Critical Reviews in Toxicology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/itxc20

Towards a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny—part III: how is substance-mediated thyroid hormone imbalance in pregnant/lactating rats or their progeny related to neurodevelopmental effects?

M. Sue Marty, Ursula G. Sauer, Alex Charlton, Rashin Ghaffari, Davy Guignard, Nina Hallmark, Bethany R. Hannas, Sylvia Jacobi, Heike-Antje Marxfeld, Stephanie Melching-Kollmuss, Larry P. Sheets, Daniel Urbisch, Philip A. Botham & Bennard van Ravenzwaay

To cite this article: M. Sue Marty, Ursula G. Sauer, Alex Charlton, Rashin Ghaffari, Davy Guignard, Nina Hallmark, Bethany R. Hannas, Sylvia Jacobi, Heike-Antje Marxfeld, Stephanie Melching-Kollmuss, Larry P. Sheets, Daniel Urbisch, Philip A. Botham & Bennard van Ravenzwaay (2022) Towards a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny—part III: how is substance-mediated thyroid hormone imbalance in pregnant/lactating rats or their progeny related to neurodevelopmental effects?, Critical Reviews in Toxicology, 52:7, 546-617, DOI: 10.1080/10408444.2022.2130166

To link to this article: https://doi.org/10.1080/10408444.2022.2130166

CRITICAL REVIEWS IN TOXICOLOGY
2022, VOL. 52, NO. 7, 546–617
https://doi.org/10.1080/10408444.2022.2130166

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

Towards a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny—part III: how is substance-mediated thyroid hormone imbalance in pregnant/lactating rats or their progeny related to neurodevelopmental effects?

M. Sue Marty[a], Ursula G. Sauer[b], Alex Charlton[c], Rashin Ghaffari[d], Davy Guignard[e]*, Nina Hallmark[f], Bethany R. Hannas[d]*, Sylvia Jacobi[g], Heike-Antje Marxfeld[h], Stephanie Melching-Kollmuss[i], Larry P. Sheets[j], Daniel Urbisch[i]*, Philip A. Botham[c] and Bennard van Ravenzwaay[k]

[a]Dow, Inc., Midland, MI, USA; [b]Scientific Consultancy–Animal Welfare, Neubiberg, Germany; [c]Syngenta, Jealott's Hill, Bracknell, UK; [d]Corteva Agriscience, Newark, DE, USA; [e]Bayer CropScience, Sophia Antipolis, France; [f]Bayer AG, Monheim, Germany; [g]Albemarle, Louvain-la-Neuve, Belgium; [h]BASF SE, Ludwigshafen, Germany; [i]BASF SE, Limburgerhof, Germany; [j]Bayer CropScience, Chesterfield, MO, USA; [k]Environmental Sciences Consulting, Altrip, Germany

ABSTRACT

This review investigated which patterns of thyroid- and brain-related effects are seen in rats upon gestational/lactational exposure to 14 substances causing thyroid hormone imbalance by four different modes-of-action (inhibition of thyroid peroxidase, sodium-iodide symporter and deiodinase activities, enhancement of thyroid hormone clearance) or to dietary iodine deficiency. Brain-related parameters included motor activity, cognitive function, acoustic startle response, hearing function, periventricular heterotopia, electrophysiology and brain gene expression. Specific modes-of-action were not related to specific patterns of brain-related effects. Based upon the rat data reviewed, maternal serum thyroid hormone levels do not show a causal relationship with statistically significant neurodevelopmental effects. Offspring serum thyroxine together with offspring serum triiodothyronine and thyroid stimulating hormone appear relevant to predict the likelihood for neurodevelopmental effects. Based upon the collated database, thresholds of ≥60%/≥50% offspring serum thyroxine reduction and ≥20% and statistically significant offspring serum triiodothyronine reduction indicate an increased likelihood for statistically significant neurodevelopmental effects; accuracies: 83% and 67% when excluding electrophysiology (and gene expression). Measurements of brain thyroid hormone levels are likely relevant, too. The extent of substance-mediated thyroid hormone imbalance appears more important than substance mode-of-action to predict neurodevelopmental impairment in rats. Pertinent research needs were identified, e.g. to determine whether the phenomenological offspring thyroid hormone thresholds are relevant for regulatory toxicity testing. The insight from this review shall be used to suggest a tiered testing strategy to determine whether gestational/lactational substance exposure may elicit thyroid hormone imbalance and potentially also neurodevelopmental effects.

ARTICLE HISTORY
Received 15 April 2022
Revised 21 September 2022
Accepted 22 September 2022

KEYWORDS
Thyroxine (T4); triiodothyronine (T3); developmental neurotoxicity (DNT) study (OECD TG 426); motor activity; acoustic startle response; cognitive function (learning and memory); extended one generation reproductive toxicity study (EOGRTS; OECD TG 443); thyroid peroxidase (TPO); sodium-iodide symporter (NIS); deiodinase (DIO); uridine diphosphate glucuronyltransferase (UGT); thyroid hormone clearance; adverse outcome pathway (AOP); mode-of-action (MoA)

Table of contents

CONTACT M. Sue Marty msmarty@dow.com Toxicology and Environmental Research and Consulting, Dow, Inc., 1803 Building, Midland, MI 48667, USA
*__Present address__: Davy Guignard, Evotec France, Toulouse, France; Bethany R. Hannas, Eli Lilly & Co., Indianapolis, IN, USA; Daniel Urbisch, Syngenta, Basel, Switzerland

Supplemental data for this article is available online at https://doi.org/10.1080/08923973.2022.2078727.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

1. Introduction

1.1. Background

The thyroid gland plays a critical role in mammalian neurodevelopment, just as generally in reproduction and development. During neurodevelopment, the brain levels of the thyroid hormones thyroxine (T4) and triiodothyronine (T3) are controlled both spatially and temporally to support normal patterns of neuronal migration, neuronal differentiation and glial myelination. If serum thyroid hormone levels are altered beyond homeostasis (i.e. if they are either too low or too high), the effect on the structure and function of the brain depends on the magnitude, duration and timing of the thyroid perturbation (Korevaar et al. 2016; Lazarus 2016; Prezioso et al. 2018; Huget-Penner and Feig 2020). The foetus is fully dependent on maternal thyroid hormone until its thyroid gland starts producing thyroid hormone, which takes place around gestational day (GD) 17 in rats (Perez-Castillo et al. 1985) and towards the end of the first trimester in humans (Thorpe-Beeston et al. 1991). Also, the rat foetus remains partially dependent until delivery (Morreale de Escobar et al. 1990; Grijota-Martinez et al. 2011).

Thyroid perturbations can result in a variety of neurodevelopmental alterations. The stage of development when the foetus or newborn is exposed to low or high serum thyroid hormone levels is critical for the onset of the type of neurodevelopmental alteration (Marty et al. 2021). In rat studies, among the most commonly measured phenotypic changes are altered learning and memory, altered motor activity, ototoxicity and neuronal migration errors. Similarly, human observational studies have shown that thyroid hormone imbalance in pregnant mothers may lead to reduced intelligence quotient and impaired learning and memory, motor deficits as well as other neurobehavioural alterations including attention deficits and hyperactivity in the child (Henrichs et al. 2010; Julvez et al. 2013; Korevaar et al. 2016; Andersen et al. 2018; Nelson et al. 2018; Spann et al. 2020). While all these deficits can also be caused by factors unrelated to thyroid function, physiological maternal serum thyroid hormone levels play an important role to ensure normal neurodevelopment.

Thus, evaluations of thyroid function during offspring development can be relevant for toxicological assessments. This is also reflected in the European Food Safety Authority (EFSA) and European Chemicals Agency (ECHA) *Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009* (EFSA and ECHA 2018). This guidance, which applies to substances regulated under the Biocidal Products Regulation (EP and Council 2012) and the Plant Protection Products Regulation (EP and Council 2009), includes an Appendix that specifically addresses thyroid hormone imbalance, i.e. *Appendix A—Additional considerations for how to assess the potential for thyroid disruption for human health*.

Various statements in this Appendix A suggest that low total or free maternal serum T3 and/or T4 with or without histopathological changes in the thyroid gland presents a concern for pre- and postnatal neurodevelopment. Hence, this approach focuses on serum thyroid hormone decreases rather than subsequent increases in thyroid stimulating hormone (TSH) levels, which may affect the structure of the thyroid gland (Huisinga et al. 2020). This focus can be explained by evidence that serum thyroid hormones can be decreased after *in utero*/lactational substance exposure, potentially affecting neurodevelopment, without corresponding changes in TSH (see e.g. Vansell and Klaassen 2001; Gilbert et al. 2021). Further, Appendix A suggests that, in the absence of substance-specific data, humans and rodents are considered to be equally sensitive to thyroid

disruption. The report of the European Commission (2017a) Thyroid Disruption Workshop is used as the main reference in Appendix A to substantiate these views. Appendix A of EFSA and ECHA (2018) very generally describes a testing scheme to determine the serum levels of thyroid hormones and the activities of specific liver enzymes whose induction may increase thyroid hormone clearance, and to exclude specific thyroid modes-of-action (MoAs). However, neither Appendix A nor the further clarifications that have since been provided by EFSA (2020a) provide broader guidance covering different thyroid-related MoAs or indicate how the data should be evaluated in a weight-of-evidence approach to reach a conclusion on whether, or not, a substance meets the European Commission (2017b, 2018) Endocrine Disruptor Criteria. In this regard, Appendix A recognises that the identification of thyroid-related hazards is currently hampered by a lack of internationally validated test methods. Overall, based upon the provisions of Appendix A of EFSA and ECHA (2018) and the further clarifications provided by EFSA (2020a), it is currently unclear how specific thyroid-related MoAs should be identified, and how the human relevance of thyroid effects and/or developmental neurotoxicity (DNT) observed in rats should be established (see also reviews by Gilbert et al. 2012, 2020; Kortenkamp et al. 2020).

1.2. Work by the ECETOC T4 Task Force

To address these uncertainties, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) has convened the T4 Task Force (TF). It is the overarching goal of this TF to review the available evidence in order to contribute to the development of a science-based tiered testing strategy to identify (1) if a substance has the ability to elicit maternal thyroid hormone imbalance and potentially also neurodevelopmental effects in the progeny; (2) if effects observed in rodents are relevant for humans in line with the European Commission (2017b, 2018) Endocrine Disruptor Criteria; and (3) if a threshold for thyroid hormone decrements can be identified below which neurodevelopmental effects would not be expected. In pursuing the different elements of this overarching goal, the ECETOC T4 TF is preparing a series of four reviews, of which the present article constitutes the third.

The first review (Sauer et al. 2020) examined the human evidence for neurodevelopmental impairment secondary to maternal thyroid hormone imbalance to identify parameters that are relevant in humans and could be adapted to toxicological assessments. In pregnant mothers, serum levels of free thyroxine (fT4) and TSH were the most frequently measured thyroid-related parameters (see e.g. Pop et al. 1999; Henrichs et al. 2010; Julvez et al. 2013; Päkkilä et al. 2015; Korevaar et al. 2016; Andersen et al. 2018; Nelson et al. 2018; Spann et al. 2020), but total T4 was also assessed (Oken et al. 2009; Lain et al. 2020). Depending on the study, neurodevelopmental assessments in the children included psychomotor and mental development (Pop et al. 1999; Julvez et al. 2013), cognitive function (Oken et al. 2009; Korevaar et al. 2016; Lain et al. 2020), expressive vocabulary (Henrichs et al. 2010) or educational attainment (Päkkilä et al. 2015; Nelson et al. 2018), and, in single studies, brain morphology assessed by magnetic resonance imaging (Korevaar et al. 2016) or clinical diagnoses of, e.g. autism or attention deficit hyperactivity disorder (Andersen et al. 2018; Spann et al. 2020). The human data are overall in support of an association between low maternal serum fT4 (and in some studies also high TSH) and increased risk for neurodevelopmental impairment. Further, the human data showed that increased maternal thyroid hormone levels can also affect child neurodevelopment (Korevaar et al. 2016). However, the available evidence did not allow identifying the most sensitive parameter(s) to assess effects in the pregnant mothers or children, nor quantitative boundaries of effects indicative of increased risks. Also, none of the human studies allowed establishing a link between substance-mediated liver enzyme induction and increased thyroid hormone clearance, and the impact of these on child neurodevelopment (Sauer et al. 2020).

The second review prepared by the ECETOC T4 TF, Marty et al. (2021), evaluated thyroid-related adverse outcome pathways (AOPs) potentially leading to adverse neurodevelopmental outcomes in mammals. The focus of that review was on AOPs that are included in the Organisation for Economic Co-operation and Development (OECD) AOP Wiki (https://aopwiki.org [accessed 2022 August]) while further considering AOPs that form part of published AOPs networks (Knapen et al. 2018; Villeneuve et al. 2018; Noyes et al. 2019). Marty et al. (2021) identified relevant molecular initiating events (MIEs), key events and key event relationships and determined the strength of evidence for the sequence of key events and for the key event relationships. The underlying hypothesis was that this information provides insight into differences in incidence, magnitude or species sensitivity of the respective neurodevelopmental adverse outcomes. Relevant MIEs included:

1. AOP 42: Inhibition of thyroid peroxidase (TPO), as a MIE that affects thyroid hormone synthesis (Crofton et al. 2019)
2. AOP 54 and AOP 134: Inhibition of the sodium-iodide symporter (NIS), as a further MIE that affects thyroid hormone synthesis (Rolaki et al. 2019)
3. AOP 152: Displacement of thyroid hormones from serum binding proteins (transthyretin, thyroid binding protein, albumin) and AOP 8: activation of pregnane X receptor leading to induction of hepatic uridine diphosphate glucuronyltransferases (UGTs) that mediate thyroid hormone metabolism, as two MIEs that affect serum levels of (free) thyroid hormones *via* enhanced thyroid hormone clearance (Villeneuve et al. 2018; Noyes et al. 2019)
4. Inhibition of deiodinases (DIOs), as a MIE that affects local concentrations of thyroid hormones in relevant tissues (Villeneuve et al. 2018; Noyes et al. 2019)
5. Inhibition of thyroid receptor transcription, as a further MIE affecting local effects of thyroid hormones in relevant tissues (Noyes et al. 2019)

Three different neurodevelopmental adverse outcomes were included in the AOPs, i.e.

1. AOP 8: Decreased cochlear function
2. AOP 42, AOP 134, AOP 152: Decreased cognitive function
3. AOP 54: Impaired learning and memory

Following up from the evidence on thyroid-related AOPs with adverse neurodevelopmental outcomes, Marty et al. (2021) then established if rodent toxicity studies conducted in accordance with standardised test protocols, such as OECD Test Guidelines (TGs; Table 1(A)), coupled with *in vitro* assays allow for identification of thyroid-related MoAs and the human relevance of effects—in line with the European Commission (2017b, 2018) Endocrine Disruptor Criteria. The standard rat toxicity studies include measurements of serum thyroid hormones, i.e. the shared key event common to most AOPs within thyroid-related AOP networks (Villeneuve et al. 2018), thyroid gland pathology and neurodevelopmental assessments (Table 1(B)). However, the standard rat toxicity studies do not directly inform on specific MoAs. Marty et al. also evaluated additional non-routine parameters reflecting critical events of AOPs that might either be added to rat studies (e.g. serum fT4, tissue levels of thyroid hormones) or addressed *in vitro* (e.g. TPO inhibition, NIS inhibition). Such non-routine parameters appear relevant to support the identification of specific thyroid-related MoAs, provided that prevailing technical limitations (e.g. with respect to the timing of measurements and method availability) are overcome. Importantly, Marty et al. showed that the current understanding of quantitative key event relationships is often weak. Nonetheless, such understanding would be needed to determine if the triggering of a MIE will ultimately result in an adverse outcome. Also, significant species differences in all processes related to thyroid hormone function are evident, but the biological implications thereof are often unknown (Marty et al. 2021).

