaluation is dominated by rat cells, although some human data is also present in the review. The endpoint category 'Stem-/Progenitor Cell Endpoints' is comprised of 'Neural Proliferation' and 'Stem-/Progenitor Cell marker Expression'. While there is only one study available for successful testing of 'Neural Proliferation' in human stem-/progenitor cells (Table 23), the other 4 cited methods evaluate progenitor cell marker expression (Table 25). This indicates an urgent **need** for chemical testing of compounds in the human species, because human and rodent cells have been shown to respond differently towards inhibitors of neural proliferation (Culbreth et al., 2012; record number 1138). Because the ReNcells used in this publication are immortalized, such a study needs to be repeated with primary human neural cells: primary or hESC-derived hNPCs.

Data availability for 'Glial Endpoints' is even thinner than for the formerly discussed endpoint categories (n= 18). And again, main data is produced with rat cells. The only human glia studies producing successful DNT evaluation with glia toxicants within the group of human DNT compounds are generated with hNPCs (record number 2313). However, for this class of compounds (PBDEs) there is no direct *in vivo* evidence for disruption of oligodendrogenesis (see above in chapter 'Oligodendrocyte Endpoints') and thus there is not only more chemical testing, but also more basic research on human oligodendrogenesis and its disturbance **needed**.

For 'Migration', 'Neuronal Network Endpoints' and 'Neuroinflammation', there are only 6, 3 and 4 studies passing the inclusion criteria of this systematic review. Thereby, the 'Migration' endpoint has mainly been studied with human stem-/progenitor cells pointing to a possibly valuable method. Also 'Neuroinflammation' data has been produced with one method type, rat primary cells, which seems to be promising to be transferred to equivalent human systems (generated from stem cell-based methods). The 'Neuronal Network Endpoints' is data-wise the weakest, yet no promising system has been identified within these studies. Yet, as mentioned earlier, a promising human ESC-based model has been established for DNT testing purposes but has not been published with compounds yet – and thus was not recognized within this systematic review (Yla-Outinen et al., 2014, Kapucu et al., 2012). Differentiation of human ESC on multi electrode arrays (MEAs) was established within this work and neuronal network activities recorded. Because the functional readout of all these endpoints discussed before integrate into proper neuronal circuits, this is a crucial endpoint which certainly **needs** more establishment work as well as chemical testing.

Summary *in vitro* part

In summary, within this systematic review we identified 1.358 individual citations for individual cell methods measuring individual endpoints with individual compounds. This data was collected out of 288 DNT *in vitro* publications. The 1.358 citations evaluated 148 individual endpoints, which were grouped into 25 different endpoint groups. These endpoint groups are comprised of 5 groups dealing with non-DNT-specific endpoint groups, 19 DNT-specific endpoint groups and the endpoint group 'Omics', which is evaluated separately due to the lack in current know-how on how to use 'omics'-studies for regulatory purposes. Each of the endpoint groups was analyzed for cell methods correctly identifying DNT positive and negative compounds. Out of this evaluation, we finally extracted 342 individual cell method citations across 23 endpoint groups where i) viability data was available in addition to the specific endpoint evaluation and ii) where DNT positive and negative compounds were evaluated correctly. When evaluating these 342 endpoint measures across the 9 different cell type categories we found these endpoints evaluated with 44 cell systems, which belong to 9 different cell type categories: 'Stem-/progenitor Cells', 'Primary Cells' and Tumor/immortalized Cells' of 3 different species: human, mouse and rat. From these, rat primary cells were the cell type category with the most citations and the highest amount of endpoint categories that it can evaluate. Human and mouse stem-/progenitor cells were also

able to evaluate a large number of endpoints. In contrast, tumor cells are rather limited in their ability to assess different DNT endpoints. While there is quite a number of data produced in primary rat cells (>90), the amount of studies assessing DNT compounds correctly in stem-/progenitor cells is approximately a third of this in each species.

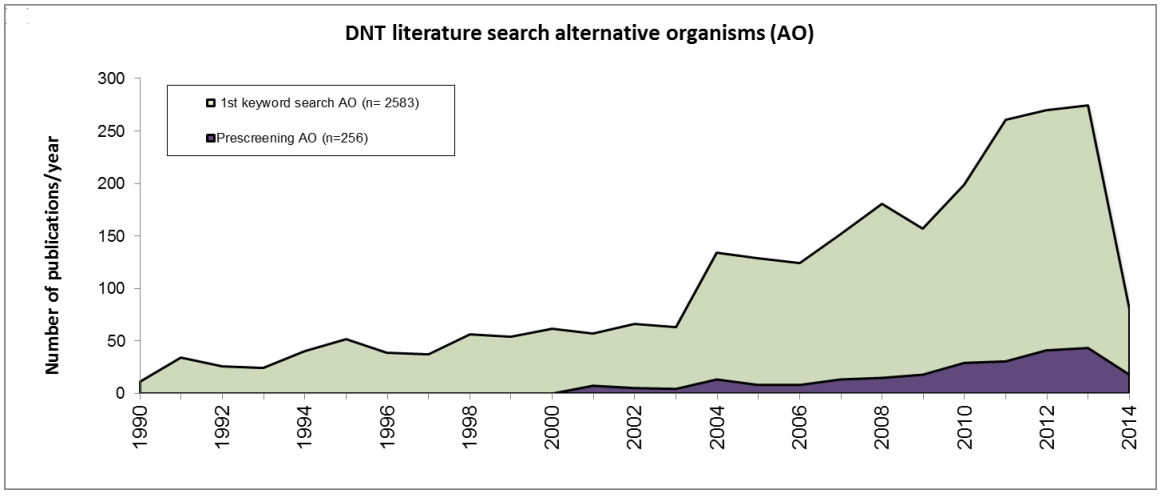
Evaluating not the performance of cell methods, but endpoint categories for identifying data gaps shows that data-rich endpoint categories are either 'Non-DNT-specific Endpoints' or 'Neuronal Endpoints'. There, again, rat primary cells are the dominant method. While some methods are published for studying 'Stem-/progenitor Cell Endpoints' and 'Glial Endpoints', this data is over all insufficient. 'Migration', 'Neuronal Networks' and 'Neuroinflammation' are very data-poor, although promising methods exist, which are currently within the status of scientific assay establishment.

3.2. Alternative Organisms

Time trend analysis of publications selected showed that, although the searches included publications from 1990 until May 2014, in the first key word selection the number each year remained below 50, but steadily increased from 2003 onwards Figure 45a.

For publications dedicated to the alternative organisms selected (zebrafish, nematode and sea urchin) it appears that there were even no DNT relevant publications before 2000 (Figure 45b). While the number of publications is of same order of magnitude from 2000 to 2010 (≤ 5/year), a significant increased number of publication with a factor 3 to 4 is apparent in past 4 years. The latter is in line with other observations in biomedical research that the zebrafish model for scientific research (human diseases, pharmaceutical drug screening,..) is gaining lot of interest (Kinth et al., 2013).

a



b

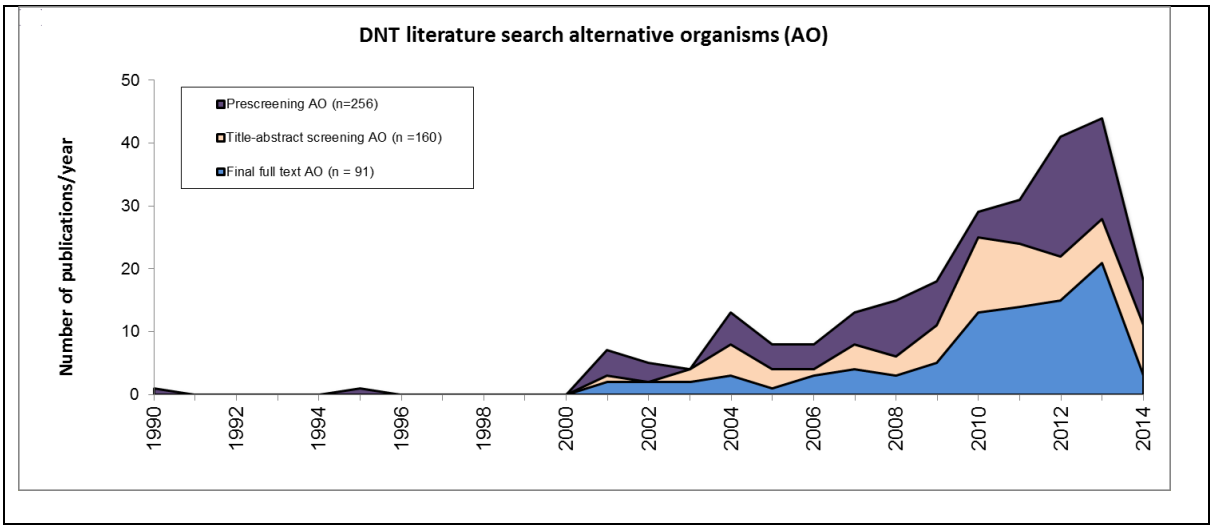


Figure 45 The number of publications in the stepwise DNT screening for alternative organisms represented in 2 graphs, respectively a) 1st key word search giving 2583 publications and pre-screening on 256 publications and b) with further selection of panel of pre-screening by title abstract screening (n=160) and final text selection (n=91).

All available 'full text' publications containing studies with alternative organisms for assessment of DNT (n = 90, 1 paper of 91 selected was not obtained) were grouped according to (I) species (zebrafish, Sea Urchin, Nematode), (II) life stage used (embryo, larvae, embryo & larvae, or transgenic embryo/larvae) and (III) endpoints as listed in the data collection sheets (Appendix K).

According to **species**, we counted 83, 5 and 2 publications employing zebrafish, sea urchin and nematode, respectively. Life stages considered were those related to effect measurement at either embryonic stage (egg), larvae stage (hatched) or both embryo & larvae stages in normal organisms, or compared to transgenic fish used at embryo or larvae stage. Of these, 42, 17, 71 and 12 studies counted across species belonged to the category of embryo, larvae, embryo & larvae or transgenic embryo/larvae life stage (Figure 46). The total number of studies using either one or more of these life stages (zebrafish - 133, sea urchin – 5 & nematode - 4) is however larger than the total number of publications identified (90) because here the life stage is the evaluated unit. Publications containing multiple life stages of organisms are counted numerous times resulting into 142 studies.

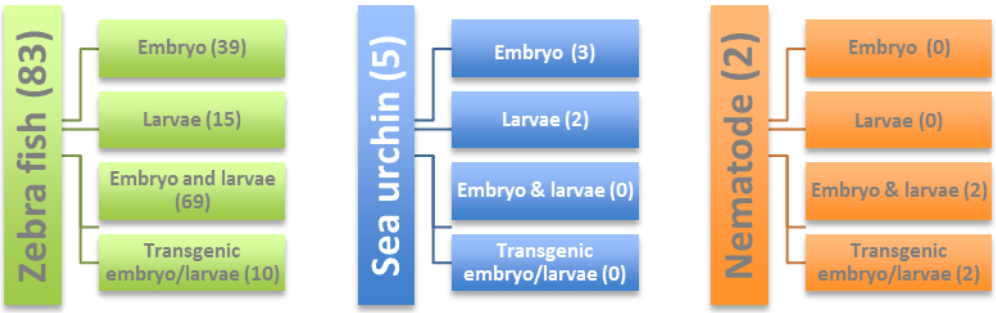
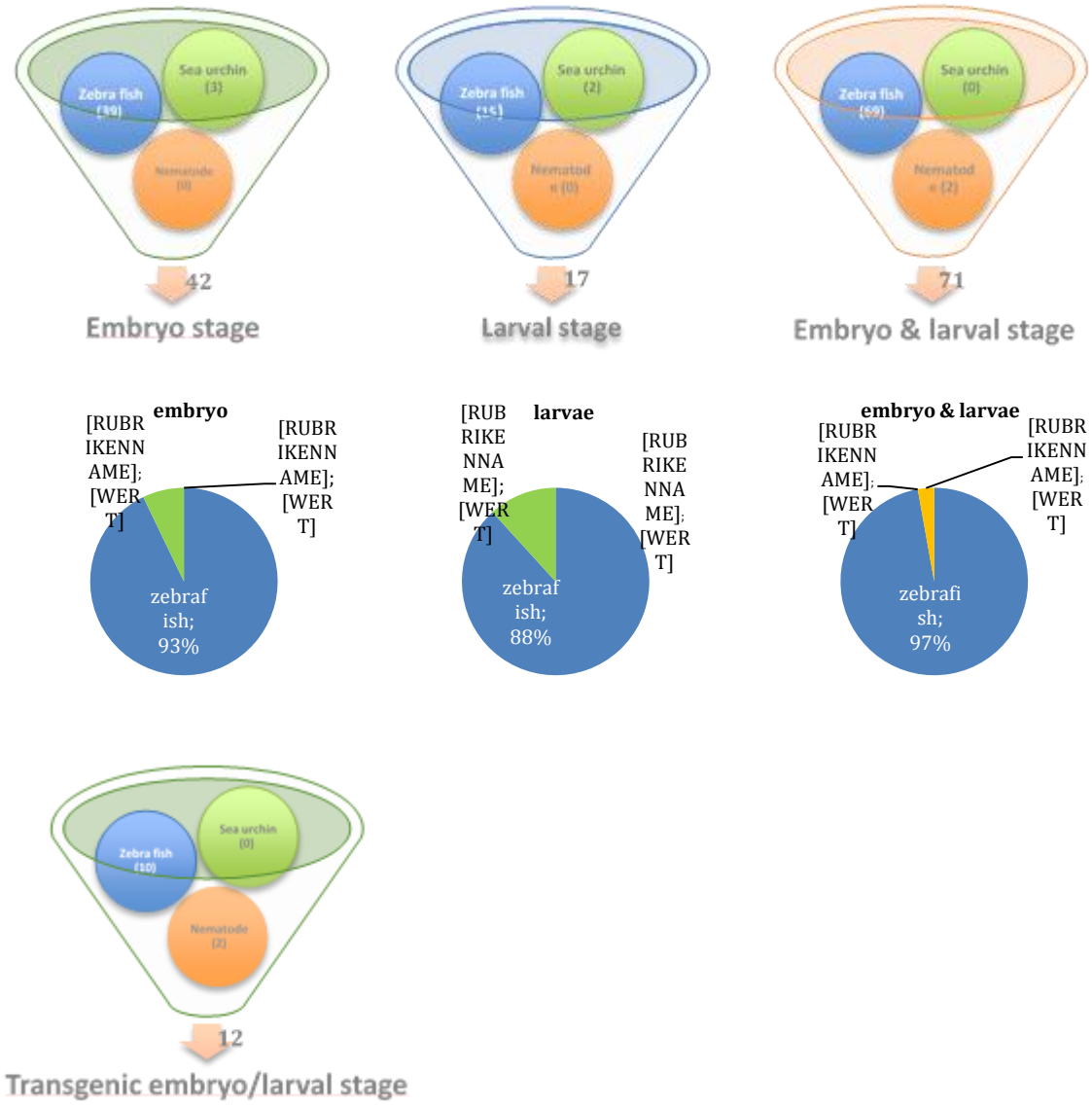


Figure 46 Within 90 DNT publications using alternative organisms (zebrafish, sea urchin and nematode), different life stages for effect evaluation comprising embryo, larvae, embryo and larvae or transgenic lines with embryo/larvae (= 12 life stage categories) were published resulting into 142 studies.

Of all alternative organisms for DNT studies, the zebrafish is used most frequently with 83 selected publications, including 133 studies on different life stages, representing respectively 92.2 % of publications or 93.7 % of studies of all species and life stages. Only a few publications (# 2, or 2.2%) on nematode were relevant for DNT, while 5 publications using sea urchin (5.6 %) were selected for

evaluation. Of all studies performed with nematode (# 4), 2 studies used embryo & larval stages while 2 other studies used a transgenic model of the nematode. Within the group of sea urchin publications (# 5), the life stages used were respectively the embryo stage (# 3) or the larval stage (# 2). Within the group of zebrafish studies, all 4 different life stages were applied giving 39 (29.3%), 15 (11.3%), 69 (51.9%), 10 (7.5%) studies for respectively embryo, larvae, embryo & larvae or transgenic embryo/larvae model (Figure 46).

According to the different **life stages** across species, most data for DNT effect measurements was generated in studies with combined embryo & larval stages (# 69), followed by the embryo stage (# 42), the larval stage (# 17) and the transgenic embryo/larval stage (# 10), as shown in Figure 47:.



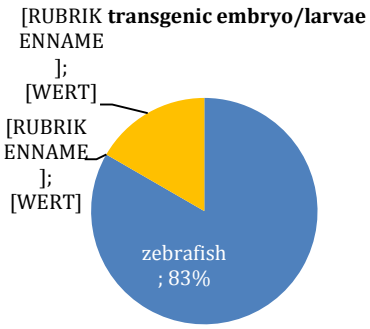


Figure 47: Number of DNT studies with alternative organisms (# 142) employing embryo, larval, embryo & larval or transgenic embryo/larval stage from zebrafish, sea urchin or nematode. Given are the distributions across species within each category of life stage. The lower pie plots show the same data in % for a specific species of all the studies from the respective life stage category.

Of the 42 studies performed with embryo stage, 39 (92.9%), 3 (7.1%) and 0 (0%) were derived from zebrafish, sea urchin and nematode, respectively. Within the group of 17 studies with only larval stages, 15 (88.2%), 2 (11.8%) and 0 (0%) originated from zebrafish, sea urchin and nematode, respectively. Within the group of 71 studies with combined embryo and larval stages, 69 (97.2%), 0 (0%) and 2 (2.9%) were derived from zebrafish, sea urchin and nematode, respectively. And within the group of studies with transgenic embryo/larval models (# 12), the species divide into 10 (83.3%), 0 (0%) and 2 (16.7%) for zebrafish, sea urchin and nematode, respectively (Figure 47:). The pie graphs generated by this data show that DNT studies performed with transgenic models mainly utilize zebrafish and few nematode (83.3% and 16.7%, respectively). Nematode studies with a combination of normal embryo & larval stages (2.8%) are a minority compared to zebrafish studies (97.2%), while no studies were selected with both embryo and larval stages of sea urchin (0%). Few studies with sea urchin at either embryo stage (7.1%) or larval stage (11.8%) were selected, which is low compared to zebrafish studies with either the embryo stage (92.9%) or the larval stage (88.2%). Overall, the majority of studies originated from zebrafish studies with combined embryo & larval stages (69 of 142 DNT studies in total, or 48.6%).

Data analysis in former sections did consider always for each of the life stages, both chorionated and dechorionated life stages. However this dechorionation step, as included in the column on life stage is an experimental variable which can have an impact on the outcome of DNT testing. This dechorionation step is regularly applied for life stages of zebrafish (mostly embryo), while not applied in studies with nematode and sea urchin. Several reasons for application of this step are put forward (1) improvement of the bioavailability of the test compound, (2) access to the embryo for stimulation upon evaluation of touch response for endpoint motor activity or (3) enhance performance of labeling methods by improving permeability for antibodies or ISH probes. Dechorionation is applied in 24 of in total 133 zebrafish studies (18%), more specific in 17 of 39 studies with embryos (43.6%), 2 of 15 studies with larvae (13.3%), 5 of 69 studies with combined embryo & larvae stage (7.2%) while not in studies with transgenic zebrafish. This gives 70.8%, 8.3%, 20.8% and 0% studies with dechorionated stages within each of the life stage categories respectively embryo, larvae, embryo and larvae, or transgenic model (Figure 48). Due to the low overall number of citations for life stage groups within endpoints, it is decided not to diversify for dechorionated versus non-dechorionated. However, it remains an important issue to be considered upon standardization of methods.

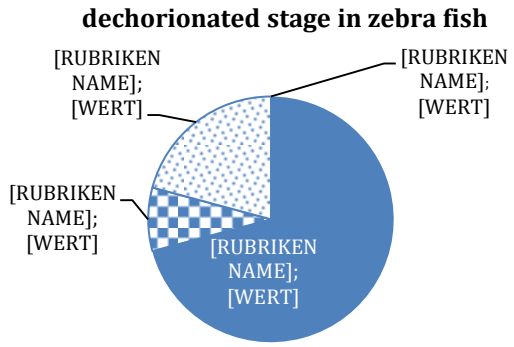


Figure 48 Percentage of DNT studies with dechorionated life stages on the total number of studies within each of the respective life stage category for zebrafish.

For these published DNT studies with 3 different species, a variety of **endpoints** were analyzed. In total, we assessed 40 different endpoints in different organisms, and different life stages. Of these, respectively 13 and 24 endpoints were grouped into 'non-DNT-specific' and 'DNT-specific' endpoints, respectively. These individual endpoints were further sorted into 'endpoint groups' as depicted in Table 33. The 'non-DNT-specific' endpoints contain 2 endpoint groups: death/development organism and general cellular function, covering respectively 6 and 7 endpoints, as specified in Table 34. These are not specific for the developing nervous system, but represent common mechanisms for toxicant-related effects on developing organisms.

In the 'DNT-specific' endpoints, 24 endpoints are listed which are exclusive for the nervous system. These include a large variety of different neuronal and neurotransmitter related endpoints or functions, such as motor behavior. These endpoints are organized in 10 endpoint groups (Table 33 & Table 35). One endpoint group, containing 3 endpoints, that does not belong to either of the categories is 'omics'. This is due to the current difficulty to handle 'omics' studies in a regulatory context. While such 'big data' is certainly useful for potential DNT biomarker identification, more research is clearly needed to confirm the validity of such biomarkers, and their specificity for DNT pathways. The complete table with all 40 endpoints, published across species and life stages, and classified into endpoint groups for respectively non-DNT specific (including omics) and specific DNT endpoints is given in Table 34 and Table 35.

Table 33 Endpoint groups for non-DNT-specific and DNT-specific endpoints. Number of specific endpoints within endpoint groups is given in brackets.

Non-DNT-specific endpoint groups	DNT-specific endpoint groups
Death/development organism (6)	Cholinergic endpoint (4)
General cellular function (7)	Catecholaminergic endpoint (1)
	CNS development (2)
	CNS gene expression (1)
	CNS protein expression (1)
	Motor activity (2)
	Neuronal endpoint (5)

	Neuronal functioning (4)
	Synaps endpoint (2)
	Serotonergic endpoint (2)
Omics (3)	

Table 34 Non-DNT specific endpoints within endpoint groups. For simplicity, omics endpoints are included in this table (similar to following figures).

Non-DNT-specific endpoints	
Endpoint groups	**Endpoints**
Death/development organism (6)	Acute toxicity
	Embryo development
	Eye development
	Growth and metamorphosis
	Muscle fibres
	Viability
General cellular function (7)	Apoptosis
	Calcium (intracellular)
	DNA breakage
	Enzyme activity esterases
	Oxidative stress
	Proliferation
	Protein expression/phosphorylation
Omics (3)	Gene expression whole organism
	miRNA expression
	Proteomics

Table 35 DNT specific endpoints within endpoint groups.

DNT-specific endpoints	
Endpoint groups	**Endpoints**
Cholinergic endpoint (4)	Acetylcholine concentration
	AChE activity
	Localisation of Ach receptors (AChRs)
	Localisation of AChE activity

Catecholaminergic endpoint (1)	Catecholaminergic neurons
CNS development (2)	Brain development
	Notochord
CNS gene expression (1)	Gene expression specific organ/tissue
CNS protein expression (1)	Protein distribution
Motor activity (2)	Behaviour
	Motor activity
Neuronal endpoint (5)	Localisation of primary & secondary motorneurons
	Localisation of sensory neurons
	Localisation of spinal motorneurons
	Neurodegeneration
	Neurogenisis
Neuronal functioning (4)	Motor axons and neuromuscular junctions
	Primary / secondary motorneurons and axon growth
	Sensory neurons & axon growth
	Axon growth
Synaps endpoint (2)	Synaptic properties
	Synaptogenesis
Serotonergic endpoint (2)	Localisation of serotonergic neurons
	Expression serotonin receptor

This grouping of endpoints is necessary for the evaluation of DNT methods, because there are too few individual endpoints measured by different organism models to evaluate them individually. This is illustrated in Figure 49 where the number of studies for each endpoint and each species (cumulating all life stages studied) is given. In Figure 50 the relative contribution of each of the endpoints in studies with across all alternative organisms (3 selected species and all life stages) is presented.

For about half of the endpoints (n = 19, 47.5%) we had only one study across all species and life stages. Within the group of zebrafish studies, there were 17 endpoints of total 40 endpoints with no or one hit (42.5%). For the other species there were 38 or 39 endpoints with no or one occurrence for respectively sea urchin and nematode. Next to this low hit rate of a majority of endpoints, the overall number of endpoints and studies with sea urchin and nematode, respectively 10 different endpoints in 13 studies and 7 different endpoints in 8 studies, are too low to further elaborate for quantitative analysis on test performance for DNT (Figure 49). Even when endpoints were grouped, as shown in Figure 51, the number of studies within DNT-specific or non-DNT specific for either sea urchin or nematode remained low (< 5). The highest hit for these 2 species was 4 studies for the non-DNT specific endpoint group

'death/development organism' for sea urchin. Figure 53 and Figure 54 reflect the relative contribution of each of the endpoints with at least 1 hit in studies of respectively sea urchin or nematode (all life stages considered). Among 13 studies with sea urchin (Figure 53), the DNT-specific endpoint groups in order of high (42.9%) to low (14.3%) contribution are cholinergic endpoints, neuronal endpoints, serotonergic endpoints and neuronal functioning, while the non-DNT specific endpoint groups were death/development organism (66.7%) and general cellular function (33.3%). For nematode studies (n=8), similar distribution of 3 endpoints groups, either within DNT-specific or non-DNT specific was seen (Figure 54). This includes the neuronal endpoint group as most important (50%) within the DNT specific category, followed by 25% of both motor activity and serotonergic endpoint groups. For the non-DNT-specific endpoint groups, the highest contribution was endpoints related to death/development of organism (50%), while omics and general cellular function endpoint groups were each 25%.

Only 9 endpoints were published 5 times or more in studies across all species (265 studies, Figure 50). It concerns following endpoints: AChE activity (n =8, 3%), apoptosis (n = 17, 6.4%), brain development (n=8, 3%), embryo development (n=68, 25.7%), gene expression in specific organ/tissue (n=15, 5.7%), gene expression in whole organism (n=24, 9.1%), motor activity (n=61, 23%), muscle fibers (n=5, 1.9%) and neurogenesis (3.4%). These numbers can mainly be attributed to the zebrafish studies, as the highest hit rate for an endpoint among sea urchin or nematode studies was 3 for the endpoint 'embryo development' in sea urchin. The highest frequency of endpoints in zebrafish studies (Figure 49) is seen for motor activity (n= 60), a DNT-specific endpoint and for embryo development (n=63), a non-DNT specific endpoint. Both endpoints show also the highest contribution to all the endpoints within studies across all species (Figure 50), resulting into 23% (n=61) for motor activity and 25.7% (n= 68) for embryo development. Next important endpoints with a relative contribution higher than 5% (for zebrafish studies, and across all species) are gene expression in whole organism (9.1%, n= 24 for all species) and apoptosis (6.4%, n= 17), both non-DNT specific endpoints followed by gene expression in specific organ/ tissue (5.7%, n= 15) as illustrated in Figure 50.

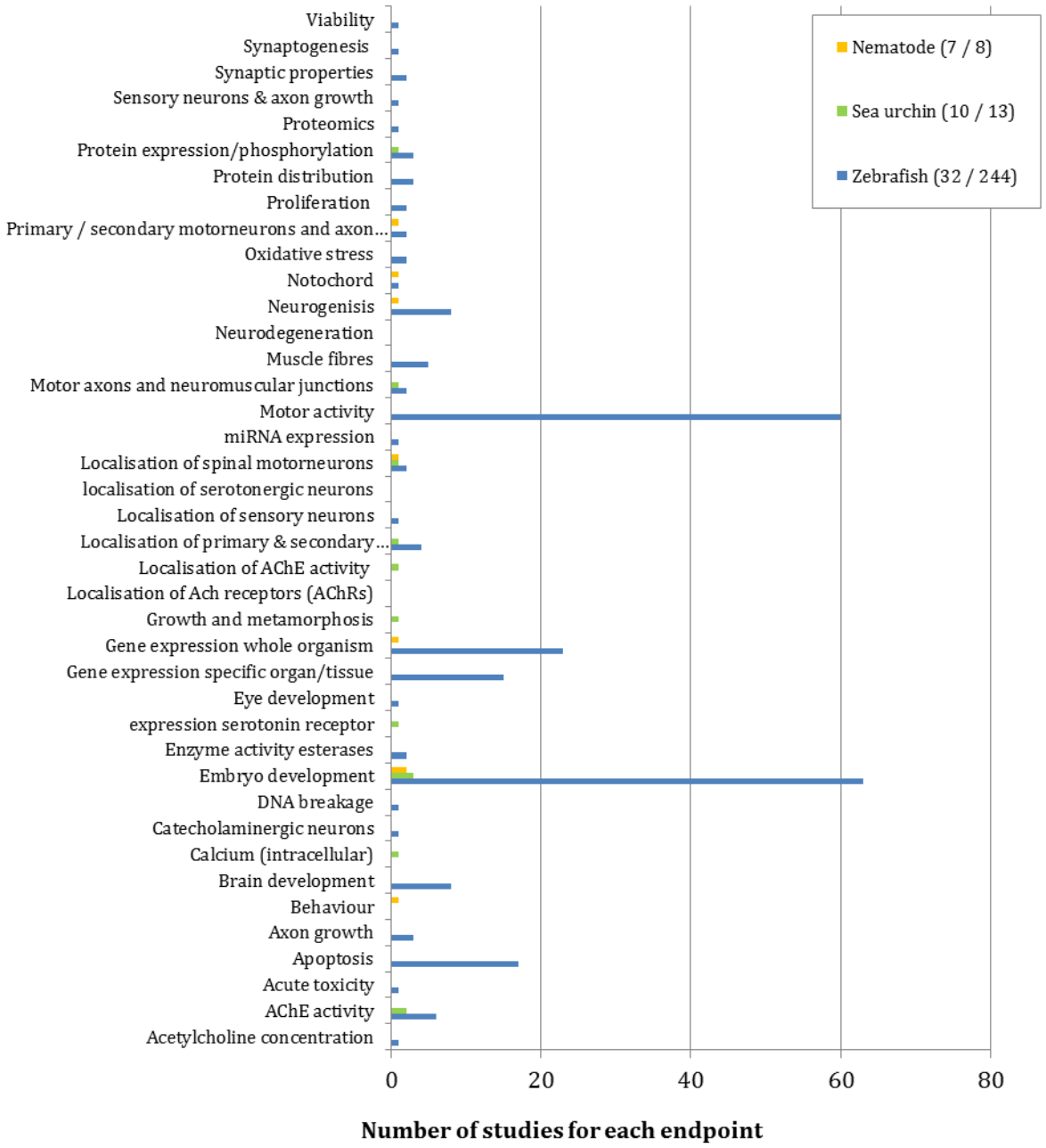


Figure 49 Individual endpoints (n = 40) published in the DNT literature using alternative whole organism approaches. For zebrafish 32 endpoints are used in 244 studies, while numbers are very low for sea urchin and nematode with respectively 10 endpoints in 13 studies and 7 endpoints in 8 studies (all life stages considered).

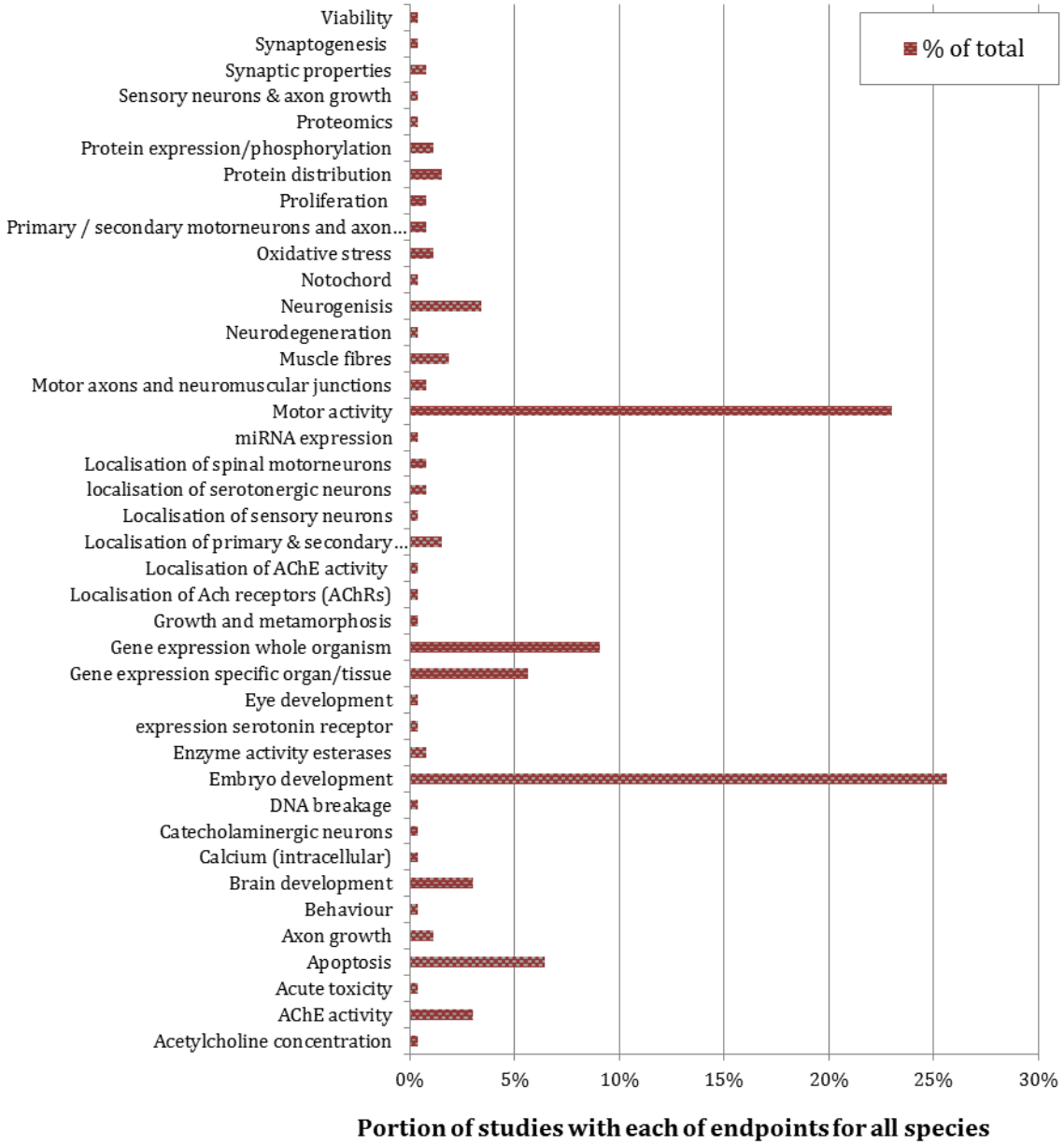


Figure 50 Relative contribution of each of single endpoints published in the DNT literature using alternative whole organism approaches (n = 265 studies for all species and life stages).

This similar pattern of occurrence of major endpoints, mainly determined by zebrafish studies (blue bar) is confirmed by organizing these into endpoint groups, as is shown in Figure 51. Endpoint groups death/development organism and motor activity for respectively non-DNT and DNT specific endpoints show highest number of studies. Next important endpoint groups are general cellular function and omics representing each more than 20 studies, followed by CNS gene expression (n= 15) and neuronal endpoints (n= 15), both DNT specific endpoint groups.

Finally, as studies with zebrafish as an alternative organism represent the majority of studies and endpoints within this review across 3 species, the most important endpoint groups for zebrafish should be considered for further data evaluation on method specificity and sensitivity. Figure 52 demonstrates that

for the DNT specific endpoint groups, motor activity covers the majority of studies (49.6%, n=60). The latter is in line with the added value of a whole organism approach, providing information on motor activity as an integrated effect measurement and apical endpoint, which cannot be derived from alternative approaches using cellular or tissue culture models. Other DNT-specific endpoints groups representing a significant part of studies (> 5%) are CNS gene expression (12.4%, n=15), neuronal endpoint (12.4%, n=15), CNS development (7.4%, n= 9), neuronal functioning (6.6%, n=8) and cholinergic endpoints (5.8%, n=7). The non-DNT specific endpoints (including omics), organized in 3 groups respectively death/development organism, general cellular function and omics contribute for 57.7% (n=71), 22.0% (n= 27) and 20.3 % (n= 25), as shown in lower part of Figure 52.