Taken together, the first two reviews by the ECETOC T4 TF (Sauer et al. 2020; Marty et al. 2021) highlighted knowledge gaps related to both human evidence and extrapolation from rodents to humans that complicate the selection and measurement of thyroid-related and/or neurodevelopmental parameters for toxicological studies and the establishment of thyroid-related MoAs. In this regard, it is also unclear whether those endpoints that are commonly addressed in standard rat toxicity studies (e.g. learning/behaviour, acoustic startle, brain morphometry, histopathology) are transferable to specific neurodevelopmental effects in humans (e.g. reduced intelligence quotient, neurobehavioural alterations). Furthermore, it is still unclear how the simultaneous action of various MoAs may affect thyroid function and hence a substance's potential to lead to adverse neurodevelopmental outcomes in rodents and/or humans.

1.3. Scope of the present review

Following the outcomes of the first two reviews, the present third review by the ECETOC T4 TF evaluates the current understanding on how substance-mediated thyroid hormone imbalance in pregnant/lactating rats and/or their pups relates to neurodevelopmental effects. Four case studies are presented, covering rat studies on a total of fourteen thyroid-active substances (and dietary iodine deficiency) that trigger the MIEs for the major thyroid-related AOPs/MoAs in mammals (Marty et al. 2021). Hence, "case study" is defined here as a group of studies evaluating substances that share one or more common (primary) MoAs.

The rat was selected as target species since it is the typical species used in mammalian toxicity studies. Accordingly, the evaluation focusses on standard, TG-compliant rat toxicity studies and non-standard investigational studies using rats, which included gestational and/or lactational exposure to substances that are known to cause thyroid hormone imbalance in rats. Both routine and non-routine endpoints are included in the evaluation with consideration of current proficiencies, feasibility and reproducibility.

The goal of this third review is to establish how thyroid-related effects in the dams and offspring are linked to neurodevelopmental toxicological profiles in rat studies with gestational and/or lactational substance exposure. Four questions were addressed in pursuing this goal:

1. Which specific neurodevelopmental effects are observed in rat studies upon gestational and/or lactational exposure

Table 1(A). Potentially relevant test guidelines.

Test guideline number/ identification	Name of test guideline	Comment
OECD TG 443	Extended one-generation reproductive toxicity study (EOGRTS)	Only formally adopted standard rat toxicity study that includes both mandatory parameters related to thyroid function and to DNT
OECD TG 426	Developmental neurotoxicity (DNT) study	Includes a spectrum of DNT endpoints but no mandatory parameters related to thyroid function; OECD TG 426 widely concordant with US EPA TG 870.6300
OECD TG 416	Two-generation reproduction toxicity study	No mandatory parameters related to thyroid function or DNT; only relevant for present review when enhanced, i.e. when endpoints related to thyroid function and/or DNT have been added
OECD TG 415	One-generation reproduction toxicity study	
US EPA (2005)	Comparative thyroid assay	Has not been adopted as a formal TG; includes thyroid parameters in adults and offspring, but no specific DNT parameters apart from clinical observations related to brain function

DNT: developmental neurotoxicity; EPA: Environmental Protection Agency; OECD: Organisation for Economic Co-operation and Development; TG: test guideline.

Table 1(B). Tests commonly used to measure developmental neurotoxicity (adapted from OECD 2004).

Parameter	Examples for commonly used tests
Motor activity	Grip strength, rotating rod, hindlimb foot splay / landing foot spread, motor activity
Sensory function and acoustic startle response	Nociception, sensory irritation, somatosensory operant discrimination task, acoustic startle response and pre-pulse inhibition, auditory discrimination procedure, visual discrimination task
Cognitive function	Habituation, ethologically based anxiety tests (e.g. elevated plus maze test), conditioned taste aversion, active avoidance, passive avoidance, spatial mazes (Morris water maze, Biel water maze, T-maze, conditional / delayed discrimination)

to substances that are known to cause thyroid hormone imbalance in rats via different MoAs?

2. When are altered thyroid hormone levels indicative of potential concern for neurodevelopmental effects in rats (and hence indicative of the need for further testing)?
   a. Is it sufficient to measure maternal serum thyroid hormone levels during gestation and/or lactation, or are foetal and/or pup serum thyroid hormone measurements necessary?
   b. Is there an added value of monitoring serum (free or total) T3, T4, and TSH in the dams and/or offspring, or is serum total T4 sufficient?
   c. Is there an added value in measuring brain thyroid hormone levels in foetuses or pups? If so, can a specific period be determined during which measurement of brain thyroid hormone is predictive of neurodevelopmental outcomes?
   d. Is it possible to identify thresholds for specific thyroid-related parameters that are indicative of neurodevelopmental or other brain-related findings?
   e. Are thyroid weight and thyroid histopathology in the dams and/or offspring sensitive parameters to indicate potential concern for neurodevelopmental effects?
3. Are similar changes in serum thyroid hormone magnitude during critical periods of development associated with the same neurodevelopmental outcomes in rats regardless of the MoA of the test compound? Is it possible to identify patterns of thyroid- and brain-related effects for substances with different MoAs?
4. How should systemic toxicity (e.g. body weight effects, organ toxicity, developmental delays, i.e. delays in the attainment of landmark structural or behavioural endpoints) be considered in rat studies in the evaluation of thyroid hormone imbalance and potential neurodevelopmental effects?

To answer these questions, this review encompasses the following sections:

- Section 2: Definition of four case studies, overview of selected case study substances and of selected rat studies, details on data extraction
- Section 3: Presentation of the brain- and thyroid-related findings from the case studies
- Section 4: Overarching discussion of the findings from the case studies
- Section 5: Conclusions: wrap-up of all findings to answer the questions presented above

It is important to note that the present review does not evaluate the human relevance of the thyroid- or brain-related findings recorded in the rat studies. This topic will be addressed in the planned fourth review of the ECETOC T4 TF review series. Therein, all findings from this review and from the reviews by Sauer et al. (2020) and Marty et al. (2021) will be consolidated to propose a science-based tiered testing strategy. This testing strategy shall serve to identify if a substance could elicit maternal/foetal/pup thyroid hormone imbalance and potentially also neurodevelopmental effects in the progeny as well as the human relevance of findings in line with the European Commission (2017b, 2018) Endocrine Disruptor Criteria and the EFSA and ECHA (2018) Endocrine Disruptor Guidance.

2. Definition of case studies, selection of case study substances, and selection of representative studies

2.1. Definition of case studies and selection of case study substances

The ECETOC T4 TF defined four case studies to cover the major thyroid-related AOPs/MoAs in mammals (Marty et al. 2021) and assigned fourteen thyroid-active substances (as well as dietary iodine deficiency) to these case studies. This assignment followed the state-of-the-science that the substances trigger the respective MIEs and hence exhibit the corresponding MoA(s) as sole or predominant MoA(s):

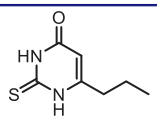
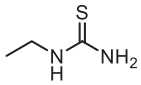
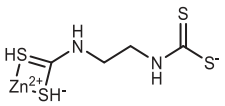
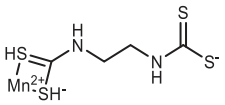
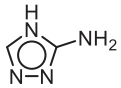
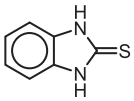
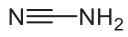
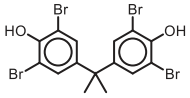
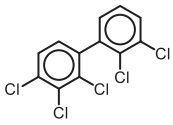
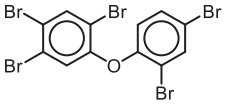
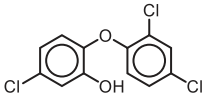
- *Case Study 1—Impairment of thyroid hormone synthesis via TPO inhibition*: propylthiouracil (PTU), methimazole (MMI), ethylene thiourea (ETU), mancozeb, mercaptobenzimidazole, amitrole and cyanamide
- *Case Study 2—Impairment of thyroid hormone synthesis via reduced iodine bioavailability in the thyroid gland:* perchlorate (NIS inhibition) and dietary iodine deficiency
- *Case Study 3—Enhancement of thyroid hormone clearance via displacement of thyroid hormone from serum binding proteins and/or via induction of liver enzymes that metabolise thyroid hormone*: tetrabromobisphenol A (TBBPA), perfluorohexane sulphonates (PFHxS), Aroclor 1254 [a polychlorinated biphenyl (PCB)], DE-71 [a mix of polybrominated diphenyl ethers (PBDEs)] and triclosan
- *Case Study 4—Other MoA(s) including DIO inhibition*: octyl methoxycinnamate (OMC), a substance whose exact MoA is still unclear but that has been observed to reduce both serum T4 levels and hepatic DIO1 activity (Schmutzler et al. 2004)

Hence, Case Studies 1 and 2 address MoAs reflecting direct action on thyroid hormone synthesis and function, and Case Studies 3 and 4 address MoAs reflecting indirect actions on thyroid hormone transport and function (*via* enhanced thyroid hormone clearance and/or local interaction at the target site).

Table 2 provides an overview of the case study substances and of their major uses. Appendix I, which is included in this paper after the bibliography, explains the rationale to define the four case studies and presents the evidence to assign each case study substance to the given case study (see Table 3 for link to the corresponding Appendix I sections). Further, the Supplementary Information *Table SI-1, Spreadsheet "Substance Phys-Chem,"* which is available online, presents some physicochemical properties of the case study substances.

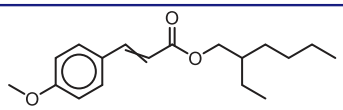
Substance-mediated inhibition of thyroid receptor binding, which is also a MIE for a thyroid-related AOP with adverse neurodevelopmental outcome in mammals (Noyes et al. 2019; Marty et al. 2021), was not considered in the present review since a comprehensive study by Paul Friedman et al. (2019) supported the conclusion that the thyroid receptor is likely not a relevant target for most environmental chemicals. To the best of the ECETOC T4 TF's knowledge, thyroid receptor transcription (as well as factors affecting it) is an area of emerging science

Table 2. Overview of the case study substances: CAS numbers, chemical structures, and major uses.

Substance	CAS number	Chemical structure	Major uses
Case Study 1			
Propylthiouracil (PTU)	51-52-5		Used therapeutically to treat hyperthyroidism, e.g. as caused by Graves' disease (Nakamura et al. 2007)
Methimazole (MMI)	60-56-0		
Ethylene thiourea (ETU)	625-53-6/		ETU is a high-production volume chemical used in the rubber industry and to produce some fungicides (US EPA 2000); ETU is a metabolite of ethylene bis-dithiocarbamate herbicides (e.g. mancozeb) in rats; therefore, studies investigating ETU or mancozeb are considered together in this review
Mancozeb	8018-01-7		
			
Amitrole	61-82-5		Terrestrial non-food herbicide used primarily in industrial outdoor areas and for non-agricultural and ornamental plants (US EPA 1996)
Mercaptobenzimidazole	583-39-1		Used to produce plastics and rubber products; https://pubchem.ncbi.nlm.nih.gov/compound/2-Mercaptobenzimidazole [accessed 2022 August]
Cyanamide	420-04-2		Used to produce, e.g. agrochemicals related to fruit crops and the preservation of fruit, colourants and drugs; https://pubchem.ncbi.nlm.nih.gov/compound/cyanamide [accessed 2022 August]
Case Study 2			
Perchlorate	14797-73-0		Commonly used as an oxidiser in solid propellants, munitions, fireworks, airbag initiators for vehicles, matches and signal flares; also used in some electroplating operations and found in some disinfectants and herbicides (US EPA 2014)
Case Study 3			
Tetrabromobisphenol A (TBBPA)	79-94-7		One of the most widely used brominated flame retardants; extensively used in electronic equipment, furniture, plastics and textiles (Zhou et al. 2020)
Perfluoro hexane sulphonates (PFHxS)	3871-99-6		Six-carbon perfluoroalkyl sulphonic acids were used as industrial surfactants or in aqueous film-forming foams (Ali et al. 2019); due their extreme persistence, the use of PFHxS and other perfluorochemicals has been discontinued (US EPA 2015)
Aroclor 1254, a PCB	52663-62-4		Formerly used, e.g. as plasticiser and adhesive (https://pubchem.ncbi.nlm.nih.gov/compound/Aroclor-1254); use of PCBs has been banned many decades ago in the European Union (https://www.efsa.europa.eu/en/topics/topic/dioxins-and-pcbs) and in the USA (https://www.epa.gov/pcbs/learn-about-polychlorinated-biphenyls-pcbs) [all: accessed 2022 August]
DE-71, a mixture of PBDEs	60348-60-9/EPA Chemical No. 32534-81-9[a]		Use of PBDEs: flame retardants in, e.g. plastics, furniture, upholstery, electrical equipment, electronic devices, textiles (US EPA 2017); European Union: use in electrical and electronic equipment restricted, maximum concentration of 0.1% by weight tolerated in homogeneous materials (EP and Council 2011); USA: most common PBDEs have exposure restrictions, some also listed as persistent organic pollutants (US EPA 2008a, 2008b, 2008c, 2008d)
Triclosan	3380-34-5		Ingredient added to many consumer products to reduce or prevent bacterial contamination; https://www.fda.gov/consumers/consumer-updates/5-things-know-about-triclosan [accessed 2022 August]

(*continued*)

Table 2. Continued.

Substance	CAS number	Chemical structure	Major uses
Case Study 4			
Octyl methoxycinnamate (OMC)	5466-77-3		UV filter

CAS: chemical abstracts service; PBDE: polybrominated diphenyl ether; PCB: polychlorinated biphenyl.
Chemical structures generated using ChemDraw v20.1.1.125; Perkin Elmer ChemOffice; https://perkinelmerinformatics.com/products/research/chemdraw [accessed 2022 August].
[a]2,2′,4,4′,5-Pentabromodiphenyl ether; depositor-supplied synonym DE-71.

Table 3. Overview of the case study substances: studies considered and cross links to Appendix I and Supplements.