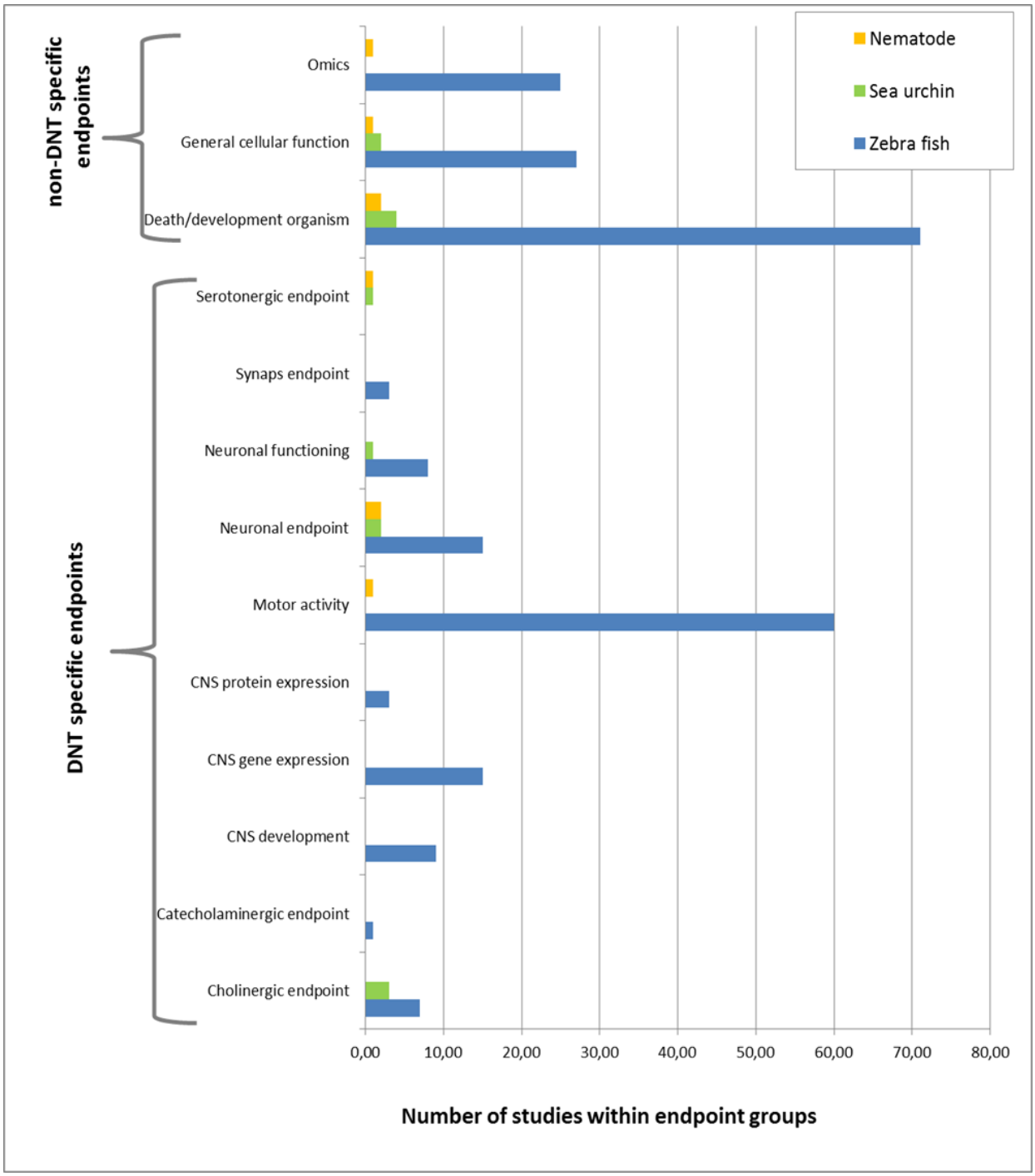
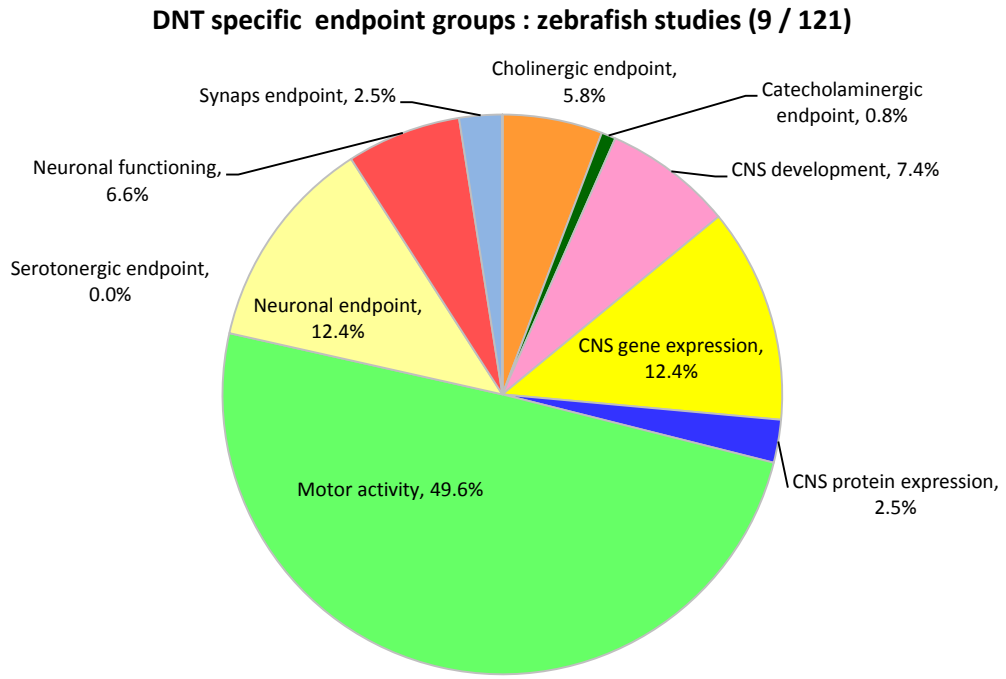


Figure 51 Number of studies within groups of endpoints for respectively DNT specific or non-DNT specific endpoints for respectively zebrafish, sea urchin and nematode (all life stages considered).



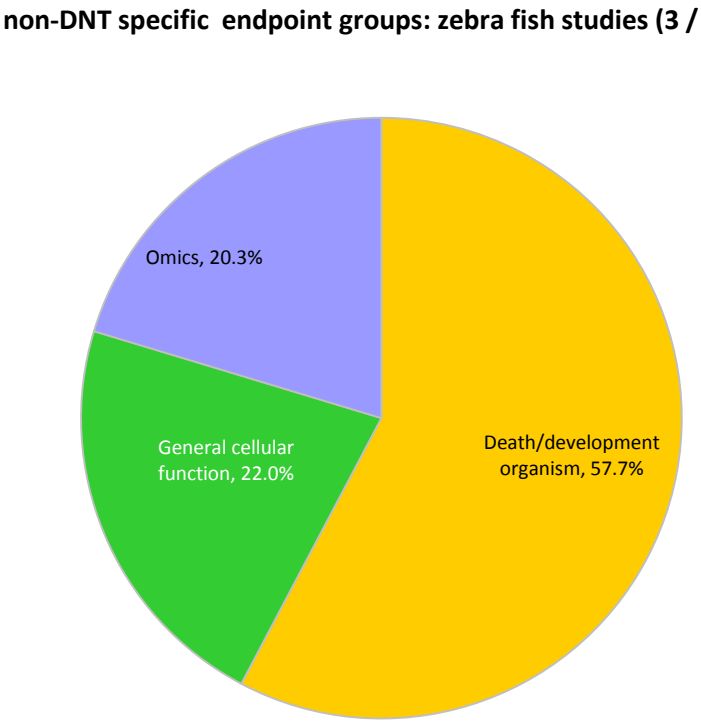


Figure 52 Relative distribution of number of studies within groups of endpoints for respectively DNT specific (9 groups*/121 studies) and non-DNT specific endpoints (3 groups*/123 studies) for zebrafish for all life stages. (* number of groups is given for those with at least 1 study).

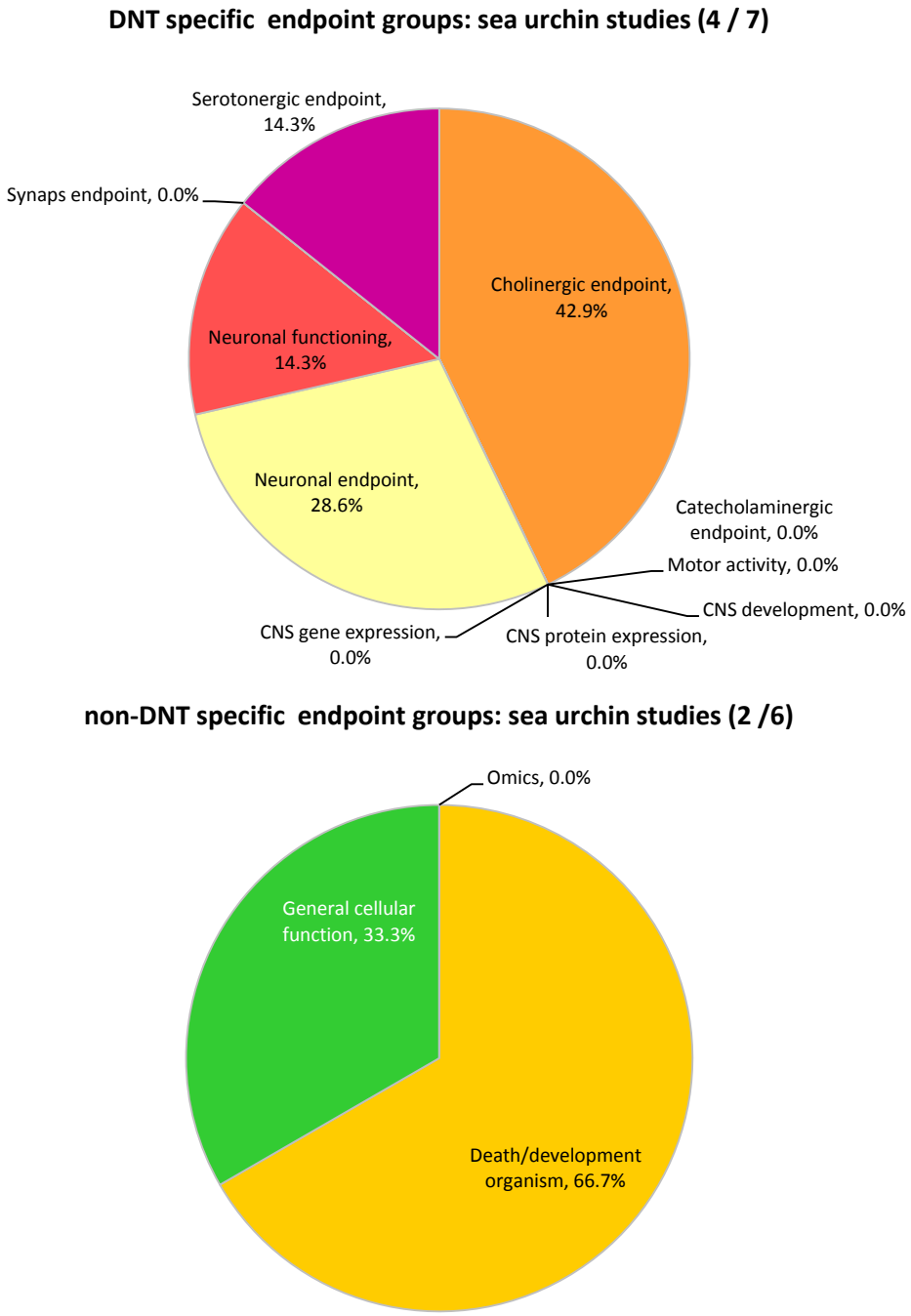
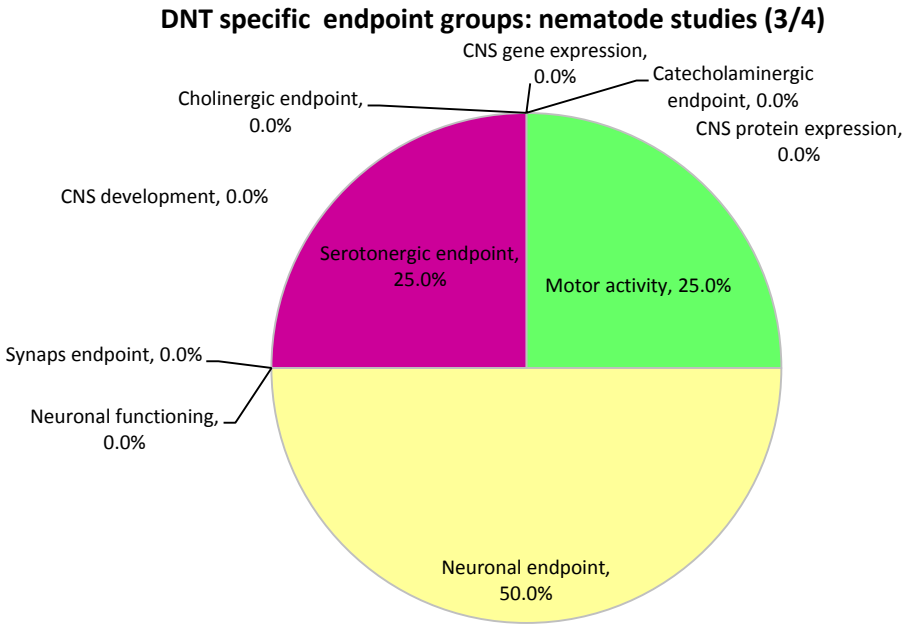


Figure 53 Relative distribution of number of studies within groups of endpoints for respectively DNT specific (4 groups*/7 studies) and non-DNT specific endpoints (2 groups*/6 studies) for sea urchin fish for all life stages. (* number of groups is given for those with at least 1 study).



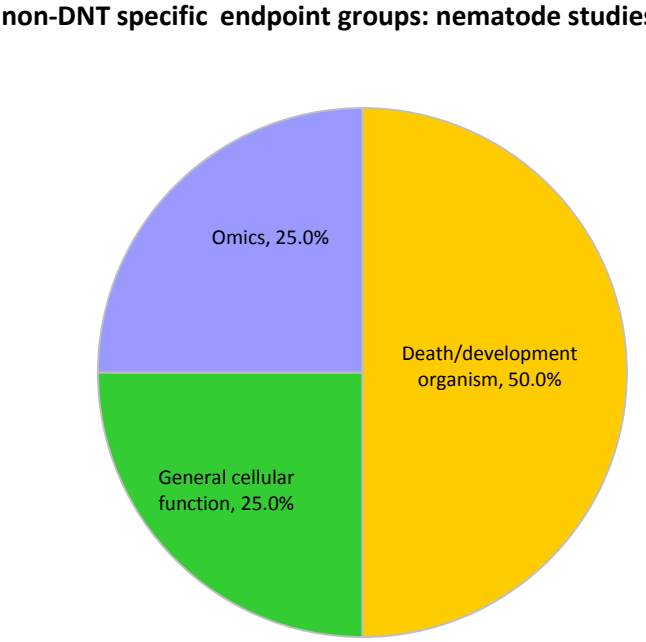


Figure 54 Relative distribution of number of studies within groups of endpoints for respectively DNT specific (3 groups*/4 studies) and non-DNT specific endpoints (3 groups*/4 studies) for nematode for all life stages. (* number of groups is given for those with at least 1 study).

In further steps of data analysis where it concerns sensitivity and specificity, the sea urchin and nematode data are not considered due to the lack or low number of hits for each of the endpoint groups (see previous section). Moreover the number of different compounds used for each of the species and endpoints is too low for a reliable analysis. Data on nematode studies with only 1 positive compound, methylmercury chloride in 2 publications are available in the data collection sheet 'AO'. For the 5 publications of sea urchin DNT assessment, only 1 positive compound chlorpyrifos is used (record nr.14528), 4 different pesticides categorized as possible positive (diazinon in Record nr. 3442 & 5514, carbaryl in record nr. 5514, primicarb in record nr. 5514 and monochrotophos in record nr. 14053) and several types of nanoparticles (non-classified, in record nr. 709).

For further evaluation the total collected data of the systematic review search of all zebrafish endpoint groups (no hit for the serotonergic endpoint group) containing data from the 4 different life stage/model was used.

One of the criteria for a reliable DNT method is to distinguish general cytotoxicity/death from specific endpoint alterations (Crofton et al., 2011). We thus compare the selected endpoint data of zebrafish from the literature search to the effects on viability for the same compounds at same concentration ranges within the same study. In many zebrafish studies, viability or assessment of survival of early life stages is often combined with evaluation of teratogenic or malformation effects, as in many cases these can cause death at a later time point. Thus the endpoint 'embryo development' as listed in the data collection sheet 'alternative organisms' includes assessment of survival, unless otherwise indicated in the column 'effect on viability' in the sheet. These analyses will enable the identification of zebrafish methods performed for each of the life stages, which are able to predict specific DNT effects distinguishable from overt acute toxicity (death) of the developing organisms.

The data were thus analyzed for effect of chemicals on viability (survival of organism) versus the (non)-DNT-specific endpoints. Thereby, the categories a) 'Specific endpoint is more sensitive towards compound exposure compared to the effect on cell viability' (green), b) 'Specific endpoint is equally or less sensitive towards compound exposure compared to the effect on cell viability' (red), c) 'Viability was not assessed' (purple), d) 'Not applicable' (grey), e) 'not clear' (orange) or f) 'Effect on endpoint and viability are compound-dependent' (turquoise; only studies with many compounds (leading to more than 20 citations) were distinguished).

Taking all measured endpoint groups and tested compounds with data for viability into consideration (n = 376 citations), 72.1%, 8.83% and 8.5% of all endpoints were more, less/equally sensitive than viability or viability was not assessed (Figure 55).

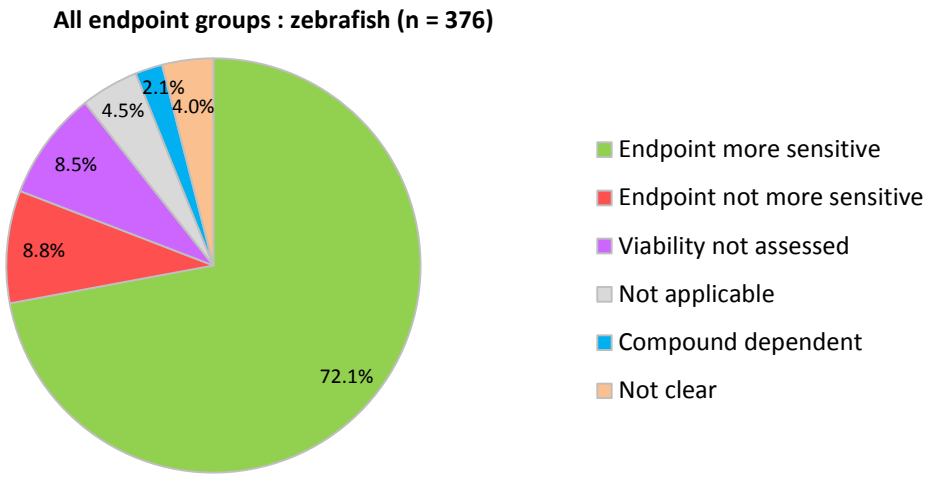


Figure 55 Relative distribution of all endpoint groups in zebrafish for their change compared to viability assessment. Thereby, a specific change is considered as an endpoint change occurring without or at levels below viability change (green). When mortality of organisms is measured at the same or lower concentrations than for that endpoint, the change is considered non-specific (red). Total number of citations (n) is given across life stages and compounds.

When we split these data into the group of non-DNT specific versus DNT specific endpoint groups a similar pattern is shown. We observed that 77.1%, 9.6% and 6.6% of all non-DNT specific endpoints were more, less/equally sensitive to viability or viability was not assessed. For DNT specific methods, 68.1% and 8.1% were respectively more and less/equal sensitive than viability, while for 10% no viability data was available (Figure 56).

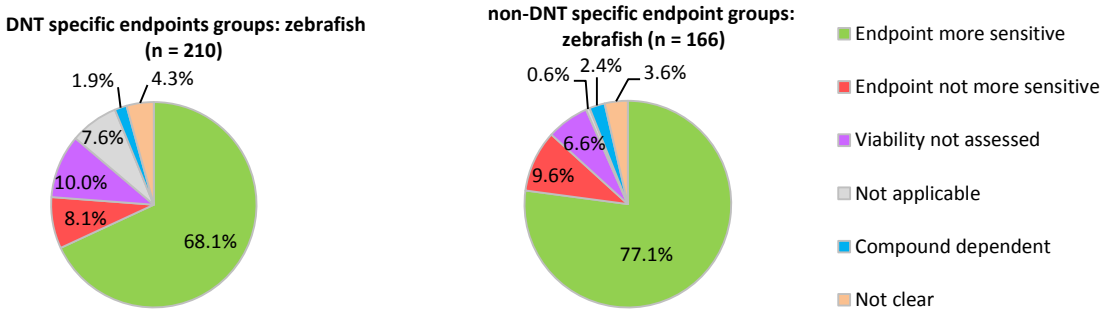


Figure 56 Comparison of DNT-specific versus non-DNT specific endpoint groups and relative distribution of their changes compared to viability assessment. Thereby, a specific change is considered as an endpoint change occurring without or at levels below viability change (green). When mortality of organisms is measured at the same or lower concentrations than for that endpoint, the change is considered non-specific (red). Total number of citations (n) is given across life stages and compounds used in zebrafish studies.

Splitting this information into the pre-defined endpoint groups for respectively non-DNT specific (Figure 57) and DNT specific endpoint groups (Figure 58) reveals that endpoint sensitivity compared to viability can significantly change among endpoint groups, while for some cases there is a lack of viability data.

There are endpoint groups where for > 10% up to 37.5% of citations corresponding viability data were not assessed. It concern omics (15.4%) of the non-DNT endpoint groups versus CNS development (14.3%), CNS gene expression (37.5%) and neuronal endpoint (18.8%) for the DNT specific group. Mainly gene expression studies, related to e.g. biomarker analysis lack viability data. It should be

warranted that such measurements need to be accompanied by survival analysis, in order to distinguish gene expression changes due to acute toxicity and general stress to toxicants from DNT specific responses.

With respect to sensitivity of endpoint groups, the catecholaminergic endpoint and the neuronal functioning scored rather high for less sensitive compared to viability with respectively 42.9% (3 citations of total n=7) and 35.7% (5 citations of total n= 14). As both are DNT-specific endpoint groups, these are less suitable for the evaluation of developmental neurotoxicity. Further studies for these relevant DNT endpoints should include more consistent an analysis of the effect of compounds on survival of the organism for accurate interpretation of DNT specific endpoints.

For the non-DNT specific endpoints (Figure 57), a score of 66.7% (omics, n=26), 80.2 % (general cellular function, n=36) and 80.6% (death/development organism, n=91) with more sensitivity than viability was obtained for the endpoint groups. For the DNT specific endpoints, except for the synapse endpoint (only few publications), the range of endpoints for more sensitive than viability goes between 50% for CNS gene expression and up to 100% for CNS protein expression. Several DNT specific endpoint groups score more sensitive than viability for more than 70% of the citations: cholinergic endpoint (90%, 9 citations of total n=10), CNS development (78.6%, 11 citations of total n=14), CNS protein expression (100%, 3 studies of total n=3) and motor activity (70.9%, 90 citations of total n=127).

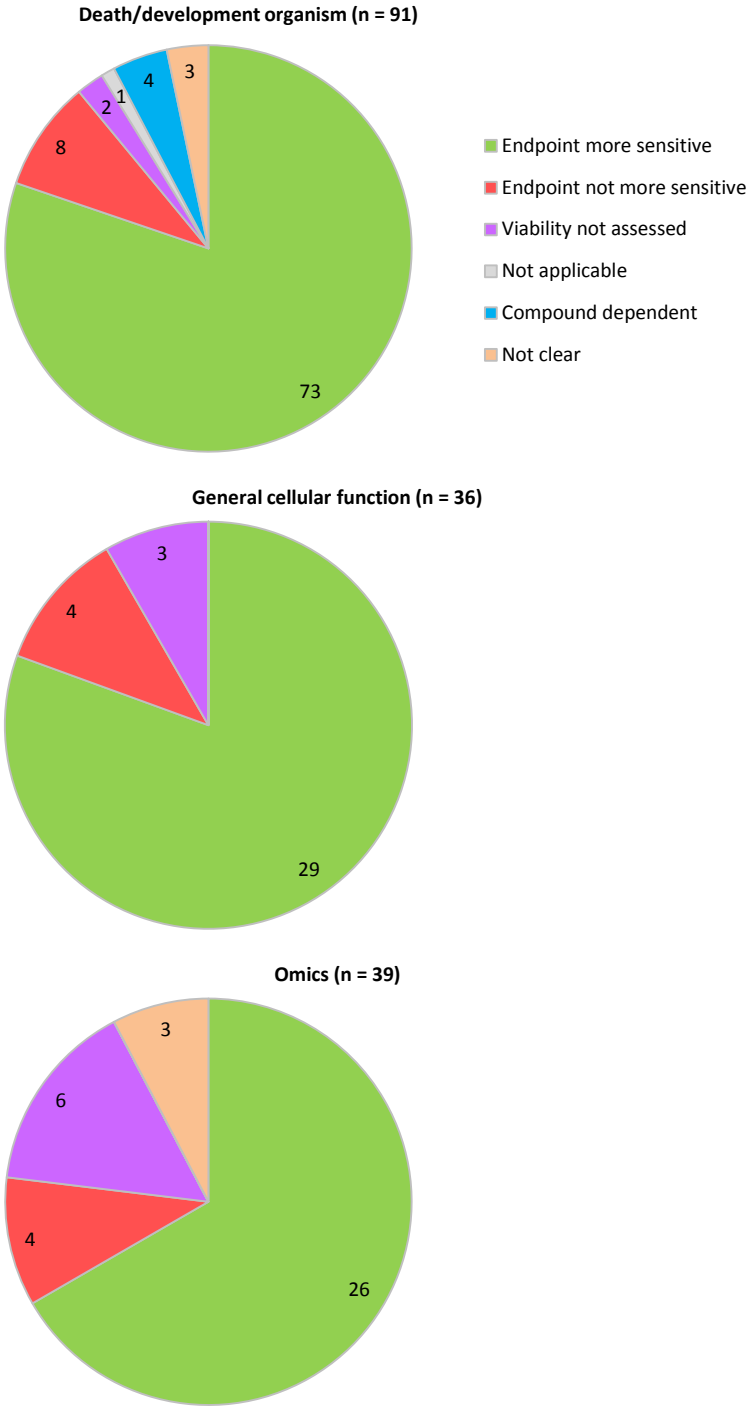


Figure 57 Distribution of DNT-specific endpoint groups for their changes compared to viability assessment. Number of citations (n) is given across life stages and compounds used in zebrafish studies. Each pie represents one endpoint group.

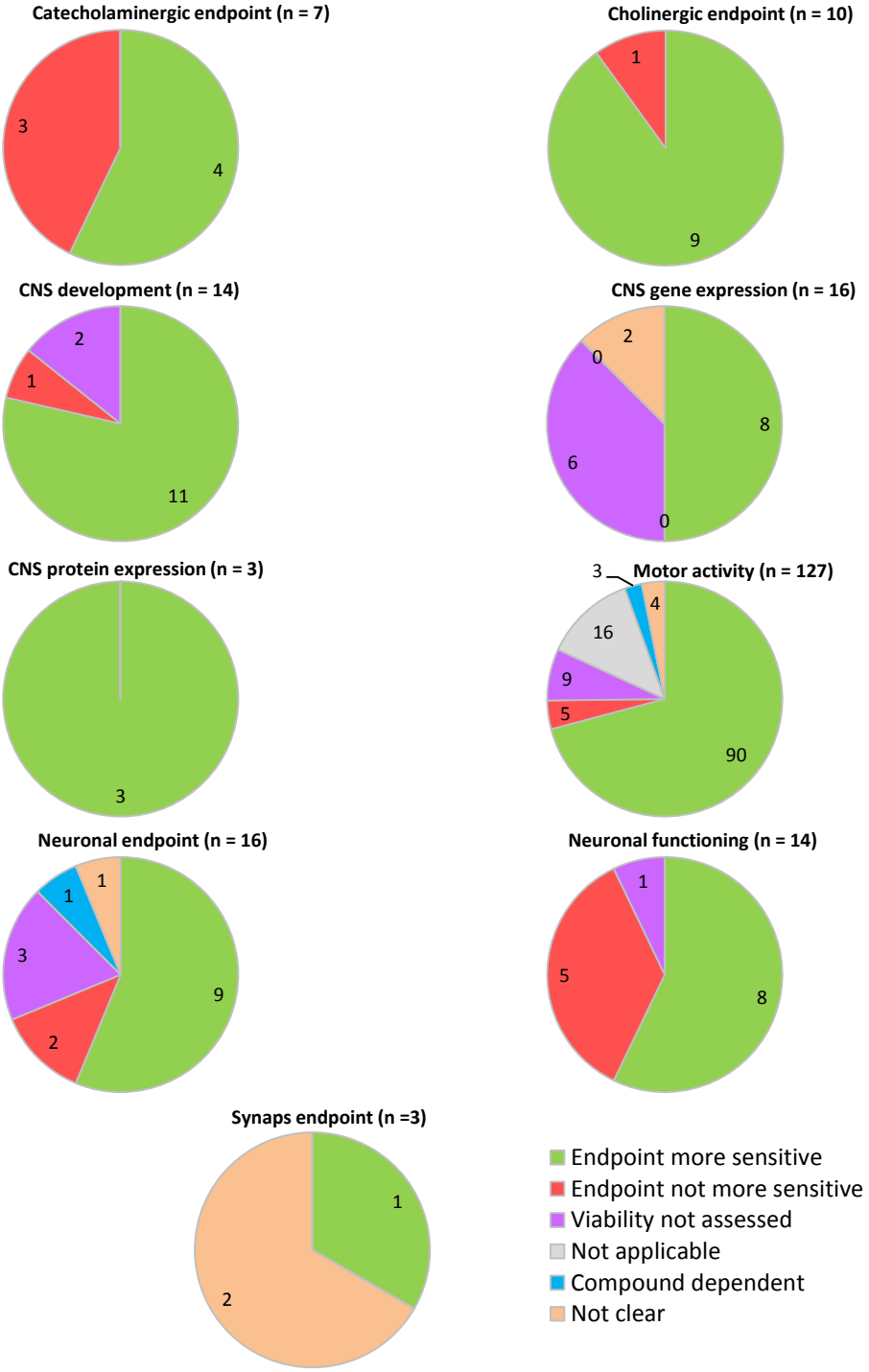


Figure 58 Distribution of DNT-specific endpoint groups for their changes compared to viability assessment. Number of citations (n) is given across life stages and compounds used in zebrafish studies. Each pie represents one endpoint group.

The endpoint evaluations shown in Figure 57 and Figure 58 were performed with all compounds tested in 376 zebrafish citations (83 publications). Of these, some are true (known) DNT positive compounds (Grandjean and Landrigan, 2006b, Grandjean and Landrigan, 2014), some are possible positives (e.g. pesticides, endocrine disrupting compounds) and some are true negatives. For a number of compounds,

no information is available and these are in the category 'not classified'. All compounds used in selected publications are listed, and distributed in their respective category based on information from literature. More explanation about classification of the compounds is given in previous section 2.3 and in Appendix H.

In a similar approach to the *in vitro* citations, we next evaluated how many of the true positives resulted in a specific DNT effect (endpoint effect more sensitive than viability) and how many of the true negatives caused either effect on viability at the same concentrations as the specific endpoint or had no effect on the DNT endpoint. Therefore, the data of the individual endpoint groups were grouped according to 'Hazard' yes/no, 'Compound Classification a priori' and 'Effects on Viability' into the categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. The category 'False Negatives' is not applicable here, because a negative endpoint evaluation of an a priori positive compound might be negative because 1) the DNT mechanism of action of this compound is not by altering this endpoint or 2) the organism, or specific life stage/model is not affected, although this compound is acting through this mode of action. The latter would be a true 'False Negative', whereas the first case would be correctly identified. However, we cannot distinguish between these two because very often the precise mode of action of a DNT compound is not known. Therefore, we call this category not 'False Negative', but 'Positives, Negative for this Endpoint'. The different groups are given in detail within Table 36 A. This evaluation will provide information that can in the next step be used for evaluating the specificity of different life stages/model used of zebrafish in assessment of DNT endpoints. In Table 36 B, compounds with a priori unclear classifications are listed. Because they do not represent clear DNT positives or negatives, one cannot judge performance of the organism and respective life stages/model system according to their responses on these compounds.

Table 36 Grouping of citations WITH available viability data (n = 342) according to 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint' (A) as well as 'Possible Positives identified as positive', 'Possible Positives identified as negative for this endpoint', 'A priori not classified, but identified as positive', A priori not classified, but identified as negative for this endpoint' and citations testing 'Multiple compounds from different categories' (B). The number of total citations in each group across all endpoints is given.

A	Hazard	Compound a priori classification	Viability	Number of citations
True Positives	'yes'	'Positive'	'Not affected' or 'Endpoint more sensitive than viability'	**97**
True Negatives	'no'	'Negative'	'Not affected' or 'Endpoint less sensitive than viability'	**4**
	'yes'	'Negative'	'Endpoint equally or less sensitive than viability'	
False Positives	'yes'	'Negative'	'Not affected' or 'Endpoint more sensitive than viability'	**3**
Positives, Negative for this Endpoint	'no'	'Positive'	'Not affected' or 'Endpoint less sensitive than viability'	**25**
	'yes'	'Positive'	'Endpoint equally or less sensitive than viability'	

B	Hazard	Compound a priori classification	Viability	Number of citations
Possible Positives	'yes'	'Possible Positive'	'Not affected' or 'Endpoint	**135**

identified as positive			more sensitive than viability'	
Possible Positives identified as negative for this endpoint	'no'	'Possible Positive'	'Not affected' or 'Endpoint less sensitive than viability'	**26**
	'yes'	'Possible Positive'	'Endpoint equally or less sensitive than viability'	
A priori not classified, but identified as positive	'yes'	'Not classified'	'Not affected' or 'Endpoint more sensitive than viability'	**30**
A priori not classified, but identified as negative for this endpoint	'no'	'Not classified'	'Not affected' or 'Endpoint less sensitive than viability'	**14**
	'yes'	'Not classified'		
Multiple compounds from different categories				**8**

These data shown in Table 36A are now evaluated for zebrafish studies with life stages/transgenic model having the ability to make correct predictions for DNT endpoints. Therefore, for each endpoint, zebrafish studies across life stages/model are analyzed for the four categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Within the category 'Positives, Negative for this Endpoint' there is a reasonable number of hits (25% of total in Table 36A), and we decided not to exclude this category. In addition, according to the definition of DNT compounds, only ONE of the multiple processes needed for brain development has to be disturbed to cause an adverse outcome. Thus, it is highly probable that some compounds affects only a small number of endpoints actually pointing to the ability of the method in zebrafish to specifically identify key events involved in particular modes of action of compounds. By this categorization, we will retrieve information on 1) whole organism/life stages/models, which are able to evaluate individual endpoints correctly, 2) whole organism/life stages/models which have the ability to assess a variety of DNT endpoints in a correct way and 3) zebrafish methods with high endpoint specificities. First, point number 1) will be assessed by walking through the endpoints individually. From this data 2) endpoint evaluations will be associated with individual methods. Finally, for 3) specificity and sensitivity analyses will then be performed in the end with the zebrafish methods identifying 'True Positives' and 'True Negatives' across individual DNT compounds and the three categories 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

- **Non-DNT-specific Endpoints**

Death/development organism

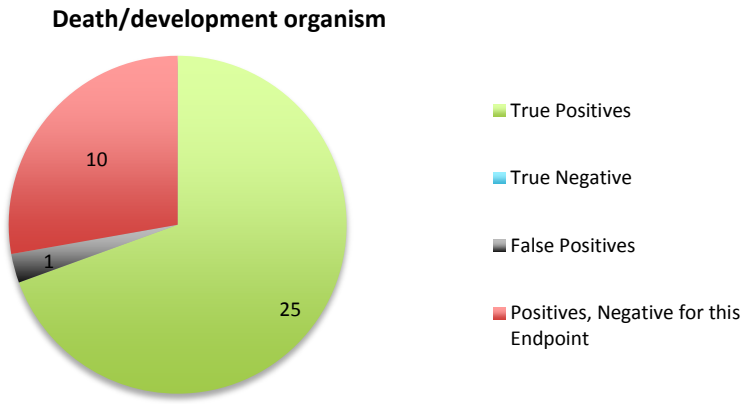


Figure 59 Distribution of citations for the endpoint group 'Death/development organism' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

For this non-DNT specific endpoint, 'death/development organism' there were 36 citations available and distribution is shown in Figure 59. True positive score was obtained for the combined embryo & larvae stage (24 citations) and the transgenic model (1), while no hits for true negative. 9 of 10 citations identified as 'positive, negative for this endpoint' refer to the endpoint 'embryo development'. For several of positive compounds such as BDE's, lead acetate, chlorpyrifos and chlorpyrifos oxon, embryo development was not affected at the selected test conditions (record nr. 1145 & 3784 for BDE-47, record nr. 1744 for BDE-153 & BDE-183, record nr. 1580 & 557 for lead acetate, record nr. 1860 & 1458 for chlorpyrifos and record nr. 1860 for chlorpyrifos oxon. This indicates the lower sensitivity of this endpoint for some of the DNT compounds, which have more specific effects on brain function rather than on embryo development at low concentrations. As most of these compounds were identified in other studies as true positives, it is likely a matter of tested concentration range. However, due to variability between exposure conditions and test methods, comparison for the same compounds between studies was not feasible.