		Details on case studies/studies	All data	Study summaries in note form
Substance	Studies considered	Appendix I	Table SI-1	SI-3
Case Study 1				
Propylthiouracil (PTU)	Axelstad et al. (2008); Kobayashi et al. (2005); Ramhøj et al. (2020a) [SOT poster]; Schneider et al. (2009); 13 EPA PTU studies[a]	Section App-I.1	CS1 PTU CS1 EPA PTU	Case Study 1.1
Methimazole (MMI)	Fegert et al. (2012); Ausó et al. (2004); Darbra et al. (2003); Ramhøj et al. (2022a)[b]		CS1 MMI	Case Study 1.2
Ethylene thiourea (ETU)	Marty et al. (2013) as cited in European Commission (2017a)		CS1 ETU Mancozeb	Case Study 1.3
Mancozeb	Beck (2008a, 2008b) as cited in European Commission (2017a); Axelstad et al. (2011a)			
Amitrole	Ramhøj et al. (2021, 2022a)[b]		CS1 Other	Case Study 1.4
Mercaptobenzimidazole	Ramhøj et al. (2021)			
Cyanamide	Ramhøj et al. (2021)			
Case Study 2				
Perchlorate	York et al. (2004); York et al. (2005a,b); Gilbert and Sui (2008); Mahle et al. (2003); Gilbert et al. (2022)[b]	Section App-I.2	CS2 Perchlorate	Case Study 2.1
Dietary iodine deficiency	Zhang et al. (2012); Gilbert et al. (2013)		CS2 Iodine Deficiency	Case Study 2.2
Case Study 3				
TBBPA	Cope et al. (2015); Saegusa et al. (2009, 2012); Van der Ven et al. (2008)/Lilienthal et al. (2008)	Section App-I.3	CS3 TBBPA	Case Study 3.1
PFHxS	Ramhøj et al. (2018, 2020b); Gilbert et al. (2021)		CS3 PFHxS	Case Study 3.2
Aroclor 1254, a PCB	Goldey et al. (1995b); Goldey and Crofton (1998); Crofton et al. (2000b); Morse et al. (1996)		CS3 Aroclor 1254	Case Study 3.3
DE-71, a mixture of PBDEs	Kodavanti et al. (2010); Zhou et al. (2002), with further information in poster abstract by Taylor et al. (2003); Ramhøj et al. (2020a [SOT poster]); de-Miranda et al. (2016); Bowers et al. (2015)/ Gill et al. (2016); Ramhøj et al. (2022b,c)[b]		CS3 DE-71	Case Study 3.4
Triclosan	Gilbert et al. (2021); Paul et al. (2012)		CS3 Triclosan	Case Study 3.5
Case Study 4				
OMC	Axelstad et al. (2011b); Ramhøj et al. (2020a) [SOT poster]	Section App-I.4	CS4 OMC	Case Study 4.1

CAS: chemical abstracts service; EPA: United States Environmental Protection Agency; OMC: octyl methoxycinnamate; PBDE: polybrominated diphenyl ether; PFHxS: perfluoro hexane sulphonates; PCB: polychlorinated biphenyl; SOT: Society of Toxicology; TBBPA: tetrabromobisphenol A.
[a]Thirteen studies that were derived from extensive research work that is being led and coordinated by the EPA and that encompass 20 publications (Goldey et al. 1995a; Sui and Gilbert 2003; Sui et al. 2005; Gilbert and Sui 2006; Gilbert et al. 2007, 2014, 2016, 2017; Sharlin et al. 2008, 2010; Gilbert 2011; Lasley and Gilbert 2011; Johnstone et al. 2013; Bastian et al. 2014; Spring et al. 2016; Hassan et al. 2017; Boyes et al. 2018; O'Shaughnessy et al. 2018a, 2018b, 2019).
[b]Ramhøj et al. (2022a), Gilbert et al. (2022) and Ramhøj et al. (2022b, c) published an amitrole / MMI study, a perchlorate study and a DE-71 study, respectively, after data evaluation for the present review was completed in September / October 2021. The data from these studies have been included in the respective CS spreadsheets, and their major findings are considered in the Section 4 discussions.

(Marty et al. 2021) with limited information addressing maternal/gestational exposure to such substances and effects on neurodevelopment.

2.2. Search queries and criteria to include/exclude publications

Since the summer of 2020, the ECETOC T4 TF continuously performed search queries in PubMed (https://pubmed.ncbi.nlm.nih.gov/ [accessed 2022 August]) to identify potentially relevant publications addressing thyroid- and brain-related effects upon *in utero*/lactational exposure to thyroid-active substances (see Supplementary Information SI-2 for details on the literature searches and on the selection of relevant publications and case study substances). The final selection of the fourteen case study substances was closely linked to the identification of relevant publications. For example, carbamazepine, perfluorooctane sulphonates (PFOS) and iopanoic acid ended up not being selected as case study substances due to the unavailability of relevant publications (see Supplementary Information *Table SI-1, Spreadsheet "Examples Excluded Studies Substances"*).

Table 4. Pre-defined publication inclusion/exclusion criteria.

Publications were included in the database if they met all of the following pre-defined criteria
Research article presenting rat study with gestational and/or lactational exposure of the dams
Study design scientifically meaningful (e.g. with respect to group size, presence of concurrent control group) and well documented
Thyroid hormone measurements in serum for pregnant and/or lactating dams and/or in serum and/or brain tissues for offspring (foetuses or pups), ideally within the window of susceptibility to the given primary neurodevelopmental endpoint(s)
Histopathological examination of the thyroid gland of the dams and/or offspring desirable but not mandatory
Assessments of neurodevelopmental or other brain-related endpoints in the offspring generally mandatory; exceptions: studies including only thyroid-related endpoints if they provided context on the relative sensitivity of these endpoints
Publications were excluded from the database if at least one of these pre-defined criteria applied
Published before 1990 (also since thyroid hormone analyses and detection levels likely have changed over time)
Review article (i.e. not primary research paper); however, review articles were evaluated for potentially relevant research articles
Unclear or scientifically inappropriate study design
Studies using multiple chemical exposures (i.e. mixtures)
Studies focussing on effects of thyroid hormone imbalance on body temperature
Electrophysiology as sole brain-related endpoint without concordant assessments of neurobehaviour or brain structure
Molecular markers (e.g. brain gene expression) as sole brain-related endpoints without concordant assessments of neurobehaviour or brain structure; however, such publications were considered in Sections 3.5 and 4.3.6 on brain gene expression (e.g. Royland and Kodavanti 2008; Royland et al. 2008)
Altered brain protein levels or brain enzyme activities as sole endpoints without concordant assessments of neurobehaviour or brain structure; exception: brain deiodinase 2 activity
Severe maternal or offspring toxicity (i.e. mortality or moribundity), poorly characterised systemic toxicity in studies using very high dose levels relative to other studies (e.g. $\geq$100 ppm propylthiouracil in drinking water) or effects only measured in the presence of severe toxicity or of $\geq$10% maternal body weight reduction Exceptions: severe signs of toxicity observed in the high-dose group(s) in preliminary studies, which then led to adjustments of the top dose in the main studies Note: this exclusion criterion stands in line with the European Commission (2017b, 2018) Endocrine Disruptor Criteria: "*Adverse effects that are non-specific secondary consequences of other toxic effects shall not be considered for the identification of the substance as endocrine disruptor*"

Some publications (just as some case study substances) were identified when referenced in other publications. Further, relevant publications co-authored by members of the ECETOC T4 TF and unpublished study reports summarised in the European Commission (2017a) Thyroid Disruption Workshop Report (which forms the scientific background to substantiate the views expressed in Appendix A of the EFSA and ECHA (2018) Endocrine Disruptor Guidance; Section 1.1) were included in the database. Since these unpublished study reports (Beck 2008a, 2008b; Marty et al. 2013) are not available to the reader, they are marked throughout as being cited in European Commission (2017a) but they are not included in the bibliography (see also Appendix I, Section App-I.1.3). Finally, two published posters from meetings of the Society of Toxicology (SOT) are included in the database since they provide relevant data, i.e. (1) the PTU study presented on the SOT poster Schneider et al. (2009), which is made available as Supplementary Information SI-7, and (2) the PTU, DE-71, and OMC studies presented on the SOT poster Ramhøj et al. (2020a), that is available at: https://files.abstractsonline.com/CTRL/8A/D/878/D3A/542/410/2A1/51F/77F/3C6/B77/68/a3101_1.pdf?display [accessed 2022 August].

Table 4 lists the pre-defined publication inclusion and exclusion criteria. Many publications retrieved included multiple brain-related endpoints. Ultimately, brain-related endpoints were selected for analysis based on sensitivity to thyroid hormone imbalance and the availability of a sufficient number of studies that included these brain-related endpoints to enable a reasonable assessment.

2.3. Assignment of publications to studies and data extraction

From amongst the publications retrieved, all publications that met the pre-defined inclusion criteria (Table 4) were included in the review. It is important to note that the ECETOC T4 TF did not aim to conduct a systematic review of all published rat studies assessing the case study substances. Instead, it was the aim to collect and review representative publications that covered the relevant thyroid- and brain-related parameters.

Table 3 lists the publications that the ECETOC T4 TF selected for each case study. These publications were assigned to individual studies as relevant. The ECETOC T4 TF defined a "study" as relating to one case study substance only. Accordingly, one publication covered more than one study if more than one substance was evaluated. For example, Gilbert et al. (2021) covered both a PFHXS study and a triclosan study. *Vice versa*, two or more publications could be assigned to one study if, for example, its thyroid-and brain-related findings had been published separately; example: the DE-71 study by Bowers et al. (2015) and Gill et al. (2016).

As available, the case studies include standard rat toxicity studies, that were conducted in accordance with (or similar to) TG-conforming test protocols (Table 1(A)), and/or non-standard investigational rat studies. In Appendix I, which is included in this paper after the bibliography, Sections App-I.1.3, App-I.2.3, App-I.3.3, and App-I.4.3 summarise the types of studies included in Case Study 1, 2, 3, and 4, respectively.

The ECETOC T4 TF extracted all thyroid- and brain-related data from the selected studies. Depending on the study considered, thyroid-related parameters included (1) maternal and/or offspring serum T4, T3, TSH, fT4, and/or free T3 (fT3); (2) maternal and/or offspring thyroid weight and histopathology; and (3) offspring brain T4 and T3 (see Li et al. 2019 and Marty et al. 2021 for technical challenges related to the measurement of thyroid-related parameters). Neurodevelopmental parameters included in the selected studies covered motor activity, acoustic startle response, cognitive function, hearing function, findings from the functional observational battery (https://www.ecfr.gov/current/title-40/chapter-I/subchapter-R/part-798/subpart-G/section-798.6050 [accessed 2022 August]), detailed clinical observations as well as brain histopathology

and morphometry. These parameters are included in TG-conforming developmental toxicity studies (Table 1(B)). Further brain-related parameters, that have not (yet) been standardised for adoption in TG-conforming studies, included electrophysiology, periventricular heterotopia and gene expression. Heterotopias are clusters of ectopic neurons in the brain indicative of altered neuronal migration; they have been observed in rats and humans and are associated with neurodevelopmental disorders like epilepsy and learning disabilities (Goodman and Gilbert 2007; Gilbert et al. 2014; O'Shaughnessy et al. 2019).

The Supplementary Information (SI) Table SI-1 presents all relevant data in the form of spreadsheets, and the Supplementary Information SI-3 presents study summaries in note form (see Box 1 for list of all Appendices and Supplements).

Box 1. Appendices (included in this paper after bibliography) and Supplements (available online only) to the present review.

Appendix I: Details on the rationale to define the four case studies and to assign the case study substances to each case study; brief overview of the types of studies included in each case study (linked to Section 2.1 and 2.3)

Appendix II: Conclusions on TPO inhibitors in the European Commission (2017a) Thyroid Disruption Workshop Report (linked to Section 4.2.1.3)

Supplement SI-1: Tabular Spreadsheets with all data (see introduction to Section 3 for details)

Supplement SI-2: Details on literature searches and selection of publications and case study substances

Supplement SI-3: Study summaries in note form

Supplement SI-4: Further details on the evaluation of the empirically set $\geq$60%/$\geq$50% offspring serum reduction thresholds and $\geq$20% and statistically significant offspring serum T3 reduction threshold (linked to Sections 3.2.1.3, 3.2.2, and 4.4.3)

Supplement SI-5: Studies in which statistically significant and/or $\geq$10% offspring body weight reduction was/was not reported (linked to Section 3.3)

Supplement SI-6: Overview panels: Pup serum T4 and T3 decrements associated with motor activity changes (linked to Section 4.3.1)

Supplement SI-7: SOT poster Schneider et al. (2009; linked to Section 2.2 and Appendix I.1.3)

In extracting data from the selected studies, the ECETOC T4 TF always presented hormone levels as reductions or increases relative to the concurrent controls (i.e. recording the fraction *by which* a measurement differed from the controls). If necessary, such data were calculated from absolute values. Further, if hormone data were only provided in graphic form, the ECETOC T4 TF manually calculated the relative hormone levels from the respective data points.

Where possible, the case studies focussed on thyroid hormone imbalance during windows of susceptibility to the respective neurodevelopmental effects (Table 5). Neurodevelopmental periods sensitive to thyroid perturbations resulting in the phenotypic changes are poorly defined in some cases. For example, for motor activity, the available evidence broadly indicates that the critical window is primarily sometime during the three weeks of lactation. Cognitive function (learning and memory) likely has pre- and postnatal critical periods. By comparison, for other parameters, the critical windows have been identified more precisely, e.g. postnatal day (PND) 4–18 for low frequency hearing loss (Goldey and Crofton 1998; Crofton et al. 2000a) and GD 19 to PND 2–6 for periventricular heterotopia, specifically below the peak of the *cingulum* of the *corpus callosum* (O'Shaughnessy et al. 2018a; Ramhøj et al. 2020a, 2021).

Further, the ECETOC T4 TF extracted information on maternal and offspring body weight changes (mostly reductions), using information from graphs and estimating the body weight changes relative to the concurrent controls, if necessary. These estimations served to determine if body weight reductions reached or exceeded 10% of concurrent controls (even if the authors of the publication had assessed these as being "not significant"). Notably, in some studies (e.g. OECD TG 426) changes in body weight gain during gestation as compared to the concurrent controls were measured. In these cases, the extent of body weight gain change was recorded together with statistical significance.

In the case studies, the information on body weight (gain) reductions in the dams is used as one parameter to determine if the maximum tolerated dose had likely been reached. The maximum tolerated dose is defined as the highest dose to produce toxic effects without causing death or significant morbidity and a $<$10% decrease in body weight relative to controls (OECD 2002). If studies yielded only minor body weight reductions and only minor T4 reductions in the dams and no neurodevelopmental findings in the offspring up to the highest dose tested, it could not be ruled out that more pronounced T4 reductions and neurodevelopmental findings could have been observable at higher doses.

The information on body weight (gain) reductions in the offspring is considered for the overall evaluation of the findings from the respective study. Pronounced body weight reduction in the offspring following *in utero*/lactational exposure to thyroid-active substances may be indicative of severe maternal toxicity (e.g. impairing maternal readiness for lactation and/or leading to insufficient quantity and quality of the milk), but it may also be indicative of thyroid disruption leading, e.g. to growth retardation (Van Wynsberghe and Klitgaard 1973). In the case studies, it was not established whether offspring body weight reduction was primarily caused by thyroid disruption or not.

Finally, some of the publications included toxicokinetics data, such as serum levels of the administered substances in the dams or offspring. Such toxicokinetics data are not considered in the present review, which focusses on thyroid- and brain-related data. Nonetheless, considerations of substance toxicokinetics are relevant to predict a substance's potential to cause hypothyroid-mediated neurodevelopmental impairment (discussed in Section 4.4.5 and 4.4.7).

3. Case studies 1–4: overarching evaluation of the findings

Key messages: The four case studies encompass 51 rat studies that include gestational and/or lactational exposure to substances that are known to cause thyroid hormone imbalance in rats *via* different direct or indirect thyroid related MoAs. In most studies, the exposure periods covered the major parts of both gestation and lactation. Differences in exposure durations are considered if findings for the same substance are inconsistent between studies.

To guide the reader through the database, Section 3 first presents the major neurodevelopmental and other brain-related

Table 5. Information on critical windows of susceptibility for major thyroid-sensitive neurodevelopmental endpoints in rats.

Endpoint	Information on critical window of susceptibility in rats
Motor activity	Clarac et al. (2004) reviewed different stages of locomotor development in the rat. On GD 15, motor neurons become excitable; central pattern generators begin to operate. Postnatally, motor maturation occurs in a rostral-caudal direction. During the immature stage (PND 0–10), central pattern generators can fire with rhythmic pattern, but postural elements for full locomotor activities develop until PND 10 (e.g. pelvis supporting body weight). During the transitory stage (PND 10–15), pups transition to more adult-like motor behaviours, including more versatile walking and rearing. Rat pups are capable of some complex behaviours at PND 16, including rotor rod performance, climbing. After PND 15, walking becomes digitigrade, and full adult locomotor behaviour is complete at ∼4 weeks of age. Forelimb swimming can be evoked in rats at PND 0, and all four limbs can contribute to swimming by PND 1; however, adult swimming patterns (hindlimb propulsion and forelimb steering) are not achieved until PND 12–15 (reviewed by Clarac et al. 2004). Motor activity testing in standard rat toxicity studies considers that locomotor ontogeny begins around PND 13 with adult patterns of locomotor activity developed around PND 21 (both neuronal and neuromuscular development required) (NAFTA 2016). Axelstad et al. (2008) reported hypoactivity in pre-weaning pups and hyperactivity in weanlings/adult offspring after PTU exposures from GD 7 to PND 17. Similarly, Akaike et al. (1991) and Kobayashi et al. (2005) reported post-weaning hyperactivity with PTU exposures from GD 18 to PND 21 and PND 0–19, respectively. Taken together, these findings indicate a likely postnatal element that further may end before PND 17. Goldey et al. (1995a) showed decreased motor activity on PND 15 with Aroclor 1254 exposure that was attenuated by 60% with T4 supplementation on PND 4–21; lack of full recovery hypothesised to be due to T4 kinetics (Goldey and Crofton 1998).
Cognitive function (learning and memory)	Tests for cognitive function in rats address learning, memory, attention processes and motivation; thus, the basis for behavioural differences may be difficult to interpret. Assessments of learning and memory are tied to the development of sensory systems (hearing, vision), vestibular system and locomotor behaviours (PND 16—weaning or possibly later with more complex tasks). It is likely that both prenatal and postnatal exposures can markedly affect cognitive function. The OECD TG 426 DNT study requires assessments of learning and memory in the rat offspring shortly after weaning and as adults. In a cross-fostering study, O'Shaughnessy et al. (2018a) observed altered fear conditioning and context learning in male offspring exposed to PTU both prenatally and postnatally (GD 6–PND 14). Akaike et al. (1991) exposed lactating dams from LD 0–19 to extremely high concentrations of 0.02% (i.e. 200 ppm) PTU in the drinking water; adult rat offspring showed altered learning and memory, i.e. increased errors in the Biel water maze and fewer correct choices in the radial arm maze at 6 and 13 weeks, respectively, indicating a likely postnatal element. Caveat: PTU doses exceeded the maximum tolerated dose. Davenport et al. (1976) exposed pregnant and/or lactating rats to dietary thiouracil and concluded that exposures beginning in late gestation (i.e. approximately GD 18 when the foetal thyroid becomes functional) and lasting through PND 14 have greater effects on learning and memory than postnatal exposures alone (i.e. greater maze deficits); by comparison, starting exposures earlier in gestation did not have greater effects on maze performance. Two caveats: Thiouracil doses exceeded the maximum tolerated dose, and toxicokinetics were not considered in this study. Eayrs (1961) reported that rats made hypothyroid at birth through either ^{131}I or methyl thiouracil treatment made significantly more errors in the Hebb-Williams closed field test (measure of cerebral capacity) than untreated littermates; the degree of impairment was decreased when treatments were initiated later in the postnatal period. When treatment was initiated on PND 25, neurobehavioural assessment did not differ from controls. T3 replacement was successful if initiated at PND 10 but not if started on PND 24.
Ototoxicity (hearing function; cochlear development)	Maturation of cochlea, opening of auditory meatus and hearing begin in the second postnatal week with adult patterns of auditory reflex present on PND 21 (NAFTA 2016). Sensitive period for effects on the development of the organ of Corti (i.e. hair cell loss in the cochlea affecting low frequency hearing) is postnatal prior to PND 21; some evidence for the sensitive period being PND 4 or PND 6 to PND 15 or PND 18 (Deol 1973; Hebert et al. 1985; Uziel 1986; Freeman et al. 1996; Goldey and Crofton 1998; Crofton et al. 2000a). The sensitive period for thyroid hormone effects on cochlear development in rats varies from GD 19 to PND 15 depending on structure (cochlear development progresses basal-to-apical), which may show as different effects on hearing (Deol 1973; Hebert et al. 1985); others estimate GD 18–PND 18 (outer hair cell synaptogenesis) (Uziel 1986; Freeman et al. 1996). A cross-fostering study verified the sensitive window as being in the first few postnatal weeks, i.e. <PND 21 (Crofton et al. 2000b). Amelioration of hearing loss with postnatal T4 replacement from PND 4 to 21 (Goldey and Crofton 1998). Upon exposure to Aroclor 1254 from GD 6 to PND 21, there was "*mild-to-moderate loss of outer hair cells in the upper-middle and apical turns*" of the organ of Corti with no effect on inner hair cells (Crofton et al. 2000a).
Neuronal migration/ periventricular heterotopia	Critical window of susceptibility: GD 19–PND 2 upon exposure to PTU (O'Shaughnessy et al. 2019). In offspring exposed to PTU only prenatally (cross-fostered on PND 2), serum PTU levels returned to control levels by PND 5 (O'Shaughnessy et al. 2018a). In some later studies examining heterotopia, offspring serum thyroid hormones were measured in the early postnatal period to PND 6, which were included in the critical window assessment (e.g. Ramhøj et al. 2020a [SOT poster], 2021). Gilbert et al. (2014) observed periventricular heterotopia (also described as subcortical band heterotopia) below the peak of the cingulum of the *corpus callosum*; location at 10 ppm PTU: between plates 33 and 39 of the Paxinos and Watson (1986) atlas, location at 1–3 ppm PTU: in the middle of this range (Gilbert et al. 2014). O'Shaughnessy et al. (2018a) observed heterotopia between the *corpus callosum*, *subiculum* and *cingulum* of the hippocampus in the posterior forebrain. Ausó et al. (2004) reported aberrant migration of single neurons in the somatosensory cortex and hippocampus of pups of dams exposed to MMI between GD 12 and 15.

GD: gestational day; LD: lactational day; MMI: methimazole; OECD: Organisation for Economic Co-operation and Development; PND: postnatal day; PTU: propylthiouracil; T3: triiodothyronine; T4: thyroxine; TG: test guideline.

findings from the studies and then the major thyroid-related findings; all findings are then jointly discussed in Section 4:

- Section 3.1: Neurodevelopmental and other brain-related findings, sorted by case study and case study substance; note: brain gene expression is considered in Section 3.5
- Section 3.2: Thyroid-related findings, sorted by thyroid-related parameter
- Section 3.3: Body weight
- Section 3.4: Liver enzyme activities
- Section 3.5: Gene expression and/or transporter levels in the offspring brain

As relevant, the different parts of Section 3 refer to spreadsheets from Table SI-1 (available online; see Table 6 for an overview of all Table SI-1 spreadsheets).

In most studies, the exposure periods began shortly after implantation, on GD 6–7, and lasted (almost) all through lactation, mostly ending on lactational day (LD) 21–22. Thus, most exposure periods covered the known windows of susceptibility to periventricular heterotopia, ototoxicity and altered motor activity (Table 5). However, some studies included considerably shorter exposure periods (e.g. lasting only a part of gestation or only a part of lactation) or considerably longer exposure periods (e.g. beginning premating and continuing with direct dosing of the weaned offspring); see *Table SI-1, Spreadsheet "Overview Exposure Duration."*

In the overarching presentation of the findings below, differences in exposure periods are considered if findings for the same substance are inconsistent between studies. Further, the evaluation of studies yielding no neurodevelopmental findings considers maternal and/or offspring serum thyroid hormone levels and whether the maximum tolerated dose was likely reached in the dams (see Section 3.3 for details on maternal body weight reduction as a parameter indicating whether the maximum tolerated dose was likely reached).

3.1. Neurodevelopmental and other brain-related findings

An overview of the neurodevelopmental and other brain-related findings recorded in the four case studies is provided in *Table SI-1, Spreadsheet "Overview DNT Details."*

3.1.1. Case Study 1: Summary of the neurodevelopmental and other brain-related findings

Key messages: Amongst the TPO inhibitors, PTU and amitrole caused neurodevelopmental effects; MMI caused neurodevelopmental effects at very high doses, and ETU/mancozeb, mercaptobenzimidazole and cyanamide did not cause neurodevelopmental effects.

All studies investigating the gestational and/or lactational effects of PTU showed alterations in motor activity, acoustic startle response and/or cognitive function (Kobayashi et al. 2005; Axelstad et al. 2008) or periventricular heterotopia (Johnstone et al. 2013; Gilbert et al. 2014, 2016; O'Shaughnessy et al. 2018b; Ramhøj et al. 2020a [SOT poster]). Neurobehavioural effects caused by PTU were observed at concentrations as low as 1.0 mg/kg body weight/day (mkd) (Kobayashi et al. 2005) and 1.6 mkd (Axelstad et al. 2008). Spring et al. (2016) and similarly O'Shaughnessy et al. (2018b) reported low incidences of periventricular heterotopia at 1 ppm (applied in the drinking water, corresponding to ~0.1–0.13 mkd) and more pronounced periventricular heterotopia with respect to incidence and volume at 3 and 10 ppm (~0.3–0.4 and 1–1.3 mkd). By comparison, Ramhøj et al. (2020a) reported statistically significant periventricular heterotopia at 2.5 mkd PTU only but not at 1.0 mkd.

Periventricular heterotopia was also observed in rats after gestational and lactational exposure to 50 mkd amitrole (Ramhøj et al. 2021).

For MMI, a study conducted similarly to an extended one-generation reproductive toxicity study (EOGRTS) showed no effects in motor activity assessments, in the functional observational battery or on brain morphometry and histopathology when testing MMI up to 3 mkd (Fegert et al. 2012). By comparison, in two investigational studies, administration of ~36 mkd MMI (calculated by the ECETOC T4 TF as 0.2 mg MMI/mL × 40 ml water consumption/day × 1/0.22 kg body weight) resulted in altered proportions of cells in different layers of the somatosensory cortex and altered responses to acoustic stimuli (Ausó et al. 2004) as well as altered locomotor activity and passive avoidance in the absence of effects on novelty-directed exploratory behaviour or anxiety behaviour (Darbra et al. 2003). Hence, the effects in these two investigational studies were observed at much higher MMI concentrations than those applied in the EOGRTS-like study. Ausó et al. applied this high concentration for a total of four days during gestation while the rat foetus is still fully dependent on maternal T4 (from GD 12 to 15) and Darbra et al. from GD 9 to LD 21. By comparison, in the EOGRTS-like study, exposure of the parental animals began premating and lasted all through gestation and lactation followed by direct exposure of the weanlings up until adulthood. Fegert et al. observed reduced body weight in the mid- and high-dose dams as well as reduced food and water consumption, increased gestational length and reduced litter size at the high dose; also, body weight of the high-dose group offspring was reduced by 15% on PND 21/22. Together with the moderate serum T4 reduction observed in both the dams and offspring (i.e. by up to 35% and 46%, respectively), these findings indicate that the maximum tolerated dose was reached in the EOGRTS-like study.

None of the studies investigating ETU or mancozeb reported any substance-mediated effects on motor activity, sensory function or cognitive function (Marty et al. 2013; Beck et al. 2008a,b), both as cited in European Commission (2017a), and Axelstad et al. (2011a). Up to 50 mkd ETU and up to 100 mkd mancozeb were applied in these studies, and these doses were established as maximum tolerated doses since severe signs of toxicity developed at higher doses.

Finally, 10 mkd mercaptobenzimidazole and up to 11.25 mkd cyanamide did not cause periventricular heterotopia (Ramhøj et al. 2021). Again, the corresponding body weight data and/or clinical observations indicate that the maximum tolerated dose was reached in these studies.

Table 6. Overview of spreadsheets included in the Supplementary Table SI-1 (available online).

Name of spreadsheet	Explanatory note	Section in manuscript
11 spreadsheets whose names begin with "Overview": these spreadsheets present the outcome of the data evaluation and are referred to in Section 3[a]		
Overview exposure duration	*Purpose of spreadsheet*: List those studies whose exposure periods were either shorter or longer than the exposure periods in the majority of studies considered (Note 1) *Note 1*: In most studies, exposure covered both the major parts of gestation and lactation (i.e. it began approximately GD 6–7, shortly after implantation, and lasted to approximately LD 20–22); differences in exposure periods are considered when thyroid- or brain-related findings were inconsistent between studies addressing the same substance	3
Overview DNT details	*Purpose of spreadsheet*: Present the details of the neurodevelopmental and other brain-related findings (Columns L–T), together with general information on study design as well as maternal/offspring body weight and serum T4	3.1 3.3
Overview T4 T3 TSH DNT	*Purpose of spreadsheet*: Present all maternal/offspring serum T4, T3 and TSH data, together with general information on study design and overview of brain-related findings *Special data evaluation*: (1) Extent of maternal/offspring T4 reductions (Column G/M); (2) trend of maternal/offspring T4 reductions (Column K/Q); (3) comparison of extents of maternal *vs.* offspring T4 reductions (Column R)	3.2.1 3.2.1.1 3.2.1.2 3.2.1.4
Overview offspring T4 at DNT dose	*Purpose of spreadsheet*: Compare extent of offspring serum T4 reduction (Column J) with lowest dose at which brain-related findings were observed (Columns M–U) *Special data evaluation*: Evaluation of $\geq$60% and $\geq$50% offspring serum T4 reduction thresholds (Column L)	3.2.1.3
Overview T4 T3	*Purpose of spreadsheet*: Present all maternal/offspring serum T4 and T3 data (Columns G–I/L–N), together with general information on study design and overview of brain-related findings (Column S) *Special data evaluation*: (1) Comparison of extent of maternal/offspring T4 *vs.* T3 reduction (Column J/O); (2) evaluation of offspring serum T3 reduction thresholds (Columns P–R)	3.2.2
Overview T4 fT4 T3 fT3	*Purpose of spreadsheet*: Present all maternal/offspring serum T4, fT4, T3, fT3 data (Columns G–J/M–Q), together with general information on study and overview of brain-related findings (Column S) *Special data evaluation*: Comparison of extent of maternal/offspring T4, fT4, T3 reduction (Column K/R)	3.2.3
Overview T4 T3 TSH	*Purpose of spreadsheet*: Present all maternal/offspring serum T4, T3 and TSH data (Columns G–J/M–P), together with general information on study design and overview of brain-related findings (Column R) *Special data evaluation*: Trend of maternal/offspring TSH alteration (Column K/Q)	3.2.4
Overview serum + brain T4 DNT	*Purpose of spreadsheet*: Present all offspring brain thyroid hormone data (Columns N–P), together with general information on study design as well as maternal/offspring serum T4, T3 data (Columns G–I/K–M) and details of the brain-related findings	3.2.5
Overview T4 thyroid WT HP	*Purpose of spreadsheet*: Compare dose level at which maternal/offspring serum TSH was statistically significant (Column H/P) with dose level at which maternal/offspring thyroid weight change was statistically significant (Column J/R) and/or thyroid histopathological findings were observed (Column K/S); also: indicate extent of maternal/offspring thyroid weight change (Column I/Q) *Special data evaluation*: Are maternal/offspring thyroid weight change and/or thyroid histopathology more sensitive than the corresponding serum T4 reduction? (Column M/U)	3.2.6
Overview liver enzymes	*Purpose of spreadsheet*: Present all data for all studies that included measurements of liver enzymes	3.4
Overview gene expression	*Purpose of spreadsheet*: Present studies that included investigations of gene expression changes in the offspring brain *Special data evaluation*: Gene expression changes (Column H) *Note 1*: This spreadsheet also includes data from studies that were not included in the main database since they did not include measurements of maternal or offspring serum thyroid hormones	3.5
13 spreadsheets whose names begin with CS1–CS4 followed by the name of the substance (or iodine deficiency)		
CS1 PTU, CS1 EPA PTU, CS1 MMI, CS1 ETU Mancozeb, CS1 Other CS2 Perchlorate, CS2 Iodine Deficiency CS3 TBBPA, CS3 PFHxS, CS3 Aroclor 1254, CS3 DE-71, CS3 Triclosan CS4 OMC	*Purpose of all these spreadsheets*: Present all data for all studies considered for the respective substance (or dietary iodine deficiency) *Note 1*: There are two spreadsheets for PTU, i.e. "CS1 PTU" and "CS1 EPA PTU" (due to differences in study protocol and brain-related endpoints considered) *Note 2*: The spreadsheet "CS1 Other" includes the amitrole, cyanamide and mercaptobenzimidazole studies, which were jointly published by Ramhøj et al. (2021) *Note 3*: The following spreadsheets also include data from studies that were published after data evaluation for the present review was completed; these data are considered in the Section 4 discussions, but not in the "Overview" Spreadsheets: - "CS1 MMI": MMI study by Ramhøj et al. (2022a) - "CS1 Other": Amitrole study by Ramhøj et al. (2022a) - "CS2 Perchlorate": Perchlorate study by Gilbert et al. (2022) - "CS3 DE-71": DE-71 study by Ramhøj et al. (2022b, 2022c)	Appendix I: Case Study 1: App-I.1.3 Case Study 2: App-I.2.3 Case Study 3: App-I.3.3 Case Study 4: App-I.4.3
Further spreadsheets		
Substance Phys-Chem	*Purpose of spreadsheet:* Present select physico-chemical data for the 14 case study substances	2.1