For the endpoint 'Death/development organism' there were no hits for either only embryo or only larvae stage (Table 37) across the 4 categories.

Table 37 For the endpoint 'Death/development organism' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Death/development organism	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	0	0	24	1
True negative	0	0	0	0
False positive	0	0	0	1

Positive, but negative for this endpoint	0	0	10	0
Sum	0	0	34	2

2. General cellular function

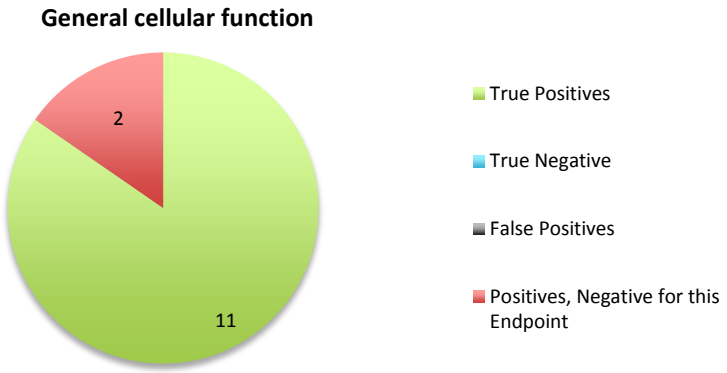


Figure 60 Distribution of citations for the endpoint group 'General cellular function' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

From all endpoint citations 'General cellular function' with viability data of the categories depicted in Table 36A are shown in Figure 60. From the total of 13 citations for the endpoint 'General cellular function' only with DNT positive and negative compounds (for classification of positive/negative/possible positive see Table 38), 11 are identified as 'True Positives', and 2 as 'Positives, Negative for this Endpoint'. These data were obtained with diverse life stage models (see Table 38). True positive results were obtained for the individual endpoints apoptosis (methods Tunel staining or acridine orange), oxidative stress (ROS assay) and protein expression (western blot) for a variety of positive compounds (0H-BDE's in record nr. 11926, ethanol in record nr. 4840 & 1316, lead acetate in record nr. 1580 & 1978, BDE's in record nr. 1027 & 1145). As there are no negative compounds tested, it is not possible to judge whether assays are able to distinguish for true positive versus true negative compounds. Next to the lack of testing with negative compounds, the overall number of citations for this endpoint group is low which makes it impossible to evaluate overall assay performance.

Table 38 For the endpoint 'General cellular function' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish General cellular function	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	6	1	4	0
True negative	0	0	0	0
False positive	0	0	0	0
Positive, but negative for this endpoint	0	0	2	0

Sum	6	1	6	0

3. Omics

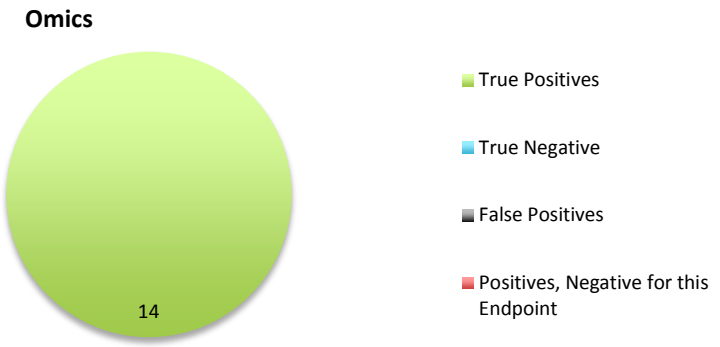


Figure 61 Distribution of citations for the endpoint group 'Omics' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

From all endpoint citations 'Omics' with viability data of the categories depicted in Table 36 A are shown in Figure 61. All 14 citations for this endpoint belong to the category 'True positive' indicating a good prediction for DNT positive compounds. The majority of citations belong to the group of embryo and larvae stage (n=12), and only 2 for embryo stage (Table 39). Despite true positive score, there is a lack of data for negative compounds. This approach, mainly covering whole organism gene expression (multiple genes by µarray of real time RT-PCR) is not expected to be very specific for DNT, but rather indicative of changes of whole organism gene profiles. This endpoint indicates changes which might point to adaptive responses, rather than direct adverse effects and are thus difficult to be used in the context of regulation, unless changes of these specific molecular events are related to a sequence of key events leading to adverse outcome. Therefore further research, especially on the development of AOPs for DNT using cellular, organ and organism model systems starting from omics data are needed.

Table 39 For the endpoint 'Omics' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Omics	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	2	0	12	0
True negative	0	0	0	0
False positive	0	0	0	0
Positive, but negative for this endpoint	0	0	0	0
Sum	2	0	12	0

DNT-specific Endpoints

Except for the endpoint group 'Motor activity' with 39 citations, including positive and negative compounds, the number of citations (n) for the other endpoint groups were always less than 10 and only positive compounds were evaluated. It concerns the 'Catecholaminergic endpoint (n=1)', 'the Cholinergic endpoint (n=6)', 'CNS development (n=6)', 'CNS gene expression (n=3), CNS protein expression (n=2), 'Neuronal endpoints (n=3)', 'Neuronal functioning (n=6) and 'Synapse endpoints (n=0).

Therefore the available data for these endpoint groups are summarized in following figures and tables except for the 'synapse endpoint group'. It can be seen that the true positive score is reasonable to good, but the lack of data does not allow to conclude about the specificity and sensitivity of these endpoints. It further appears that if data available at all, the embryo & larvae stage is mostly covered. The latter is likely most relevant in DNT perspective as different stages of brain development are included and thus not restricted to certain windows of development.

1. Catecholaminergic endpoint

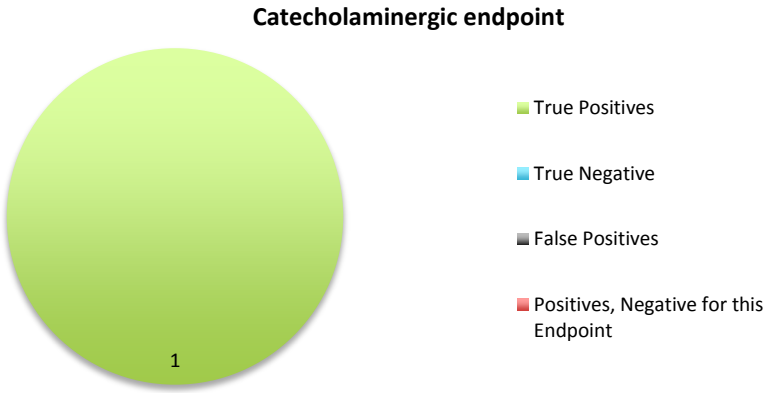


Figure 62 Distribution of citations for the endpoint group 'Catecholaminergic endpoint' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

For the pesticide dieldrin, positive for DNT, it was shown by Ton et al. (2006) that the number of catecholaminergic neurons, including dopaminergic neurons (tyrosine hydroxylase labeling) in zebrafish larvae brains at 96 hpf was decreased (record nr. 4025).

Table 40 For the endpoint 'Catecholaminergic endpoint' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Catecholaminergic endpoint'	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	0	0	1	0

True negative	0	0	0	0
False positive	0	0	0	0
Positive, but negative for this endpoint	0	0	0	0
Sum	0	0	1	0

2. Cholinergic endpoint

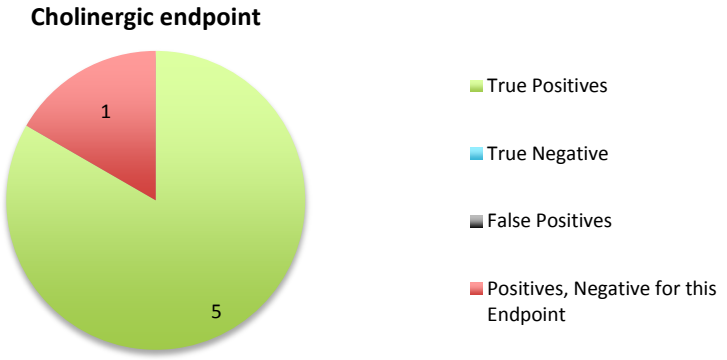


Figure 63 Distribution of citations for the endpoint group 'Cholinergic endpoint' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

All endpoints studied in this endpoint group refer to the measurement of acetylcholine esterase activity (5 citations) and concentration of acetylcholine (1 citation) in zebrafish, life stage embryo larvae with exposure periods up to 72 hpf or up to 120 hpf. Effects of DE-71 (technical mixture of PBDE) in range of µg/l give true positive results with increased activity of AChE, and corresponding decrease of acetylcholine concentration (Chen et al., 2012, record nr. 1027). Other citations were all studies with the organophosphate chlorpyrifos, and its metabolite chlorpyrifos-oxon. Similar to *in vitro* studies, chlorpyrifos-oxon is shown to be more potent. Jacobson et al. (2010) (record nr. 13959) demonstrated inhibition of AChE activity in zebrafish exposed for 72 hpf to nM concentrations of chlorpyrifos-oxon (record nr. 13959). Another study using chlorpyrifos exposure up to µM range demonstrated similar AChE inhibition but exposure lasted up to 120 hpf (record nr. 1458, Yen et al., 2011). Comparison of effects of both compounds for AChE activity in zebrafish was presented by Yang et al. (2011) (record nr. 1860). Inhibition of AChE activity was observed at 72 hpf exposure to chlorpyrifos-oxon as low as 30 nM, and was associated with decreased axon growth. However, this cholinergic effect was not observed upon 24 – 72 hpf exposure to the parent compound, chlorpyrifos (record nr. 1860) which resulted in this citation with score 'positive, but negative for this endpoint'. It was suggested by the authors that lack of effects could be attributed to the absence or less metabolic activity at early life stages, as was confirmed by HPLC measurements of parent compound and metabolites in zebrafish embryo and larvae at different time points. It is known that zebrafish embryo stages (24-48 hpf) are not metabolic competent, but liver becomes functional from 72 hpf onwards (Alderton et al., 2010). The latter is in line with observed CPF effects on AChE activity at later time points (record nr. 1458).

Table 41 For the endpoint 'Cholinergic endpoint' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Cholinergic endpoint'	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	0	0	5	0
True negative	0	0	0	0
False positive	0	0	0	0
Positive, but negative for this endpoint	0	0	1	0
Sum	0	0	6	0

3. CNS development

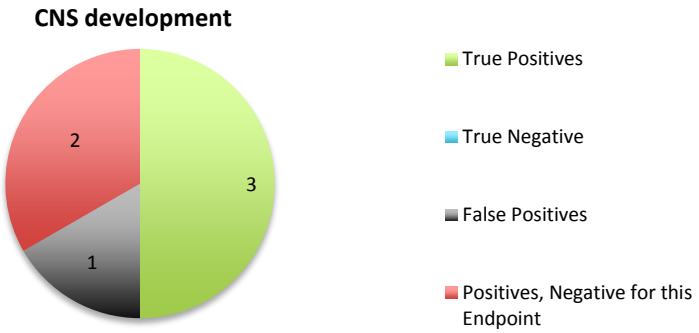


Figure 64 Distribution of citations for the endpoint group 'CNS development' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

Morphometric and immunohistochemistry methods were used to evaluate brain development as a measure of CNS development in embryo and larvae stages. Cowden et al. (2012) reported effects on CNS development by measurements of the retino-tectal area (whole mount anti-acetylated tubulin staining) in 48 hpf exposed zebrafish to ethanol and valproate. Both scored true positive (Cowden et al., 2012, record nr. 1351). One true positive citation (ethanol) and false positive (acetaminophen) were reported using the transgenic model Tg(nkx2.2a:mEGFP), with GFP expression for axon growth of motor neurons (Zhang et al., 2013, record nr. 705). Two publications with respectively BDE-47 (record nr. 3784, Lema et al. 2007) and lead acetate (record nr. 557, Peterson et al., 2013) exposure gave score 'positive, but negative for this endpoint'. Flow of cerebrospinal fluid (fluorescence microscopy) and brain length (light microscopy) were the measured, but less usual endpoints by respectively Lema et al. (2007) and Peterson et al. (2013), and showed to be respectively less sensitive than zebrafish viability or not affected by the compound.

Table 42 For the endpoint 'CNS development' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish CNS development'	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	2	0	0	1
True negative	0	0	0	0
False positive	0	0	0	1
Positive, but negative for this endpoint	0	0	2	0
Sum	2	0	2	1

4. CNS gene expression

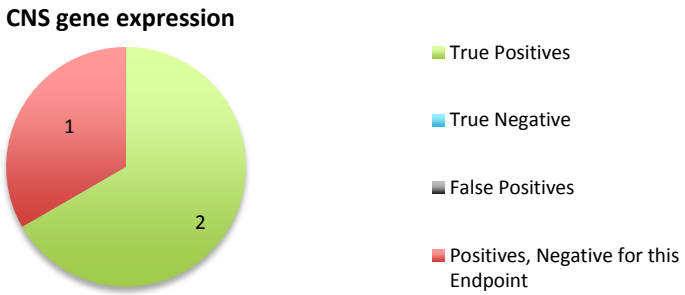


Figure 65 Distribution of citations for the endpoint group 'CNS gene expression' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

Zebrafish embryo and larvae stages were evaluated for CNS gene expression by studying specific gene expression in the brain using in situ hybridization (ISH) methods. The latter method does allow to allocate changed gene expression to specific regions of the brain or the central nervous system (dedicated areas), which is different (though mainly qualitative to semi-quantitative) from less specific whole body gene expression as is often used for omics approaches. Multiple genes involved in brain patterning and neurogenesis were studied by ISH, for further confirmation by real time RT-PCR. Ho et al. (2013, record nr. 662) demonstrated effects of methyl mercury (30-60 µg/l MeHg) giving a true positive score. Similarly, effects of lead acetate (0.2 mM) were shown in zebrafish embryo by Dou et al. (2011). Throughout the brains, gfap and huC gene expression patterns decreased significantly, particularly in the diencephalon region. RT-PCR further confirmed the down regulation of these 2 genes. The TUNEL assay demonstrated also the toxic effect of lead acetate by increased apoptosis of neuron and glia cells. However some other genes, e.g. ngn1 and crestin gene expression patterns were similar in both the Pb-treated embryos. The neurotoxic effects of lead acetate was also accompanied by decreased motor activity at sublethal conditions (record nr. 1580).

Reelin expression, a glycoprotein involved in neuronal development is measured by Peterson et al.(2013) after exposure to lead acetate but was not affected upon in situ hybridization analysis and resulted into 'positive, but negative for this endpoint' (see Table 43). However, this method might be less sensitive as the whole body real time RT-PCR analysis (grouped –omics endpoint) on the other hand resulted into true positive result, demonstrating higher sensitivity than effects on apoptosis and brain length (record nr. 557).

Table 43 For the endpoint 'CNS gene expression' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish CNS gene expression	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	0	0	2	0
True negative	0	0	0	0
False positive	0	0	0	
Positive, but negative for this endpoint	0	0	1	0
Sum	0	0	3	0

5. CNS protein expression

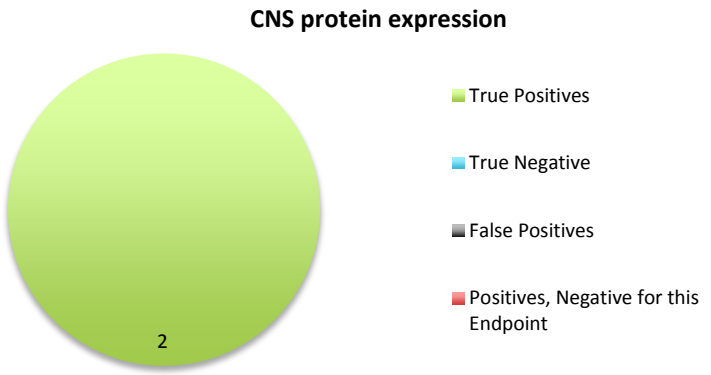


Figure 66 Distribution of citations for the endpoint group 'CNS protein expression' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

Immunohistochemical methods demonstrated specific protein expression, relevant for neurological functions in target areas. Two citations (Ho et al., 2013 , record nr. 662 and Jacobson et al., 2010, record nr. 13959) demonstrate true positive results for respectively MeHg and chlorpyrifos oxon exposure.

Table 44 For the endpoint 'CNS protein expression' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish CNS protein expression	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	1	0	1	0
True negative	0	0	0	0
False positive	0	0	0	
Positive, but negative for this endpoint	0	0	0	0
Sum	1	0	1	0

6. Neuronal endpoints

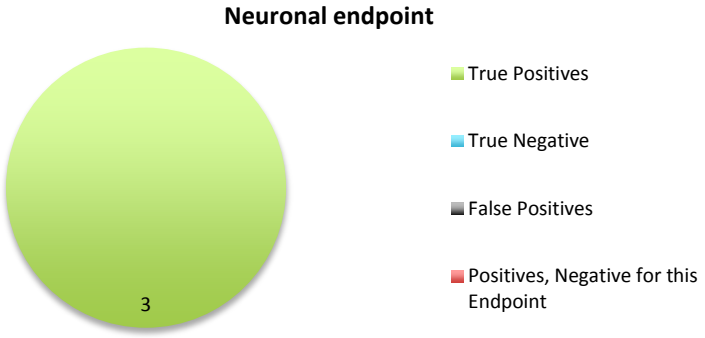


Figure 67 Distribution of citations for the endpoint group 'Neuronal endpoint' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

Significant effects of positive compounds, respectively BDE-47 (Chen et al., 2012, record nr. 1145, for larvae) and ethanol (Sylvain et al., 2010, record nr. 14229, for embryo & larvae stage) on axonal growth of primary and secondary neurons were demonstrated by immunohistochemical labelling of zebrafish early life stages up to hpf. These structural changes were shown to be accompanied by effects on motor activity (see later paragraph) as could be related to disturbance of neuronal connectivity. Furthermore Hassan et al. (2011) showed true positive result for a study with methylmercury exposed zebrafish giving a significant decrease of proliferating cells in the zebrafish neural tube, after labeling with proliferating cell nuclear antigen (PCNA) at MeHg concentrations down to 10 µg/l (record nr. 11748).

Table 45 For the endpoint 'Neuronal endpoint' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish	Embryo stage	Larvae stage	Embryo & larvae	Transgenic model

Neuronal endpoint		stage		
True positive		1	2	
True negative	0	0	0	0
False positive	0	0	0	
Positive, but negative for this endpoint	0	0		0
Sum		1	2	

7. Neuronal functioning

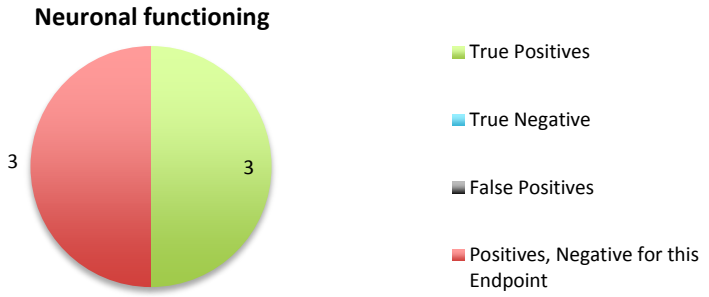


Figure 68 Distribution of citations for the endpoint group 'Neuronal functioning' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

True positive citations for neuronal functioning refer to immunocytochemistry labelling of acetylated α-tubulin as a measure for axon branching. Sylvain et al. (2010) showed typical abnormalities such as curved axons and aberrant branching patterns for dorsal and ventral motor neurons after 72 hpf ethanol exposure (2-2.5%) of zebrafish embryo & larvae (record nr. 14229). This method was also used by Ton et al. (2006) to visualize axons tracts, both in the brain and the caudal embryo after 48 hpf exposure to dieldrin. In the brain, the pattern and size of axon tracts is used as an indicator of affected brain regions. Exposure to 100 µM dieldrin caused a reduction of the optic tectum and reduced projections to the posterior commissure (record nr. 4025).

Transgenic zebrafish models were used by Yang et al (2011) to study sensory and motor neurons upon exposure to chlorpyrifos, and its metabolite chlorpyrifos oxon. Exposure of chlorpyrifos oxon from 24 to 72 hpf altered patterns of axonal growth in transgenic neurog1 zebrafish that express GFP in Rohon-Beard and DRG sensory neurons (Yang et al., 2011, record nr. 1860). These effects were not evident for chlorpyrifos exposure, the latter resulting into the 'positive, but negative for this endpoint score'. Other transgenic zebrafish models, expressing GFP under the control of either the NBT or the islet-1 promoter were used to assess axonal growth patterns in primary and secondary motor neurons, respectively. Similar to observations for sensory neurons, exposure from 24 to 72 hpf to chlorpyrifos at concentrations ranging from 0.01 to 1 µM had no effect on GFP-positive axons in transgenic zebrafish models. However, chlorpyrifos oxon inhibited axonal growth of GFP-positive neurons in both transgenic lines. Significantly

decreased lengths of both ventral and dorsal neurons of primary motor neurons were observed in NBT zebrafish exposed to chlorpyrifos oxon at 1µM In islet-1 zebrafish, chlorpyrifos oxon similarly inhibited axonal growth but significant effects were observed at the lower concentration of 0.1 µM, which did correspond to the same concentration range observed to impair touch-induced swimming behavior. These results again confirm the importance to use zebrafish at later stages (up to 120 hpf) in order to include metabolic competence resulting into metabolisation of the parent compound into the more potent metabolite, and which could turn results 'Positives, Negative for this Endpoint' into 'True positives'. The latter needs further investigation for other DNT relevant chemicals.

Despite these results for chlorpyrifos, which appeared inherent to the metabolic competence of the organism rather than to the endpoint, methods on neuronal endpoints (neuron differentiation) and functioning (axon patterning) appear valuable tools to assess DNT effects in zebrafish and often are linked to other relevant endpoints (see next paragraph). There is need for extended studies using positive as well as negative compounds, next to overall standardization of procedure.

Table 46 For the endpoint 'Neuronal functioning' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Neuronal functioning	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive		0	2	1
True negative	0	0	0	0
False positive	0	0	0	
Positive, but negative for this endpoint	0	0		3
Sum	0	0	2	3

8. Motor activity

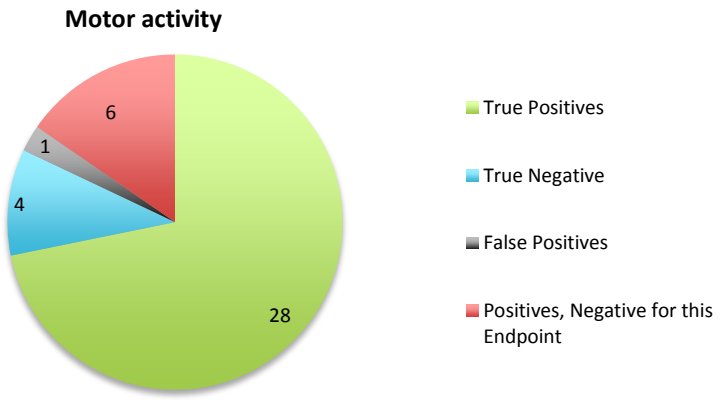


Figure 69 Distribution of citations for the endpoint group 'Motor activity' across the 4 categories 'True Positives', 'True Negatives', 'False Positives' and 'Positives, Negative for this Endpoint'. Each citation represents a zebrafish life stage/model study with one compound, thus numerous citations might result from one publication with multiple compounds and/or life stages.

From all endpoint citations 'Motor activity' with viability data of the categories depicted in Table 36A are shown in Figure 69. From the total of 39 citations for the endpoint 'Motor activity' only with DNT positive and negative compounds, 28 are identified as 'True Positives', 4 as true negative, 1 as 'False Positive' and 6 as 'Positives, Negative for this Endpoint'. These data were obtained with diverse life stage models. Embryo and embryo & larvae stage are equally important for detection of true positives (9 and 11 citations), while for 2 negative compounds methods for respectively the embryo and the larval stage each give 2 citations for true negative. (see Table 47). On the other hand, respectively 1 citation for embryo and 4 citations for larvae give 'Positive, but negative for this endpoint'.

Table 47 For the endpoint 'Motor activity' grouping of life stages/transgenic model in zebrafish, number of citations are given across 'True Positives', 'True Negatives' and 'Positives, Negative for this Endpoint'.

Zebrafish Motor activity	Embryo stage	Larvae stage	Embryo & larvae stage	Transgenic model
True positive	9	7	11	1
True negative	2	2	0	0
False positive	0	0	0	1
Positive, but negative for this endpoint	1	4	1	0
Sum	12	13	12	2

It appears thus that for zebrafish studies, the DNT-specific endpoint group, 'motor activity' is most data rich, representing both true positive and negative.

We now further elaborate a bit more in this evaluation of motor activity in general, considering the specific behavioral types during early zebrafish development (see test methods in data collections sheets; Appendix K) which are applied to assess motor activity in each of different life stages. Therefore some explanation on development of behavioral patterns in relation to neuronal network development is given first (Figure 70; (Brustein et al., 2003).

Background on behavioral profiles in zebrafish in relation to key events during early brain development

Within the endpoint group 'Motor activity', different types of behavioural responses are monitored in embryo, larval or embryo & larval stages of zebrafish which mainly refer to sensorimotor behaviour (behaviour patterns representing spontaneous activity and motoric and sensory functionality). Other behavioural outcomes, as monitored in mammals (autonomic-affective behaviour, based on the organismal self-regulation and emotion; cognitive behaviour patterns, related to attention, memory, evaluation, reasoning, problem solving and decision making), reproductive (behaviour patterns considered to be sex-specific) have not been part of this systematic review. This is because (1) such behavioural patterns are mainly related to juvenile and adult life stages, rather than being manifest in the selected early developmental stages in zebrafish as the model should be compliant with 3R or (2) adequate methods have not yet been developed for these early developmental stages, though recent developments with respect to habituation of larvae (~5 dpf) are reviewed by (Roberts et al., 2013).

The stereotypic motor activity of the developing zebrafish includes three sequentially appearing behaviours (Figure 70): a transient period of alternating **tail coiling's**, followed by **responses to touch** and the appearance of **organized swimming (Brustein et al., 2003)**.

Zebrafish embryos show their **first motor activity at 17 hpf**. This immature motor behaviour consists of **spontaneous repeating, alternating coils of the tail that persist over the course of several hours**. The frequency of these coils in dechorionated embryos peaks at 1 Hz at 19 hpf and slowly decreases to 0.1 Hz by 26 hpf. The spontaneous tail coilings are shown to be neural in origin and the substrate essential for the appearance of spontaneous coils is located completely within the spinal cord. Morphological observations suggest that a maximum of six cell types located wholly within the immature embryonic spinal cord are responsible for the appearance of spontaneous motor activity (Saint-Amant and Drapeau, 1998, Saint-Amant, 2006).

Periodic depolarisations (PDs) are observed in all ipsilateral caudal and ventral lateral descending interneurons and in most of the commissural primary ascending interneurons from 19 to 24 hpf. PDs are synchronous in all active neurons and these are electrically coupled to each other by gap junctions. It has been shown that gap junctions have a critical role in the propagation of PDs, but the mechanism by which PDs are initiated is still unknown. Voltage dependent sodium channels are necessary for the generation of PDs, suggesting an important role for sodium influx. Further, there is significant calcium entry during PDs and calcium-activated potassium channels may play a role in PD termination (Saint-Amant and Drapeau, 2000, Saint-Amant and Drapeau, 2001).

The next step in the progression of motor behaviours in the zebrafish embryo is the appearance of the **touch response at 21 hpf**. This new behaviour suggests a change in the early gap junction-mediated motor network and implies at the very least, new functional connections with the sensory system. Hindbrain projections and chemical neurotransmission are already present by 21 hpf in the zebrafish embryo and are required for normal touch responses. The morphology of the spinal cord is changing rapidly at the stage when touch responses first appear. After 21 hpf, at least four more types of interneurons are added to the spinal cord (Bernhardt et al., 1990). The Mauthner neuron is presumably involved in the first manifestations of touch responses as its axon is already at the third spinal segment by 21 hpf (Mendelson, 1986a, Mendelson, 1986b). Whole cell patch clamp studies show that over 80% of primary motor neurons display a new type of spontaneous rhythmic event in addition to the PDs that are already present. These events were named synaptic bursts because they are composed of rapid depolarization's that quickly return to baseline as opposed to the sustained drive during PDs (Saint-Amant and Drapeau, 2000).

Embryonic zebrafish can be induced to swim as early as 28 hpf the 3rd major motility pattern. The cycle (two alternating tail beats) frequency of the first observed **swimming** bouts is 8 Hz and the duration of the episodes is short lived. Slightly later in development, at 36 hpf, the embryos can swim at a 30 Hz cycle frequency (Saint-Amant and Drapeau, 1998). Two-day larvae swim in long uninterrupted bursts with tail beats that can reach a frequency of over 100 Hz and last tens of seconds. After 3 days of development, swimming switches to an intermittent mode of swimming. These motor events are composed of a string of a few cycles of tail beats punctuated by periods of inactivity (Buss and Drapeau, 2001, Muller and van Leeuwen, 2004). After 4 days, larvae have been shown to exhibit four main types of motor behaviours. Two of these motor gaits were classified as turning behaviours, namely routine turns and escape turns. The other two motor gaits were classified as swimming behaviours and have been designated slow swims and burst swims (Saint-Amant, 2006). Swimming requires integration of both hindbrain reticulospinal neurons, spinal cord interneurons and secondary motor neurons. The exact network interactions are not fully understood but the synaptic drive to motor neurons comprises rhythmic glutamatergic and tonic glycinergic components (Brustein et al., 2003).

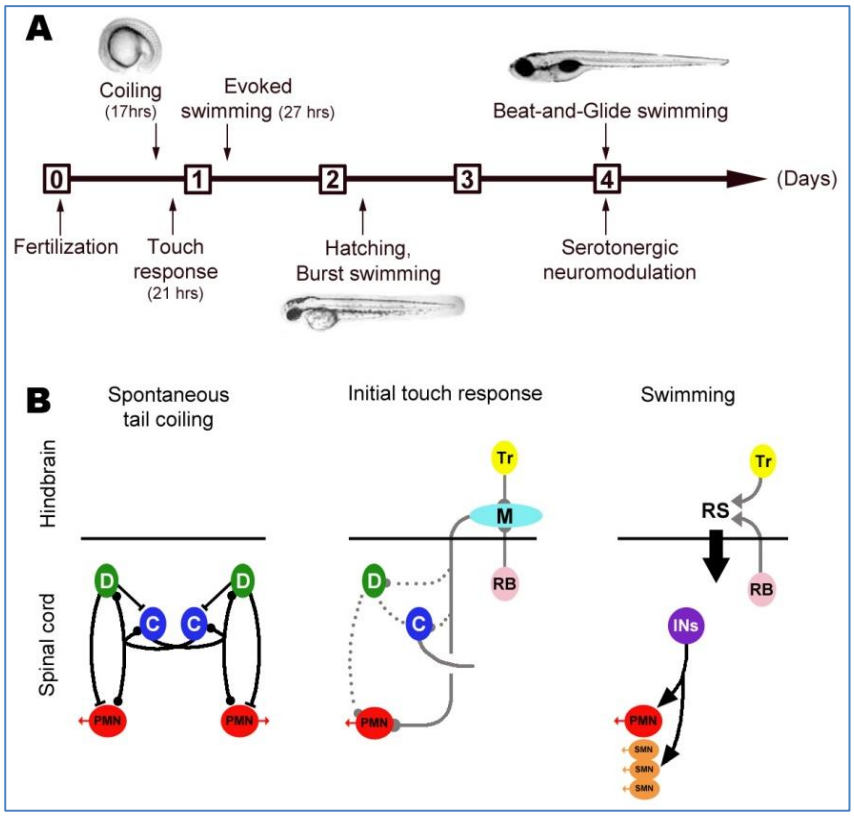


Figure 70 The chronological sequence of the appearance of motility patterns during development of the zebrafish. Images emphasize key stages of zebrafish embryos development before and after hatching. (B) Schematic diagrams of the possible neural building blocks underlying early motility patterns. *Left*: the neural network active during spontaneous tail coilings is limited to the spinal cord and includes only primary motor neurons and a restricted number of interneurons (D = descending, C= commissural). The network activity is based on electrical coupling. *Center*: the initial touch response requires the activation of both hindbrain neurons (M = Mauthner, Tr= Trigeminal) and spinal neurons including the sensory Rohon-Beard (RB) neurons. Dashed lines represent hypothetical connectivity. At this time the synaptic response is mediated partly by glutamatergic synapses. *Right*: at 27 hpf the embryo can swim in response to touch. Swimming requires integration of both hindbrain reticulospinal neurons (RS) and spinal cord interneurons (Ins) and secondary motor neurons (SMN). The exact network interactions are not fully understood but the synaptic drive to motor neurons comprises rhythmic glutamatergic and tonic glycinergic components (according to Brustein et al., 2003).

These 3 main behavioural patterns are related to specific time window of development (general defined in data sheets as life stages) and methods have been developed throughout the past years to detect (mostly by automatic video tracking) changes in response to chemical exposure. Inventory of the literature during the systematic review demonstrated a lack of standardisation of methods with respect to evaluation, time of exposure, either of not the use of a stimulus and the type, the detection method,…. as well as terminology for the behavioural method.

In the next table (Table 48), the citations for motor activity of zebrafish (n = 125) are distributed across life stages/transgenic model with details on the specific method used, as mentioned in the data collection sheet_AO. These life stages represent citations from both chorionated and dechorionated organisms. It should be clear that the dechorionation step is e.g. inherent to the test method 'touch response' for the embryo stage (in egg) includes.

Table 48: Details on specific behavioral test methods, as used for each of the life stages of zebrafish to detect motor activity (terminology used in publications, and given in data collection sheet) with corresponding number of

citations. Test methods are categorized using (1), (2) & (3) into main behavioral patterns for further evaluation (see next table).

Test methods for endpoint 'motor activity'	Total for all stages/model	Embryo	Larvae	Embryo & larvae	Transgenic model
Infrared beam swimming, detecting transient signal fluctuation upon larval swimming (3)	1			1	
Light-dark challenge startle response (2)	1			1	
Light-dark challenge swimming (3), (*)	34		13	21	
Optomotor response, swimming after visual stimuli (3)	1			1	
Photomotor response (3)	1	1			
Spontaneous coiling (1), (*)	26	25			1
Spontaneous swimming (3), (*)	45	1	25	15	6
Startle response (2), (*)	2			2	
Swimming-several parameters (3), (*)	2			2	
Touch response (2), (*)	12	3		9	
Sum 'motor activity'	125	30	30	52	7

*: only for these methods, true positive or true negative citations were available, for further analysis in next table.

It is clear that the embryo & larval stage is most used for a diversity of methods to detect motor activity (n=52), while behavioural patterns related to swimming, either spontaneous or after stimulus (light-dark challenge) are mostly used across life stages.

In order to evaluate the performance of these methods, we selected only those endpoint methods (marked * in Table 49) for citations which scored either 'true positive' or 'true negative' (n=28 or n=4 for motor activity). These methods were grouped according to the 3 main defined patterns [(1) tail coiling, (2) touch response, (3) swimming] for motility in zebrafish early life stages as given in Figure 70 and evaluated across life stages, within the endpoint group 'motor activity'. Due to the low number of citations (and the lack of standardisation of methods) it is not relevant to distinguish in this analysis for e.g. spontaneous versus stimulated swimming.

Table 49: Number of citations according to test performance (true positive: TP, true negative: TN), for grouped methods of behavioural patterns at different life stages, for those with data for true positive (total n=28) or true negative (total n=4) for endpoint group motor activity.

Main motility pattern (see figure)	Response	Embryo	Larvae	Embryo & larvae	Transgenic model
1. Spontaneous tail coiling	TP	8			
	TN	2			
2. Touch response	TP	1		5	
	TN				
3. Swimming	TP		7	6	1

	TN	2

The number of citations is highest for methods measuring swimming activity (either or not after stimulation), with 14 true positive and 2 true negatives . Next is spontaneous tail coiling, applicable to the early embryo stage with 10 hits including 8 true positives and 2 true negatives, while touch response gives for evaluated positive compounds 1 true positive for the embryo stage and 5 true positives in the embryo & larval stage. Each of these methods are highly relevant as integrated endpoint of early brain development and should be further evaluated for their suitability to be included in integrated DNT test strategy. Therefore, these is a clear need for method standardisation (include challenge either or not, time point of exposure and evaluation) and evaluation of standard methods for extended panel of both positive and negative compounds.

The observed behaviour pattern, related to specific life stages as described above is clearly linked to early cellular events (neuron differentiation & migration) and neural network establishment (synapse, neurotransmission, ..) in the developing nervous system of zebrafish. Several of these endpoints (neuronal endpoint group, neural functioning group, synapse endpoint,..) have been evaluated in this systematic review and are compared for their association with motor activity response for known DNT positive chemicals.

Citations for true positives were evaluated for simultaneous analysis of endpoints for motor activity (apical) and other DNT-specific endpoints such as related to neuron development, axon growth,…, scoring true positive and all important in the AOP approach, providing evidence for early events resulting into changed behavioural activity. Relevant publications are listed in the next table (Table 50).

Table 50 True positives for the endpoint 'Motor activity', related DNT-specific endpoints with effects in the same model system (life stage zebrafish) at same exposure condition (reported within same publication) is given to show the associated specificity of both endpoint groups.

True positive for Motor activity	Zebrafish Life stage	Apical Endpoint Motor activity	Early event DNT specific endpoint	Reference
BDE-47	Larvae	Light/dark challenge swimming response & Spontaneous swimming	Neuronal: axonal growth of primary & secondary motor neurons, neuromuscular junction	Chen 2012 (Record Nr. 1145)
DE-71	Embryo & larvae	Light/dark challenge swimming response	Cholinergic: acetylcholine concentration, AChE enzyme activity	Chen 2012 (Record Nr. 1027)
Chlorpyrifos	Embryo & larvae	Touch response	Cholinergic: AChE enzyme activity	Yen 2011 (Record Nr. 1458)
Chlorpyrifos-oxon	Embryo & larvae	Spontaneous swimming	Cholinergic: AChE enzyme activity	Yang 2011 (Record Nr. 1860)
Ethanol	Embryo & larvae	Touch response	Neuronal: axonal growth of primary & secondary motor neurons	Sylvain 2010 (Record Nr. 14229
Ethanol	Transgenic model (embryo-larvae)	Spontaneous movement & touch response	Neuronal: axon length	Zhang 2013 (Record Nr. 705)

For the **embryo stage** (including dechorionated embryo), no hits 'true positive' for motor activity were evaluated in parallel to DNT specific endpoints in the same study.

For larvae, only 1 publication scored as true positive for motor activity with other DNT-specific measurements in the same study (Table 50). Motor behavior development and swimming behavior in response to light-to-dark photoperiod stimulation were studied at various developmental stages (6-90 hpf). BDE-47 exposure significantly affected spontaneous movement, and altered larvae swimming behavior in response to light stimulation in developing zebrafish. Consistent with these motor deficits, BDE-47 significantly inhibited axonal growth of primary and secondary motor neurons during the early developmental stages at low BDE-47 concentrations, suggesting the functional relevance of structural changes. Apoptotic effects and muscular defects (non-specific DNT effects) were only seen at the higher BDE concentrations (Chen et al., 2012b).

Several studies on **embryo & larvae stage**, with 'true positive' for motor activity showed in same publications true positive response for cholinergic endpoints. Examples are given in Table 50 where exposure to DNT positive compounds disrupts the neurobehavior of zebrafish larvae and affect cholinergic neurotransmission. In work by Chen et al. (2012a) it is shown that low DE-71 concentration caused hyperactivity, whereas higher concentrations decreased activity during the dark period. During the light period, larval activity was significantly reduced in a concentration-dependent manner. In the cholinergic system, acetylcholinesterase activity significantly increased in the 68.7 and 227.6 µg/L exposure groups, respectively, and acetylcholine concentration accordingly decreased in the 227.6 µg/L exposure group. In zebrafish exposed to chlorpyrifos over the first five days of embryonic and larval development, survival, acetylcholinesterase (AChE) activity and behaviour were affected. However, effects are DNT specific as both behaviour and AChE inhibition occur at non-lethal levels of chlorpyrifos (Yen et al. 2001). Similar observations were reported by Yang et al. (2011) for exposure to the metabolite, chlorpyrifos-oxon. Sylvain et al. (2010) investigated the effects of embryonic alcohol exposure (1.5%, 2% and 2.5% EtOH) on motor neuron and muscle fiber morphology in 3 days post fertilization (dpf) larval zebrafish. EtOH treated fish exhibited fewer bouts of swimming in response to touch, compared with untreated fish. At 2 and 2.5% EtOH higher rates of motor neuron axon defects were seen after immune labelling for primary and secondary motor neurons. In addition, red and white muscle fibres revealed that fish exposed to EtOH had significantly smaller fibres compared with controls. Thus, both specific (motor neuron development) and non-specific (muscle fibre morphology) DNT endpoints are affected by early alcohol exposure in zebrafish and are related to deficits in locomotion.

Motor activity disturbance was observed in a **transgenic zebrafish model** after exposure to ethanol. It appeared in that same publication that GFP-labeled ventral axons from trunk motor neurons, which easily are observed and quantified in transgenic larvae Tg(nkx2.2a:mEGFP) were highly sensitive to neurotoxins. Ethanol exposure (3-120 hpf) resulted in significant effects on axon length, a DNT specific endpoint at levels below those for general toxicity, measured as total body length (non-DNT specific, embryo development) in the transgenic model (Zhang and Gong, 2013).

Summary alternative organisms

Within this systematic review for DNT methods using alternative organisms we identified 432 citations for individual methods, comprising early life stages of zebrafish, sea urchin and nematode measuring individual endpoints for individual compounds. This dataset of citations represented a total of 40 individual endpoints, where 32 endpoints were represented in 244 zebrafish studies for 83 selected publications, 10 endpoints in 13 sea urchin studies for 5 publications and 7 endpoints in 8 nematode studies in 2 publications. Due to the low number of hits for about half of the individual endpoints (19 endpoints with only 1 citation across species and life stage), and similar to the data analysis for *in vitro*

studies, endpoints were grouped into 3 groups for non-DNT specific endpoints and 9 groups for DNT specific endpoints. Due to too low number of studies within endpoint groups, respectively 7 studies on 4 DNT specific endpoints and 6 studies on 2 non-DNT specific endpoints in sea urchin, and 4 studies on 3 DNT specific endpoints and 4 studies on 3 non-DNT specific endpoints in nematode, it was decided that only zebrafish data were further evaluated for sensitivity and specificity of DNT methods. Considering all the zebrafish data with viability registration, 376 citations (in 83 publications) were retained and about 72.1% of the studies showed (non-)DNT specific endpoint changes more sensitive than viability assessment. Within the group of non-DNT specific endpoints, the group 'death/development organism' was best presented (91 studies in zebrafish across life stages), while the endpoint group 'motor activity' with 127 studies represented the major part in the DNT-specific endpoints. With respect to assay performance, studies with viability data were grouped according their response to classes of compounds, respectively positives, negatives, possible positive and not-classified. For each of the endpoint groups, and for each of the life stages of zebrafish (embryo, larvae, embryo & larvae, transgenic model), the ability of the assay to correctly predict the response of the positive and negative compounds was assessed. However significant data gaps became clear: the number of citations (n=7) performed with negative compounds (n=2) in 2 distinct publications was extremely low across all endpoints. For the majority of DNT-specific endpoints the number of citations representing measurements for both positive and negative compounds was overall too low. Only the endpoint groups 'motor activity' (39 citations), the non-specific DNT endpoints 'death/development organism' (36 citations), 'General cellular function (13 citations) and 'Omics' (14 citations) were reasonably represented. The performance within the endpoint group motor activity (39 citations) gave 28 true positive studies, 4 true negative studies , 1 false positive, and 6 studies with positive compounds with negative outcome for motor activity. The measurements of motor activity across life stages were almost equally distributed (embryo, n=12; larvae, n = 13; embryo & larvae, n=12), while the transgenic model represented only 2 citations on motor activity. Despite low numbers of studies for the other DNT specific endpoints, several studies with a true positive response for motor activity showed associated true positive responses for individual endpoints related to e.g. neuronal development and functioning (axon growth, primary and secondary motor neurons), or cholinergic function (acetylcholine concentration, AChE enzyme inhibition), pointing to early effects on brain development which integrate into changes into the apical endpoint 'motor activity'. With respect to metabolic competence, it is demonstrated that exposure of zebrafish beyond 72 hpf is important, and that observations should be extended until the 120 hpf (period for ethical acceptance) in order to include activity of metabolised compounds as was shown for chlorpyrifos and its metabolite chlorpyrifos oxon.

3.3. In vivo

A total of 78 original primary research publications containing *in vivo* data (73) or *in vivo* and *in vitro* data (5) for DNT evaluation remained in the study for full text evaluation after the successive rounds of exclusion (see 2.2; selection).

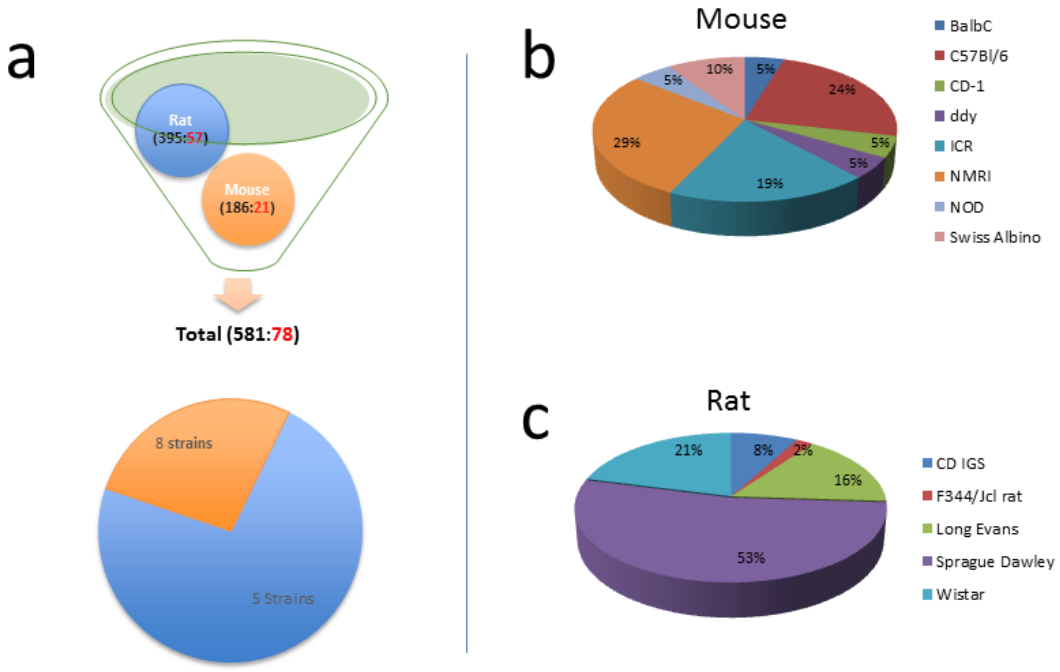


Figure 71 Number of publications in the dataset. a) A total of 78 publications have been retained for full text evaluation in this project. 57 publications used rat models (covering 5 strains of rats) and 21 publications used mouse models (covering 8 strains of mice). Black numbers in funnel is the total number of OAs the publications (see text). b) Mouse strains used in the included publications. NMRI-mice are used in 29% of the publications and is the most frequently used mouse strain. c) Rat strains used in the included publications. The Sprague Dawley strain has been used in the majority of the papers (53%).

They have all been thoroughly scrutinized, and the relevant data has been extracted and systematized in the data collection sheet (Appendix K). All included publications were written in English, were based on rodent studies and were published between 1990 and May 2014 (Figure 71). In the 78 publications there are often more than one "relation" published. In order to annotate publications that report investigation on more than one brain region, outcome endpoint and/or exposure design, the primary DNT *in vivo* data set (n=78) was used to generate a greater set of so called Outcome-Agent of toxicity relations (OA-relations; Figure 72). Each OA-relation represents one unique exposure and outcome-endpoint setting with the outcomes ranging from individual molecular endpoints to behavioral endpoints. When we used the data to generate OA-relations, we found a total of 581 in the 78 publications (Figure 73b).

Figure 72 Outcome-Agent relation or OA-relation. An OA-relation is composed of an empirically determined outcome that is preceded by an experienced agent of toxicity.

Publication year

Although the searches included publications from 1990 until May 2014, a majority of the included publications have been published over the past 15 years (Figure 73b).

There is of course a possibility that we have created a bias in our selection where we unintendedly have selected newer publications. However, the same trend is also present when we look at the graphs of the total number of curated publications (Figure 73, n=3466) and also for the unfiltered initial search (n=13 235) suggesting that the past 15 years have indeed witnessed an increase in DNT publication rate. The average number of OAs per paper is twice as many between 2003 and 2014 (n=7.9) compared to 1990-2002 (n=4.7). Even so, counting the papers we would still expect more included publications from 1990-2002 if we compare the selected data with the unfiltered data. Although there may be many reasons for this bias, one likely explanation is that over time (and with the introduction of the U.S. EPA DNT and OECD TG 426 guidelines) there has been a trend to include more experimental groups in the studies to enable dose-response modelling (e.g. benchmark dosing). Thus, the decision to exclude all the publications without at least three dose levels and a concurrent control (in accordance with the OECD TG 426 criterion) has likely created a bias towards newer publications. However, this decision (to only include publications with >3 dose groups) has also created a dataset which is useful as a starting-point for *in vivo-in vitro* comparisons. It is imperative to be able to model dose-dependent relations if we want to identify *in vitro* assays that in a meaningful way reflect DNT *in vivo*.

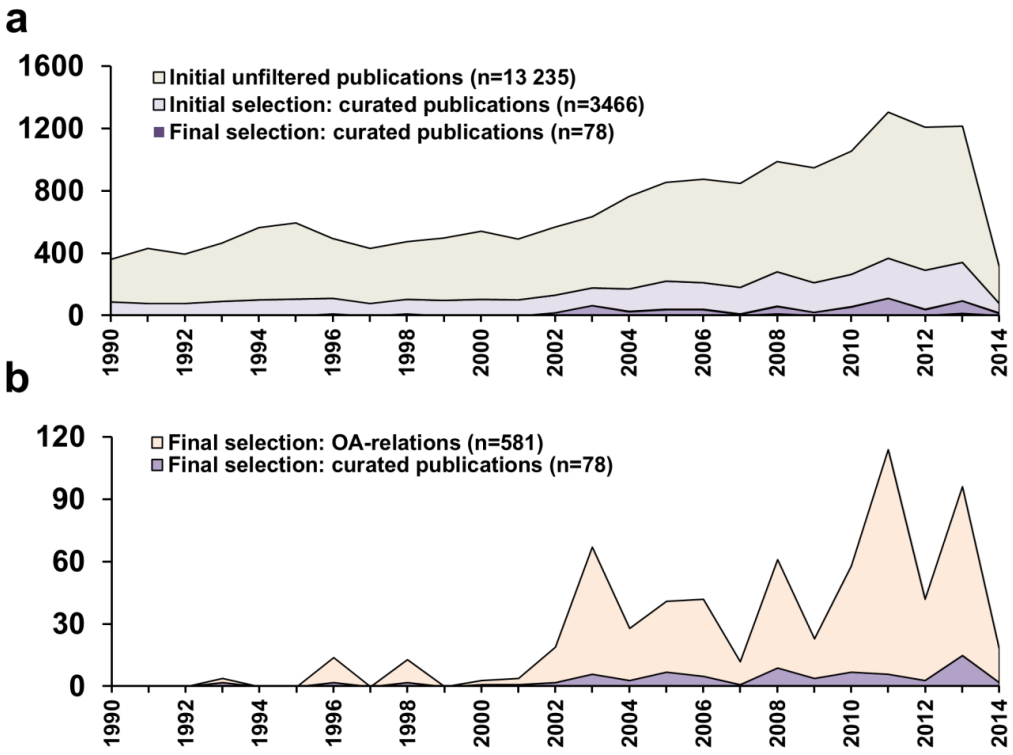


Figure 73 The number of publications in the DNT *in vivo* data set by year of publication between 1990 and May 2014. a) An overview of the relation between the initial unfiltered literature (KI after group I+II, n=13 235). An initial selection round led to a final DNT *in vivo* data set of n=3466 publications (DNT 1st selection KI). A final curation-based filtering led to a primary DNT *in vivo* data set of n=78 publications which was then classified more in depth and put on the data selection sheets. From 2001, there is a general increase in DNT publications b) Based on OA-relations, we can see an increase over time in publications meeting the inclusion criteria and a greater increase in corresponding OA-relations.

Species

All the 78 publications included in this project are rodent studies (mouse and rat). Although there are aspects of nervous system development that have been primarily investigated in other species (such as the development of ocular dominance in cat models (Hubel and Wiesel, 1962), such models are rarely used for DNT studies. A total of 57 publications used rat strains and 21 publications used mouse strains (Figure 74). For the mouse models, NMRI mice (outbread strain) and C57/Bl6 (inbread strain) are used in 52% of the publications. Some strains (such as the ddy and the BalBc mouse) are only used in one publication each, and for more specific research questions. For the rat model (which is the preferred species in the OECD TG 426 guideline), the Sprague Dawley rat (outbread strain) is the most commonly used strain in the included publications (n=35) followed by the wistar rat (outbread, n=12). Although the use of specific species and strains of animals to a large extent is based on tradition rather than scientific reasoning, some strains are overrepresented in certain contexts. For behavioural studies, for instance, the Sprague Dawley rat is the most commonly used strain (Figure 74), which to at least some extent explains the dominance of that particular strain in the included studies. Regarding time-trends of model strain usage, it is not possible to see any time-trend in the data (Appendix K).

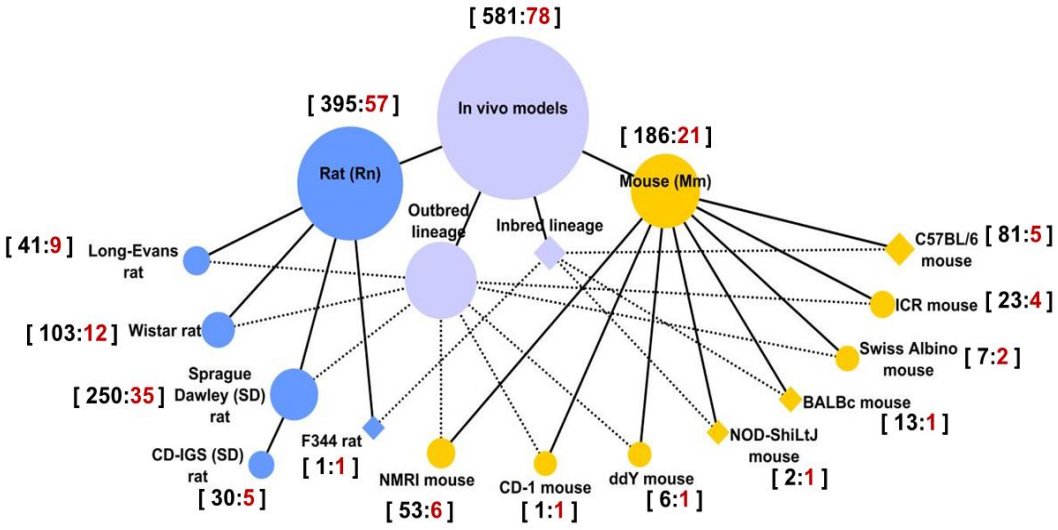


Figure 74 A network representation of the animal models used in the included publications. The DNT *in vivo* data set was composed of a total of 581 OA-relations corresponding to 78 unique publications. The animals used were primarily rats (blue colour; OA-relations: 395, Publications: 57) and secondarily mice (orange colour; OA-relations: 186, Publications: 21). Rats and mice were in turn sub-classified into several strains and stocks (connected by solid lines; Rats, n=4 and Mice, n=8). The outbred and inbred status is shown by dotted lines. Sprague Dawley rats were the most commonly used rats (OA-relations: 250, Publications: 35) whereas C57BL/6 mice (OA-relations: 81, Publications: 5) and NMRI mice (OA-relations: 53, Publications: 6) were the most commonly used mice. The size of the nodes in the network represents the number of OA-relations. Number of OA-relations (black colour) and corresponding number of unique publications (red colour) are shown within brackets.

Brain parts used

A total of 21 brain parts have been assayed in the 78 publications. However, this count is biased in that many papers use less defined anatomical brain parts. Six of the 21 brain parts are major lobes or even "whole brain" (Figure 75a). If we exclude behaviour OA-relations (where no individual brain part is

assayed) six brain parts (forebrain, whole brain, hippocampus, striatum, cerebral cortex and cerebellum) make up 85% of the total. If we exclude forebrain and whole brain, the most popular anatomical brain parts used are in descending order hippocampus, cerebral cortex, striatum and cerebellum.

One preliminary hypothesis of ours was that one would be able to detect that the biosampling of more general brain regions over time would become complemented by an increase in reports of OA-relation focusing on smaller, more distinct anatomical regions. However, no such trend is seen in the data. Thus, the majority of the studies focus on the same brain parts, while significant parts of the brain are not included. Whether this is a bias in our dataset, or if this truly represents the bulk of the data is an important issue. If the currently available data (i.e. all the publications covering various aspects of DNT regardless of inclusion/exclusion criteria) evaluate either big chunks of the brain or with large overlaps in the choice of anatomical region, then this is a major knowledge gap.

Force feeding (oral, oral gavage, oral injection) is the major exposure paradigm used across all brain regions when the exposure was initiated postnatally, followed by intraperitoneal injection (Figure 75b).

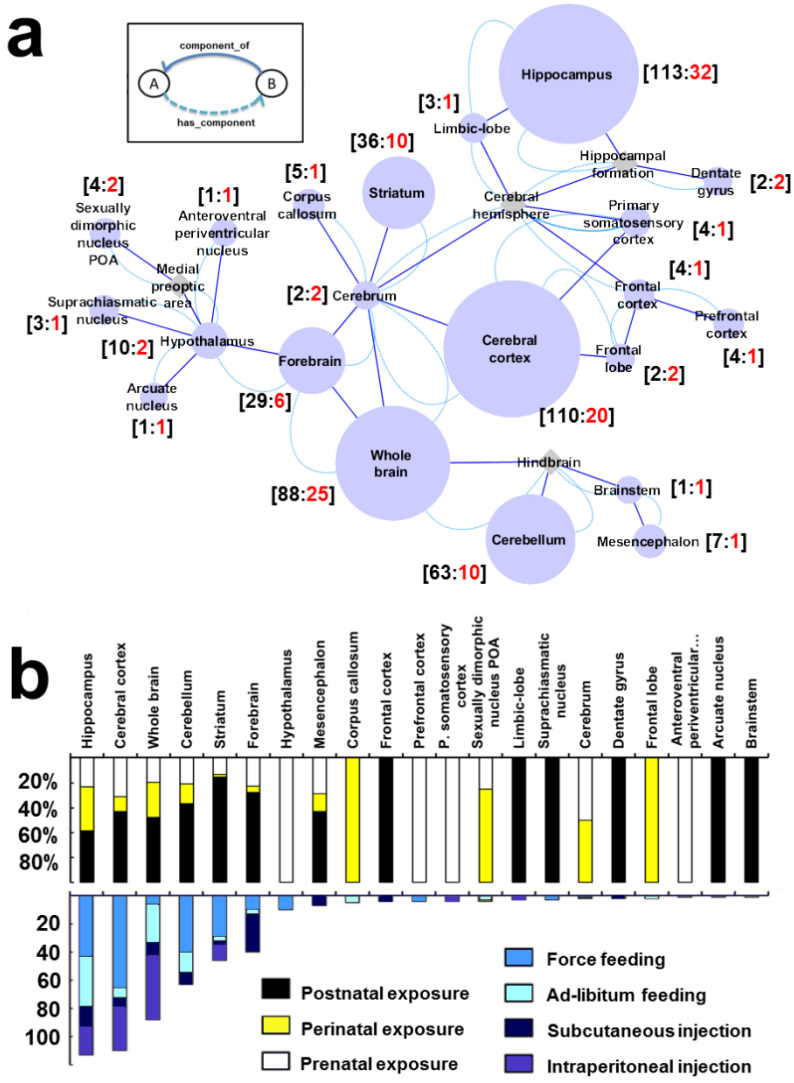


Figure 75 An overview of the relation between terms used to represent brain regions in the data set. a) Brain region terminology visualized as semantic network based on the Foundational Model for Anatomy (FMA) and the Neurolex terminology. The most

commonly investigated brain regions are the *Hippocampus*, '*Cerebral cortex'*, '*Whole brain*', and *Cerebellum*. Holonym and meronym relations are represented by 'component_of' (dark blue solid line) and has_component (light blue dotted line). The size of the nodes reflects the number of OA-relations in the data set. Number of OA-relations (black colour) and corresponding number of unique publications (red colour) are shown within square brackets. b) Brain region terms listed in descending order of OA-relations from left to right. Histograms below show the percent of exposure durations (Prenatal, white colour; Perinatal, yellow colour; Postnatal, black colour) and total numbers of types of exposure (Force feeding, blue colour, Ad-libitum intake of toxicant, light blue; Subcutaneous injection, dark blue; Intraperitoneal injection, purple colour).

Length of exposure and developmental period

In order to make the annotation of OA-relations and publications more manageable, we divided them into three different exposure paradigms. Prenatal exposure is here defined as exposure before birth (i.e. exposure between gestational day (GD) 0 and GD 19 (mouse) or between GD0 and GD21 (rat). A total of 24 publications (142 OA-relations) used prenatal exposure. Perinatal exposure is here defined as exposure both during gestation and postnatal.

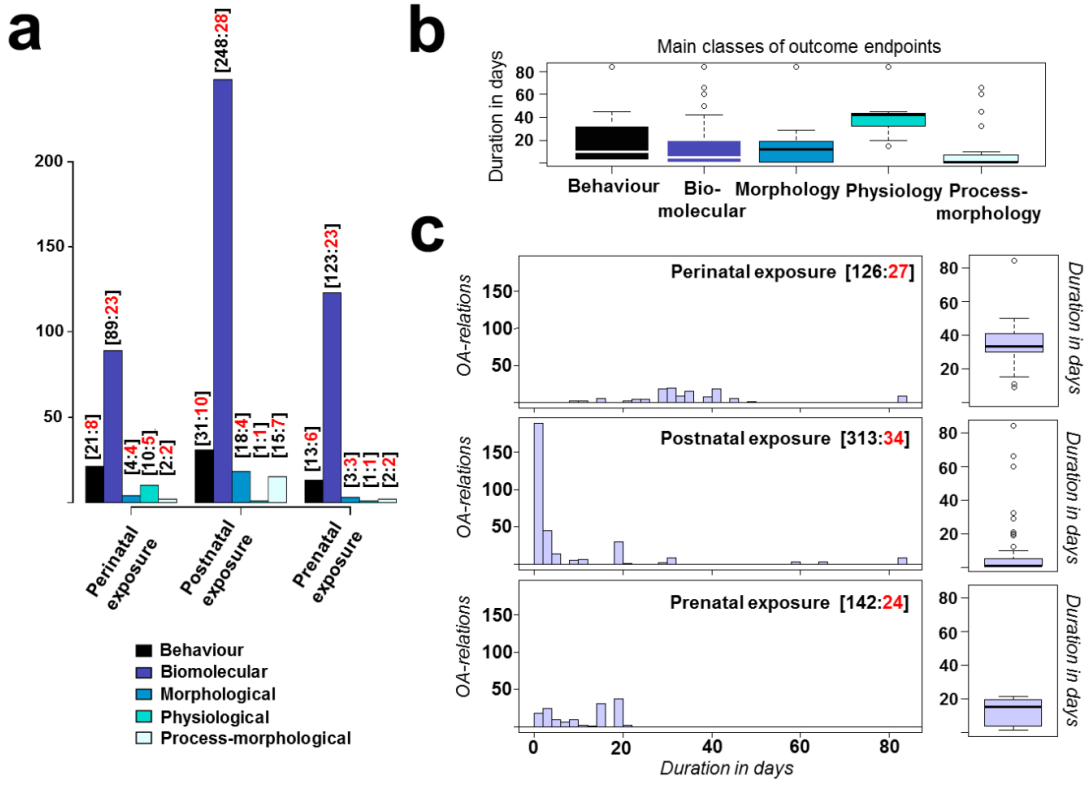


Figure 76 An overview of the exposure durations in the data set. The overall duration length for each OA-relation is calculated by converting the gestation period to n=19 in mice and n=21 in rats, converting the reported days of exposure initiation and termination into this scale, and subsequently subtracting the day of initiation from the day of termination. a) Bar plots with the OA-relation proportions of different outcome endpoint classes in relation to the different exposure paradigms. b) A set of boxplots with the duration in days for the different main classes of outcome endpoints (Behaviour mean=17.3, median=10; Biomolecular mean=14.3, median=5; Morphology mean=12.2 median=12; Physiology mean=40.3 median=42, and Process-morphology mean=12.4 median=1). c) A set of histograms and corresponding boxplots for the different exposure paradigms. Perinatal exposure (OA-relations: 126, Publications: 27) is any exposure to an agent of toxicity that is initiated during the prenatal/embryogenesis period in mammals (i.e. in utero exposure) and terminated post-partum. Postnatal exposure (OA-relations: 313, Publications: 34) is any exposure to an agent of toxicity that is initiated post-partum but before weaning (nominally post-partum day 21 in rats and mice). Prenatal exposure (OA-relations: 142, Publications: 24) is any exposure to an agent of toxicity that is both initiated and terminated during the prenatal/embryogenesis period of mammals. Boxplots of the

spread in days of exposure duration are shown to the right of the histograms (Perinatal mean=36.4 median=33; Postnatal mean=8.1 median=1; Prenatal mean=11.2 median=14.5).

A total of 34 publications (313 OA-relations) used perinatal exposure. Postnatal exposure is defined as exposure from birth and on. A total of 25 publications (142 OA relations) used postnatal exposure. 7 publications used more than on exposure paradigm (i.e. different sets of animals were used in the publication). For the postnatal exposure paradigm, the acute (24h or shorter) exposure was dominant with a mean exposure length of 8.1 days and a median exposure length of 1 day. Prenatal exposure is by definition up to 21 days (for rat), but there is a quite even distribution of exposure length across the publications. As a result, the mean and median values are quite close (mean=11.2; median 14.5 days). Perinatal exposure has a mean exposure length of 40 days, which may be explained by use of the suggested OECD guideline exposure paradigm from GD7 to postnatal day (PND) 21 (Figure 76).

Classification of OA-relation endpoints

The terminology representing outcome endpoints was formalized and categorized into five classes, four phenotypic classes and one biomolecular class. The four phenotypic classes are morphological (structure; size and shape of cells, tissues, organs and bodies), physiological ('structure over time' endpoints such as cellular and organ functionality and electrophysiology), behaviour ('body over time' such as cognitive behaviour) and process-morphology (outcome endpoints representing changes in morphology and physiology over time such as developmental processes and cellular processes). The fifth outcome class represents empirical measurements of specific biomolecular targets alternatively multiplex characterizations of full biomolecular complements ('omics' measurements). Outcome endpoints representing a mix of biomolecular and morphological outcome such as those observed and/or measured with immunohistochemistry methods were considered primarily biomolecular as the antibodies used target specific biomolecular epitopes. When needed, the outcome endpoints were classified across several hierarchical levels with increasing level of detail for each level. The full extent of the classifications can be seen in the data collection sheet (Appendix K). An example for a behaviour endpoint would be the relation between sensorimotor behaviour and the use of the Rotarod test (Figure 77).

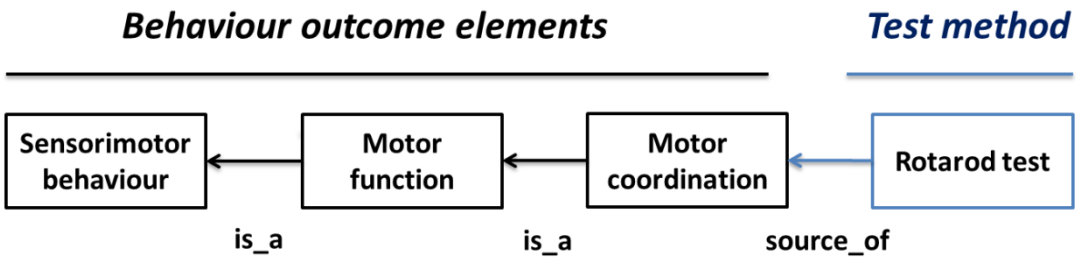


Figure 77 Example of relation between behaviour outcome endpoints at different levels of resolution and the test method used to observe and/or measure the outcome.

Using the data in these five principal outcome endpoint classes, we asked how the endpoint levels divided across the three exposure regimes. It is clear that biomolecular endpoints make up the vast majority of endpoints across all three exposure periods. For perinatal exposure, biomolecular OAs make up 70% of the total, whereas the corresponding figures for postnatal and prenatal are 79% and 86% respectively. A

considerable number of these OAs are mRNA and polypeptide level measurements. Almost no morphological endpoints are found in the prenatally exposed group (Figure 76a).

Behaviour endpoints

The behaviour outcome endpoints were divided into four main classes: autonomic-affective (behaviour patterns based on the organismal self-regulation, feeling or emotion), cognitive (behaviour patterns related to the expression of knowledge, attention, memory, evaluation, reasoning, problem solving and decision making), reproductive (behaviour patterns considered to be sex-specific) and sensorimotor (behaviour patterns representing spontaneous activity and motoric and sensory functionality). The behaviour-associated OA-relations were annotated into the different classes based on the primary purpose of the reported behaviour tests. A total of 26 papers (66 OA-relations) were included that were classified as behaviour. 9% of these were autonomic-affective, 30% were cognitive, 3% were reproductive and 58% were sensorimotor (Figure 78). Sensorimotor behaviour endpoints were the overall more well-studied endpoints followed by cognitive behaviour endpoints.

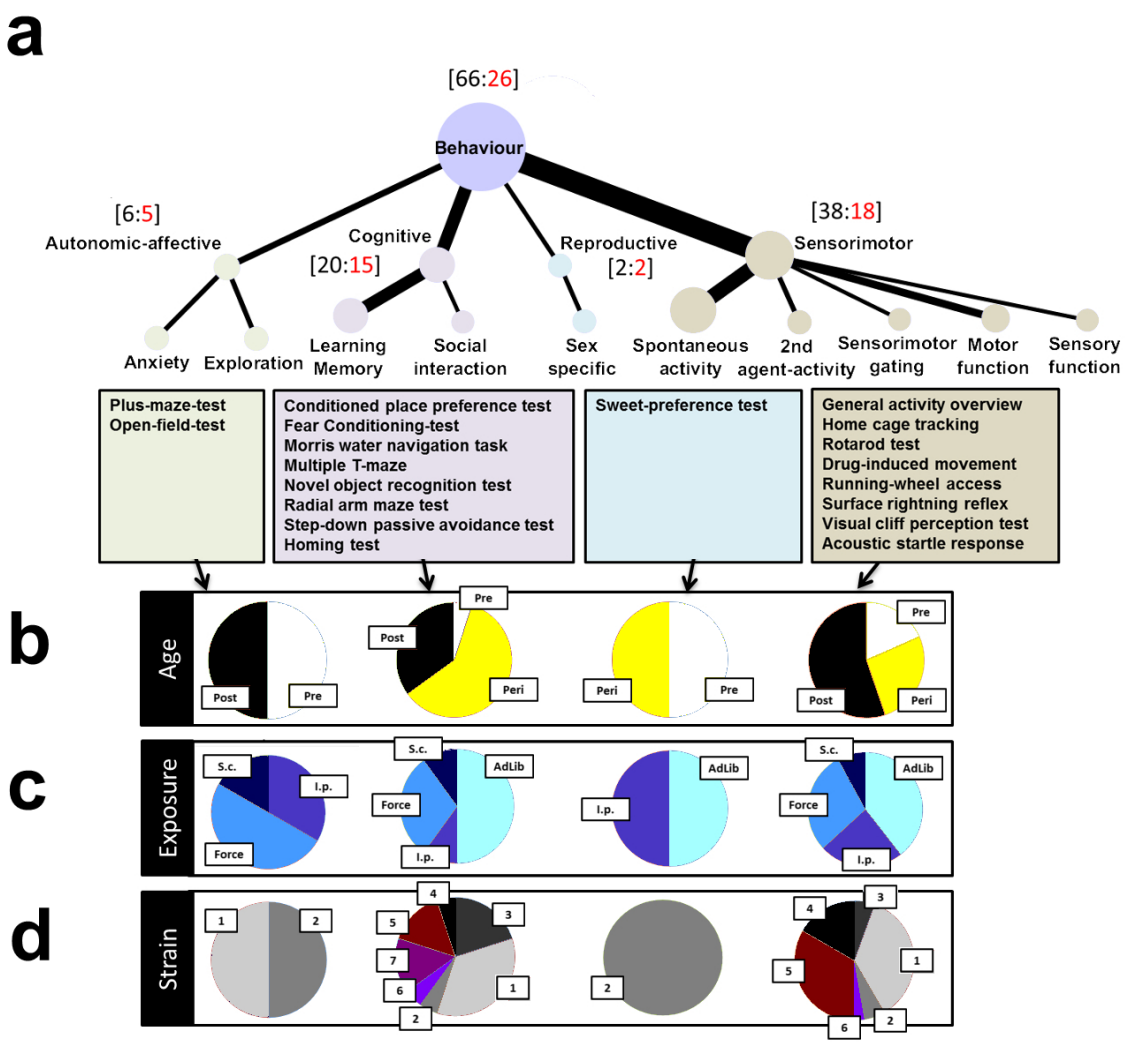


Figure 78 An overview of the subset of the DNT *in vivo* data set that uses behaviour observations as outcome endpoints. a) The behaviour studies were classified into four main categories: *Autonomic-affective* (OA-relations: 6, Publications: 5), *Cognitive* (OA-relations: 20, Publications: 15), *Reproductive* (OA-relations: 2, Publications: 2), and *Sensorimotor* (OA-relations: 38, Publications: 18). Each main class had several sub-classes (*Anxiety*, *Exploration*, *Leaning & Memory*, *Social interaction*, *Sex-specific*, *Spontaneous activity*, *Secondary agent induced motor activity*, *Sensorimotor gating*, *Motor function* and *Sensory function*) which in turn were connected to the main purposes of the behaviour tests mentioned in the behaviour data subset (named in the boxes). The relations are visualized as a network with the node size and edge width representing the number of OA-relations for the mentioned/studied behaviour types. Number of OA-relations (black colour) and corresponding number of unique publications (red colour) are shown within brackets. b) Pie charts of the OA-relations from the different behaviour main classes with their respective proportion of the three different exposure-duration types (Age: Prenatal, white colour; Perinatal, yellow colour; Postnatal, black colour). c) Pie charts of the OA-relations from the different behaviour main classes with their respective proportion of the four different administration types (Exposure: Force feeding, blue colour, Ad-libitum intake of toxicant, light blue; S.c. or subcutaneous injection, dark blue; I.p. or intraperitoneal injection, purple colour). d) Pie charts of the OA-relations from the different behaviour main classes with their respective proportion of the different rat and mouse strains and stocks (Strain 1: Sprague Dawley rats, 2:Wistar rats, 3: Long-Evans rats, 4: C57Bl/6J mice, 5: NMRI mice, 6: ddY mice, and 7: ICR mice).