(*continued*)

Table 6. Continued.

Name of spreadsheet	Explanatory note	Section in manuscript
Examples Excluded Studies Substances	*Purpose of spreadsheet*: Examples for publications that were excluded after detailed evaluation of the full text (many publications were excluded as out of scope after review of the titles and abstracts)	2.2, 4.1
Abbreviations for all sheets	*Purpose of spreadsheet:* Explanations for all acronyms used in the other spreadsheets	Not linked

CS: case study; DE-71: a mix of polybrominated diphenyl ethers; DNT: developmental neurotoxicity; EPA: Environmental Protection Agency; ETU: ethylene thiourea; fT4: free thyroxine; fT3: free triiodothyronine; GD: gestational day; LD: lactational day; MMI: methimazole; HP: histopathology; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PTU: propylthiouracil; T4: thyroxine; T3: triiodothyronine; TBBPA: tetrabromobisphenol A; TSH: thyroid stimulating hormone; WT: weight.

[a]Please see heading in the respective overview spreadsheets for database considered in that spreadsheet.

3.1.2. Case Study 2: Summary of the neurodevelopmental and other brain-related findings

Key messages: For the NIS inhibitor perchlorate, the only brain-related findings are electrophysiological alterations. One dietary iodine deficiency study showed altered cognitive function, whereas the other one did not, despite more pronounced iodine deficiency.

The NIS inhibitor perchlorate did not cause significant effects on motor activity or any effects on the acoustic startle response when tested at up to 30 mkd (York et al. 2004, 2005a, 2005b). Also, gestational and lactational exposure to perchlorate did not alter cognitive function in the water maze and passive avoidance tests when tested at up to 30 mkd (York et al. 2004) or in the water maze and fear conditioning tests when tested at up to 140 mkd (Gilbert and Sui 2008). Gilbert and Sui (2008) did observe electrophysiological alterations, i.e. reduced baseline synaptic transmission in hippocampal field potentials at 4.5–140 mkd perchlorate, reduced inhibitory function at 44 and 140 mkd and an augmentation in long-term potentiation in the population spike at 140 mkd.

In the two dietary iodine deficiency studies, Zhang et al. (2012) observed longer escape latency in the Morris water maze in the PND 40–44 offspring of the iodine-deficient group (1.2 μg iodine/day). By comparison, Gilbert et al. (2013) did not observe any effects in the Morris water maze, fear conditioning or acoustic startle response tests in the PND 60–180 offspring of the iodine deficient groups (0.2–0.3 and 0.7–1.1 μg iodine/day), but these groups did show reduced synaptic transmission in the dentate gyrus. In both studies, the treatment period began premating and lasted all through gestation and lactation. Zhang et al. (2012) did not provide any data on the body weight of either the dams or offspring, and Gilbert et al. (2013) reported that "*body weights measured postnatally in dam and offspring did not differ across diet groups*."

3.1.3. Case Study 3: Summary of the neurodevelopmental and other brain-related findings

Key messages: TBBPA, PFHxS and triclosan did not elicit statistically significant neurodevelopmental effects. One TBBPA study reported dose-dependently altered brainstem auditory evoked potentials (BAEPs). This study did not analyse statistical significance but applied a benchmark method with critical effect size to demonstrate risk (discussed in Section 4.2.3.1). Aroclor 1254 altered motor activity and the acoustic startle response and caused low frequency hearing loss. DE-71 altered cognitive function in some studies, but not in others, despite similar exposure periods and dosimetry. Also, DE-71 elicited some effects on motor activity and the acoustic startle response but did not induce periventricular heterotopia.

Of the substances that enhance thyroid hormone clearance, TBBPA did not alter motor activity, the acoustic startle response or cognitive function (passive avoidance, water maze) in an enhanced two-generation study testing up to the limit dose of 1000 mkd (Cope et al. 2015), nor did it affect brain histopathology (Saegusa et al. 2009, 2012; Cope et al. 2015). Also, Lilienthal et al. (2008) and Van der Ven et al. (2008) reported no effects on cognitive function (fear conditioning, sweet preference), but they did report dose-dependently altered BAEPs at 3–3000 mkd (exposure from premating all through lactation). Lilienthal et al. evaluated the BAEP results using a benchmark method with critical effect size analysis for demonstrating a risk, and they reported a benchmark dose lower limit of the 90% confidence interval for a 5% increase in hearing thresholds in female rats in the range of 1–40 mkd, depending on frequency (discussed in Section 4.2.3.1).

PFHxS elicited no effects on motor activity or learning and memory (radial arm maze) when tested at up to 25 mkd (Ramhøj et al. 2018, 2020b), nor did it alter the acoustic startle response or trace fear conditioning, and it also did not elicit periventricular heterotopia when tested at 50 mkd (Gilbert et al. 2021). In both studies, the body weight of the dams and offspring was reported as being unaffected in the PFHxS-treated groups. Nonetheless, dosing was likely sufficiently high to elicit effects. In the study by Ramhøj et al. (2018, 2020b), the top dose had been reduced to 25 mkd PFHxS after application of 45 mkd in a preliminary study. Ramhøj et al. provide no further explanation for this reduction. The ECETOC T4 TF assumes that the top dose was reduced in the main study because ~20% litter size reduction was observed in the range finding study in a further test group, i.e. the 45 mg "PFHxS plus endocrine disruptor mixture" group, which is not considered in this review that excludes mixture evaluation.

Aroclor 1254 caused transient reductions in motor activity (PND 13–15) and altered acoustic startle responses (PND 23–24) at 8 mkd (Goldey et al. 1995b; Goldey and Crofton 1998) but not at 6 mkd (Crofton et al. 2000b). Four and 8 mkd Aroclor 1254 caused low-frequency hearing deficits in the adult offspring (Goldey et al. 1995b; Goldey and Crofton 1998)

with the critical period of exposure being postnatal (Crofton et al. 2000b). Goldey et al. (1995b) reported that 25% and 50% of the offspring from the 8 mkd-group had died by PND 12 and 21, respectively, and that these pups exhibited early eye opening. While Goldey and Crofton (1998) applied this same dose, they did not report any offspring mortality, but again early eye opening.

Further, gestational exposure (GD 10–16) to 5 and 25 mkd Aroclor 1254 resulted in transiently altered DIO2 activity in the forebrain. On GD 20, DIO2 activity in the forebrain was increased by 35% and 100% in the low- and high-dose groups, respectively; on PND 4, it was reduced by ~48% in the high-dose group only, but it was not altered at ≥ PND 90 (Morse et al. 1996). DIO2 activity was the only brain-related parameter considered by Morse et al. The ECETOC T4 TF suggests that the increased DIO2 activity likely reflects a compensatory response to the decreased brain T4 levels (to maintain brain T3 levels) rather than a direct effect of Aroclor 1254, as a similar response was seen in the PTU study by Sharlin et al. (2010).

DE-71 caused no effects on motor activity (Zhou et al. 2002; Kodavanti et al. 2010; Bowers et al. 2015; de-Miranda et al. 2016), apart from a decrease in rearing on PND 110 and 450 at the mid and high doses of 3 and 30 mkd (Bowers et al. 2015). This effect was interpreted as a potential alteration in hippocampal function or anxiolytic behaviour (Lever et al. 2006) rather than altered ambulation. Also, no effects were evident in the functional observational battery (Kodavanti et al. 2010) or on emergence latency or pre-pulse inhibition (Bowers et al. 2015). Exposure to 30 mkd DE-71 decreased the acoustic startle response in PND 90 offspring (Bowers et al. 2015). As regards effects on cognitive function, Bowers et al. reported no findings when the PND 230–235 offspring were tested in the water maze, whereas de-Miranda et al. (2016) found that the PND 70–100 female offspring made more errors in the radial arm maze at this same dose level (30 mkd) but did not require more time to accomplish tasks. Finally, gestational and lactational exposure to up to 40 mkd DE-71 did not cause periventricular heterotopia (Ramhøj et al. 2020a [SOT poster]).

With respect to the question of whether the maximum tolerated dose was likely reached in the dams in these DE-71 studies, exposure to 30 mkd DE-71 did not cause significant maternal body weight reductions (Zhou et al. 2002; Kodavanti et al. 2010; Bowers et al. 2015), but maternal serum T4 was reduced by up to 58% (Kodavanti et al. 2010). Further, significant offspring body weight reductions by >10% were recorded at least at one timepoint during lactation (Kodavanti et al. 2010; Bowers et al. 2015; Ramhøj et al. 2020a [SOT poster]). The ECETOC T4 TF suggests that, since 58% maternal serum T4 decrements are associated with offspring thyroid hormone imbalance and neurodevelopmental effects in other studies, these findings indicate that the top dose was sufficiently high to elicit neurodevelopmental effects even in the absence of a traditional maximum tolerated dose or of histopathological effects on the offspring brain.

Triclosan did not cause any statistically significant effects on the acoustic startle response or any effects on trace fear conditioning, and it also did not cause periventricular heterotopia (Gilbert et al. 2021). With respect to the question whether the maximum tolerated dose was likely reached in the dams, Gilbert et al., referring to Paul et al. (2010a, 2012), reported that the applied dose of 300 mkd triclosan had been chosen based on previous work to maximally reduce serum T4 without causing overt toxicity in the dams.

3.1.4. Case Study 4: Summary of the neurodevelopmental and other brain-related findings

Key messages: OMC reduced motor activity of the female offspring but did not cause any other statistically significant neurodevelopmental effect.

Gestational and lactational exposure to OMC caused some effects on motor activity, which were mostly not dose dependent (Axelstad et al. 2011b). However, in the high-dose group female offspring (1000 mkd), motor activity was statistically significantly reduced at both 9 and 17 weeks (Axelstad et al. 2011b); therefore, the ECETOC T4 TF evaluates this finding as biologically relevant (for definition, see EFSA 2011a, 2017). In the radial arm maze, a reduction in the number of errors was reported in both the 500 and 1000 mkd groups at 17 weeks, whereas acoustic startle and auditory function were not altered at 28 and 31 weeks, respectively (Axelstad et al. 2011b). In the OMC study by Ramhøj et al. (2020a [SOT poster]), altered gene transcription in the brain cortex of the PND 16 offspring was reported at 375 and 500 mkd OMC but no periventricular heterotopia at up to 500 mkd OMC.

3.1.5. Overarching evaluation of the neurodevelopmental and other brain-related findings

Key messages: Motor activity, acoustic startle response and periventricular heterotopia appear more sensitive parameters than cognitive function (i.e. they were observed at lower dose levels). It was not possible to establish patterns of brain-related effects either for specific substances or for specific MoAs. However, the spectrum of brain-related parameters included in the individual studies, and the timepoints of measurement, varied considerably.

Forty-seven of the 51 studies considered in the four case studies included neurodevelopmental and other brain-related assessments, i.e. all studies except for Mahle et al. (2003), Beck (2008a) in European Commission (2017a), Schneider et al. (2009) [SI-7], and Paul et al. (2012). The spectrum of parameters included in the individual studies as well as the timepoints of measurement varied considerably (Sections 3.1.1–3.1.4):

- *Cognitive function*: assessed in 22 studies; generally, in adult animals and sometimes also around weaning; assessments included water maze, radial arm maze, fear conditioning, passive avoidance tests and sweet preference
- *Motor activity*: assessed in 20 studies; thereof: in 16 studies at more than one timepoint (once each during mid-lactation and/or around weaning and in adulthood), in

three studies at one timepoint only, and in one study it was unclear how often motor activity was assessed; finally, in one study, motor activity was assessed by radiotelemetry
- *Acoustic startle response*: assessed in 17 studies; thereof: in nine studies at more than one timepoint (mostly once each around weaning and in adulthood)
- *Hearing function and auditory brainstem responses:* hearing function assessed in five studies; auditory brainstem responses assessed in two studies
- *Periventricular heterotopia*: assessed in 15 studies, including three standard rat toxicity studies, in which periventricular heterotopia was not specifically mentioned, but brain histopathology performed around PND 16, or later up until adulthood in accordance with the respective TG, would have informed on the presence of periventricular heterotopia (provided that the histopathological sections included the respective brain locations)
- *Brain histopathology and morphometry* (apart from explicit assessments of periventricular heterotopia): assessed in 13 studies
- *Detailed clinical observation and functional observational battery*: performed in 5 studies
- *Expression of genes/levels of brain-related proteins* assessed in 18 studies (see Section 3.5 for details)
- *Electrophysiology of hippocampal tissues* assessed in seven studies (which were all conducted by the US EPA; discussed in Section 4.3.5)
- *DIO2 activity in forebrain* assessed in one study

Statistically significant neurodevelopmental or other brain-related findings were recorded in all 13 EPA PTU studies (highlighting the toxic potential of this substance) as well as in 18 of the 34 further studies that included brain-related assessments. Tables 7 and 8 provide an overview of the statistically significant brain-related findings recorded in the different studies. For this overview, the 13 EPA PTU studies are presented in Table 8, i.e. separately from the 18 further studies (Table 7), due to differences in their study design. Specifically, in the EPA PTU studies, PTU was applied in the drinking water, whereas it was applied *via* oral gavage in the other PTU studies. The brain-related investigations in the EPA PTU studies focussed on electrophysiology, heterotopia and gene expression whereas they focussed on more traditional brain-related investigations (e.g. cognitive function, motor activity, acoustic startle response) in the other PTU studies.