On a higher resolution level, the top behaviour class were learning and memory followed by spontaneous behaviour. Only two studies are classified as reproductive behaviour (using a test method for determining sweet preference) which does not allow one discern any general trends. A range of different rat and

mouse strains was used to observe and measure sensorimotor and cognitive outcomes, the most common in both cases being Sprague Dawley rats (Figure 78b-d). For sensorimotor studies (n=38 OA-relations) NMRI-mice were second most common animal model after Sprague-Dawley rats. A proportional difference in exposure period settings is seen between these two main behaviour outcome classes. Cognitive outcome OA-relations were dominated by perinatal exposure with very little prenatal exposure while sensorimotor behaviour OA-relations had a greater ratio of postnatal exposures. Across all annotated behaviour outcome classes, seven only have one OA-relation (Figure 79a). No trend with regard to exposure paradigm and endpoint level (EL) can be seen.

Biomolecular endpoints

The biomolecular group of OA-relations is by far the most extensive. It includes 460 OA-relations of which polypeptide expression levels is the largest of the 31 groups, followed by message-RNA expression level (Figure 79b). It is not surprising that these groups have many OAs, since western blots and PCRs are often performed in assay designs with multiple targets. Measurements of monoaminergic signalling is also a relatively large group, but is only performed in the pre- and postnatal exposure regimens in this dataset. If we take all the individual gene products (in the widest sense, including different RNA types, oligopeptides and polypeptides (proteins)) and rank them according to number of OA-relations, we see that Sp1 (n=12), GFAP (n=10), BDNF (n=10) and CaMKII (n=10) are the top four gene products (Figure 80a). If we instead look for persistency across papers (Figure 80b), GFAP and Gap43 are found in 7 publications each. If we were to plot the gene data against individual authors, we would see that only a few research groups have published the bulk of these papers (data not shown). This trend would be even more apparent if we would look to the omics studies in this way. The green bars in Figure 80a show the relatively few omics papers that have been included in the dataset. Although there are more omics publications available, many of them have used less than three exposure groups due to the resource intensive nature of these techniques.

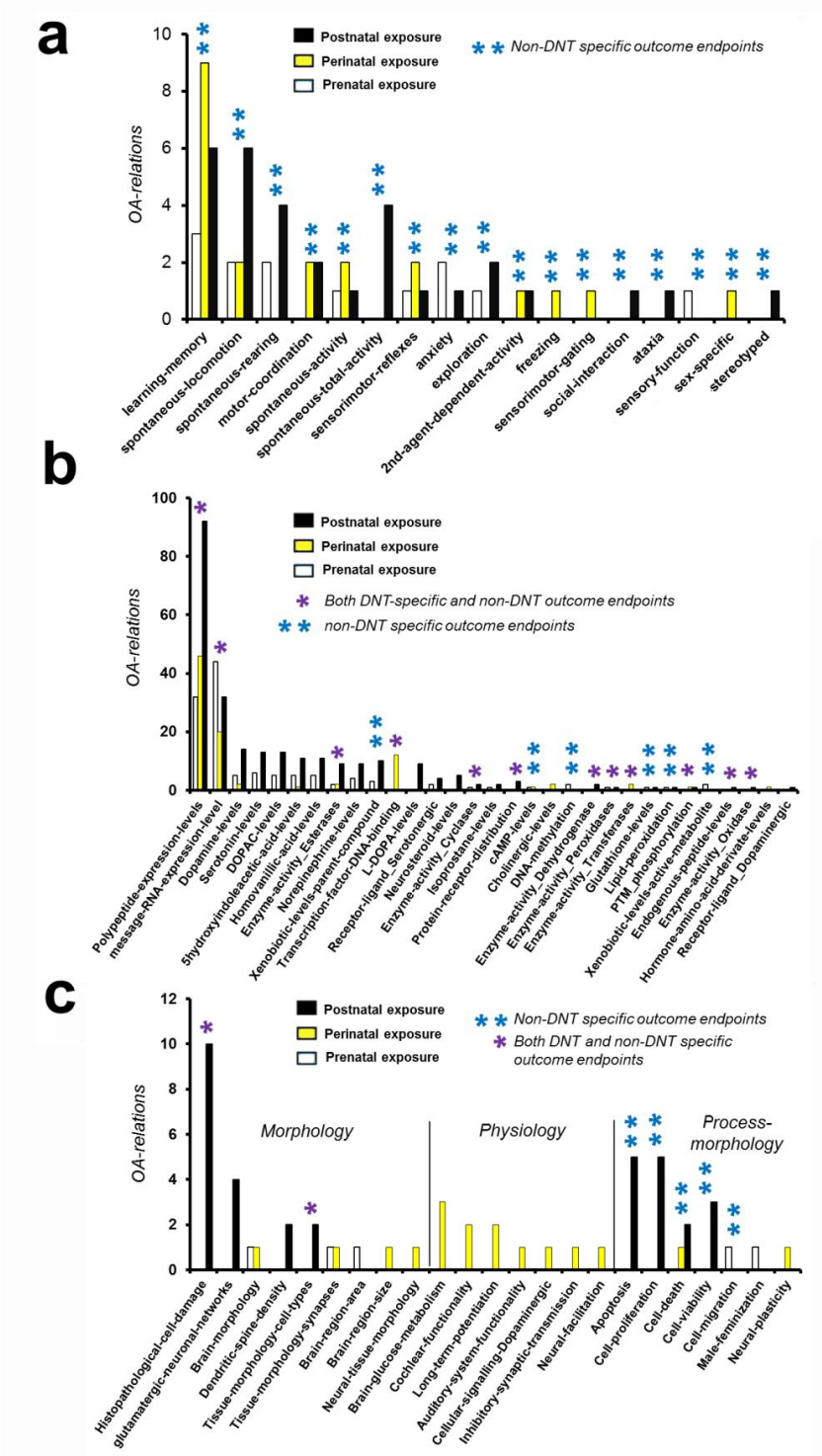


Figure 79 An more in depth overview of the different outcome endpoints in relation to exposure paradigms and to what extent the outcome sub-classes can be considered to be DNT outcome specific. a) Ranked histogram with OA-relations for behaviour outcome endpoint subclasses with the most commonly used endpoint to the left. b) Ranked histogram with OA-relations for biomolecular outcome endpoint subclasses with the most commonly used endpoint to the left. Behaviour sub-classes are considered to be DNT non-specific (two blue asterisks). Subclasses such as 'polypeptide expression levels'(protein expression) or 'messenger-RNA expression level' (mRNA expression) may represent measurements of gene products that are both specific and non-specific for DNT (single purple asterisk). c) Ranked histogram with OA-relations for morphology, physiology and process-morphology endpoints in descending order.

Morphology, physiology and process-morphology endpoints

These outcome endpoint classes all have very few OAs (n<10). For morphological endpoints, this is partly dependent on that immunohistochemistry OA-relations were considered to have biomolecular outcomes rather than morphological outcomes. Histopathological cell death is the largest class in the morphology endpoints. This includes necrosis and apoptosis measured by both various fluorescent stains (e.g. Fluoro-jade and TUNEL stains) used for microscopy evaluations. Histopathological cell-death OA-relations are only found in publications using postnatal exposure. They are with one exception (brain glucose metabolism) covering aspects of electrophysiology. All of them are assayed in animals receiving perinatal exposure.

Compound-specific evaluations

In the present project, a total of 12 DNT-positive compounds have been identified and evaluated in the included publications (Figure 81a). If we look at the number of OA-relations linked to individual compounds, we see that the top-ranked substance is BDE-47 with a total of 63 OA-relations (Figure 81a). However, all these OAs come from the same publication (Figure 80b). Bisphenol-A (BPA) is the compound used in the most publications (n=9), which is surprising considering that it is (DNT publication-wise) a relatively new compound; on the other hand BPA is a compound which has received much attention due to scientific as well as policy debate. Other top-ranked compounds such as lead, ethanol and methyl mercury are well-established developmental neurotoxicants and have been so for decades. 10 compounds have been used in more than one publication, and of these only four (ethanol, lead, methyl mercury and manganese) are considered to be DNT-positive compounds (ref). The candidates (according to the criteria set above) emerging from the five substances where we have the most publications included in the dataset is found in Figure 81.

When we sorted the biomolecular outcome endpoints after their developmental exposure settings (period of exposure and compound), we could see that 14 publications used postnatal exposure, whereas only 4 used prenatal exposure (Figure 82a). 7 compounds were used in publications with more than one period of exposure and 7 compounds were used in perinatal exposure settings. From the compounds in Figure 82a it is not possible to discern any trend regarding chemical type and period of exposure. Bisphenol-A was used in publications with 5 different behaviour study types using prenatal, perinatal as well as postnatal exposures (Figure 82b). Biomolecular endpoints measured for BPA include Gap43, CamKII and ERβ. For ethanol, we did not include any publications that assayed both behaviour and molecular endpoints, and no publications with perinatal exposure was included (Figure 82c). Publications with 8 different behaviour paradigms were included for chlorpyrifos (Figure 82d). None of the included lead publications used prenatal exposure, but used 3 different behavioural paradigms (Figure 82e). The included MeHg-studies only assayed behaviour with Morris water navigation task but had studies that included all three exposure paradigms (Figure 82f).

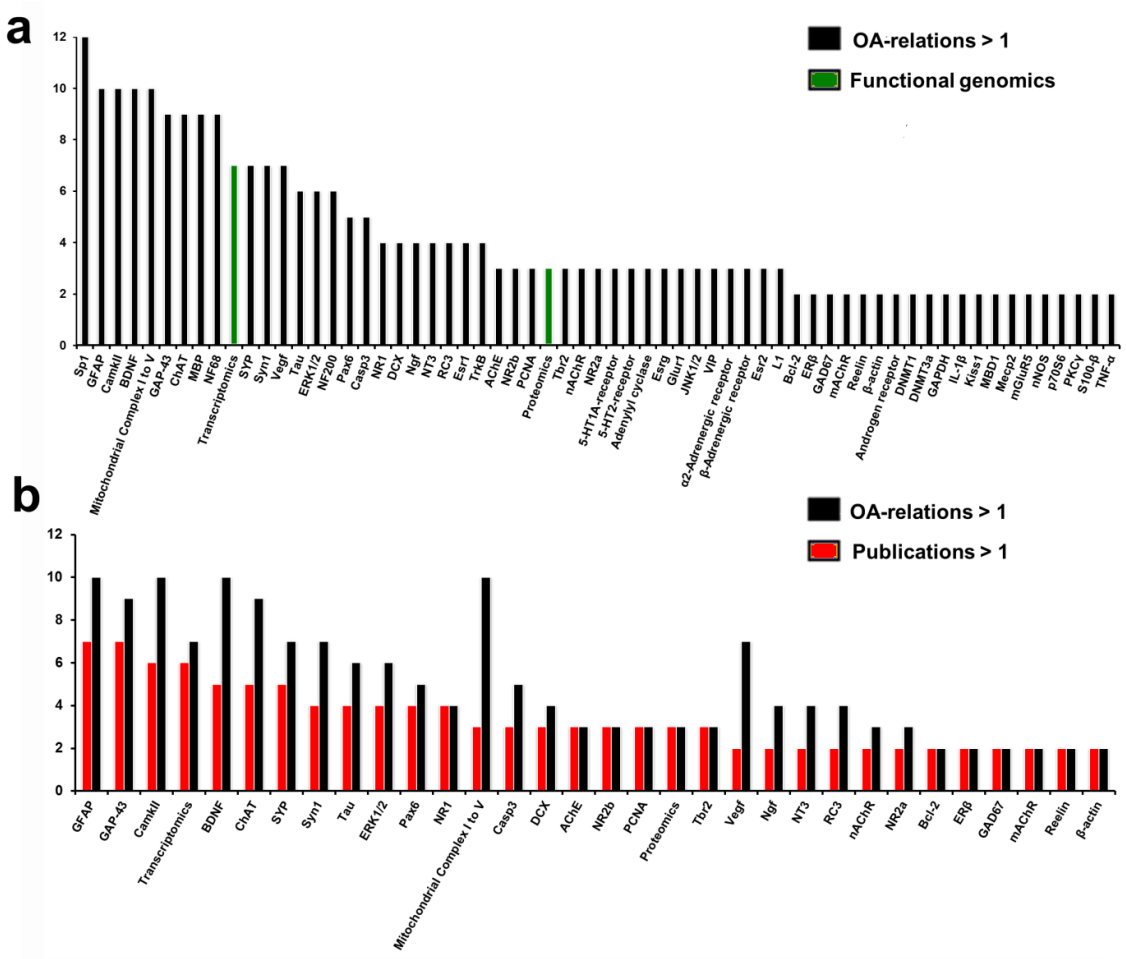


Figure 80 Biomolecular gene product endpoints in the data set. a) Biomolecular endpoints in descending order with the most common OA-relation endpoint to the left. The histogram includes all endpoints with OA-relations n>1. Functional genomics endpoints are coloured green. The gene products are used as targets in different test methods (e.g. protein expression of a given gene product may be measured both with immunoblotting, immunohistochemistry, radiolabelling of ligand-receptor interactions followed by scintillation analysis). b) Biomolecular outcome endpoints ranking based on number of unique publications (red colour bars). The histogram includes endpoints with Publications n>1 and OA-relations n>1.

The OA-relations classified as physiology endpoints are quite few (n=11, Figure 79). They are also quite limited in that they with one exception (brain glucose metabolism) are covering aspects of electrophysiology. All the included publications in this class use studies that have been exposed perinatally (Figure 78c). Process-morphology endpoints are representing changes in morphology and physiology over time, such as developmental processes and cellular processes. A total of 19 such processes have been classified in our publications. They include apoptosis (n=4), cell proliferation (n=4), cell viability (n=4), cell-death (n=3). Cell migration, male feminization and neural plasticity all have only one OA-relation (i.e. only one publication) in this dataset.

Omics publications

Only 7 transcriptomics publications and 3 proteomics publications are included in the dataset (Figure 80a). Although there are more omics publications available, very few have passed the (at least for omics publications) quite strict criterion of three dose groups plus a control group. In the past very few groups used their resources on several dose groups in their omics studies, and probably chose to include more replicates instead. The oldest omics-publication included is from 2005, which underlines the fact that really useful studies with techniques for genome- and proteome-wide studies have only started to emerge in the DNT literature. Consequently, to date, functional genomics outcome publications of DNT are too few to fully evaluate their relevance in a dose-response setting – both as a means for *in vivo* regulatory studies and as a source of candidates for *in vitro* test system development.

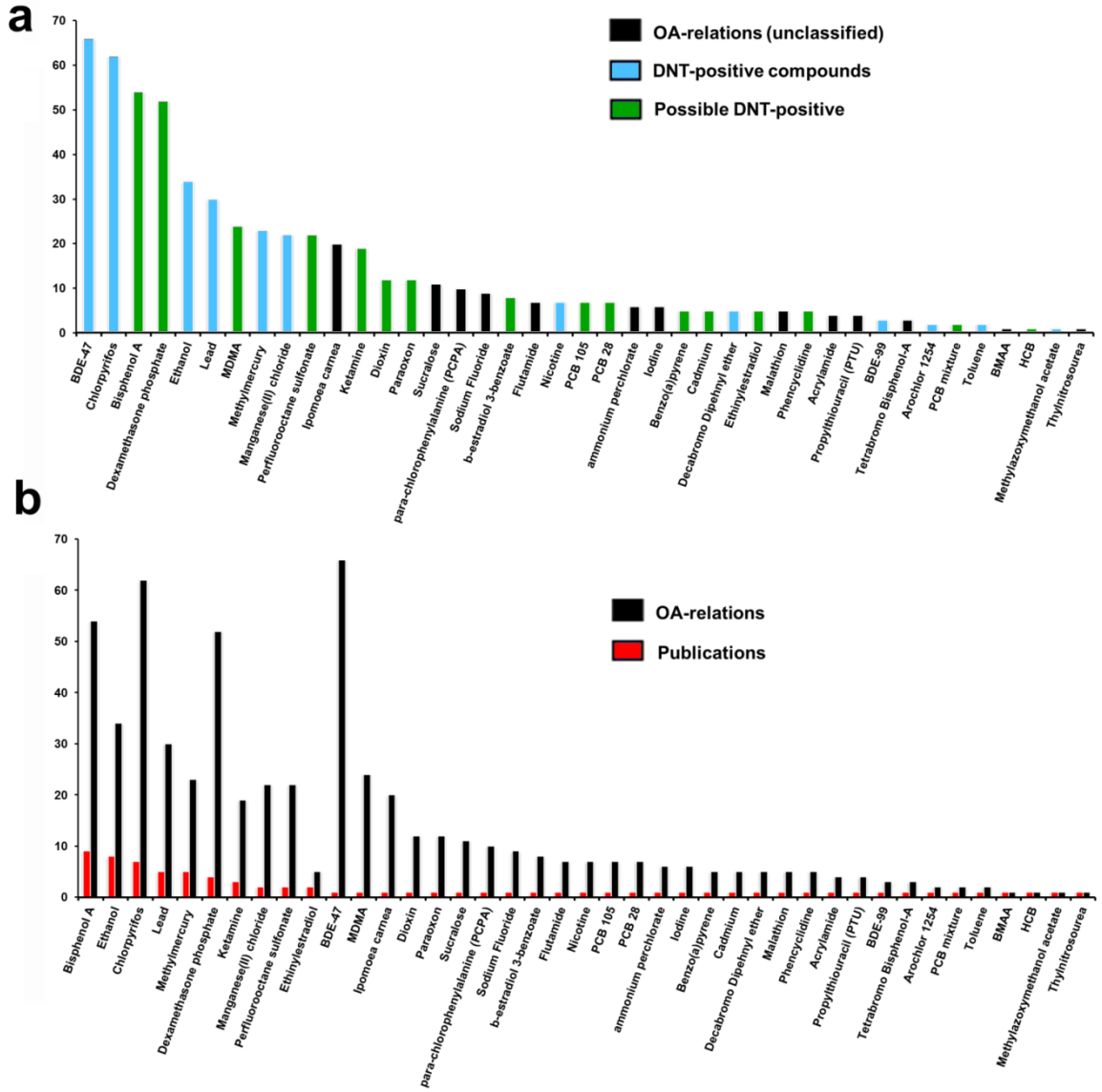


Figure 81 The number of compounds in the DNT *in vivo* data set. a) Compounds with the most OA-relations to the left. Compounds that have been classified as known DNT-positive are coloured light blue and possible DNT-positive compounds are coloured green. b) Compounds are ranked according to presence in unique publications (red coloured bars).

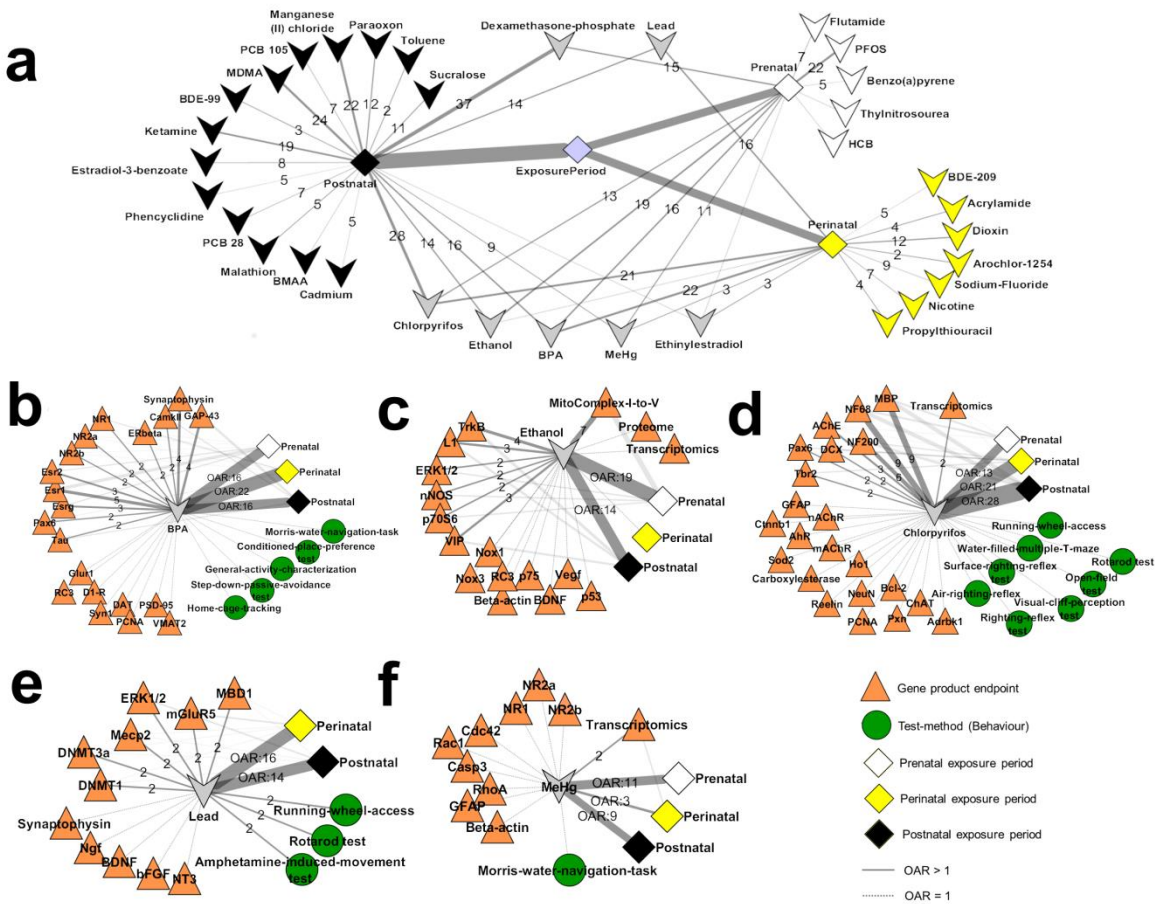


Figure 82 The number of compounds associated with biomolecular endpoints in the DNT *in vivo* data set. a) Compounds with biomolecular outcome endpoints sorted after their developmental exposure settings (prenatal exposure, white nodes; perinatal exposure, yellow nodes, and postnatal exposure, black nodes). Compounds tested in more than one exposure setting are in grey colour. The edges between nodes represent the total number of OA-relations for each given binary relation. b-f) The five top compounds in the *in vivo* data set that occur in more than one unique publication (Bisphenol A or BPA, Ethanol, Chlorpyrifos, Lead, and Methyl mercury or MeHg). The measured biomolecular endpoints (orange nodes), exposure setting (white, yellow and black nodes) and any behaviour tests (green nodes) for each compound are shown in the form of networks. Edges represent the number of OA-relations (solid edges have OA-relations >1 and dotted-edges have only one OA-relation). The edge labels represent the number of OA-relations for solid edges.

Endpoints for further evaluation

One of the strengths with the dataset in this publication is that all of the included publications have at least three dose groups, and a control. Thus, if we use the publications that have multiple OA-relations, we can compare and deduce which endpoints are most sensitive in the dose-response relation for a given design. The question is which endpoint to use as "template" or denominator in such a relation. *In vitro* studies may well use viability to compare and see whether the endpoint evaluated is more or less sensitive for the exposure. The rationale for this would then be that any effect that is less sensitive than viability is not specific for the endpoint evaluated. Although this approach may have its merits *in vitro*, it is not feasible for *in vivo* studies, partly because the relation between *in vivo* endpoints and cell viability is less obvious, and consequently viability is not assayed to the extent needed for such analyses.

A more suitable approach for *in vivo* endpoints would be to select the papers where behaviour studies have been used in conjunction with one of the other three phenotypic classes or the biomolecular class for the same experimental design. If we combine this approach with exposure agent used, we can get a shortlist of (1) the endpoints used in (2) dose-response studies with (3) a positive DNT agent during (4) a sensitive developmental period that is (5) combined with a positive effect on behaviour. This would generate a short-listing of candidates for further *in vivo* and *in vitro* development and refinement. A total of 12 DNT-positive compounds have been used in the included publications (Figure 81a). If we look at the number of OA-relations linked to individual compounds, we see that the top-ranked substance is BDE-47 with a total of 63 OA-relations (Figure 81a). However, all these OAs come from the same publication (Figure 81b). Bisphenol-A (BPA) is the compound used in most publications (n=9), which is surprising considering that it is (DNT publication-wise) a relatively new compound. Other top-ranked compounds such as lead, ethanol and methyl mercury are well-established developmental neurotoxicants and have been so for decades. Ten compounds have been used in more than one publication, and of these only four (ethanol, lead, methyl mercury and manganese) are considered to be DNT-positive compounds. The candidates (according to the criteria set above) emerging from the five substances where we have the most publications included in the dataset are found in Figure 82.

There is large variation in chosen biomolecular endpoints both across individual compounds and different exposure settings. Due to aspects of differences in animal species, strains and stocks and the impact of developmental sensitivity periods on the final outcome in the OA-relation, it is difficult to discern which endpoints are most promising for further *in vivo* and *in vitro* test development. Different exposure paradigms are likely to generate different outcomes and only a minority of compounds (Figure 82a) have been reported for all three exposure settings. The molecular endpoints outlined in Figure 82b-f may serve as potential starting points for further analyses, but limiting the studies to just singular molecules will not likely be enough to complement the guideline studies. Rather, a concerted effort of comparative *in vivo*-*in vitro* studies which also includes adjusting the guideline animal studies is likely needed.

3.4. In silico

Chemical substances have to pass two barriers before they can reach the developing brain: the placental barrier and the blood-brain barrier of the developing foetus. Substances can only exert DNT effects if they are able to pass these barriers. Therefore, the permeation through the blood brain or the placental barrier and the possibility to monitor this permeation *in silico* was one subject in this review and was covered in the *in silico* section.

The selection procedure described in the methodology section, resulted in 224 publications. The title-abstract screening of these publications resulted in only two publications and one review paper on QSAR for placental barrier permeation. The two publications were selected for the data collection sheet (record number 448 and 162), while the review paper was selected as grey literature (record number grey lit 6). A large number of publications on the estimation of blood-brain barrier permeation were found, but were excluded from the data collection sheet because they do not mention the early life stages of the developing brain. Passage of biological membranes relevant for the early life stage was an inclusion criterion during the title-abstract screening. Results of research on the adult brain barrier may not be protective enough for the foetal brain, as there is physiological evidence that the transport systems are functionally more active in the developing brain (Ek et al., 2012).

The existing QSARs relevant for DNT are based on the physical-chemical and structural properties of the substance enabling placental transfer, and not on the DNT effect itself. The existing QSARs available in the two publications are developed from data on drugs.

If a substance is not able to cross the placental barrier, it will not be a DNT substance. With this concept in mind, (Hewitt et al., 2007) developed five QSAR screening models for placental transfer from five different data sets of drugs. Placental transfer is expressed as clearance index (dataset 1, 4 and 5) or a transfer index (dataset 2 and 3), i.e. the ratio between the clearance (resp. transfer) of a test substance and the reference substance antipyrine. Antipyrine is a small lipophilic compound that is known to be transported across the placenta via **passive diffusion** and is not bound to protein. The ability of the five models to predict placental transfer ranges from being very satisfactory for data set 5 with data from one source and for one particular therapeutic use ($r^2 = 0.858$, $Q^2 = 0.810$) to relatively poor for data set 1 with data from different sources and on heterogeneous compounds ($r^2 = 0.635$, $Q^2 = 0.567$). The models based on a heterogeneous set of compounds (data set 1, 2 and 3) lead to more robust models with extended applicability domain, compared to models based on structurally homogeneous substances for specific therapeutic uses (data set 4 and 5). Hence the consideration was put forward by the authors whether it is recommended to build several models with a high performance for a narrow group of substances or one model with a wider applicability domain but with a lower performance.

Analysis of the model equations indicate that the descriptors generally relate to hydrophobicity, hydrogen bonding ability, and molecular size. Placental transfer is positively correlated with lipophilicity, expressed by the octanol-water partition coefficient of compounds. If the substance is highly hydrophilic (log Pow ≤ 0.5 according to (Giaginis et al., 2011) it may not enter the phospholipid structure of the placental membrane. Hydrogen bonding is also an important factor in predicting placental transfer. If a drug is able to form easily hydrogen bonds in water, this will increase its water solubility hence possibly reducing membrane penetration. Increasing mass is associated with decreased transfer rate. This is consistent with experimental work, showing that molecules with MW > 500 Da may have reduced transport across the placental barrier, their clearance index is <0.1, reflecting incomplete placental transfer (Pacifici, 2005). The models developed by Hewitt et al. (2007) do not consider active transport. It is known that substances may cross the placental barrier by active transport. Hence it is acknowledged by the authors that the real placental transfer may be underestimated by models considering only passive diffusion. Potential metabolisation before transfer is also not considered. . In case a substance is metabolised before transfer, the model is making estimations on the wrong substance. However this concern may be overcome by using models on metabolism (e.g. TIME). Details on algorithms and statistics are provided in Appendix K (data collection sheet '*in silico*'). The models comply with the five OECD principles for QSARs for regulatory use.

The QSAR model developed by Hewitt et al. (2007) from dataset 1 was validated with an original test set of 76 pesticides and an extended test set of 134 pesticides (Worth A., 2011). All substances had experimental results on developmental toxicity. Of the original test set 37 substances tested positive for developmental toxicity and 39 substances tested negative. The original test set was extended with 59 substances that tested positive. The average clearance predicted by the QSAR model was 0.66 for the positives and 0.70 for the negatives (Table 51). Since the average clearance for the positives was about the same as the clearance for the negatives, it seems that the model was not able to discriminate between pesticides that tested positive or negative for developmental toxicity. However, it must be indicated that the model was not developed specifically for developmental toxicity, and that it was developed from data on drugs.

Table 51 Distribution of predicted placental clearance values for developmental toxicants and non-developmental toxicants (from Worth, 2011)

Clearance statistics	Original test set		Extended test set	
	Positives (37)	Negatives (39)	Positives (95)	Negatives (39)
Average	0.66	0.70	0.66	0.70

Maximum	1.39	3.30	2.56	3.30
Minimum	0.15	0.14	-0.43	0.14
Standard deviation	0.25	0.54	0.39	0.54

The clearance index was also used in a multivariate data analysis of a set of 84 structurally diverse drugs and chemicals in order to model placental transfer and to identify those descriptors that had a relevant impact on the transfer (Giaginis et al., 2009). A robust QSAR model ($r^2 = 0.73$, $q^2 = 0.71$) based on 16 original descriptors (out of 82 descriptors) was developed. The descriptors reflecting the polarity (e.g. polar surface area, hydrogen bonding capability, dipole moment and number of ionizable groups) of the compounds proved to be the most important with negative correlation. Lipophilicity and to a lesser extent molecular size exerted a positive contribution to the predicted permeation. Details on statistics are provided in Appendix K(data collection sheet *in silico*). The model complies with the five OECD principles for QSARs for regulatory use.

An *in silico* model on the prediction of Adsorption, Distribution, Metabolism and Excretion (ADME) processes such as passive biomembrane penetration has been reported (Mälkiä, 2004), but it remains unclear to which extent the placental membrane is comparable to other biomembranes with regard to permeation.

Recently, another approach was proposed by the POEMS Network Maths Study Group of the National Centre for the Replacement, Refinement & Reduction of Animals in Research, who aims to apply mathematics to 3R problems (Currie, 2014). The approach is based on the knowledge that thyroid hormones are essential for the control of metabolism and development, especially nervous system development. Altered thyroid hormone levels during critical periods of development might result in adverse outcome in the developing foetus. The proposal of the group is to develop a mathematical model to determine whether the no-effect level of drugs and chemicals on thyroid hormones in male rats are suitable protective for human foetus and neonates.

QSAR models are being developed for various endpoints of physical-chemical properties human toxicity, ecotoxicity and environmental fate. Therefore, we explored a number of websites providing free or commercial QSAR models, listed in Appendix D. Of the 22 websites on QSAR models or on experimental toxicity data, we explored and contacted, none contained a model specific for DNT or placental barrier permeation. The main reason for the lack of a model for DNT is that building a model requires a large amount of measured data, which are not available yet for DNT of placental barrier permeation (Crofton et al., 2012).

Besides using QSAR models, another technique for filling data gaps is read-across. With read-across, the endpoint information for one or more source chemicals, is used to make a prediction of the endpoint for the target chemical. Read-across is based on the identification of similar compounds (ECHA, 2008b). Within the grey literature search, five web sites providing a read-across tool were found that could be used to identify structurally related substances (record number grey lit 1 to grey lit 5). The tools are listed in Appendix D These read-across tools can be used to search for analogues of known DNT substances or to find out whether a substance has an analogue that exerts DNT properties. None of these tools is specific for the DNT endpoint or placental barrier permeation. The tools leave it to the user to determine when a specific analogue is suitable for DNT, as the determination of what structure is appropriate can vary depending on the endpoint assessed.

Besides a structural similarity, a common mode of action may also be a means of identifying groups of similar substances from which to build categories and support data bridging through read-across or other types of predictions. The approach of the adverse outcome pathway (AOP) is built on the mode of action. The QSAR toolbox has started to implement adverse outcome pathways, which may in the future enhance the applicability of the tool for grouping (potential) DNT substances based on similar mode of action. Initiatives such as the AOP-Wiki are welcomed and may in the future turn out to be valuable tools for the prediction of DNT properties.

Regarding ongoing new *in silico* developments for DNT or placental barrier permeation, e-mail answers from QSAR and/or DNT experts indicated that they were not aware of such recent developments (answers from M. Cronin, C. Giaginis and E. Benfenati). None of the model builders that provide their models via websites indicated to have an intention for developing a QSAR model for DNT effects or placental permeation.

4. Discussion and Recommendations

The paradigm for regulatory testing of developmental neurotoxicology (DNT) has been the subject of much scientific debate and has been reviewed by expert groups (Raffaele et al., 2010, Makris et al., 2009). On-going discussions concern matters such as whether or not the OECD Test Guideline (TG) 426 for DNT testing (OECD, 2007) is sensitive and/or reliable enough to serve as a basis for the risk assessment of DNT in humans, and if corresponding guidance documents (OECD, 2008, OECD, 2004) are detailed enough. The scientific background to these concerns stems from the growing public and societal concern, 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. Areas of possible improvements of the OECD TG 426 and the corresponding guidance documents in terms of their reliability and usability for scientific and regulatory judgements and decisions have been identified (Beronius A, 2013).