While Table 7 is sorted by case study substances, Table 8 only includes EPA PTU studies, i.e. studies from Case Study 1. A comparison of the patterns of brain-related effects recorded between studies evaluating the same substance shows no regularities. Hence, it was not possible to establish patterns of brain-related effects either for specific substances, for specific MoAs or across case studies. However, the spectrum of parameters included in the individual studies (and the timepoints of measurement) varied considerably (discussed in Section 4.1).

In Tables 7 and 8, the statistically significant brain-related findings recorded in a particular study are sorted by the lowest dose level at which the respective finding was statistically significant. Further, for each study, the grey shading in Tables 7 and 8 indicates the lowest dose level at which any brain-related finding was recorded as statistically significant. Thus, for each study, it is shown which brain-related parameter was more "sensitive" than the other brain-related parameters included in that study. In this context, "sensitivity" implies that the given effects were observed at lower doses than effects observed in other brain-related assessments. Generally, assessments of brain gene expression or electrophysiology proved very sensitive. Further, motor activity, acoustic startle response and periventricular heterotopia were sensitive parameters even though the nature of the finding was not necessarily consistent across the substances (e.g. decreased motor activity vs. increased motor activity *vs.* altered habituation). Parameters related to cognitive function were never the most sensitive parameters among the neurodevelopmental parameters examined in the respective studies.

3.2. Thyroid-related findings

Key messages: For a balanced evaluation of serum T4, T3, and TSH data, Section 3.2 includes two select EPA PTU studies (also since the 13 EPA PTU studies generally yielded widely comparable T4, T3, and TSH data) and all further 38 studies, i.e. a total of 40 studies. All available data are considered in the evaluation of serum fT4/fT3, brain T4/T3, and thyroid weight and histopathology.

The subsections below summarise the thyroid-related findings from the four case studies. Parameters considered are serum levels of T4 in the dams and offspring (Section 3.2.1), serum levels of T3, fT4/fT3, and TSH in the dams and offspring (Sections 3.2.2–3.2.4), brain T4 and T3 levels in the offspring (Section 3.2.5) and thyroid weight and thyroid histopathology in the dams and offspring (Section 3.2.6). Hence, the findings are presented here separately for each thyroid-related parameter. A combined evaluation of the parameters, i.e. patterns of thyroid-related effects observed for the case study substances, is presented and discussed in Section 4.2.

Below, the overarching evaluation of the findings related to maternal and offspring serum T4, T3, and TSH levels covers two select EPA PTU studies and all further 38 studies, i.e. a total of 40 studies. The decision to include only two of the 13 EPA PTU studies considered that the EPA studies had very similar study design as regards dose setting and exposure period and that they were generally conducted in the same laboratory thereby generally yielding very similar serum T4, T3 and TSH levels. This decision was also taken in order not to create an imbalance towards PTU studies when evaluating findings across substances (all other case studies include up to five studies per substance).

Specifically, the two EPA PTU studies by (1) Goldey et al. (1995a) and (2) Johnstone et al. (2013), Gilbert et al. (2014, 2016), and O'Shaughnessy et al. (2018b) with further details in Hassan et al. (2017) were selected, because they included very high doses of 25 and 10 ppm, respectively,

Table 7. Sensitivity of neurodevelopmental and other brain-related assessments in those studies that yielded positive findings (see Table 8 for data from the EPA PTU studies).

Substance	Reference	Dose (mkd) Exposure period	Neurodevelopmental / brain-related assessments included in the respective study: lowest dose at which statistically significant effect was observed			No statistically significant effect relative to control
			Low dose	Mid dose	High dose	
PTU	Axelstad et al. (2008)	0.8, 1.6, 2.4 GD 7 – LD 17	No significant effects	**16 weeks: motor activity**	PND 14, 23: motor activity 5-6 months: radial arm maze, hearing function	PND 17: motor activity 8 / 9 weeks: water maze
	Kobayashi et al. (2005)	0.4, 1.0 2.5 GD 18 – LD 21	**9 weeks: auditory startle response**	PND 21: auditory startle PND 63: motor activity 9 weeks: swimming maze, mRNA expression	PND 21: motor activity: no habituation	PND 16: motor activity PND 21-22: swimming maze, mRNA expression
	Ramhøj et al. (2020a) [a]	1, 2.5 GD 7 – LD 22	**Lower of two doses: PND 16: gene expression**		PND 16: heterotopia	Not applicable
MMI	Ausó et al. (2004)	Approx. 36 GD 12 – GD 15	**One dose only: PND 40: wild runs following acoustic stimulus, some seizures; altered proportions of cells in somatosensory cortex**			Not applicable
	Darbra et al. (2003)	Approx. 36 GD 9 – LD 21	**One dose only: PND 21, 40, 60: effects on motor activity PND 80: passive avoidance tasks**			PND 21: Novelty-directed exploratory behaviour, anxiety behaviour
Amitrole	Ramhøj et al. (2021)	25, 50 GD 7 – LD 22	**PND 16: heterotopia, altered neuronal migration (assessed in top dose)**			Not applicable
Perchlorate	Gilbert and Sui (2008)	4.5, 44, 140 GD 6 – LD 30	**5-9 months: ↓ baseline synaptic transmission in hippocampal field potentials [b]**	5-9 months: ↓ inhibitory function (electrophysiology)	5-9 months: augmentation in long-term potentiation in population spike	3 months: water maze 7-8 months: fear conditioning 13 months: motor activity
Dietary iodine deficiency	Zhang et al. (2012)	1.2 µg I/day Premating – LD 21	**One dose only: PND 40-44: longer escape latency in Morris water maze PND 7, 45: altered gene expression in hippocampus and levels of brain-related proteins**			Not applicable
	Gilbert et al. (2013)	0.7-1.1, 0.2-0-3 µg I/day Premating – LD 21	**PND 120: ↓ synaptic transmission in dentate gyrus [b]**	Only two iodine deficient groups	Not applicable	PND 60: acoustic startle response PND 70: fear conditioning PND 100: water maze
TBBPA	Lilienthal et al. (2008)	3-3,000 Premating – LD 21	Data not presented relative to dose level: PND 50-110: females: dose-related ↑ in BAEP thresholds & in wave IV latency in low frequency range up to 4 kHz; males: ↑ of absolute latency of wave IV and interpeak latencies II–IV at low frequencies			PND 50-110: fear conditioning; males: BAEP thresholds; females and males: click thresholds PND 100-150: sweet preference
Aroclor 1254	Goldey et al. (1995b)	1, 4, 8 GD 6 – LD 21	No significant effects	**PND 15: motor activity PND ≥ 85: hearing deficits**	PND 24: acoustic startle amplitudes	PND 13, 17, 19, 21, 30, 60, 90: motor activity PND ≥ 85: acoustic startle response
	Goldey and Crofton (1998)	8 GD 6 – LD 21	**One dose only: PND 13, 15: motor activity; PND 23: acoustic startle response; PND 365: hearing deficits, baseline startle amplitudes**			PND 12, 19, 21, 29, 45, 66, 102, 121: motor activity
	Crofton et al. (2000)	6 GD 6 – LD 21	**One dose only: PND 85-95: hearing deficit**			PND 28, 70: acoustic startle response
Substance	**Reference**	**Dose (mkd) Exposure period**	**Neurodevelopmental / brain-related assessments included in the respective study: lowest dose at which statistically significant effect was observed**			**No statistically significant effect relative to control**
			Low dose	**Mid dose**	**High dose**	
DE-71	De-Miranda et al. (2016)	30 LD 5 – LD 22	**One dose only: PND 70, 100: females more errors in radial arm maze**			PND 40-42, 70, 100: motor activity
	Bowers et al. (2015)	0.3, 3, 30 GD 1 – LD 21	No significant effects relative to control	**PND 100 (110): motor activity (rearing)**	PND 90: acoustic startle response	PND 12, 15: grip strength PND 16, 55, 100, 230: motor activity / habituation PND 33, 60: beam test PND 35, 80: emergence latency PND 235: water maze PND 450: nicotine-induced motor activity
OMC	Axelstad et al. (2011b)	500, 750, 1,000 GD 7 – LD 17	No significant effects relative to control	No treatment-related significant effects relative to control	**9, 17 weeks: females: motor activity**	17 weeks: radial arm maze 28 weeks: acoustic startle response 31 weeks: oto-acoustic emissions, auditory brainstem response
	Ramhøj et al. (2020a) [a]	375, 500 GD 7 – LD 22	**Lower of two doses: PND 16: cortical gene expression (one gene)**		PND 16: cortical gene expression (five genes)	PND 16: heterotopia PND 22: relative brain weight

BAEP: brainstem auditory evoked potential; EPA: Environmental Protection Agency; GD: gestational day; LD: lactational day; mkd: mg/kg body weight/day; MMI: methimazole; mRNA: messenger RNA; PND: postnatal day; PTU: propylthiouracil; TBBPA: tetrabromobisphenol A.

Grey shading and bold text: Lowest dose level at which statistically significant neurodevelopmental or brain-related effects were observed.

See Appendix I for details on study design of these (mostly) investigational studies and see the respective SI-1 Case Study (CS-) Spreadsheets for details of the brain-related findings. This table does not consider the brain-related findings from the DE-71 study by Taylor et al. (2003) since this poster abstract does not refer to timepoints of measurement or to dose levels at which effects were observed.

[a]SOT poster.

[b]Electrophysiological findings were also recorded at 3 and 10 ppm PTU in the studies by Sui et al. (2005), Gilbert and Sui (2006), and Gilbert et al. (2007); see Table 8.

corresponding to ~2.5–3.3 and 1–1.3 mkd, respectively, and thus ranged close to the top dose of 2.4–2.5 mkd applied in the four studies in which PTU was administered *via* oral gavage (Kobayashi et al. 2005; Axelstad et al. 2008; Schneider et al. 2009 [SOT poster; SI-7]; Ramhøj et al. 2020a [SOT poster]). Notably, due to the differences in PTU administration (drinking water *vs.* oral gavage), the resulting internal doses (reflected by the areas under the curve) could differ between PTU studies even though the levels of the administered doses are comparable.

By contrast, as regards serum fT4/fT3, brain T4/T3, and thyroid weight and histopathology, Sections 3.2.3, 3.2.5, and 3.2.6, respectively, consider all studies in which these parameters were addressed.

Table 8. Sensitivity of neurodevelopmental and other brain-related assessments in the EPA PTU studies.

Study	Dose groups (ppm)	Lowest dose at which statistically significant effect was observed			No statistically significant effect relative to control
		Low dose	Mid dose	High dose	
Goldey et al. (1995a)	1, 2, 5, 25	No significant effects at two lowest doses	**PND 21: hearing deficits; PND 21, adult: Altered acoustic startle response**	PND 21: motor activity	Not applicable
Sui et al. (2005), Sui and Gilbert (2003)	3, 10	**Lower of two doses: PND 21-30: electrophysiological findings**		Not applicable	Basal levels of proteins implicated in synaptic plasticity unaffected
Sharlin et al. (2008)	1, 2, 3	**Total cell density in CC**	Density of GFAP protein positive astrocytes in CC and AC; density of MAG-positive oligodendrocytes in CC	Total cell density in AC; density of MAG-positive oligodendrocytes in AC	Not applicable
Sharlin et al. (2010)	1, 2, 3	No significant effects	**PND 15: DIO2 activity, DIO2 mRNA expression in retro-splenial cortex; RC3 mRNA expression in hippocampus**	PND 15: MCT8 mRNA in hippocampus (CA1, CA3)	Not applicable
Gilbert (2011)	1, 2, 3	**LTP of EPSP slope impaired**	EPSP slope amplitude decreased	↓ integrity of inhibitory synaptic processing; trace fear conditioning	Not applicable
Bastian et al. (2014), Spring et al. (2016)	1, 3, 10	**Heterotopia in some pups (incidence)**	Pronounced heterotopia (incidence and volume)	Not applicable	Not applicable
Gilbert et al. (2017)	3, 10	Lower of two doses: PND 21: no significant effects		**PND 30: smaller hippocampal volume, less neuronal differentiation**	Not applicable
Boyes et al. (2018)	1, 2, 3	No significant effects	**Adult: Altered pattern-elicited visual evoked potentials**	Adult: ↑ a-wave amplitude; green flicker ERG amplitude	Adult: b-wave parameters; amplitude of UV flicker ERGs
O'Shaughnessy et al. (2018b), Johnstone et al. (2013), Gilbert et al. (2014, 2016)	0.1, 0.5, 1, 2, 3	Three lowest doses: no significant effects	**GD 20: cortical gene expression**	Not applicable	Not applicable
	1, 2, 3, 10	**Adult: heterotopia**	PND 14: ≥ 2 ppm: alterations in cortical structure, synaptic physiology, water maze; gene expression	Adult: 10 ppm: duration and severity of clonus seizures significantly elevated	1 year: motor activity (radiotelemetry)
O'Shaughnessy et al. (2018a)	3	**One dose only, cross fostering study: PND 14, 18: heterotopia both after *in utero* treatment only and after *in utero* and lactational treatment (in this group: impaired context conditioning in males)**			No heterotopia after lactational treatment only
O'Shaughnessy et al. (2019)	10	**One dose only; PND 14: heterotopia; ↓ Shh expression, abnormal cell adhesion, altered radial glia morphology**			Not applicable

AC: anterior commissure; CA1, CA3: areas of the hippocampus; CC: *Corpus callosum*; EPSP: excitatory post-synaptic potential; ERG: electroretinogram; GFAP: glial fibrillary acidic protein; GD: gestational day; LD: lactational day; LTP: long-term potentiation; MAG: myelin-associated glycoprotein; MCT: monocarboxylate transporter; mkd: mg/kg body weight/day; mRNA: messenger RNA; PND: postnatal day; Shh: sonic hedgehog; UV: ultraviolet.
Grey shading and bold text: Lowest dose level at which statistically significant neurodevelopmental or brain-related effects were observed.

3.2.1. Serum levels of T4 in the dams and offspring

Findings from the four case studies related to maternal and offspring serum T4 levels are summarised in *Table SI-1, Spreadsheet "Overview T4 T3 TSH DNT."*

3.2.1.1. Alterations of serum T4 levels in the dams and offspring and timepoints of measurement.

Key messages: Maternal serum T4 levels were generally reduced, and the reductions were statistically significant at least at one timepoint of measurement. Offspring serum T4 levels were statistically significantly or non-significantly altered at least at one timepoint.

Maternal serum T4 levels were measured in 27 of the 40 studies (*Column G in Spreadsheet "Overview T4 T3 TSH DNT"*). Amongst these, 20 studies included maternal serum T4 measurements at more than one timepoint (often: once each during gestation and lactation). In all 27 studies that included maternal serum T4 levels, this parameter was altered at least at one timepoint. (The case study substances were selected as causing thyroid hormone imbalance.) Generally, the maternal serum T4 levels were reduced, and the decrements were statistically significant at least at one timepoint. Exceptions are a few studies, in which maternal serum T4 was non-significantly increased. Such non-significant increases in maternal serum T4 were observed in two PTU studies after the exposure period (Axelstad et al. 2008; Johnstone et al. 2013) and in the mercaptobenzimidazole and cyanamide studies (Ramhøj et al. 2021), where this finding coincided with non-statistically significant offspring thyroid hormone imbalance and absence of neurodevelopmental findings.

Offspring serum T4 levels were measured in 38 of the 40 studies (*Column M in Spreadsheet "Overview T4 T3 TSH DNT"*). Amongst these, 31 studies included offspring serum T4 measurements at more than one timepoint. Mostly, offspring serum T4 was measured at different timepoints during lactation; in 12 studies, it was additionally measured at the end of gestation (GD 20–21), and in 15 studies (additionally) on ≥ PND 40. In eight studies, the earliest timepoint of offspring serum T4 measurement was at the end of lactation (PND 20/21) or post-weaning (PTU: Kobayashi et al. 2005; MMI: Ausó et al. 2004; Darbra et al. 2003; TBBPA: Cope et al. 2015; Saegusa et al. 2009, 2012; Van der Ven et al. 2008/Lilienthal et al. 2008; DE-71: de-Miranda et al. 2016; Bowers et al. 2015/Gill et al. 2016).

In 37 of the 38 studies that included offspring serum T4 measurements, this parameter was altered at least at one timepoint. The only exception is the MMI study by Ausó et al. (2004) who found serum T4 levels to be unaffected in juvenile rats on PND 40 after four-day *in utero* exposure to MMI from GD 12 to GD 15.

Just as in the dams, serum T4 levels in the offspring were also generally reduced. Exceptions are the mercaptobenzimidazole study by Ramhøj et al. (2021), where offspring serum T4 was statistically significantly increased by 50% on PND 6. A few other studies reported non-statistically significant increases in offspring serum T4 levels (Saegusa et al. 2009,

2012; Axelstad et al. 2011b). The ECETOC T4 TF considers the non-significant findings to reflect normal data variability.

3.2.1.2. Extent of serum T4 reduction in the dams and offspring.

Key messages: Based on available gestational and neonatal timepoints (GD 20–PND 0), maternal serum T4 decrements tended to be associated with offspring serum T4 decrements, however, not to the same magnitude. During or at the end of lactation, serum T4 reduction was often more pronounced in the offspring than in the dams, and this trend was observable across case studies.

The comparison of the extent of maternal *vs.* offspring serum T4 reduction (*Column R in Spreadsheet "Overview T4 T3 TSH DNT"*) was impaired by the circumstance that, in many studies, T4 was measured at different timepoints in the dams and offspring. Considering serum T4 pairs (maternal *vs.* offspring) only if they had been collected on the same day or one day apart yielded the following result:

- T4 reduction at the end of gestation:
  - *4 studies—more pronounced in dams than in offspring*: one PTU study, two perchlorate studies, one iodine deficiency study
  - *6 studies—more pronounced in offspring than in dams*: one PTU study, one PFHxS study, two Aroclor 1254 studies, one DE-71 study, one triclosan study
  - *Almost equal serum T4 reduction*: one triclosan study except for one datapoint where a non-statistically significant serum T4 decrement was slightly more pronounced in the dams than in the offspring
- T4 reduction during or at the end of lactation:
  - *3 studies—more pronounced in dams than in offspring*: one MMI study, one ETU study, one triclosan study, one OMC study (in the latter, up to 100% maternal serum T4 reduction; further discussed in Section 4.2.4)
  - *10 studies—more pronounced in offspring than in dams*: two PTU studies, one mancozeb study, two perchlorate studies, one iodine deficiency study, two Aroclor 1254 studies, two DE-71 studies

Hence, across all studies considered, serum T4 reduction at the end of gestation was more pronounced in the dams than in the offspring in one of two studies for the TPO inhibitor PTU (Case Study 1) as well as for the NIS inhibitor perchlorate and for dietary iodine deficiency (both: Case Study 2). By comparison, in the recent perchlorate study by Gilbert et al. (2022), that was published after data evaluation for the present review was complete, offspring serum T4 decrements were more pronounced than maternal serum T4 decrements on GD 20 (data not shown in Figure 1); in this study, the dams were fed with iodine-restricted diet to prevent a masking of effects of the NIS inhibitor (Section 4.2.2.1).

Serum T4 reduction at the end of gestation was generally more pronounced in the offspring than in the dams (or almost equal) for the Case Study 3 substances PFHxS, Aroclor 1254, DE-71, and triclosan, which are assumed to enhance thyroid hormone clearance (Figure 1). Serum T4 reduction

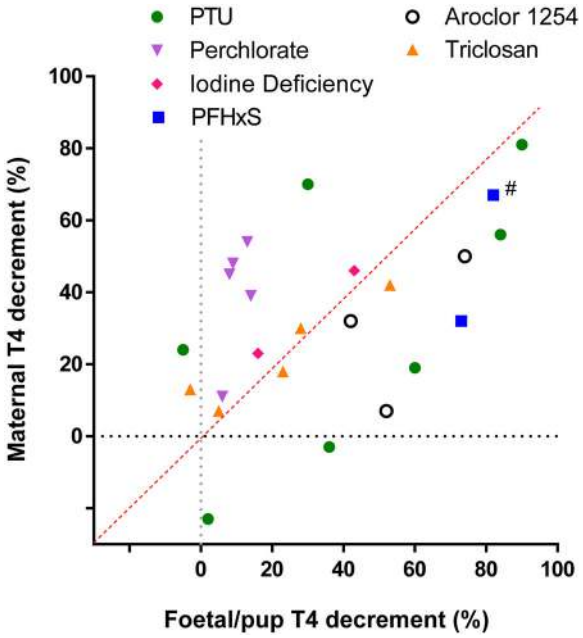


Figure 1. Relationship between maternal serum T4 decrements measured on GD 20–GD 21 and foetal/pup serum T4 decrements measured on GD 20–PND 0. GD: gestational day; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T4: thyroxine. This graph compares all (statistically significant or non-significant) data recorded at all dose levels from the studies that included maternal serum T4 measurements on GD 20–GD 21 and offspring serum T4 measurements on GD 20–PND 0. All data expressed as percent of concurrent control group for the respective study. [#]Datapoint for 25 mkd PFHxS as published in the Supplementary Table 1 to Gilbert et al. (2021) referring to a study by Ramhøj et al. Values below the dotted vertical and/or horizontal lines represent increased T4 levels. The dotted red line is an arbitrary line with a slope of 1 that was used as visual aid but not for any specific statistical analysis.

during or at the end of lactation was often more pronounced in the offspring than in the dams, and this trend was observable across case studies.

Some studies included separate T4 measurements for male and female offspring. A comparison of the serum T4 data available for the two sexes showed no clear trend for increased sensitivity of males or females with respect to the extent of serum T4 reduction (*Column M in Spreadsheet "Overview T4 T3 TSH DNT"*).

To further evaluate the extent of maternal *vs.* offspring serum T4 reduction, a threshold of ≥60% serum T4 reduction (in the dams or offspring) as compared to the concurrent controls was empirically set based upon the dataset from the four case studies. In this comparison, the most pronounced maternal and/or offspring serum T4 reduction observed in the respective study was applied regardless of timepoint of measurement (Table 9; see Section 3.2.1.3 for rationale to set the ≥60% threshold).

Maternal serum T4 reductions ≥60% were recorded in seven studies. These pronounced maternal serum T4 reductions only coincided with ≥60% offspring serum T4 reductions in the PTU study by Schneider et al. (2009) [SOT poster; SI-7] and in the amitrole study by Ramhøj et al. (2021; Table 9). However, offspring serum T4 levels were often reduced by ≥60% without concordant ≥60% maternal serum T4 decrements (provided that maternal serum T4 data were available). For example, ≥60% offspring serum T4 reductions were recorded in all PTU studies, all Aroclor 1254 studies and in

Table 9. Studies with ≥60% maternal and/or offspring serum T4 decrements.

Case study	Substance	≥ 60% serum T4 decrement				Reference
		Dose level (mkd)	Maternal serum T4	Dose level (mkd)	Offspring serum T4	
CS1	PTU	2.4	59%	2.4	84%	Axelstad et al. (2008)
		NA	NA	2.5	71%	Kobayashi et al. (2005)
		NA	NA	2.5	91%	Ramhøj et al. (2020a) [SOT poster]
		2.5	70%	2.5	81%	Schneider et al. (2009) [SOT poster; SI-7]
		NA	NA	Approx. 7.5-10 [a]	98%	Goldey et al. (1995a)
		Approx. 0.9-1.2 [b]	54%	Approx. 3-4 [b]	92%	Johnstone et al. (2013)
	MMI	NA	NA	Approx. 36 [c]	78%	Darbra et al. (2003)
	ETU	Approx. 10-17 [d]	76%	Approx. 10-17 [d]	53%	Marty et al. (2013) as cited in European Commission (2017a)
	Amitrole	50	85%	50	64%	Ramhøj et al. (2021)
CS2	Perchlorate	140	60%	140	21%	Gilbert and Sui (2008)
	Dietary iodine deficiency	[e]	69%	[e]	42%	Zhang et al. (2012)
		[f]	51%	[f]	62%	Gilbert et al. (2013)
CS3	PFHxS	25	65%	25	43%	Ramhøj et al. (2018, 2020b)
		50	32%	50	75%	Gilbert et al. (2021)
	Aroclor 1254	NA	NA	8	89%	Goldey et al. (1995b)
		NA	NA	8	80%	Goldey and Crofton (1998)
		6	40%	6	82%	Crofton et al. (2000b)
		25	50%	25	74%	Morse et al. (1996)
	DE-71	30.6	58%	30.6	80%	Kodavanti et al. (2010)
		30	48%	30	66%	Zhou et al. (2002)
		NA	NA	40	87%	Ramhøj et al. (2020a) [SOT poster]
		30	Reduced in HDG	30	93%	Bowers et al. (2015) / Gill et al. (2016)
CS4	OMC	1,000	100% [g]	1,000	35%	Axelstad et al. (2011b)

CS: case study; DE-71: a mix of polybrominated diphenyl ethers; ETU: ethylene thiourea; HDG: high-dose group; mkd: mg/kg body weight/day; MMI: methimazole; NA: not available; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PTU: propylthiouracil; SOT: Society of Toxicology.
Grey shadings indicate ≥60% serum T4 decrements. If the respective study included more than one timepoint of serum T4 measurement, the most pronounced T4 decrement is recorded in this table; see *Supplementary Table SI-1, Spreadsheet "Overview T4 T3 TSH DNT"* for details.
[a]Applied dose: 25 ppm in the drinking water.
[b]Applied dose: 10 ppm in the drinking water.
[c]Applied dose: 0.2 mg/mL.
[d]Applied dose: 140 ppm in the drinking water.
[e]Low iodide diet: 1.2 µg I/day.
[f]Low iodide diet: 0.2–03 µg I/day.
[g]See Section 4.2.4 for discussion.

most DE-71 studies (Table 9). In the DE-71 study by de-Miranda et al. (2016), 52% and 48% serum T4 reduction in the male and female offspring, respectively, were recorded on PND 23 as only timepoint of T4 measurement; therefore, it is unclear if T4 reduction was more pronounced—and possibly exceeded the ≥60% threshold—around mid-lactation.

3.2.1.3. Are ≥60%/≥50% offspring serum T4 reductions in the top-/lower-dose groups indicative of statistically significant neurodevelopmental findings?

Key messages: The ECETOC T4 TF empirically set thresholds of ≥60%/≥50% offspring serum T4 reduction in the top-/lower-dose groups and evaluated whether T4 decrements meeting these thresholds are indicative of statistically significant neurodevelopmental findings. The present database showed some association between extent of offspring serum T4 decrement and statistically significant neurodevelopmental effects in rats.

The ECETOC T4 TF aimed to identify thresholds of offspring serum T4 reduction that were indicative of statistically significant neurodevelopmental findings in the present database (*Table SI-1 Spreadsheet "Overview Offspring T4 at DNT dose"*). Neurodevelopmental findings considered in this evaluation included altered motor activity, acoustic startle response, cognitive function, hearing and periventricular heterotopia, as well as electrophysiological findings as indicative of transient perturbations of neuronal function. By comparison, altered brain DIO2 activity or altered brain gene expression as sole brain-related findings were not assessed as neurodevelopmental findings since they do not *per se* reflect neurodevelopmental impairment. Therefore, the Aroclor 1254 study by Morse et al. (1996) was excluded from this part of the evaluation since DIO2 activity was the only brain-related parameter considered therein. Also, the OMC study by Ramhøj et al. (2020a [SOT poster]) was assessed as showing no adverse neurodevelopmental effects since the only statistically significant brain-related findings were changes in gene expression whereas periventricular heterotopia was not observed.

In striving to identify which level of offspring serum T4 reduction was indicative of statistically significant neurodevelopmental findings, the ECETOC T4 TF evaluated the present database to empirically set one or more thresholds that yielded the fewest possible numbers of false negative and false positive outcomes; see Table 10 for decision matrix. This evaluation showed the need to empirically set two thresholds. The empirically set threshold of ≥60% offspring serum

Table 10. Decision-matrix applied for the further evaluation of offspring serum T4 reduction.

		Offspring serum T4 reduction ≥60% in top-dose group and ≥50% in lower-dose groups?	
		Yes	No
Statistically significant neurodevelopmental findings?	Yes	True positive	False negative
	No	False positive	True negative

The ECETOC T4 TF set the ≥60% and ≥50% offspring serum T4 reduction thresholds empirically to yield the fewest numbers of false negative and false positive findings in the database collated for the present review. The assessment schemes were then used for an evidence-based and pragmatic analysis to further evaluate whether the mentioned thresholds are indicative of statistically significant neurodevelopmental findings in this database (Sections 3.2.1.3 and 4.4.3).

T4 reduction introduced in Section 3.2.1.2 was only applied for the top-dose groups. For the lower-dose groups, the ECETOC T4 TF empirically set a further threshold of ≥50% offspring serum T4 reduction. The ≥60% threshold is associated with the level of offspring serum T4 seen at (or close to) the maximum tolerated dose, whereas the ≥50% threshold reflects the extent of offspring serum T4 reduction seen at the lowest dose level at which statistically significant neurodevelopmental findings were observed. Accordingly, the two thresholds relate to two different experimental settings, and application of both thresholds yielded the fewest numbers of false negative and false positive results for the respective dose groups (preliminary evaluations that yielded the ≥60%/≥50% thresholds not shown). It is important to note that the ≥60%/≥50% thresholds were not set in view of conducting a statistically driven analysis but to facilitate an evidence-based pragmatic analysis of the present database. Further, this evaluation does not aim to define one or more thresholds for use in regulatory toxicity testing. Rather, it aims to determine phenomenological thresholds based upon the present database, whereas their relevance and reliability for regulatory toxicity testing remains to be established (Section 4.4.7).