Support for causal relationships for DNT in humans is particularly strong for exposures to lead, methyl mercury, or polychlorinated biphenyls (PCBs) (Grandjean and Landrigan, 2006a). 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 demonstrate that DNT occur at doses that are lower than those which affect adult brain function. Despite strong links between DNT and human exposure to lead, methyl mercury, or PCBs, there are still knowledge gaps between molecular initiating events (MIE), cellular and organ effects and the adverse outcome (AO) in human DNT for any of these compounds (Bal-Price et al., 2015b).

Grandjean and Landrigan (2006a) 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 providing data on a spectrum of endpoints in order to improve the ability to detect the potential for DNT-properties.

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 as well as 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 OECD TG 426 for DNT testing. 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. The review recognized that the DNT guideline has 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. Overall, the review concluded that the OECD TG 426 represents the best available science for assessing the potential for DNT in human health risk assessment, that DNT-data generated by OECD TG 426 studies are relevant and reliable for this assessment, and that continued research is needed within this field (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. 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).

Based on DNT case studies for four chemicals Beronius et al. (2013) performed a detailed comparative investigation, which analysed how well these studies followed and reported on 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 that were raised during the compound evaluations. The study concluded that neurodevelopment and neurotoxicity are inherently very complex areas, and, in particular, there are massive gaps in knowledge about normal brain development on the functional, structural and molecular levels, which complicates both OECD TG 426 DNT testing and risk/safety evaluations.

However, a recent workshop report involving different stakeholders from regulatory agencies, industry and academia (International Stakeholder NETwork, ISTNET) for DNT states that clearly DNT constitutes an area in critical need of test method development (Crofton et al. ALTEX 2014; Bal-Price et al. Arch Toxicol 2015). The stakeholders also recognized that due to the resource-intensity of the current DNT OECD TG 426 and also for ethical reasons there is need of novel and alternative non-mammalian *in vivo* models as well as *in vitro* and *in silico* testing-models. Such models are expected to facilitate rapid and cost effective screening of large numbers of chemicals for their potential to cause DNT. They can also be used to clarify DNT-mechanisms and modes-of-action to provide information on human relevance. Alternative testing methods are also needed to complement or replace OECD TG 426 in order to fulfil present requirements of chemicals regulation, including tiered testing approaches to advice regulatory decision making processes (Bal-Price et al., 2015a, Crofton et al., 2011).

On this basis, this systematic review collected all data published on DNT effects of chemicals since 1990. After the first keyword search (for keywords see Appendix B) the search retrieved 11125 records distributed across DNT *in vivo* (5019), *in vitro* (2725), alternative organisms (2583) and *in silico* (798)

publications. Applying defined inclusion/exclusion criteria (2.2) diminished these publications to 452 selected publications (DNT *in vivo* 73, *in vitro* 288, alternative organisms 90 and *in silico* 2) included in the final evaluation of this review.

After the first keyword search the majority of publications came out to be rodent *in vivo* publications. However, the *in vivo* publications ended up being the minority across the three wet lab-based pillars of this systematic review. This selection was based on the criteria that only *in vivo* publications, which contained three dose groups in addition to the control, were included, and, in addition, a number of chemicals exposure categories were omitted. This decision was based on the positions that only a reasonable number of publications could be accommodated in the project and furthermore, that it is imperative to be able to model dose-dependent relations if we want to identify critical effects that in a meaningful way reflect DNT *in vivo*. An experimental design based on fewer dose groups (A vs B dosing for instance) will give very limited information about the endpoint under examination. Within the selected 78 *in vivo* publications, this review intended to identify new testing models and *in vivo* endpoints, which might be predictive for human DNT in addition to the OECD TG 426. The 78 publications (corresponding to 581 unique OA-relations) were included for full-text evaluation.

Rat models were used in 57 publications (covering 5 strains of rats) and 21 publications used mouse models (covering 8 strains of mice). Although we included publications from 1990 to May 2014, only five publications were older than year 2000, suggesting that our inclusion criteria were mainly met by newer publications. 21 brain parts have been assayed in the 78 publications, but 6 brain parts (forebrain, whole brain, hippocampus, striatum, cerebral cortex and cerebellum) make up 85% of the publications and hippocampus, striatum, cerebral cortex and cerebellum make up 57% of the publications. If this bias towards certain brain parts holds true for all available DNT research publications, it suggests that our current knowledge about DNT is based on a functionally very small part of the brain. Outcome endpoints were formalized and categorized into five classes, four phenotypic classes and one biomolecular class. The biomolecular group of OA-relations was by far the most extensive, and also includes the best available biomarker candidates which however need considerably more validation *in vivo* as well as otherwise. In general, the available published *in vivo* data is very scattered and heterogeneous, making comparisons across exposure paradigms and compounds very difficult. Thus, to make more useful evaluations, publications with more consistent experimental designs are needed, as well as more focused *in vivo-in vitro* comparisons.

Few animal based models alternative to OECD TG 426 have appeared in this *in vivo* data-set. One alternative DNT-model in mice that has been consistently published, using a growing set of chemical substances but mainly using the same design in terms of exposure setting, animal strain, length of exposure and evaluation characteristics is the neonatal mouse model focused around the brain growth spurt in mice around post-natal day 10, which has been devised by Per Eriksson and colleagues at Uppsala University, Sweden. Their mouse *in vivo* model may serve as an example of an alternative mammalian model focusing on the cholinergic system that can be further considered for evaluation as a complementary model to the methods specified in the current TG 426 guideline. Zebrafish represents a non-mammalian alternative *in vivo* model for OECD TG 426, which has been evaluated for its use during alternative life stages i.e. up to 5 days post fertilization (as embryo, larvae, embryo & larvae) in the present study. It can be expected that the increasing use of zebrafish, including the adult life stage in biology and medical research, will have an impact also on toxicity testing paradigms, so that adult zebrafish also may become an alternative *in vivo* DNT model for OECD TG 426.

The change in toxicological testing paradigms in general proposes chemical hazard assessment *in silico/in vitro/*with model organism for decision-making and/or prioritization for rodent *in vivo* testing (Collins et al. 2007). Therefore, besides supplementary rodent *in vivo* endpoints, DNT publications on *in silico/in vitro/*with model organisms were another focus of this systematic review. Because of the international consensus that chemical testing should move to human-based methods to minimize species effects in

responses to compound exposure (Krewski et al., 2010b, Seidle and Stephens, 2009a, Krewski et al., 2010a) one focus of this systematic review was put onto human stem-/progenitor cell-based methods, while for *in vivo* alternative organisms due to data-richness the zebrafish received the main attention.

Concerning the *in vitro* part, a large amount of very valuable DNT publications performed in primary rodent cells indicate that cells taken out of the *in vivo* context maintain certain cellular and molecular functions they hold *in vivo* by e.g. showing similar responses to xenobiotics (e.g. Wang et al. 2014, record number 366; Wang et al. 2013, record number 654; Yeo et al. 2013, record number 675; Bose et al. 2012, record number 967; Robinson et al. 2011; record number 1701; Hicks and Miller 2011; record number 1827; Burke et al. 2006, record number 3980; Turner et al. 2007, record number 4033; Wan Norhamidah et al. 2013, record number 10818; Go et al. 2012, record number 11518; Wayman et al. 2012a,b record numbers 11907, 11908; L'Episcopo et al., 2013). However, in some instances these *in vitro-in vivo* correlations failed on the level of toxicant-dependent protein expression and protein kinase activation (Alm et al. 2008, record number 3270; Cremin and Smith 2002, record number 20658) possibly due to responses of cells being present *in vivo* but not *in vitro* (e.g. astrocytes (record number 3270) endothelial, microglia cells) or due to the 2-dimensionality of cultures, which are known to change cellular behavior *in vitro* (Alepee et al., 2014). With the knowledge on species-specificities of cell and organ responses mainly from pharmaceutical exposures (rev. in Hanke, 2006, Leist and Hartung, 2013) and the experience from animal cells being useful when taken into culture as indicated above, the logical consequence would be *in vitro* testing on human primary cells. For compound testing, such cells should be in a multi cell type context, just as they appear in developing brains *in vivo*, because e.g. pure neuronal cultures might respond differently to chemical compounds than neuron-glia co-cultures (Giordano et al., 2008, Giordano et al., 2009; record numbers 3205 and 2964).

Human primary cells in form of cortical 3-dimensional (3D) hNPC have been commercially available for many years (supplier Lonza) and have been identified within this systematic review to measure the DNT-specific endpoints NPC proliferation, migration, neuron and glia differentiation (record numbers 64, 2682, 2205, 2313, 2686). In addition, Gulisano et al. (2009) established a primary long-term cell culture (FNC-B4) obtained from the human fetal olfactory neuroepithelium, which also showed effectively measuring NPC proliferation, cytoskeleton alterations, neuronal and glial differentiation (record number 3150).

Besides human primary cells, within the general scientific development of the stem cell field, differentiation of hESCs to neural cells has become more and more prominent in recent years. Thereby, hESCs have been successfully developed into neural crest cells (NCC; Zimmer et al., 2012)), which later in embryonic development give rise to neuroectodermal tissue including peripheral nerves and glia cells (Le Douarin et al., 2008). Moreover, different sources of hESC were differentiated towards neural rosettes with diverse differentiation protocols mimicking neural plate and neural tube formation (record numbers 878, 1118, 1532, 2963, 13548). Formation of neural plate and neural tube as well as NCC formation and migration are prerequisites for neural development and happen during embryogenesis in the first trimester of gestation. Differentiation towards the neuroepithelial lineage either through neural rosette formation or directly from hESC results in the generation of NS/PCs (record numbers 9, 660, 769, 796, 1077, 1176, 13548). To what extent the hESC-derived NS/PC represent primary hNS/PC is clearly something that **needs** to be investigated. NS/PC proliferation is an important neurodevelopmental process ultimately driving the pool of cells for neurogenesis and can be assessed with any neural stem-/progenitor cell-based system. From either primary hNPC or hESC-derived NS/PC, neural cells can be differentiated. Neuronal differentiation by loss of stem cell marker expression and gain of neuronal markers was shown by multiple studies on the mRNA or protein level (record numbers 769, 796, 2963, 10516, 1176, 13548). However, while some studies also show expression of some glial markers, mainly via RT-PCR, glial differentiation into radial glia, astrocytes and oligodendrocytes from hESC has not yet been studied intensively and **needs** more elaborate work. Glia differentiation, however, seems to be an earlier event in primary hNPCs than in hESC differentiation (record numbers 2313, 2682). While migration via the

neurosphere assay has been studied with primary hNPCs (2205, 2313, 2682), the assay has not yet been shown to function with hESC. In the primary neurosphere assay, migration is based on radial glia migration. If these cells also migrate primarily out of hESC-derived hNS/PCs, growing in 3D conditions (such as neurospheres) **needs** to be evaluated. However, (Hoelting et al., 2014) nicely showed that neurosphere generation in 3D is possible with such hESC-derived hNS/PCs (record number 796). NPC-formation and -proliferation, neural migration as well as neuronal and glia differentiation are necessary for neuronal subtype differentiation, synaptogenesis and network formation, which ultimately steers brain function. In this systematic review, only very few publications were identified measuring electrical network activity using rodent, mainly rat primary cells (record numbers 1901, 2021, 2319, 20775, 23876). However, although it seems that this important endpoint can also be measured with hESC-derived neuronal cultures (Kapucu et al., 2012, Yla-Outinen et al., 2014) it is necessary to establish standardized protocols for human-based network formation with NS/PC-based methods. Moreover, there is the urgent **need** to show that current human DNT compounds (Grandjean and Landrigan, 2014) are indeed able to disturb human network formation. Because the literature cited here for the hESC- and NS/PC-based assays were identified within this systematic review due to their abilities to evaluate developmental neurotoxicants, these citations are far from being comprehensive when looking at the cell methods within the entity of the scientific literature. However, method evaluation beyond DNT was beyond the scope of this report.

In addition to primary and hESC-based NS/PCs, this systematic review identified human umbilical cord blood (hUCB)-derived NS/PCs as additional DNT methods (record numbers 1064, 2712, 10075, 10516). These differentiated hUCB-NS/PC also express neural markers. However, there is no study with these cells using human DNT compounds, which were identified as 'True Positives' i.e. by distinguishing specific DNT effects from decreases in viability. More compounds **need** to be tested to assess performance of hUCB-NS/PC for DNT testing. Moreover, similarity to the human developing brain *in vivo* or primary hNPCs needs to be evaluated to unravel the developmental timing these cells belong to. This is especially important as hUCB-NS/PCs express the Arylhydrocarbon Receptor (AhR) and react towards exposure with AhR ligands with induction of AhR-dependent Cytochrome P450 gene expression. Primary hNPCs do not express the AhR (Gassmann et al. 2010, record number 2205) and thus represent lack of human fetal brain AhR expression *in vivo* (Jiang et al., 2010a). At which developmental stage human brain starts to express components of AhR signaling is to the best of our knowledge not known.

Cellular models or advanced cellular networks on a chip, even if based on human cells and in a 3D configuration are still artificial systems lacking the surrounding organism covering ADME properties. DNT testing is a major challenge due to the critical time windows, as well as the physiological and morphological complexity of the central and peripheral nervous system (CNS & PNS). DNT testing of unknown compounds with only '*in vitro'* methods may not be sufficient to replace current rodent *in vivo* testing, especially when (1) the initiating event or primary site of action is unclear or unknown, (2) multiple target sites are hit by chemical compounds, (3) sensitivity for chemicals is dependent of the time window of exposure during early brain development and (4) ADME properties significantly determine biological response.

With respect to the developing organism and functioning of the nervous system, one can argue that the whole is greater than the sum of its parts. Therefore, simple whole organism systems which enable assessment of integrative effects have been suggested to be ultimately powerful approaches, especially in the case of developmental neurotoxicity (Lein et al., 2007). To make DNT testing feasible by saving time and money as well as following the 3R principle the zebrafish model used at embryo and larvae stages up to 5 days post fertilization, for compliance with animal legislation (EU Commission, 2012) is a suitable organism, with added value to other alternatives (*in vitro*) especially at the level of behavioral testing.

In this systematic review, next to zebrafish, two other non-mammalian model organisms were initially evaluated: sea urchin and nematode. Another model organism, the fruit fly *Drosophila* was excluded from

this systematic review for reasons of lower level of biological organization (insect group). It appeared after full text evaluation that only 2 nematode publications and 5 sea urchin publications were compliant with inclusion/exclusion criteria, compared to 83 publications using the zebrafish model. Based on data richness, endpoint-based evaluation of test performance was further focused to the zebrafish model.

Except for advantages of lower organism, related to their small size, rapid embryonic development and short life cycle, which greatly reduce space and time costs, the zebrafish is a vertebrate with fundamental processes of neurodevelopment and signal transduction, with high similarity to those that occur in humans (Guo, 2009, Kaslin et al., 2013, Rico et al., 2011). Zebrafish possess 26206 protein-coding genes, more than any previously sequenced vertebrate, and comparison to the human reference genome shows that approximately 70% of human genes have at least one obvious zebrafish orthologue (Howe et al., 2013). Genes involved in controlling and mediating locomotion behavior in zebrafish embryo and larvae have been identified (Granato et al., 1996) through functional genomics studies (mutant models).

Only a few papers of our systematic review on zebrafish were dedicated to the study of gene expression, specific for DNT and including viability assessment. In situ hybridization studies by Dou et al. (2011, record number 1580) and Ho et al. (2013, record nummer 662) demonstrated brain specific gene expression upon exposure to respectively lead acetate, and methylmercury chloride both resulting into a true positive score. Several other 'omics' studies are reported (39 citations) but as these are mostly whole body analysis, studying multiple genes by real time RT-PCR or microarray analysis, information on specificity for DNT relevant changes is lacking. While changes of genes or proteins, specific for brain development, can serve as early biomarkers with links to processes in other model systems, the systematic review clearly demonstrated that there is a need for brain specific gene expression analysis at sub lethal levels in zebrafish models which could support analysis of DNT outcome. Similar observations were applicable for several other DNT specific endpoint groups (e.g. catecholaminergic and cholinergic endpoints, neuronal endpoints and functioning): a general lack of data with respect to the number of tested positive and negative compounds, as well as lack of assessment of viability needed to decide upon the sensitivity of the DNT specific endpoint. The evaluation of endpoint group 'motor activity' provided by far the most data for the DNT specific endpoints (60 publications across all life stages/transgenic model and test compounds). Based on the results of the systematic review, the strength of the endpoint 'motor activity', especially for the zebrafish model, as an integration of early key events (specific DNT-endpoints), is shown. For a number of true positive compounds, changes of motor behavior were supported by evidence for effects at primary and secondary motor neuron populations and axon growth, as well as disturbed cholinergic function (Chen et al 2012a, Chen et al., 2012b, Yen et al. 2011, Yang et al. 2011, Sylvain et al. 2010; record numbers 1027, 1145, 1458, 1861 and 14229). However, assay performance with respect to specificity and sensitivity needs to be further evaluated with an extension of the panel of test compounds. The systematic review showed only 2 publications testing motor activity with negative compounds) giving 4 citations with true negative score (Selderslaghs et al. 2013; record number 717) and 1 false positive score, the latter using a transgenic model (Zhang et al. 2013; record number 705). But also for positive compounds, more data need to be obtained, especially at the level of diversification between life stages. Different behavioral patterns, defined as spontaneous tail coiling, touch response, and swimming occur as a function of early development associated with specific events related to neuronal development and network formation, and these were covered to certain extent in the systematic review. However, next to the limited number of citations for each of positive and negative test compounds at each of life stages, a general lack of method standardization became evident. Both at the level of time window and duration of exposure as well as time point of effect measurements, huge variety was observed between publications. Finally, comparative work with the organophosphate chlorpyrifos (parent compound) and chlorpyrifos oxon (metabolite) demonstrated the need to include life stages (from 72 hpf to 120 hpf) which have functional liver, demonstrated by measurement of metabolites in the liver and occurrence of true positive DNT effects. Furthermore, behavioral patterns were either or not studied after stimulation (e.g. light/dark challenge) and the significane of these mythological variables for the

final outcome should be assessed. Thus an overall need of protocol standardization and an extension of the dataset of test compounds (especially negatives) is required to evaluate assay performance to assess motor activity patterns in most DNT relevant, but ethically accepted zebrafish early life stages.

The systematic review on *in silico* models related to DNT was focused to evaluation of barrier function, both the placental and the blood-brain barrier. No QSAR or read-across models on DNT itself have been found during the literature search. Published QSAR models for DNT properties consider the ability of substances to pass the placental barrier (record number 448 and 162). They give an indication of which physical-chemical and structural properties influence placental transfer (expressed as clearance index i.e. the ratio between the clearance of a test substance and the reference substance). The QSAR descriptors generally relate to lipophilicity, hydrogen bonding ability, and molecular size. If the substance is highly hydrophilic (log Pow $\leq$ 0.5 according to (Giaginis et al., 2011)) it may not enter the phospholipid structure of the placental membrane. If a drug is able to form easily hydrogen bonds in water, this will increase its water solubility hence possibly reducing membrane penetration. Increasing mass is associated with decreased transfer rate. This is consistent with experiments, showing that molecules with MW > 500 Da may have reduced transport across the placental barrier, their clearance index is <0.1, reflecting incomplete placental transfer (Pacifici, 2005). These cut-off values for log Pow, MW and clearance index can be used as exclusion criteria in the DNT screening. A validation with pesticides and chemicals of the QSAR model developed by Hewitt et al. (2007) from a heterogeneous set of drugs (data set 1) showed that the model was not able to discriminate between positive or negative responses for developmental toxicity (Worth A., 2011).

The clearance index was also used in a multivariate data analysis of a set of structurally diverse drugs and chemicals (Giaginis et al., 2009). The descriptors reflecting the polarity (e.g. polar surface area, hydrogen bonding capability, dipole moment and number of ionizable groups) of the compounds proved to be the most important with negative correlation. Lipophilicity exerted a positive contribution to the predicted permeation.

Active transport is not considered in the published QSAR models, hence these models may underestimate the real placental transfer. Potential metabolisation before transfer is also not considered. In case a substance is metabolized before transfer, the model is making estimations on the wrong substance. However this problem may be overcome by models on metabolism.

A public website with a QSAR model on DNT or placental barrier permeation is not available. The main reason for this lack is that building a model requires a large amount of measured data, which are not available yet. One reason for this lack of data may be that DNT is not an obligatory endpoint in the registration of chemical substances under REACH (unless DNT evidence exists). It should be encouraged to measure the clearance index for the placental barrier for a high number of chemicals and pesticides. Once data are available, development of QSAR models for placental permeation will surely follow, hereby creating a tool for high throughput screening for prioritization. It has to be indicated that in general QSAR models are not suitable for inorganics and organometallics. On top of developing new QSARs, there may also the option to extend the applicability domain of the QSAR models developed by Hewitt (2007) for pesticides and chemicals. In the long run QSARs based on similar mode of action may be developed to predict DNT properties, when more insight will be gained on the adverse outcome pathway of DNT mechanisms.

Within an intelligent testing strategy, at this moment the best available *in silico* technique for DNT screening seems to be read-across. With read-across, the endpoint information for one or more source chemicals is used to make a prediction of the endpoint for the target chemical. Five read across tools (e.g. OECD Toolbox and the Analog Identification Methodology (A.I.M) of US-EPA) are available that could be used to search for analogues of known DNT substances or to find out whether a substance has an analogue that exerts DNT properties (record number grey lit 1 to grey lit 5). None of these tools is

specific for the DNT endpoint or placental barrier permeation. It is up to the user to determine when a specific analogue is suitable for DNT, as the determination of what structure is appropriate can vary depending on the endpoint assessed. Development of read-across tools specific for DNT would be welcomed.

These data from hESC differentiation up to zebrafish motor behavior described within these last paragraphs sums up to a DNT alternative testing strategy covering major key events of neurodevelopment (Figure 83). Although there is not a tremendous amount of data on DNT chemicals and their effects on different neurodevelopmental key events available, there is a substantial basis on which data gaps can be closed by **specific, problem-driven research**.

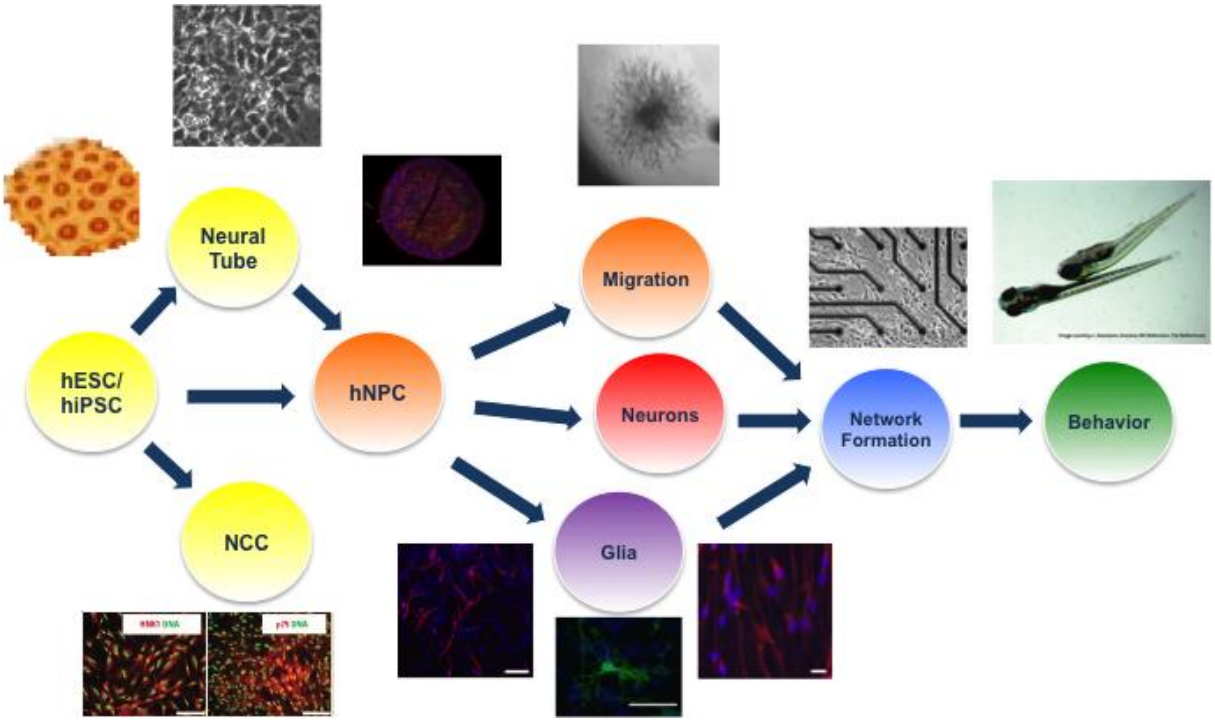


Figure 83 Potential alternative DNT testing strategy. Each circle represents a major key event having to be tested by a human/zebrafish-based assay.

For identification of data gaps, the outcome of this systematic review is very helpful. First of all, as indicated above, there are diverse hESC models available, which are differentiated towards the neural lineage with the purpose of chemical testing. What is needed at this point is a harmonization of protocols and the usage of ethically sound hiPSC-based neurodevelopmental assays. Usage of hiPSC has the advantage that basically every lab in the world can work with such cells without needing special permissions. For chemical effects of NCC and neural tube formation, more chemical testing data is needed. While the proof-of-principle is brought that the assays work in general, now more data on assay performance (sensitivity/specificity) is needed with human DNT compounds. The same is true for the key events migration, neuron and glia differentiation, as well as specific glia endpoints like astrocyte maturation and myelin formation (oligodendrocyte maturation and function) as well as network formation and electrical activity. While data has been generated across different species and models, future initiatives should focus on human-based NS/PC methods, preferably ethically non-confronting methods, which are scientifically sound. Therefore, there is also the need for human *in vivo* validation of any NS/PC-based method, regardless if it is a primary hNPC or a hESC- or hiPSC-based NPC. One possible approach of such a validation would be through transcriptomic profiles of human existing *in vivo* data in comparison to different *in vitro* methods.

Within the establishment of an alternative test strategy for DNT, the complementary value of the zebrafish model (embryo/larvae stage up to 120 hpf) lies especially at the level of testing for sensori-motor activity endpoints, in order to address behavioral endpoints which are essential within (developmental) neurotoxicity assessment. It is shown that early cellular & physiological events, as covered into more details by the battery of *in vitro* assays, are also evident in the zebrafish models studied and associated with motor activity, demonstrating the relevance of this model to predict human DNT (proof of concept). Further studies to extend the data set, standardization of protocols and evaluation of data is needed to demonstrate for both positive and negative compounds the complementarity of *in vitro* methods and the whole organism zebrafish approach for a DNT test strategy. Both embryo and larval stage, covering different time-related steps in neuronal development and neurotransmission, as well as metabolic competent life stage (72-120 hpf) should be included for reliable assessment of motor responses integrating compound-specific effects on molecular initiating events and the sequence of early key events.

For testing compounds, one would call a compound a 'positive hit' when either in one of the human cell-based methods or in the zebrafish assay there is an impact at a concentration reflecting human exposure (as has to be determined by PBPK modeling). This 'either or principle' makes sense because a positive hit in the human assays and a negative in the zebrafish could be caused by species differences, yet the human data would obtain the 'species truth'. Or a negative hit in the human assays is accompanied by a positive hit in the zebrafish, which could be due to additional, indirect targets of the chemical in the whole organism not present in the human-based assays or ADME properties of the chemical (e.g. metabolisation of parent compound into DNT active metabolite) which are more likely to be represented in the whole organism approach. In case the assays of both species are negative, the compound is a true negative and accordingly if both species are positives, the compound is also a true positive. For a selected panel of known positive and negative human DNT compounds, the validation of this proposed test strategy should get priority.

For compounds identified as 'Positives' in such a testing strategy, a rodent *in vivo* biomarker will be helpful that predicts *in vivo* toxicity in a mammal. However, with all the species apart from humans, one always has to take species differences into consideration.

The establishment of a human NS/PC-based *in vitro* testing strategy could in the medium run be embedded into e.g. a 'developing-organism-on-a-chip' project. As the 'Human/organs-on-a-chip'-initiatives on both sides of the Atlantic are already pretty successful such human stem cell-based methods could also gain scientific attention for designing pure human systems in a highly sophisticated manner (Marx et al., 2012).

Final conclusion

In summary, this systematic review identified a variety of methods covering early and later stages of neurodevelopment that have the ability to predict DNT of chemicals. Few *in vivo* models alternative to OECD TG 426 were identified. In general, the available published *in vivo* data is scattered and heterogeneous, making comparisons across exposure paradigms, compounds and chemicals very difficult. Thus, to make more useful evaluations, publications with more consistent experimental designs are needed, and more focused *in vivo-in vitro* comparisons are needed to establish useful effect biomarkers and testing models. From the alternative methods a testing strategy covering early and later neurodevelopmental stages (from stem cell to zebrafish larvae motor behaviour) can be assembled. For ultimately gaining regulatory acceptance, definition of biological application domains of human NS/PC-based methods by performing *in vitro - in vivo* validation is needed. Moreover, protocols for cell-based and zebrafish assays need international standardization. With such standardized protocols, the test battery

needs to be tested for its sensitivity and specificity by testing concentration-responses of known DNT positive and negative compounds across the different assays. Such data might then be used either for regulatory decision-making or compound priorisation for rodent guideline *in vivo* testing. Based on the critical need of progress in the DNT area by regulators and knowledge gathered from the present survey of the available literature in the field, it is quite clear that more explorative as well as dedicated validation research, i.e. by testing chemicals in an alternative testing strategy, is urgently needed. In addition, there might also be room for improvements in terms of procedural development of the regulatory process, including development of text in guidelines and guidance documents. There is also need of a form or forum for continuous contact between knowledge-building and knowledge-using capacities in the DNT-area, like the ISTNET consortium for DNT (Bal-Price et al., 2015a, Crofton et al., 2014), which focus on regulatory needs, as it is expected that new information and models will become available on a continuous basis.

5. Summary of Recommendations

5.1. In vitro

- There is consensus that there is a need for alternative methodologies that can more rapidly and cost effectively screen large numbers of chemicals for their potential to cause DNT or investigate mechanisms to provide information on **human relevance** (Crofton et al., 2011).
- Cells taken out of the *in vivo* context maintain certain cellular and molecular functions they hold *in vivo.* Cellular composition and dimension of *in vitro* models are crucial.
- Primary human cells identified within this systematic review: hNPC (Lonza) and FNC-B4 cells (human fetal olfactory neuroepithelium).
- Human embryonic stem cells (hESC) were employed, yet posing an ethical issue on their use and differences in national laws for working with such material.
- Human induced pluripotent stem cells (hiPSC)/hNPC are ethically without concern and therefore a useful alternative, yet despite technology development no DNT study has been performed with such cells.
- Endpoint groups and categories resemble 'Key Events'.
- For neither hESC or hiPSC comparison to human *in vivo* data exists: NEED for definition of the biological application domain.
- There is a NEED for evaluation of glia differentiation of hNPC/hiPSC because there seem to be differences.
- There is a NEED to establish neural migration also in hiPSC because data on migration assays is very sparse.
- Diverse protocols for hESC with distinct outcomes were identified: NEED for hNPC/hiPSC-based harmonized neural induction and differentiation protocols with identification of neurodevelopmental stages to standardize DNT testing procedures.

5.2. In vivo

- I*n vivo* studies (OEDC TG 426, US-EPA DNT guideline) are the current state-of-the-art for assessing and regulating DNT.
- The current Systematic Review collected information to allow for detailed analyses and detection of endpoints/biomarkers and methods that are available beyond present state-of-the-art in the DNT area of public research.

- Few in vivo models alternative to OECD TG 426 were identified. The protocol developed by the group of Per Eriksson at Uppsala University, Sweden, is one model that can be considered for further evaluation in relation to the OECD TG 426. Zebrafish beyond larval stage might have the potential for use as a non-mammalian in vivo alternative to OECD TG 426. Evaluation of this model can be recommended.
- Initial analyses of the collected data revealed that there is need of more consistency in experimental designs and reporting from *in vivo* studies in the open DNT literature. *In vivo* data published in the open literature is scattered and heterogeneous making comparisons across exposure paradigms and compounds difficult, and thus to meet regulatory requirements.
- Initial analyses of the collected data suggest that more focused studies on *in vivo-in vitro* comparisons have the potential to improve mechanistic understanding and to allow for the establishment of causal relationships in the DNT area of research. The starting point for such work will be to build on and compare already available rodent in vivo data (from industry and academia) to newly generated rodent in vitro data. There is very limited human in vivo data available to allow for such comparative studies.
- There is room for improving training and capacity building in the DNT area and for procedural developments of the regulatory process, including further development of text in DNT guidelines and guidance documents.

5.3. Alternative organisms

- Whole organism alternative approaches have added value to *in vitro* methods as key events are integrated into adverse apical effects, while ADME properties are included
- Zebrafish, being a vertebrate and representing highly conserved genes with rodents and human, appeared most data rich model considering early life stages.
- Changes of motor behaviour, associated with neurological key events should be studied in early and late developmental stages covering both time series of embryo and brain development, as well as metabolic competence.
- Methods for behavioural analysis among life stages and endpoints observed (e.g. spontaneous coiling, touch response, spontaneous swimming, light/dark challenge,;.) should be further harmonized and standard protocols need to be evaluated for DNT test performance with more positive and negative compounds

5.4. In silico

- *In silico* part considered systematic review of methods for assessment of chemicals passing 2 barriers: the placental barrier and the blood-brain barrier of the developing foetus.
- Selected 2 QSARs publications, relevant for DNT are based on the physical-chemical and structural properties of the substance enabling placental transfer, and not on the DNT effect itself. The QSARs available in the two publications were developed from data on drugs.
- Analysis of the model equations indicated that the descriptors generally relate to hydrophobicity, hydrogen bonding ability, and molecular size. This is consistent with experimental work, showing that molecules with MW > 500 Da may have reduced transport across the placental barrier and their clearance index is <0.1 (the ratio between the clearance, resp. transfer of a test substance and the reference substance, e.g. antipyrine a small lipophilic molecule), reflecting incomplete placental transfer.
- The models do not consider active transport, neither metabolites of compounds.
- There is an overall lack of experimental data for chemicals and pesticides on placental transfer, and specific properties of fetal blood brain barrier function in relation to permeability are not dealt with in publications.

5.5. Common recommendations

- While the proof-of-principle is brought that *in vitro* assays work in general for assessing NCC and neural tube formation, migration, differentiation, glia endpoints and network formation with electrical activity, now more data on assay performance (sensitivity/specificity) is needed with human positive/negative DNT compounds (chemical testing).
- Human *in vitro* - *in vivo* validation of any NS/PC-based/zebrafish embryo method is needed, e.g. by comparison of transcriptomic profiles of human existing *in vivo* data to such of *in vitro* methods. This will help defining the biological application domain of individual assays.
- Zebrafish larvae and embryo were identified as suitable model organisms for DNT testing, especially sensory-motor function tests as behavioral outcomes. More data on assay performance (sensitivity/specificity) is needed with human positive/negative DNT compounds (chemical testing).
- Human *in vitro* cell-based methods combined with zebrafish whole organism might provide a DNT testing battery covering early and late neurodevelopmental processes. Protocols for these methods need standardization. Test battery performance has to be tested with human DNT positive and negative compounds (chemical testing).
- There is the NEED to obtain experimental data on the clearance index for the placental barrier for a high number of chemicals and pesticides. Once data are available, development of QSAR models for placental permeation can follow, hereby creating a tool for high throughput screening for prioritization of DNT chemicals as part of strategy.
- Adverse Outcome Pathways (AOPs) for DNT are urgently needed. AOPs help (i) in identification of knowledge gaps. Moreover, (ii) they help in determination if the models (*in vivo* or *in vitro*) used for AOP building are –due to their biology- suitable for Key Event evaluation (biological application domain/species differences). Over all, (iii) AOPs will help regulators in using data from alternative approaches in the risk assessment process by creating more certainty.
- Based on the critical need of progress in the DNT area by regulators it is quite clear that more explorative as well as dedicated validation research, including testing chemicals in an alternative testing strategy, is urgently needed.
- Over all, there is the NEED of a form or forum, like the ISTNET (International Stakeholder NETwork) consortium for DNT, for continuous contact between knowledge-building and knowledge-using capacities in the DNT-area, which focus on regulatory needs, as it is expected that new information and models will become available on a continuous basis.

5.6. Possible EFSA follow-up activities

- Workshop and project sponsorship on ‘Harmonization of protocols for human neural S/PC-based assays and zebrafish assays contributing to DNT testing’.
- Project sponsorship on compound testing across a DNT testing battery.
- Follow-up on ISTNET (International Stakeholder Network for creating a (Developmental) Neurotoxicity Testing Roadmap)

REFERENCES

Abbott B, Held G, Wood C, Buckalew A, Brown J and Schmid J, 1999. AhR, ARNT, and CYP1A1 mRNA quantitation in cultured human embryonic palates exposed to TCDD and comparison with mouse palate in vivo and in culture. Toxicological sciences, 47, 62-75.

Abel T and Zukin RS, 2008. Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. Current opinion in pharmacology, 8, 57-64.

Adams J, Vorhees CV and Middaugh LD, 1990. Developmental neurotoxicity of anticonvulsants: human and animal evidence on phenytoin. Neurotoxicology and teratology, 12, 203-214.

Alderton W, Berghmans S, Butler P, Chassaing H, Fleming A, Golder Z, Richards F and Gardner I, 2010. Accumulation and metabolism of drugs and CYP probe substrates in zebrafish larvae. Xenobiotica, 40, 547-557.

Alepee N, Bahinski T, Daneshian M, De Wever B, Fritsche E, Goldberg A, Hansmann J, Hartung T, Haycock J, Hogberg H, Hoelting L, Kelm JM, Kadereit S, McVey E, Landsiedel R, Leist M, Lubberstedt M, Noor F, Pellevoisin C, Petersohn D, Pfannenbecker U, Reisinger K, Ramirez T, Rothen-Rutishauser B, Schafer-Korting M, Zeilinger K and Zurich MG, 2014. State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. Altex, 31, 441-477.

Allegrucci C and Young L, 2007. Differences between human embryonic stem cell lines. Human reproduction update, 13, 103-120.

Andersen ME, Clewell HJ, Bermudez E, Dodd DE, Willson GA, Campbell JL and Thomas RS, 2010. Formaldehyde: Integrating dosimetry, cytotoxicity and genomics to understand dose-dependent transitions for an endogenous compound. Toxicological sciences, kfq303.

Andersen ME and Krewski D, 2009. Toxicity testing in the 21st century: bringing the vision to life. Toxicol Sci, 107, 324-330.

Bal-Price A, Crofton KM, Leist M, Allen S, Arand M, Buetler T, Delrue N, FitzGerald RE, Hartung T, Heinonen T, Hogberg H, Bennekou SH, Lichtensteiger W, Oggier D, Paparella M, Axelstad M, Piersma A, Rached E, Schilter B, Schmuck G, Stoppini L, Tongiorgi E, Tiramani M, Monnet-Tschudi F, Wilks MF, Ylikomi T and Fritsche E, 2015a. International STakeholder NETwork (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. Arch Toxicol, 89, 269-287.

Bal-Price A, Crofton KM, Sachana M, Shafer TJ, Behl M, Forsby A, Hargreaves A, Landesmann B, Lein PJ, Louisse J, Monnet-Tschudi F, Paini A, Rolaki A, Schrattenholz A, Sunol C, van Thriel C, Whelan M and Fritsche E, 2015c. Putative adverse outcome pathways relevant to neurotoxicity. Crit Rev Toxicol, 45, 83-91.

Bal-Price A, Crofton, Kevin M., Shafer, Timothy, Sachana, Magdalini, Behl, Mamta, Forsby, Anna, Hargreaves, Alan, Landesmann, Brigitte, Lein, Pamela J., Louisse, Jochem, Monnet-Tschudi, Florianne, Paini, Alicia, Rolaki, Alexandra, Schrattenholz, André, Suñol, Cristina, van Thriel, Christoph, Whelan, Maurice, Fritsche, Ellen 2014. Workshop Report: Adverse Outcome Pathways (AOP) Relevant to Neurotoxicity - Report in press.

Bal-Price AK, Coecke S, Costa L, Crofton KM, Fritsche E, Goldberg A, Grandjean P, Lein PJ, Li A and Lucchini R, 2012. Advancing the science of developmental neurotoxicity (DNT): testing for better safety evaluation. Altex, 29, 202-215.

Baumann J, Barenys M, Gassmann K and Fritsche E, 2014. Comparative human and rat "neurosphere assay" for developmental neurotoxicity testing. Curr Protoc Toxicol, 59, 12.21.11-12.21.24.

Bernhardt RR, Chitnis AB, Lindamer L and Kuwada JY, 1990. Identification of spinal neurons in the embryonic and larval zebrafish. J Comp Neurol, 302, 603-616.

Beronius A HA, Heimeier R, Håkansson H, 2013. Risk assessment of developmental neurotoxicity: Evaluation of the OECD TG 426 test guideline and guidance documents. IMM/Rapport 1/2013.

Boekelheide K and Andersen ME, 2010. A mechanistic redefinition of adverse effects - a key step in the toxicity testing paradigm shift. Altex, 27, 243-252.

Bose R, Onishchenko N, Edoff K, Lang AMJ and Ceccatelli S, 2012. Inherited effects of low-dose exposure to methylmercury in neural stem cells. Toxicological sciences, kfs257.

Broccoli V, Giannelli SG and Mazzara PG, 2014. Modeling physiological and pathological human neurogenesis in the dish. Frontiers in neuroscience, 8,

Brustein E, Saint-Amant L, Buss RR, Chong M, McDearmid JR and Drapeau P, 2003. Steps during the development of the zebrafish locomotor network. J Physiol Paris, 97, 77-86.

Burke K, Cheng Y, Li B, Petrov A, Joshi P, Berman RF, Reuhl KR and DiCicco-Bloom E, 2006. Methylmercury elicits rapid inhibition of cell proliferation in the developing brain and decreases cell cycle regulator, cyclin E. Neurotoxicology, 27, 970-981.

Burry M, Guizzetti M, Oberdoerster J and Costa LG, 2003. Developmental neurotoxicity of toluene: in vivo and in vitro effects on astroglial cells. Developmental neuroscience, 25, 14-19.

Buss RR and Drapeau P, 2001. Synaptic drive to motoneurons during fictive swimming in the developing zebrafish. J Neurophysiol, 86, 197-210.

Chen C-N, Liang C-M, Lai J-R, Tsai Y-J, Tsay J-S and Lin J-K, 2003. Capillary electrophoretic determination of theanine, caffeine, and catechins in fresh tea leaves and oolong tea and their effects on rat neurosphere adhesion and migration. Journal of agricultural and food chemistry, 51, 7495-7503.

Chen L, Huang C, Hu C, Yu K, Yang L and Zhou B, 2012a. Acute exposure to DE-71: Effects on locomotor behavior and developmental neurotoxicity in zebrafish larvae. Environmental Toxicology and Chemistry, 31, 2338-2344.

Chen X, Huang C, Wang X, Chen J, Bai C, Chen Y, Chen X, Dong Q and Yang D, 2012b. BDE-47 disrupts axonal growth and motor behavior in developing zebrafish. Aquatic Toxicology, 120, 35-44.

Coecke S, Goldberg AM, Allen S, Buzanska L, Calamandrei G, Crofton K, Hareng L, Hartung T, Knaut H and Honegger P, 2007. Workgroup report: incorporating in vitro alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. Environmental health perspectives, 115, 924.

Establishing a common format for the

submission of the information pursuant to Directive 2010/63/EU of the European Parliament and of the

Council on the protection of animals used for scientific purposes,

Corvino V, Marchese E, Michetti F and Geloso MC, 2013. Neuroprotective strategies in hippocampal neurodegeneration induced by the neurotoxicant trimethyltin. Neurochemical research, 38, 240-253.

Cowan WM, Jessell TM and Zipursky SL, 1997. Molecular and cellular approaches to neural development. Editor. Oxford University Press,

Crofton K, Fritsche E, Ylikomi T and Bal-Price A, 2014. International STakeholder NETwork (ISTNET) for creating a developmental neurotoxicity testing (DNT) roadmap for regulatory purposes. Altex, 31, 223-224.

Crofton KM, Foss JA, Hass U, Jensen KF, Levin ED and Parker SP, 2008. Undertaking positive control studies as part of developmental 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 A, Knaut H, Buzanska L and Goldberg A, 2011. Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. Altex, 28, 9-15.

Crofton KM, Mundy WR and Shafer TJ, 2012. Developmental neurotoxicity testing: a path forward. Congenital anomalies, 52, 140-146.

Culbreth ME, Harrill JA, Freudenrich TM, Mundy WR and Shafer TJ, 2012. Comparison of chemical-induced changes in proliferation and apoptosis in human and mouse neuroprogenitor cells. Neurotoxicology, 33, 1499-1510.

Currie R, 2014. NC3Rs/POEMS Network Maths Study Group - Applying mathematics to 3Rs problems

Mathematical modelling to reduce animal use in neurodevelopmental safety assessment in humans.

Dare E, Fetissov S, Hokfelt T, Hall H, Ogren SO and Ceccatelli S, 2003. Effects of prenatal exposure to methylmercury on dopamine-mediated locomotor activity and dopamine D2 receptor binding. Naunyn Schmiedebergs Arch Pharmacol, 367, 500-508.

de Esch C, Slieker R, Wolterbeek A, Woutersen R and de Groot D, 2012. Zebrafish as potential model for developmental neurotoxicity testing: a mini review. Neurotoxicology and teratology, 34, 545-553.

De Groot DM, Hartgring S, Van de Horst L, Moerkens M, Otto M, Bos-Kuijpers MH, Kaufmann WS, Lammers JH, O'Callaghan JP and Waalkens-Berendsen ID, 2005. 2D and 3D assessment of neuropathology in rat brain after prenatal exposure to methylazoxymethanol, a model for developmental neurotoxicty. Reproductive Toxicology, 20, 417-432.

Deng W, McKinnon RD and Poretz RD, 2001. Lead exposure delays the differentiation of oligodendroglial progenitors in vitro. Toxicol Appl Pharmacol, 174, 235-244.

Dimou L and Gotz M, 2014. Glial cells as progenitors and stem cells: new roles in the healthy and diseased brain. Physiol Rev, 94, 709-737.

Dorman DC, Allen SL, Byczkowski JZ, Claudio L, Fisher Jr JE, Fisher JW, Harry GJ, Li AA, Makris SL and Padilla S, 2001. Methods to identify and characterize developmental neurotoxicity for human health risk assessment. III: pharmacokinetic and pharmacodynamic considerations. Environmental health perspectives, 109, 101.

Drobic B, Dunn KL, Espino PS and Davie JR, 2006. Abnormalities of chromatin in tumor cells. In: Cancer: Cell Structures, Carcinogens and Genomic Instability. Springer, 25-47.

ECHA, 2008a. Guidance on information requirements and chemical safety assessment Chapter R.7a: Endpoint specific guidance.

ECHA M, 2008b. Guidance on information requirements and chemical safety assessment. Chapter R, 8,

Ek CJ, Dziegielewska KM, Habgood MD and Saunders NR, 2012. Barriers in the developing brain and< i> Neurotoxicology</i>. Neurotoxicology, 33, 586-604.

Evans MJ and Kaufman MH, 1981. Establishment in culture of pluripotential cells from mouse embryos. nature, 292, 154-156.

Fahrion JK, Komuro Y, Li Y, Ohno N, Littner Y, Raoult E, Galas L, Vaudry D and Komuro H, 2012. Rescue of neuronal migration deficits in a mouse model of fetal Minamata disease by increasing neuronal Ca2+ spike frequency. Proceedings of the National Academy of Sciences, 109, 5057-5062.

Falk-Filipsson A, Hanberg A, Victorin K, Warholm M and Wallén M, 2007. Assessment factors—applications in health risk assessment of chemicals. Environmental research, 104, 108-127.

Farina M, Rocha JB and Aschner M, 2011. Mechanisms of methylmercury-induced neurotoxicity: evidence from experimental studies. Life sciences, 89, 555-563.

Fitzpatrick J, Mendez E and 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.

Foti SB, Chou A, Moll AD and Roskams AJ, 2013. HDAC inhibitors dysregulate neural stem cell activity in the postnatal mouse brain. International Journal of Developmental Neuroscience, 31, 434-447.

Fritsche E, 2014. Human Stem/Progenitor Cell-Based Asseys for Neurodevelopmental Toxicity testing.

Gassmann K, Abel J, Bothe H, Haarmann-Stemmann T, Merk HF, Quasthoff KN, Rockel TD, Schreiber T and Fritsche E, 2010. Species-specific differential AhR expression protects human neural progenitor cells against developmental neurotoxicity of PAHs. Environ Health Perspect, 118, 1571-1577.

Gassmann K, Baumann J, Giersiefer S, Schuwald J, Schreiber T, Merk HF and Fritsche E, 2012. Automated neurosphere sorting and plating by the COPAS large particle sorter is a suitable method for high-throughput 3D in vitro applications. Toxicol In Vitro, 26, 993-1000.

Gerdin MJ and Eiden LE, 2007. Regulation of PC12 cell differentiation by cAMP signaling to ERK independent of PKA: do all the connections add up? Science Signaling, 2007, pe15.

Giaginis C, Tsantili-Kakoulidou A and Theocharis S, 2011. Assessing drug transport across the human placental barrier: from in vivo and in vitro measurements to the ex vivo perfusion method and in silico techniques. Current pharmaceutical biotechnology, 12, 804-813.

Giaginis C, Zira A, Theocharis S and Tsantili-Kakoulidou A, 2009. Application of quantitative structure–activity relationships for modeling drug and chemical transport across the human placenta barrier: a multivariate data analysis approach. Journal of Applied Toxicology, 29, 724-733.

Gimenez-Llort L, Ahlbom E, Dare E, Vahter M, Ogren S and Ceccatelli S, 2001. Prenatal exposure to methylmercury changes dopamine-modulated motor activity during early ontogeny: age and gender-dependent effects. Environ Toxicol Pharmacol, 9, 61-70.

Giordano G, Kavanagh TJ and Costa LG, 2008. Neurotoxicity of a polybrominated diphenyl ether mixture (DE-71) in mouse neurons and astrocytes is modulated by intracellular glutathione levels. Toxicology and applied pharmacology, 232, 161-168.

Giordano G, Kavanagh TJ and Costa LG, 2009. Mouse cerebellar astrocytes protect cerebellar granule neurons against toxicity of the polybrominated diphenyl ether (PBDE) mixture DE-71. Neurotoxicology, 30, 326-329.

Go HS, Kim KC, Choi CS, Jeon SJ, Kwon KJ, Han S-H, Lee J, Cheong JH, Ryu JH and Kim C-H, 2012. Prenatal exposure to valproic acid increases the neural progenitor cell pool and induces macrocephaly in rat brain via a mechanism involving the GSK-3β/β-catenin pathway. Neuropharmacology, 63, 1028-1041.

Granato M, van Eeden FJ, Schach U, Trowe T, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Odenthal J and Nusslein-Volhard C, 1996. Genes controlling and mediating locomotion behavior of the zebrafish embryo and larva. Development, 123, 399-413.

Grandjean P and Landrigan PJ, 2006a. Developmental neurotoxicity of industrial chemicals. Lancet, 368, 2167-2178.

Grandjean P and Landrigan PJ, 2006c. Developmental neurotoxicity of industrial chemicals. The Lancet, 368, 2167-2178.

Grandjean P and Landrigan PJ, 2014. Neurobehavioural effects of developmental toxicity. The Lancet Neurology, 13, 330-338.

Gulisano M, Pacini S, Punzi T, Morucci G, Quagliata S, Delfino G, Sarchielli E, Marini M and Vannelli GB, 2009. Cadmium modulates proliferation and differentiation of human neuroblasts. J Neurosci Res, 87, 228-237.

Guo S, 2009. Using zebrafish to assess the impact of drugs on neural development and function. Expert Opin Drug Discov, 4, 715-726.

Hanke T, 2006. Lessons from TGN1412. Lancet, 368, 1569-1570; author reply 1570.

Hansen JM, Carney EW and Harris C, 1999. Differential alteration by thalidomide of the glutathione content of rat vs. rabbit conceptuses in vitro. Reproductive Toxicology, 13, 547-554.

Harrill JA, Freudenrich TM, Robinette BL and Mundy WR, 2011. Comparative sensitivity of human and rat neural cultures to chemical-induced inhibition of neurite outgrowth. Toxicology and applied pharmacology, 256, 268-280.

Heaton MB, Moore DB, Paiva M, Madorsky I, Mayer J and Shaw G, 2003. The Role of Neurotrophic Factors, Apoptosis-Related Proteins, and Endogenous Antioxidants in the Differential Temporal Vulnerability of Neonatal Cerebellum to Ethanol. Alcoholism: Clinical and Experimental Research, 27, 657-669.

Hewett JA, 2009. Determinants of regional and local diversity within the astroglial lineage of the normal central nervous system. J Neurochem, 110, 1717-1736.

Hewitt M, Madden J, Rowe P and Cronin M, 2007. Structure-based modelling in reproductive toxicology:(Q) SARs for the placental barrier†. SAR and QSAR in Environmental Research, 18, 57-76.

Ho NY, Yang L, Legradi J, Armant O, Takamiya M, Rastegar S and Strahle U, 2013. Gene responses in the central nervous system of zebrafish embryos exposed to the neurotoxicant methyl mercury. Environ Sci Technol, 47, 3316-3325.

Hoelting L, Leist M and Stoppini L, 2014. Using Pluripotent Stem Cells and Their Progeny as an In Vitro Model to Assess (Developmental) Neurotoxicity. Predictive Toxicology: From Vision to Reality, 279-320.

Holson RR, Freshwater L, Maurissen JP, Moser VC and 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.

Hou Z, Zhang J, Schwartz MP, Stewart R, Page CD, Murphy WL and Thomson JA, 2013. A human pluripotent stem cell platform for assessing developmental neural toxicity screening. Stem cell research & therapy, 4, S12.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P,

Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Urun Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberlander M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nusslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J and Stemple DL, 2013. The zebrafish reference genome sequence and its relationship to the human genome. nature, 496, 498-503.

Huang F and Schneider JS, 2004. Effects of lead exposure on proliferation and differentiation of neural stem cells derived from different regions of embryonic rat brain. Neurotoxicology, 25, 1001-1012.

Hubel DH and Wiesel TN, 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol, 160, 106-154.

Ibarrola N and Rodriguez-Pena A, 1997. Hypothyroidism coordinately and transiently affects myelin protein gene expression in most rat brain regions during postnatal development. Brain Res, 752, 285-293.

Incardona JP, Gaffield W, Kapur RP and Roelink H, 1998. The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. Development, 125, 3553-3562.

Irmady K, Zechel S and Unsicker K, 2011. Fibroblast growth factor 2 regulates astrocyte differentiation in a region-specific manner in the hindbrain. Glia, 59, 708-719.

Ishido M and Suzuki J, 2010. Inhibition by rotenone of mesencephalic neural stem-cell migration in a neurosphere assay in vitro. Toxicology in vitro, 24, 552-557.

Jenkins L, Harries N, Lappin JE, MacKenzie AE, Neetoo-Isseljee Z, Southern C, McIver EG, Nicklin SA, Taylor DL and Milligan G, 2012. Antagonists of GPR35 display high species ortholog selectivity and varying modes of action. Journal of Pharmacology and Experimental Therapeutics, 343, 683-695.

Jessell TM, Sanes, J.R., , 2000. The development of the nervous system. McGraw-Hill, New York, 1019–1114.

Jiang Y-z, Wang K, Fang R and Zheng J, 2010a. Expression of aryl hydrocarbon receptor in human placentas and fetal tissues. Journal of Histochemistry & Cytochemistry, 58, 679-685.

Jiang YZ, Wang K, Fang R and Zheng J, 2010b. Expression of aryl hydrocarbon receptor in human placentas and fetal tissues. J Histochem Cytochem, 58, 679-685.

Johansson C, Castoldi AF, Onishchenko N, Manzo L, Vahter M and Ceccatelli S, 2007. Neurobehavioural and molecular changes induced by methylmercury exposure during development. Neurotox Res, 11, 241-260.

Kabat K, Buterbaugh GG and Eccles CU, 1985. Methylazoxymethanol as a developmental model of neurotoxicity. Neurobehavioral Toxicology & Teratology,

Kadereit S, Zimmer B, van Thriel C, Hengstler JG and Leist M, 2012. Compound selection for in vitro modeling of developmental neurotoxicity. Front Biosci, 17, 2442-2460.

Kakita A, Inenaga C, Sakamoto M and Takahashi H, 2002. Neuronal migration disturbance and consequent cytoarchitecture in the cerebral cortex following transplacental administration of methylmercury. Acta neuropathologica, 104, 409-417.

Kapucu FE, Tanskanen JM, Mikkonen JE, Yla-Outinen L, Narkilahti S and Hyttinen JA, 2012. Burst analysis tool for developing neuronal networks exhibiting highly varying action potential dynamics. Front Comput Neurosci, 6, 38.

Kaslin J, Kroehne V, Benato F, Argenton F and Brand M, 2013. Development and specification of cerebellar stem and progenitor cells in zebrafish: from embryo to adult. Neural Dev, 8, 9.

Kaufmann W, 2003. Current status of developmental neurotoxicity: an industry perspective. Toxicology letters, 140, 161-169.

Kinth P, Mahesh G and Panwar Y, 2013. Mapping of Zebrafish Research: A Global Outlook. Zebrafish, 10, 510-517.

KISBY GE, KABEL H, HUGON J and SPENCER P, 1999. DAMAGE AND REPAIR OF NERVE CELL DNA IN TOXIC STRESS 1*. Drug metabolism reviews, 31, 589-618.

Kodavanti PR, Coburn CG, Moser VC, MacPhail RC, Fenton SE, Stoker TE, Rayner JL, Kannan K and Birnbaum LS, 2010. Developmental exposure to a commercial PBDE mixture, DE-71: neurobehavioral, hormonal, and reproductive effects. Toxicol Sci, 116, 297-312.

Krewski D, Acosta D, Jr., Andersen M, Anderson H, Bailar JC, 3rd, Boekelheide K, Brent R, Charnley G, Cheung VG, Green S, Jr., Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J and Zeise L, 2010a. Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health B Crit Rev, 13, 51-138.

Krewski D, Acosta Jr D, Andersen M, Anderson H, Bailar III JC, Boekelheide K, Brent R, Charnley G, Cheung VG and Green Jr S, 2010d. Toxicity testing in the 21st century: a vision and a strategy. Journal of Toxicology and Environmental Health, Part B, 13, 51-138.

L'Episcopo F, Tirolo C, Testa N, Caniglia S, Morale MC, Impagnatiello F, Pluchino S and Marchetti B, 2013. Aging-induced Nrf2-ARE pathway disruption in the subventricular zone drives neurogenic impairment in parkinsonian mice via PI3K-Wnt/β-catenin dysregulation. The Journal of Neuroscience, 33, 1462-1485.

Le Douarin NM, Calloni GW and Dupin E, 2008. The stem cells of the neural crest. CELL CYCLE-LANDES BIOSCIENCE-, 7, 1013.

Lein P, Locke P and Goldberg A, 2007. Meeting report: alternatives for developmental neurotoxicity testing. Environ Health Perspect, 115, 764-768.

Lein P, Silbergeld E, Locke P and Goldberg AM, 2005. In vitro and other alternative approaches to developmental neurotoxicity testing (DNT). Environmental toxicology and pharmacology, 19, 735-744.

Leist M and Hartung T, 2013. Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. Archives of toxicology, 87, 563-567.

Luo J and Miller MW, 1998. Growth factor-mediated neural proliferation: target of ethanol toxicity. Brain Research Reviews, 27, 157-167.

Makris SL, Raffaele K, Allen S, Bowers WJ, Hass U, Alleva E, Calamandrei G, Sheets L, Amcoff P, Delrue N and 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.

Mancini JD, Autio DM and Atchison WD, 2009. Continuous exposure to low concentrations of methylmercury impairs cerebellar granule cell migration in organotypic slice culture. Neurotoxicology, 30, 203-208.

Marta CB, Adamo AM, Soto EF and Pasquini JM, 1998. Sustained neonatal hyperthyroidism in the rat affects myelination in the central nervous system. J Neurosci Res, 53, 251-259.

Martin GR, 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A, 78, 7634-7638.

Marx U, Walles H, Hoffmann S, Lindner G, Horland R, Sonntag F, Klotzbach U, Sakharov D, Tonevitsky A and Lauster R, 2012. 'Human-on-a-chip'developments: a translational cutting-edge alternative to systemic safety assessment and efficiency evaluation of substances in laboratory animals and man? Alternatives to laboratory animals: ATLA, 40, 235-257.

Mendelson B, 1986a. Development of reticulospinal neurons of the zebrafish. I. Time of origin. J Comp Neurol, 251, 160-171.

Mendelson B, 1986b. Development of reticulospinal neurons of the zebrafish. II. Early axonal outgrowth and cell body position. J Comp Neurol, 251, 172-184.

Moors M, Rockel TD, Abel J, Cline JE, Gassmann K, Schreiber T, Schuwald J, Weinmann N and Fritsche E, 2009. Human neurospheres as three-dimensional cellular systems for developmental neurotoxicity testing. Environ Health Perspect, 117, 1131-1138.

Muller UK and van Leeuwen JL, 2004. Swimming of larval zebrafish: ontogeny of body waves and implications for locomotory development. J Exp Biol, 207, 853-868.

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

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

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

OECD/OCDE 2007. Guideline 426 for the testing of chemicals-Developmental Neurotoxicity study.

Ornoy A, 2009. Valproic acid in pregnancy: how much are we endangering the embryo and fetus? Reproductive Toxicology, 28, 1-10.

Pacifici GM, 2005. Transfer of antivirals across the human placenta. Early Hum Dev, 81, 647-654.

Parran D, Mundy W and Barone S, 2001. Effects of methylmercury and mercuric chloride on differentiation and cell viability in PC12 cells. Toxicological sciences, 59, 278-290.

Perreault M, Feng G, Will S, Gareski T, Kubasiak D, Marquette K, Vugmeyster Y, Unger TJ, Jones J and Qadri A, 2013. Activation of TrkB with TAM-163 results in opposite effects on body weight in rodents and non-human primates. PloS one, 8, e62616.

Purchase IF, 1980. Inter-species comparisons of carcinogenicity. Br J Cancer, 41, 454-468.

Radio NM, Breier JM, Shafer TJ and Mundy WR, 2008. Assessment of chemical effects on neurite outgrowth in PC12 cells using high content screening. Toxicological sciences, 105, 106-118.

Raffaele KC, Fisher JE, Jr., Hancock S, Hazelden K and 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 KC, Rowland J, May B, Makris SL, Schumacher K and Scarano LJ, 2010. The use of developmental neurotoxicity data in pesticide risk assessments. Neurotoxicol Teratol, 32, 563-572.

Rico EP, Rosemberg DB, Seibt KJ, Capiotti KM, Da Silva RS and Bonan CD, 2011. Zebrafish neurotransmitter systems as potential pharmacological and toxicological targets. Neurotoxicol Teratol, 33, 608-617.

Roberts AC, Bill BR and Glanzman DL, 2013. Learning and memory in zebrafish larvae. Front Neural Circuits, 7, 126.

Rodriguez-Pena A, Ibarrola N, Iniguez MA, Munoz A and Bernal J, 1993. Neonatal hypothyroidism affects the timely expression of myelin-associated glycoprotein in the rat brain. J Clin Invest, 91, 812-818.

Rovida C, Longo F and Rabbit RR, 2010. How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? Altex, 28, 273-294.

Saint-Amant L, 2006. Development of motor networks in zebrafish embryos. Zebrafish, 3, 173-190.

Saint-Amant L and Drapeau P, 1998. Time course of the development of motor behaviors in the zebrafish embryo. J Neurobiol, 37, 622-632.

Saint-Amant L and Drapeau P, 2000. Motoneuron activity patterns related to the earliest behavior of the zebrafish embryo. J Neurosci, 20, 3964-3972.

Saint-Amant L and Drapeau P, 2001. Synchronization of an embryonic network of identified spinal interneurons solely by electrical coupling. Neuron, 31, 1035-1046.

Seidle T and Stephens M, 2009a. Bringing toxicology into the 21st century: A global call to action. Toxicology in vitro, 23, 1576-1579.

Seidle T and Stephens ML, 2009e. Bringing toxicology into the 21st century: a global call to action. Toxicol In Vitro, 23, 1576-1579.

Seiler AE and Spielmann H, 2011. The validated embryonic stem cell test to predict embryotoxicity in vitro. Nature protocols, 6, 961-978.

Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H and Hennessy L, 2013. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proceedings of the National Academy of Sciences, 110, 3507-3512.

Sharif A, Duhem-Tonnelle V, Allet C, Baroncini M, Loyens A, Kerr-Conte J, Collier F, Blond S, Ojeda SR, Junier MP and Prevot V, 2009. Differential erbB signaling in astrocytes from the cerebral cortex and the hypothalamus of the human brain. Glia, 57, 362-379.

Simpson KL, Weaver KJ, de Villers-Sidani E, Lu JY-F, Cai Z, Pang Y, Rodriguez-Porcel F, Paul IA, Merzenich M and Lin RC, 2011. Perinatal antidepressant exposure alters cortical network function in rodents. Proceedings of the National Academy of Sciences, 108, 18465-18470.

Slotkin TA, 2004. Cholinergic systems in brain development and disruption by neurotoxicants: nicotine, environmental tobacco smoke, organophosphates. Toxicology and applied pharmacology, 198, 132-151.

Stansfield KH, Pilsner JR, Lu Q, Wright RO and Guilarte TR, 2012. Dysregulation of BDNF-TrkB signaling in developing hippocampal neurons by Pb(2+): implications for an environmental basis of neurodevelopmental disorders. Toxicol Sci, 127, 277-295.

Stebbings R, Findlay L, Edwards C, Eastwood D, Bird C, North D, Mistry Y, Dilger P, Liefooghe E and Cludts I, 2007. "Cytokine storm" in the phase I trial of monoclonal antibody TGN1412: better understanding the causes to improve preclinical testing of immunotherapeutics. The Journal of Immunology, 179, 3325-3331.

Strasser A, Wittmann H-J, Buschauer A, Schneider EH and Seifert R, 2013. Species-dependent activities of G-protein-coupled receptor ligands: lessons from histamine receptor orthologs. Trends in pharmacological sciences, 34, 13-32.

Sylvain NJ, Brewster DL and Ali DW, 2010. Zebrafish embryos exposed to alcohol undergo abnormal development of motor neurons and muscle fibers. Neurotoxicology and teratology, 32, 472-480.

Theunissen P, Schulpen S, Van Dartel D, Hermsen S, van Schooten F and Piersma A, 2010. An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. Reproductive Toxicology, 29, 383-392.

Thomas JD, Goodlett CR and West JR, 1998. Alcohol-induced Purkinje cell loss depends on developmental timing of alcohol exposure and correlates with motor performance. Developmental Brain Research, 105, 159-166.

Tran TT, Chowanadisai W, Crinella FM, Chicz-DeMet A and Lönnerdal B, 2002. Effect of high dietary manganese intake of neonatal rats on tissue mineral accumulation, striatal dopamine levels, and neurodevelopmental status. Neurotoxicology, 23, 635-643.

Tyl RW, Crofton K, Moretto A, Moser V, Sheets LP and 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.

Valles S, Sancho-Tello M, Minana R, Climent E, Renau-Piqueras J and Guerri C, 1996. Glial fibrillary acidic protein expression in rat brain and in radial glia culture is delayed by prenatal ethanol exposure. J Neurochem, 67, 2425-2433.

van Dartel DA, Pennings JL, de la Fonteyne LJ, Brauers KJ, Claessen S, van Delft JH, Kleinjans JC and Piersma AH, 2011. Concentration-dependent gene expression responses to flusilazole in embryonic stem cell differentiation cultures. Toxicol Appl Pharmacol, 251, 110-118.

Vemuri MC and Chetty CS, 2005. Alcohol impairs astrogliogenesis by stem cells in rodent neurospheres. Neurochemistry international, 47, 129-135.

Wayman GA, Bose DD, Yang D, Lesiak A, Bruun D, Impey S, Ledoux V, Pessah IN and Lein PJ, 2012. PCB-95 modulates the calcium-dependent signaling pathway responsible for activity-dependent dendritic growth. Environmental health perspectives, 120, 1003.

Worth A. MF-G, S. Lapenna and R. Serafimova., 2011. Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. . submitted to EFSA.

Yamanaka S, 2012. Induced pluripotent stem cells: past, present, and future. Cell stem cell, 10, 678-684.

Yang D, Lauridsen H, Buels K, Chi L-H, La Du J, Bruun DA, Olson JR, Tanguay RL and Lein PJ, 2011. Chlorpyrifos-oxon disrupts zebrafish axonal growth and motor behavior. Toxicological sciences, 121, 146-159.

Yla-Outinen L, Joki T, Varjola M, Skottman H and Narkilahti S, 2014. Three-dimensional growth matrix for human embryonic stem cell-derived neuronal cells. J Tissue Eng Regen Med, 8, 186-194.

Young C, Klocke BJ, Tenkova T, Choi J, Labruyere J, Qin YQ, Holtzman DM, Roth KA and Olney JW, 2003. Ethanol-induced neuronal apoptosis in vivo requires BAX in the developing mouse brain. Cell Death Differ, 10, 1148-1155.

Zhang X and Gong Z, 2013. Fluorescent Transgenic Zebrafish Tg (nkx2. 2a: mEGFP) Provides a Highly Sensitive Monitoring Tool for Neurotoxins. PloS one, 8, e55474.

Zhou T, Taylor MM, DeVito MJ and Crofton KM, 2002. Developmental exposure to brominated diphenyl ethers results in thyroid hormone disruption. Toxicol Sci, 66, 105-116.

Zimmer B, Lee G, Balmer NV, Meganathan K, Sachinidis A, Studer L and Leist M, 2012. Evaluation of developmental toxicants and signaling pathways in a functional test based on the migration of human neural crest cells. Environmental health perspectives, 120, 1116-1122.

APPENDICES

Appendix A. Search strings

DNT search 'Web of Science':

#1: ("DNT") OR ("Developmental Neurotoxi*") OR ("Neurodevelopmental Toxi*")

#2: ("neonatal" OR "neo natal" OR "Immature" OR "prenatal" OR "pre natal" OR "postnatal" OR "post natal" OR "perinatal" OR "peri natal" OR "in utero") AND ("neurotoxi*" OR "toxi*") AND ("development*")

#3: ("brain" OR "CNS" OR "nervous system") AND ("Developmental toxi*")

#4: ("Developing brain" OR "fetal Brain" OR "young brain") AND ("toxi*" OR "neurotoxi*")

#5: ("neonatal exposure" OR "neo natal exposure" OR "prenatal exposure" OR "pre natal exposure" OR "postnatal exposure" OR "post natal exposure" OR "perinatal exposure" OR "peri natal exposure" OR "in utero exposure" OR "fetal exposure") AND ("behavior" OR "behaviour" OR "neurobehaviour*" OR "neurobehavior*")

#6: ("locomotor activity" OR "motor activity" OR "neurotoxi*") AND ("embryo*")

Final string: #1 OR #2 OR #4 OR #5 OR #6

DNT search 'Pubmed':

(((((((((DNT[Title/Abstract] OR Developmental neurotoxi*[Title/Abstract] OR Neurodevelopmental toxi*[Title/Abstract])))) OR (((neonatal[Title/Abstract] OR neo natal[Title/Abstract] OR immature[Title/Abstract] OR prenatal[Title/Abstract] OR pre natal[Title/Abstract] OR postnatal[Title/Abstract] OR post natal[Title/Abstract] OR perinatal[Title/Abstract] OR peri natal[Title/Abstract] OR in utero[Title/Abstract] OR fetal[Title/Abstract]) AND (neurotoxi*[Title/Abstract] OR toxi*[Title/Abstract]) AND development*[Title/Abstract]))) OR ((Developmental toxi*[Title/Abstract] AND (brain[Title/Abstract] OR CNS[Title/Abstract] OR nervous system[Title/Abstract])))) OR (((Developing brain[Title/Abstract] OR fetal Brain[Title/Abstract] OR young brain[Title/Abstract]) AND (toxi*[Title/Abstract] OR neurotoxi*[Title/Abstract])))) OR (((neonatal exposure[Title/Abstract] OR neo natal exposure[Title/Abstract] OR prenatal exposure[Title/Abstract] OR pre natal exposure[Title/Abstract] OR postnatal exposure[Title/Abstract] OR post natal exposure[Title/Abstract] OR perinatal exposure[Title/Abstract] OR peri natal exposure[Title/Abstract] OR in utero exposure[Title/Abstract] OR fetal exposure[Title/Abstract]) AND (behaviour[Title/Abstract] OR behavior[Title/Abstract] OR neurobehaviour*[Title/Abstract] OR neurobehavior*[Title/Abstract])))) OR (((motor activity[Title/Abstract] OR locomotor activity[Title/Abstract] OR neurotoxi*[Title/Abstract]) AND embryo*[Title/Abstract]))) AND "1990/01/01"[PDat] : "2014/06/30"[PDat]

BB search Web of Science:

("Blood brain barrier*" OR "Brain blood barrier*" OR "Placental barrier*") AND (“*in silico*” OR “computational” OR “prediction*” OR “(Q)SAR” OR “(Q)SARs” OR “QSAR” OR “QSARs”)

BB search Pubmed:

((((((((((((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND QSAR[Title/Abstract])))))))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat])))) OR (((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND prediction*[Title/Abstract]))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat]))) OR ((((((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND QSARs[Title/Abstract])))))))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat]))) OR ((((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND (Q)SARs[Title/Abstract])))))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat]))) OR ((((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND (Q)SAR[Title/Abstract])))))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat]))) OR (((((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND computational[Title/Abstract]))) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat]))) OR (((((((Blood brain barrier*[Title/Abstract]) OR Brain blood barrier*[Title/Abstract]) OR Placental barrier*[Title/Abstract]) AND *in silico*[Title/Abstract])) AND ("2014/01/04"[PDat] : "2014/30/06"[PDat])))

Appendix B. Keywords for the keyword search

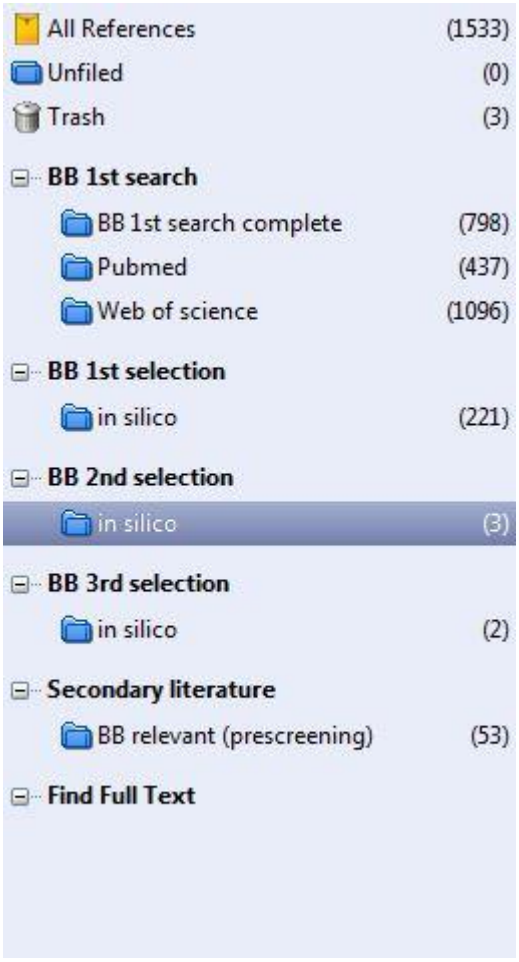
Partner	I. Method related	OR	II. Target system/organ related	AND	III. Toxicological endpoint related	OR	IV. Study design related
in vitro	alternative method, throughput, *in vitro*, non animal alternatives, non-animal alternatives,	OR	3D culture, 3D-culture, cell culture, cell-culture, cell line, cell-line, neural cell, progenitor cell, neuronal cell, neurospheres, primary cell, stem cell, precursor cell, neuroblastoma, neural crest cell, pheochromocytoma, neural culture, neuronal culture, primary culture, cell model, cultured cell, teratocarcinoma, cell system, immortalized cell,	AND	action potential, adverse outcome, astrocyte, axon growth, biomarker, calcium channel, proliferation, cholinergic, cholinesterase, dendrite growth, differentiation, dopaminergic, expression, glia, glutamate receptor, glutaminergic, migration, electrode array, muscarine receptor, myelination, network formation, neurite outgrowth, neurodegeneration, neuron, neuronal network neurotransmitter, nicotine receptor, oligodendrocyte, pathway, signalling, sodium channel, synaptogenesis, thyroid effect, voltage gated, voltage-gated, nuclear receptor, cell devision, signaling	OR	Validation, Prediction, key event,mechanism, metal,initiating event, pesticide,pharmaceutical, screening,toxicity pathway,

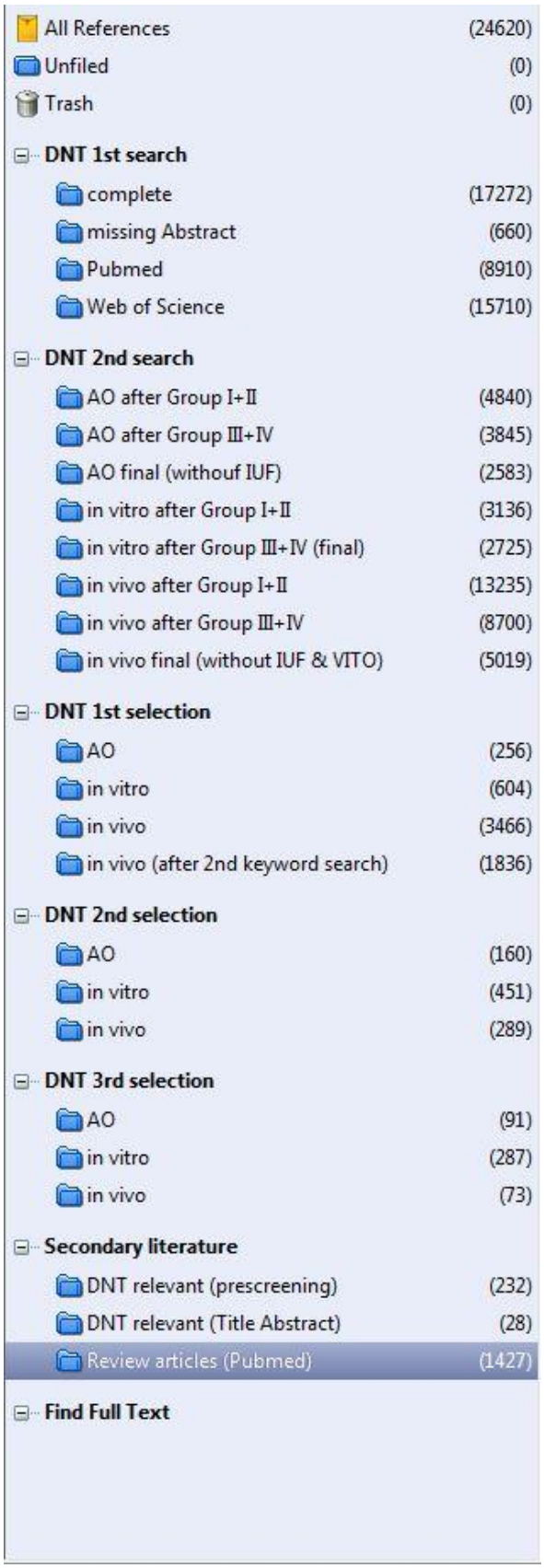
AO	(Q)SAR, alternative method, analogue approach, chemical screening, computational, drug transport, grouping, *in silico*, *in vivo*, integrated test strategy, model, non animal alternative, non-animal alternative, non-testing method, physico chemical propert, physicochemical propert, physico-chemical propert, predicting, prediction, QSAR, read across, structural alert, structure activity relationship, structure-activity relationship, structure-activity-relationship, test battery, throughput, tool, whole organism,	OR	BBB, blood brain barrier, blood-brain barrier, brain blood barrier, brain slice, brain-blood barrier, C. elegans, Caenorhabditiselegans, D. rerio, Daniorerio, nematode, non mammalian, non-mammalian, placental barrier, roundworm, sea urchin, zebrafish, zebrafish,	AND	adverse outcome, axon growth, behavio, biomarker, calcium channel, chemical parameter, chemical propert, cholinergic, cholinesterase, cognition, dopaminergic, expression, glutamate receptor, glutaminergic, light-dark response, locomotor, log BB, motor activity, motor response, motorneuron, movement, network formation, neurite outgrowth, neuro-degeneration, neuro-endocrine, neuronal network, neurotransmitter, nuclear receptor, pathway, permeability, permeation, rest-wake response, signal-ling, startle response, swimming activity, synaptogenesis tail, contraction, thyroid effect, touch response, transgenic, locomotion activity, motor neuron, sodium channel, neurogenesis, apoptosis, cell devision, differentiation, proliferation, signaling,	OR	See *in vitro*

in vivo	in vivo	OR	brain slice, human, mouse, mice, murine, rodent, 'rat', 'rats'	AND	adverse outcome, behavio, brain histopathology, cognition, expression, light-dark response, light dark response, motor activity, motor response, neurodegeneration, pathway, rest-wake response, rest wake response, startle response, swimming activity, tail contraction, tail movement, thyroid effect, anxiety, neuroendocrine, nuclear receptor,	OR	See IUF
Additional Keyword search for *in vivo*							
5-HT, 5HT, acetyl, adrenalin, biomarker, dopamin, epigen, estrogen, expression, genom, methylat, modification, molecular, mRNA, norepinephrine, oestrogen, pathway, phosphoryl, protein, proteome, regulated, serotonin, thyroid,							

Appendix C. Screenshot of EndNote library

'Systematic review on DNT search Results' **'Systematic review on biological barrier search results'**





Appendix D. Grey literature search

Grey literature search.xlsx

Appendix E. Contact mail for 'grey' literature search

DNT search:

Dear Madam or Sir:

The Leibniz Research Institute on Environmental Medicine (IUF), Germany, in cooperation with the Karolinska Institute (KI) for Environmental Medicine (IMM) in Sweden and the Flemish Institute for Technological Research (VITO), Belgium, is doing a systematic literature review on *in vitro* and alternative Developmental Neurotoxicity (DNT) testing methods on behalf of the European Food and Safety Authority (EFSA). The review question is summarized as:

'Which test methods or approaches are available to evaluate developmental neurotoxic effects of chemical exposure?'

The main goal of the project is a systematic and comprehensive literature search and analysis on state of the art of DNT testing methods currently available or under development that should support the EFSA Pesticides Unit with respect to the peer review of active substances under Reg. 1107/2009.

As this search includes publically available peer reviewed research publications and publicly available 'grey literature', we performed a keyword search on the website of your institution/society/organization addressing unpublished DNT methods. Additionally we would like to ask you if you can provide any information concerning unpublished DNT methods which might have been developed at your institute.

In order to achieve a comprehensive understanding of any existing test methods on developmental neurotoxic effects it is extremely important for us to collect all kind of information and your cooperation is highly appreciated.

We thank you in advance for your cooperation.

Kind regards

BB search:

Dear Madam or Sir:

The Leibniz Research Institute on Environmental Medicine (IUF), Germany, in cooperation with the Karolinska Institute (KI) for Environmental Medicine (IMM) in Sweden and the Flemish Institute for Technological Research (VITO), Belgium, is doing a systematic literature review on *in vitro* and alternative Developmental Neurotoxicity (DNT) testing methods on behalf of the European Food and Safety Authority (EFSA). The review question is summarized as:

'Which test methods or approaches are available to evaluate developmental neurotoxic effects of chemical exposure?'

The main goal of the project is a systematic and comprehensive literature search and analysis on state of the art of DNT testing methods currently available or under development that should support the EFSA Pesticides Unit with respect to the peer review of active substances under Reg. 1107/2009.

As substances not reaching the bloodstream are considered to be of less toxicity concern the blood-brain and placental barrier permeability may be an important parameter in the strategy for DNT testing. This is why we performed a complementary search for publications on *in silico* methods that predict this parameter.

As this search includes publically available peer reviewed research publications and publicly available 'grey literature', we performed a keyword search on the website of your institution/society/organization addressing unpublished *in silico* methods that model biological barriers. Additionally we would like to ask you if you can provide any information concerning unpublished methods which might have been developed at your institute. In order to achieve a comprehensive understanding of any existing test methods on developmental neurotoxic effects it is extremely important for us to collect all kind of information and your cooperation is highly appreciated.

We thank you in advance for your cooperation.

Kind regards

Personal contacts:

Dear 'Personal contact':

The Leibniz Research Institute on Environmental Medicine (IUF), Germany, in cooperation with the Karolinska Institute (KI) for Environmental Medicine (IMM) in Sweden and the Flemish Institute for Technological Research (VITO), Belgium, is doing a systematic literature review on *in vitro* and alternative Developmental Neurotoxicity (DNT) testing methods on behalf of the European Food and Safety Authority (EFSA). The review question is summarized as:

'Which test methods or approaches are available to evaluate developmental neurotoxic effects of chemical exposure?'

The main goal of the project is a systematic and comprehensive literature search and analysis on state of the art of DNT testing methods currently available or under development that should support the EFSA Pesticides Unit with respect to the peer review of active substances under Reg. 1107/2009.

As this search includes publically available peer reviewed research publications and publicly available 'grey literature', we would like to ask you if you can provide any information concerning unpublished DNT methods or methods to model biological barriers which might have been developed at your institute.

In order to achieve a comprehensive understanding of any existing test methods on developmental neurotoxic effects it is extremely important for us to collect all kind of information and your cooperation is highly appreciated.

We thank you in advance for your cooperation.

Kind regards

Appendix F. Publication selection based on title and abstract

Study selection_Title_Abstr

Appendix G. Publication selection based on full text

Study selection_Fulltext sc

Appendix H. Comprehensive list of compounds used in this Systematic review

Positive Compounds	Netative Compound	Possible positive	not classified
3-OH-BDE 47	Acetaminophen	17a-Estradiol	1H-[1,2,4]oxadiazolo-[4,3-α]quinoxalin-1-one
5-OH-BDE 47	Amoxicilin	17ß-Estradiol	2,3-Dioxo-6-nitro-7-sulfamoylbenzo(f)quinoxaline
6-OH-BDE 47	Aspartame	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)	3-Amino-1-hydroxy-pyrrolidin-2-one
Aroclor1254 (PCB)	Aspirin	2, 4-dichlorophenoxy-acetic acid (2,4-D)	ammonium perchlorate
Arsenic	Caffeine	2,3-Dimethoxy-1,4-Naphtoquinone (DMNQ)	atenolol
Arsenic tiroxide	Diclofenac	2,4-Dinitrophenol	BAPTA-AM
BDE-100	Diphenhydramine	2-Isopropyl-6-methyl-4-pyrimidinol	Bisbenzimide H
BDE-153	Paracetamol	3,6,5-Trichloro-2-Pyridinol	Bisindolylmaleimide
BDE-183	Penicillin	3'-Methoxy-4'-Nitroflavone	Bismuth sodium tartrate
BDE-209	Penicillin G	3-Methylcholanthrene	Blebbistatin
BDE-28	Pentanoic acid	Abamectin	Brilliant blue
BDE-47	Puromycin	Acetamiprid	C18-Ceramide
BDE-49	Saccharin	Acrylamide	CGP-37849
BDE-99	Sorbitol	Aldicarb	Dimethylsulfoxide (DMSO)
Chlorpyrifos	Sucralose	Aluminium	Dopamine
Chlorpyrifos oxon		Aluminiumchlorid	Ethylnitrosourea
Cyclopamine		Amphetamine	
DE-71		Antimycin A	Flavopiridol
Decabromo Dipehnyl ether		atrazine	GSK4716 (synthetic ERRβ/ϒ ligand)

dichlorodiphenyltrichloroethane (DDT)		Benzidine	H1152
Dieldrin		Benzo[a]pyrene	HA-1077
Dimethyhl mercury		Bifenthrin	Hydrogen peroxide
Diphenylhydantoin		Bisphenol A	Ifenprodil
Ethanol		BMAA	Indirubin
Lead		Brefeldin A	IPA-3
Lead acetate		Buprenorphine	K252a
Lead chloride		Cadmium	Kainic Acid
Manganese(II) chloride		Cadmium chloride	L701324
MeHgCl		Carbamazepine	mefenamic acid
Mercury(II)chloride		Carbaryl	Menadione
Methylazoxymethanol		cartap	Mevastatin
Methylazoxymethanol acetate		Chlorpromazine	MK-801
Methylmercury		Cisplatin	Nanoparticles: AgNP
Methylmercury chloride		cobalt chlorid	Nanoparticles: AgNP, citrate coated
Methymercury hydroxide		Cocaine	Nanoparticles: AgNP, polyvinylpyrrolidone coated
Monomethylarsonic acid		Cocaine hydrochloride	Nanoparticles: AuNP, - charged with MES
Nicotine		Colchicine	Nanoparticles: AuNP, + charged with TMAT
OH-BDE-47		Cycloheximide	Nanoparticles: AuNP, neutral charge with MEE
PCB 105		Cyproconazole	Nanoparticles: CdTe quantum dots
PCB 126		Cytochalasin	Nanoparticles: CoNP
PCB 28		Danofloxacin	Nanoparticles: SiNP
PCB153		Desflurane	Nanoparticles: TiO2NP
PCB180		Dexamethasone	NMDA
PCB47		Dexmedetomidine	Norepinephrine
PCB77		D-glutamate	ODQ
PCB-95		Diazinon	PE-nanoparticles

Sodium arsenite		Diazoxon	Piericidin
Sodium Fluoride		Dibutyltin dichloride	Protein tyrosine phosphatase inhibitor IV
Sodium valproate		Dichlorvos	Quinoline Yellow
Toluene		Diethanolamine	S1P
Valproate		diethyl phthalate (DEP)	Saponin
		diethylstilbesterol (DES)	Serotonin
		Dimethyl phthalate	Silicon dioxide (nano)
		Dimethyltin dichloride	Silver nanoparticles (citrate-coated)
		di-n-butyl phthalate (DBP)	Sodium dodecyl sulfate
		Diquat	Sodium orthovanadate
		Domoic acid	Sodium selenite
		Doramectin	Sodium vanadium oxide
		Doxorubicin	SP600125
		Endosulfan	Sphingosine
		Endosulfan sulfate	Tert-butyl hydroperoxide
		Estradiol	thiazovivin
		Ethinylestradiol	TNFalpha
		Ethylmethanesulfonate	Tween-20
		Ethylnitrosourea	U0126
		Fentanyl	U1026
		Fipronil	Y-27632
		Fluoxetine hydrochloride (Prozac)	
		Flusilazole	
		Flutamide	
		Glutamate	
		Glyphosate	

		Haloperidol	
		Hexaconazole	
		Honokiol	
		Hydroxyurea	
		Imidacloprid	
		Iminodipropionitrile	
		Ipomoea carnea aqueous fraction	
		Isoflurane	
		Ketamine	
		L-glutamic acid	
		lindane	
		Liquiritigenin	
		Lithium	
		Lovastatin	
		Malaoxon	
		Maneb	
		MDMA	
		Mechlorethamine	
		Methadone	
		Methamphetamine	
		methimazole	
		Methylazoxymethanol	
		Methylisothiazolinone	
		Methylmethanesulfonate	
		Methylnitrosourea	
		Methyltin trichloride	

		Midazolam	
		Mipafox	
		mono-(2-ethylhexyl) phthalate	
		Monochrotophos	
		Monocrotophos	
		MPP+	
		multiple compounds	
		Narciclasine	
		Nickel	
		Nickel chloride	
		Nimodipine	
		Nitrogen mustard (HN2)	
		Nitrous oxid	
		Nocodazole	
		nonylphenol	
		Ochratoxin A	
		Okadaic acid	
		Oligomycin	
		omega-Agatoxin	
		omega-Conotoxin	
		Omeprazole	
		p-aminophenylarosonic acid	
		Paranitrophenol	
		Paraoxon	
		Paraquat	
		Parathion	

		Pentachlorphenol	
		Perfluordecanonic acid (PFDA)	
		Perfluorobutane sulfonate	
		Perfluorobutane sulfonic acid (PFBS)	
		Perfluorobutyric acid (PFBA)	
		Perfluoroctanic acid (PFOA)	
		Perfluorononanoic acid (PFNA)	
		Perfluorooctane sulfonate	
		Perfluorooctanesulfonamide	
		Perfluorooctanoic acid	
		Phencyclidine	
		Phenytoin	
		Physostigmine	
		Pirimicarb	
		Potassium chromate	
		Propofol	
		Propylthiouracil (PTU)	
		Retinoic Acid	
		Rotenone	
		Sevoflurane	
		Silver nitrate	
		Simvastatin	
		Sodium benzoate	
		Sodium metam	
		ß-estradiol 3-benzoate	
		Staurosporine	

		TCDD	
		Tetrahydrocannabinol	
		Thallium chloride	
		Thapsigargin	
		Thimerosal (Hg)	
		Thiopental	
		Thyroxine (T4)	
		t-retinoic acid	
		Tributyltin	
		trichloro-2-pyridonol (TCPy)	
		Trichlorphenol	
		Triethyltin chloride	
		Trimethyl tinn chloride	
		Trimethyltin	
		Tri-ortho-cresyl phosphate	
		Tris (1,3-dichlorisopropyl) phosphat	
		Tris (1,3-dichloroisopropyl) phosphat	
		Tris (1-chloropropyl) phosphat	
		Tris (2-chloroethyl) phosphat	
		Tris(2,3-dibromo-1-propyl) phosphate	
		Vincristine	
		Xenon	

Appendix I. Comprehensive list of all *'in vitro'* endpoints

Apoptosis/cell death	Apoptosis
	Cell Death
	DNA breakage
	Cytochrome c localization
	TNF alpha secretion
	NF-κB activity
	DNA damage
	Autophagy
	Necrosis
Astrocyte endpoints	Astrocyte morphology
	Astrocyte differentiation
	Astrocyte differentiation
	Astrocyte growth
	Astrocyte death
	Number of astrocytes
	Glutamine synthetase activity
	Glial differentiation
	Hyaluronic acid secretion
	Concanavalin binding properties of Astrocytes
	Distribution of concanavalin binding
Axonal endpoints	Axonal growth cone collape
	Axonal Guidance
	Axonal Branching
	Axon length
	Axon growth
	Axon specification
	number of axons per neuron
	Neuronal differentiation
	Expanded growth cones
Catecholaminergic function	Monoamine and metabolite levels
	Catecholamine synthesis, turnover, metabolism
Chlorine signalling	Intraneuronal chlorine concentration
Cholinergic function	AchE activity
	Choline acetyltransferase
	Cholinesterase activity
Cytoskeleton	Microtubule disruption
	Actin cytoskeleton

Dendritic endpoints	Dendrite length
	number of dendrites per neuron
	Number of dendritic branches per neuron
	Dendritic morphology
Dopaminergic function	Tyrosine hydroxylase
	Neuronal differentiation
	Gene expression
	Dopaminergic neuron count
	Dopamine transporter
Energy metabolism	ATP generation
	Mitochondrial transmembrane potential
	Mitochondrial function
	Mitochondrial axonal transport
	Mitochondrial length
	Mitochondrial number
Epigenetics	miRNA expression
	DNA methylation
	Histone acetylation
Gabaergic function	Neuronal differentiation
	Glutamic acid decarboxylase
General cellular function	Protein expression/phosphorylation
	Calcium (intracellular)
	Gene expression
	Protein localization
	Cell growth
	Enzymatic activity
	Protein Colocalization
	Protein synthesis
	Protein kinase C
	Adenylate cyclase activity
	cAMP level
	Receptor activation
	Sp1 DNA binding
	AP1 DNA binding
	Phospholipase C activity
	RNA synthesis
	Phosphoinositide metabolism
	Calcium channel activity
	Choline metabolism
	DNA content

	LDL receptor activity
	Cholesterol synthesis
General neuronal endpoints	Neuronal differentiation
	Cytoskeletal network organization
	Neuron density
	Neuronal apoptosis
	Neuronal death
	Neuronal differentiation
	Neuronal growth
	Neuronal maturation
	Neuronal morphology
	Neuronal number
	Neuronal polarity
	NTE activity
	Number of Neurons
	Purkinje cell toxicity
Glutamatergic function	Protein expression/phosphorylation
	glutamate concentration (extracellular)
	NMDAR channel activity
	Gene expression
	AMPA/Kainate receptor functionality
	NMDA receptor binding
	Neuronal differentiation
Migration	Adhesion
	Migration
Neural proliferation	Proliferation
	Cell Cycle Analyses
	Cell number
Neurite endpoints	Neurite length
	Neuronal process formation
	Neurite complexity
	Neurite initiation
	Neurite outgrowth
	Number of branch points
	Number of filopodia
	Number of neurites
	Number of neurites per neuron
	number of primary branches
	number of secondary branches
	process formation

	total number of branches
Neuroinflammation	Microglial death
	Microglial activity
Neuronal functioning / electrophysiology	Network activity
	Neuronal activity
	Electrophysiological response
	Network formation
Neuronal trophic factors	NT3 secretion
	BDNF secretion
	NFG binding
	NGF secretion
NS/PC endpoints	NPC Apoptosis
	Neural differentiation
	Sphere formation
	Number of neurite-forming embryoid bodies
	Neural rosette formation
	Neural outgrowth of EBs
	Aggregate formation
	Notch signalling
Oligodendrocyte endpoints	Oligodendrocyte differentiation
	Oligodendrocyte toxicity
	CNP-activity
	Oligodendrocyte arborization
Omics	Gene expression
	Proteomics
	Metabolomics
Oxidative stress	Oxidative stress
	Lipid peroxidation
	Glutathion level (intracellular)
	Glutathion peroxidase activity
	SOD activity
	Oxidative DNA damage
	γ-Glutamyl-cysteine-synthetase activity
	Redox state
Synapse endpoints	Puncta per neuron
	Puncta per cell body
	Puncta per dendrite
	Synapse Formation

Appendix J. Comprehensive list of all individual cell models

Cell model	Species	Cell type category	Brain region
ASCh-7 (human immortalized astrocytes)	Human	Tumor/Immortalized cell lines	n.a.
C17.2 murine-derived mutipotent neural stem cell line	Mouse	Tumor/Immortalized cell lines	Cerebellum
C6	Rat	Tumor/Immortalized cell lines	n.a.
hES derived embryoid bodies (Ebs)	Human	Stem/Progenitor cells	n.a.
hES derived neural progenitors	Human	Stem/Progenitor cells	n.a.
hES derived neurons	Human	Stem/Progenitor cells	n.a.
hESC	Human	Stem/Progenitor cells	n.a.
hESC (H1 cell line)-derived hNSC	Human	Stem/Progenitor cells	n.a.
hESC (H9 line WA09)	Human	Stem/Progenitor cells	n.a.
hESC (H9 line WA09)-derived embroid bodies	Human	Stem/Progenitor cells	n.a.
hESC (H9 line WA09)-derived neural rosettes	Human	Stem/Progenitor cells	n.a.
hESC (H9 line WA09)-derived neuroepithelial precursors	Human	Stem/Progenitor cells	n.a.
hESC (H9 line WA09)-derived NPC differentiation	Human	Stem/Progenitor cells	n.a.
hESC (H9)	Human	Stem/Progenitor cells	n.a.
hESC (HUES-1 cell line)	Human	Stem/Progenitor cells	n.a.
hESC line KhES-3 (xy genotype)	Human	Stem/Progenitor cells	n.a.
hN2	Human	Tumor/Immortalized cell lines	n.a.
hNPC	Human	Stem/Progenitor cells	whole Brain
Human astrocytes	Human	Primary cells	whole Brain
Human cortical	Human	Primary cells	Cortex

neurons			
Human neuroblast long term cell line	Human	Primary cells	Olfactory neuroepithelium
Human primary astrocytes	Human	Primary cells	whole Brain
Human reaggregated brain cells	Human	Primary cells	whole Brain
human umbilical cord blood neural stem cells	Human	Stem/Progenitor cells	n.a.
human umbilical cord blood stem cells-derived neuronal cells	Human	Stem/Progenitor cells	n.a.
LUHMES cells	Human	Tumor/Immortalized cell lines	Mesencephalon
mES cell line E14Tg2a	Mouse	Stem/Progenitor cells	n.a.
mESC (B6G-2)	Mouse	Stem/Progenitor cells	n.a.
mESC line B6G-2 (xy genotype), RIKEN	Mouse	Stem/Progenitor cells	n.a.
mESC line Royan B1	Mouse	Stem/Progenitor cells	n.a.
mESC NVRQS-11F	Mouse	Stem/Progenitor cells	n.a.
mESC, C57BL/6 line Royan B1	Mouse	Stem/Progenitor cells	n.a.
mESC, C57BL/6-C2	Mouse	Stem/Progenitor cells	n.a.
mESC, CGR8	Mouse	Stem/Progenitor cells	n.a.
mESC, D3	Mouse	Stem/Progenitor cells	n.a.
mNPC	Mouse	Stem/Progenitor cells	whole Brain
mNPC ICR mice telencephalon	Mouse	Stem/Progenitor cells	Telecephalon
Mouse cereballar astrocytes	Mouse	Primary cells	Cerebellum
Mouse cereballar granule neurons	Mouse	Primary cells	Cerebellum
Mouse cerebellar granule cells	Mouse	Primary cells	Cerebellum
Mouse cerebellar microwell cultures	Mouse	Primary cells	Cerebellum
Mouse cerebellar neurons	Mouse	Primary cells	Cerebellum
Mouse cerebellar slice cultures	Mouse	Primary cells	Cerebellum
Mouse cortical	Mouse	Primary cells	Cortex

astrocytes			
Mouse cortical astroglia	Mouse	Primary cells	Cortex
Mouse cortical glia cells	Mouse	Primary cells	Cortex
Mouse cortical neural precursor cells - C57BL/6J	Mouse	Stem/Progenitor cells	Cortex
Mouse cortical neural stem cells (Millipore)	Mouse	Stem/Progenitor cells	Cortex
Mouse cortical neurons	Mouse	Primary cells	Cortex
mouse cortical neurons, E17	Mouse	Primary cells	Cortex
Mouse cortical slice cultures	Mouse	Primary cells	Cerebral Cortex
Mouse hippocampal astrocytes	Mouse	Primary cells	Hippocamus
Mouse hippocampal cells	Mouse	Primary cells	Hippocampus
Mouse hippocampal neurons	Mouse	Primary cells	Hippocamus
Mouse hippocampal slice cultures	Mouse	Primary cells	Hippocamus
Mouse neocortical neurons	Mouse	Primary cells	Neocortex
Mouse neural crest derived cells	Mouse	Stem/Progenitor cells	Neural crest
Mouse neuronal and glial cocultures	Mouse	Primary cells	Cerebral Cortex
Mouse reaggregate cultures	Mouse	Primary cells	Striatum and mesencephalon
mouse SFME cells induced to astrocytes	Mouse	Primary cells	n.a.
Mouse striatal neurons	Mouse	Primary cells	Striatum
Mouse ventral midbrain cells	Mouse	Primary cells	Midbrain
Mouse whole embryo culture	Mouse	Primary cells	n.a.
N2a	Mouse	Tumor/Immortalized cell lines	n.a.
NB69	Human	Tumor/Immortalized cell lines	n.a.

NS-1 cells (subclone of PC-12)	Rat	Tumor/Immortalized cell lines	n.a.
NS-5 cells generated from 1290la mouse ESC	Mouse	Stem/Progenitor cells	n.a.
NT-2	Human	Tumor/Immortalized cell lines	n.a.
NT-3	Human	Tumor/Immortalized cell lines	n.a.
Oligodendrocyte progenitor cells	Rat	Primary cells	Optic nerves
P19 cells	Mouse	Tumor/Immortalized cell lines	n.a.
PC-12	Rat	Tumor/Immortalized cell lines	n.a.
Rat astroglia cultures	Rat	Primary cells	Forebrain
Rat brain-derived progenitor cells	Rat	Stem/Progenitor cells	cerebrum
Rat brainstem astrocytes	Rat	Primary cells	Brainstem
Rat brainstem cells	Rat	Primary cells	Brainstem
Rat cerebellar astrocytes	Rat	Primary cells	Cerebellum
Rat cerebellar granule precursor cells	Rat	Primary cells	Cerebellum
Rat cerebellar neurons	Rat	Primary cells	Cerebellum
Rat cerebellar slice cultures	Rat	Primary cells	Cerebellum
Rat cerebral cortex microwell cultures	Rat	Primary cells	Cerebral Cortex
Rat CG-4 cells	Rat	Tumor/Immortalized cell lines	n.a.
Rat CGC	Rat	Primary cells	Cerebellum
Rat cortical astrocytes	Rat	Primary cells	Cortex
Rat cortical cells	Rat	Primary cells	Cortex
rat cortical neurons	Rat	Primary cells	Cortex
Rat cortical neurospheres	Rat	Stem/Progenitor cells	Cortex
Rat cortical oligodendrocytes	Rat	Primary cells	Cortex
Rat cortical precursor cells	Rat	Stem/Progenitor cells	Cortex

rat cortical slice cultures	Rat	Primary cells	Cortex
Rat dorsal root ganglia neurons	Rat	Primary cells	Dorsal Root Ganglion
Rat embryonic cortical NSC	Rat	Stem/Progenitor cells	Cortex
Rat fetal midbrain neurons	Rat	Primary cells	Midbrain
rat forebrain slices	Rat	Primary cells	Forbrain
Rat frontal cortex	Rat	Primary cells	Cortex
Rat ganglionic neurospheres	Rat	Stem/Progenitor cells	ganglionic eminence
Rat hippocampal astrocytes	Rat	Primary cells	Hippocampus
Rat hippocampal neurons	Rat	Primary cells	Hippocamus
Rat hippocampal neurons/cortical astrocytes coculture	Rat	Primary cells	multiple
Rat hippocampal slice cultures	Rat	Primary cells	Hippocamus
Rat hippocampus microwell cultures	Rat	Primary cells	Hippocamus
Rat hypothalamic cells	Rat	Primary cells	Hypothalamus
Rat locus coeruleus neurons	Rat	Primary cells	Locus Coeruleus
Rat medial septal Neurons	Rat	Primary cells	Medial septum
Rat midbrain cells	Rat	Primary cells	Midbrain
Rat mixed cortical cultures	Rat	Primary cells	Cortex
Rat neocortical coronal slices	Rat	Primary cells	Neocortex
Rat neocortical neurons	Rat	Primary cells	Neocortex
Rat neonatal forebrain cultures	Rat	Primary cells	Forbrain
Rat neopallium cells	Rat	Primary cells	Neocortex
Rat NSC	Rat	Stem/Progenitor cells	Hippocampus
Rat Oligodendrocyte Progenitor cells	Rat	Stem/Progenitor cells	Cerebral hemispheres
Rat primary glial cell culture	Rat	Primary cells	Cortex

Rat primary oligodendrocytes	Rat	Primary cells	whole Brain
Rat re-aggregate cultures	Rat	Primary cells	Telecephalon
Rat striatal neurospheres	Rat	Stem/Progenitor cells	Striatum
Rat substantia nigra neurons	Rat	Primary cells	Locus Coeruleus
Rat subventricular zone cells	Rat	Primary cells	Subventricular zone
Rat superior cervical ganglia neurons	Rat	Primary cells	Cervical Ganglia
Rat ventral mesencephalic neurosphers	Rat	Stem/Progenitor cells	Mesencephalon
Rat ventral mesencephalon cells	Rat	Primary cells	Mesencephalon
Rat whole brain	Rat	Primary cells	whole Brain
ReN CX cells (Millipore)	Human	Tumor/Immortalized cell lines	Cortex
rNPC	Rat	Stem/Progenitor cells	Cortex
SH-SY5Y	Human	Tumor/Immortalized cell lines	n.a.

Appendix K. Data collection sheet


DCS in vitro.xlsx


DCS in vivo.xlsx


DCS AO.xlsx


DCS in silico.xlsx

Appendix L. Definition of term 'validation'

The term 'validation' can be used in different contexts. For this project, mainly distinguishing formal EURL-ECVAM validation from other types of scientific validation is necessary. EURL-ECVAM formal validation (http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvams-validation-process) is at the interface between test method development/optimisation and regulatory acceptance/international recognition and ensures a science-based and conscientious evaluation of test methods and approaches (e.g. Integrated Testing Strategies, ITS), independent of specific interests, establishing their overall performance and fitness for a given purpose. Therefore, a time- and labour-intensive four-step procedure has to be followed: 1) Assessment of test method submissions taking stakeholder and ICATM (International Collaboration on Alternative Methods) input into account; 2) Planning and conduct of validation studies, possibly in collaboration with ECVAM NETVAL (Network of Validation Laboratories) laboratories; 3) Coordination of independent scientific peer review by the EURL ECVAM Scientific Advisory Committee (ESAC); 4) Development of EURL ECVAM recommendations on the validity status of test methods taking stakeholder and ICATM input into account. For DNT testing, there is no existing method that has undergone formal EURL-ECVAM validation.

However, the term 'validation' can also be applied in a different context, i.e. 'scientific validation'.This type of validation approaches validity of an alternative test method from a different perspective by addressing important questions and gaps frequently arising by employing the adverse outcome pathway (AOP) concept. Questions, which might be addressed by scientific experiments, include the topics of *in vitro – in vivo* extrapolation, species differences, cell – organ transfer and concentration/dose extrapolation: Which signalling pathways are involved in neurodevelopment/DNT with regard to specific neurodevelopmental processes (e.g. Stansfield et al., 2012)? Do the *in vitro* methods thereby reflect physiology *in vivo*? Are there species differences involved or are these pathways conserved across species (Gassmann et al., 2010, Jiang et al., 2010b)? How do the pathways and adverse outcomes on the cellular level translate to the organ and the organism? At what internal doses is adversity observed *in vivo* and how can this be transferred into *in vitro* concentrations? By specifically addressing such questions, trust of data generated by applying alternative methods will be enhanced. This will lead to a reduction in uncertainty in the regulatory context as trust in a method is based on valid scientific criteria.