After the empirical setting of the ≥60%/≥50% thresholds, the evaluation scheme presented in Table 10 was applied to assign all relevant studies as either true positive, false positive, true negative or false negative. If a study included more than one dose level, both thresholds were applied; if a study included only one dose level, only the ≥60% threshold was applied. The evaluation aimed at answering the question: Are offspring serum T4 levels that meet the empirically set ≥60%/≥50% thresholds in the top-dose group/lower-dose groups indicative of statistically significant neurodevelopmental findings? This evaluation follows the approach taken in the EFSA and ECHA (2018) Endocrine Disruptor Guidance: "*A decrease in T4 (total or free) in the absence of adverse histological changes should act as a trigger for further studies*," i.e. neurodevelopmental assessments. Thereby, this evaluation is based upon the assumption that observed neurodevelopmental effects are the variables that depend on the extent of offspring thyroid hormone imbalance. Nonetheless, substances can also cause neurodevelopmental effects by non-thyroid pathways (Section 4.2).

The application of the ≥60%/≥50% offspring serum T4 reduction thresholds yielded these results (Table 11 and *Column L* in *Spreadsheet "Overview Offspring T4 at DNT Dose"*):

- 10 studies true positive, i.e. ≥60%/≥50% offspring serum T4 reduction in the top-/lower-dose groups and statistically significant neurodevelopmental findings: four PTU studies, one MMI study, one amitrole study, three Aroclor 1254 studies, one DE-71 study
- Three studies false positive, i.e. ≥60%/≥50% offspring serum T4 reduction in the top-/lower-dose groups but no statistically significant neurodevelopmental findings: one PFHxS study, two DE-71 studies
- 11 studies true negative, i.e. <60%/<50% offspring serum T4 reduction in the top-/lower-dose groups and no statistically significant neurodevelopmental findings: one study each for MMI, ETU, mancozeb, mercaptobenzimidazole, cyanamide, PFHxS, triclosan and OMC and two studies each for perchlorate and TBBPA
- Four studies false negative i.e. <60%/<50% offspring serum T4 reduction in the top-/lower-dose groups but statistically significant neurodevelopmental findings: two dietary iodine deficiency studies, one perchlorate study, and one OMC study

For two of the four studies that were evaluated as "false negative," the observed brain-related effects were electrophysiological findings (Gilbert and Sui 2008; Gilbert et al. 2013). For the third study (the OMC study by Axelstad et al. 2011b), the pattern of maternal and offspring hormone-related effects does not give a clear picture for an association between altered thyroid function and neurodevelopmental effects (further discussed in Section 4.2.4). The fourth study is the dietary iodine deficiency study by Zhang et al. (2012) who observed altered learning and memory but only moderate offspring serum T4 reduction (39% on PND 7). These findings do not match the findings from the other dietary iodine deficiency study, Gilbert et al. (2013), where more pronounced iodine deficiency than applied by Zhang et al. resulted in pronounced offspring serum T4 reduction without concordant alterations in learning and memory (further discussed in Section 4.2.2.2).

Five studies were excluded from the evaluation of the ≥60%/≥50% thresholds either because the ECETOC T4 TF questioned the suitability of the respective timepoints of offspring serum T4 measurement to determine the maximum extent of T4 reduction or because this evaluation was not possible based upon the information provided (Table 12). As described above, the Aroclor 1254 study by Morse et al. (1996) was also excluded from the evaluation of the ≥60%/≥50% offspring serum T4 reduction thresholds since brain DIO2 activity was the only brain-related parameter included therein. Interestingly, this study is nonetheless "true positive." Brain DIO2 activity was altered in both dose groups (5 and 25 mkd). The ≥60%/≥50% thresholds are met on GD 20

Table 11. Application of the ≥60%/≥50% offspring serum T4 reduction thresholds to assign studies as either true positive, false positive, true negative, or false negative (see decision-matrix in Table 10).

Substance	Study	Comment
True positive: ≥60%/≥50% offspring serum T4 reduction in top-/lower-dose groups AND statistically significant neurodevelopmental findings		
PTU	Axelstad et al. (2008) Ramhøj et al. (2020a)[a] Goldey et al. (1995a) Johnstone et al. (2013), Gilbert et al. (2014, 2016), O'Shaughnessy et al. (2018b), Hassan et al. (2017)	Following the T4 data provided by Johnstone et al. (2013), the ≥60%/≥50% thresholds are applicable at 2, 3, and 10 ppm PTU, but not at 1 ppm PTU since offspring serum T4 was non-significantly reduced by 29% on PND 4 at this dose level whereas statistically significant periventricular heterotopia was observed (Gilbert et al. 2014). However, foetal serum T4 decrements were pronounced at 1 ppm (61% on GD 20) following the T4 data provided by Hassan et al. (2017) and O'Shaughnessy et al. (2018b) for this same study (foetal serum T4 data not provided by Johnstone et al.). The ECETOC T4 TF considers the pronounced foetal T4 decrement overriding to the moderate T4 decrement observed on PND 4 and thus suggests that the periventricular heterotopia observed at 1 ppm (which was not reproduced in other studies; Section 4.3.4) was associated with pronounced offspring serum T4 reduction (≥50%) during the window of susceptibility.
MMI	Darbra et al. (2003)	
Amitrole	Ramhøj et al. (2021)	The brain-related finding is periventricular heterotopia in the HDG (50 mkd amitrole). The ≥60% threshold was met on PND 16 (64%), but it was just undercut on PND 6 (56%). Hence, the most pronounced T4 reduction does not fall within the (presumed) window of susceptibility (GD 19 to approximately PND 2–6 upon exposure to PTU; O'Shaughnessy et al. 2019; Ramhøj et al. 2020a).[a]
Aroclor 1254	Goldey et al. (1995b) Goldey and Crofton (1998) Crofton et al. (2000b)	
DE-71	Bowers et al. (2015), Gill et al. (2016)	Offspring serum T4 was only measured on PND 21, and it was reduced by 52/51% and 93/91% in the male/female offspring of the MDG and HDG, respectively. These T4 decrements coincided with reduced rearing in the MDG and HDG offspring on PND 100 (as only motor activity-related effect at this timepoint) and with altered acoustic startle responses in the HDG offspring on PND 90 (but not on PND 20). There were no (or no dose-dependent) motor activity-related effects at any other timepoint (PND 12–450), and different tests for cognitive function were not altered on PND 35, 80, and 235.
False positive: ≥60%/≥50% offspring serum T4 reduction in top-/lower-dose groups BUT NO statistically significant neurodevelopmental findings		
PFHxS	Gilbert et al. (2021)	Offspring serum T4 reduction by 73%, 67%, 75%, 54% on PND 0, 2, 6, 14, respectively, but no periventricular heterotopia and no effect on acoustic startle response or trace fear conditioning (suggested window of susceptibility for periventricular heterotopia upon exposure to PTU: GD 19 to PND 2–6; suggested window of susceptibility to altered cognitive function: during lactation up until weaning; Table 5).
DE-71	Kodavanti et al. (2010)	Pronounced offspring serum T4 reduction on PND 14 and 21 but no statistically significant effects on functional observational battery or motor activity on PND 24, 58–60, or 273.
	Ramhøj et al. (2020a)[a]	Pronounced offspring serum T4 reduction on GD 21 and PND 16 but no periventricular heterotopia, no statistically significant effects on cortical gene expression and no effects on brain weight on PND 16.
True negative: <60%/<50% offspring serum T4 reduction in top-/lower-dose groups AND NO statistically significant neurodevelopmental findings		
MMI	Fegert et al. (2012)	
ETU	Marty et al. (2013) in European Commission (2017a)	
Mancozeb	Axelstad et al. (2011a)	
MBI	Ramhøj et al. (2021)	
Cyanamide	Ramhøj et al. (2021)	
Perchlorate	York et al. (2004) York et al. (2005a, 2005b)	
TBBPA	Cope et al. (2015)	
	Saegusa et al. (2009, 2012)	Only evaluated morphological alterations in the brain, while discussing that these are indicative of DNT.
PFHxS	Ramhøj et al. (2018, 2020b)	
Triclosan	Gilbert et al. (2021)	
OMC	Ramhøj et al. (2020a) [a]	No periventricular heterotopia, some alterations in cortical gene expression as sole brain-related findings.
False negative: <60%/<50% offspring serum T4 reduction in top-/lower-dose groups BUT statistically significant neurodevelopmental findings		
Perchlorate	Gilbert and Sui (2008)	21% T4 reduction on PND 21 (non-statistically significant T4 reduction on PND 4 and 14) and electrophysiological findings at 5–9 months (reduced baseline synaptic transmission in hippocampal field potentials, reduced inhibitory function and augmentation in long-term potentiation in the population spike), but no effects on motor activity at 13 months, on spatial learning in water maze at 3 months or fear conditioning at 7–8 months.
Dietary iodine deficiency	Zhang et al. (2012)	39% and 42% T4 reduction on PND 7 and PND 45, respectively, but decreased spatial learning ability in the low iodine offspring.
	Gilbert et al. (2013)	Electrophysiological findings in both iodine-deficient groups (at PND 120; after surgical implantation of electrodes), but the ≥60%/≥50% thresholds are only met in the "excessive iodine deficiency" group and only on PND 21 (62% T4 reduction), but not on GD 20 or PND 14 (43% and 49% T4 reduction, respectively); in the "iodine deficiency group," T4 reductions were 16, 19% (non-significant), 19% on GD 20, PND 14, PND 21, respectively.
OMC	Axelstad et al. (2011b)	Up to 35% serum T4 reduction in the male offspring on PND 16, non-significant T4 reduction in the female offspring; reduced motor activity in the HDG females at 9 and 17 weeks (but up to 100% serum T4 reduction in the dams); reduced errors in the males in radial arm maze but no (or no statistically significant) effects on acoustic startle response, oto-acoustic emissions, or auditory brainstem response.

DE-71: a mix of polybrominated diphenyl ethers; DNT: developmental neurotoxicity; ETU: ethylene thiourea; GD: gestational day; HDG: high-dose group; MBI: mercaptobenzimidazole; MDG: mid-dose group; MMI: methimazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; SOT: Society of Toxicology; T4: thyroxine; TBBPA: tetrabromobisphenol A.

[a]SOT poster, see bibliography for weblink.

Table 12. Studies that the ECETOC T4 TF excluded from the evaluation of the ≥60%/≥50% offspring serum T4 reduction thresholds.

Substance	Study	Comment
PTU	Kobayashi et al. (2005)	Offspring serum T4 was measured only at PND 21 and 9 weeks. The ≥60%/≥50% thresholds were met in the HDG and MDG, and most brain-related findings were recorded either only in the HDG or in the HDG and MDG. However, at 9 weeks, the acoustic startle response was also altered in the LDG, and at this dose level, serum T4 in the offspring was non-significantly reduced (by 27%) on PND 21 and non-significantly increased (by 7%) at 9 weeks. The ECETOC T4 TF suggests that offspring serum T4 decrements may well have been ≥50% in the LDG in this study earlier during lactation.
MMI	Ausó et al. (2004)	This study included four days of exposure to 0.2 mg/mL (36 mkd) MMI during gestation (GD 12–15), but measurements of T4 on PND 40 where T4 levels were unaffected. By comparison, Darbra et al. (2003) measured 78% T4 reduction on PND 21 upon application of an equally high dose of MMI from GD 9 to PND 21. As regards neurodevelopmental findings, Ausó et al. reported that, on PND 39, "*52% of the pups born to the goitrogen-treated dams responded to an acoustic stimulus with wild runs, followed in some by seizures*." Ausó et al. also reported single ectopic cells in the primary somatosensory cortex and the hippocampal CA1 area and specified these as heterotopic. The ECETOC T4 TF suggests that offspring serum T4 decrements may well have been ≥60% during the presumed window of susceptibility for periventricular heterotopia, i.e. GD 19 to PND 2–6 (O'Shaughnessy et al. 2019), with increased seizure susceptibility reported in Gilbert et al. (2014).
TBBPA	Van der Ven et al. (2008), Lilienthal et al. (2008)	Offspring serum T4 was only measured at 14 weeks, where serum T4 was reduced by 48/46% in the male/female offspring. Tests for cognitive function (fear conditioning, sweet preference) yielded no findings, but BAEPs were dose-dependently altered (further discussed in Section 4.2.3.1). Thus, the moderate offspring serum T4 decrement is consistent with the absence of findings on cognitive function. Also, the ECETOC T4 TF suggests that offspring serum T4 decrements may well have been ≥60% during the first three postnatal weeks, i.e. during lactation, as window of susceptibility to most neurodevelopmental effects.
DE-71	de-Miranda et al. (2016)	Offspring serum T4 was reduced by 52% on PND 23 as only timepoint of T4 measurement. As regards effects on neurodevelopment, the PND 70–100 female offspring made more errors in the radial arm maze; however, motor activity was not altered in males or females on PND 40–42 or PND 70–100. By comparison, in the DE-71 studies by Zhou et al. (2002), Kodavanti et al. (2010), and Ramhøj et al. (2020a),[a] the most pronounced T4 decrement was recorded on PND 6 or 14 (Section 3.2.1.4), and it did exceed 60%. Thus, considering that offspring T4 was reduced by 52% on PND 23 in the study by de-Miranda et al., it might well have been reduced by ≥60% on PND 6–14.
DE-71	Zhou et al. (2002), Taylor et al. (2003)	Evaluation of the ≥60%/≥50% thresholds not possible since the only brain-related findings available for this study are provided as high-level information on the Taylor et al. (2003) poster abstract and do not include information on the dose level at which effects were observed.

BAEPs: brainstem auditory evoked potential; DE-71: a mix of polybrominated diphenyl ethers; GD: gestational day; HDG: high-dose group; LD: lactational day; LDG: low-dose group; MDG: mid-dose group; MMI: methimazole; PND: postnatal day; PTU: propylthiouracil; SOT: Society of Toxicology; T4: thyroxine; TBBPA: tetrabromobisphenol A.
[a]SOT poster, see bibliography for weblink.

(T4 reductions by 52% and 74% at 5 and 25 mkd, respectively) but not on PND 4, 21, or 90. As substance exposure was from GD 10 to GD 16, the ECETOC T4 TF considers serum T4 measurement on GD 20 most relevant to determine maximum T4 reduction.

Finally, the empirically set threshold of ≥60% offspring serum T4 reduction was re-evaluated, but this time considering (1) only the most pronounced offspring serum T4 decrement measured in that study regardless of dose level or timepoint of measurement; and (2) whether any statistically significant neurodevelopmental findings were recorded or not (yes/no). In this evaluation, 13 studies were established as true positive, two studies as false positive, 11 studies as true negative, and three studies as false negative (Supplementary Information SI-4.1). Hence, the evaluation was similar to that considering both offspring serum T4 reduction thresholds, but three further studies were assessed as true positive:

- The dietary iodine deficiency study by Gilbert et al. (2013), which was false negative in the evaluation of the ≥60%/≥50% thresholds
- The PTU study by Kobayashi et al. (2005) and the DE-71 study by Zhou et al. (2002) and Taylor et al. (2003), which were both excluded from the evaluation of the ≥60%/≥50% thresholds (Table 12)

3.2.1.4. Trends in offspring serum T4 reduction over the course of the exposure period.

Key messages: For PTU, ETU, amitrole, and perchlorate exposures, offspring serum T4 decrements aggravated over the course of the exposure period. For MMI (in the EOGRTS-like study), cyanamide, PFHxS, triclosan, and OMC exposures, offspring serum T4 decrements attenuated over the course of the exposure period. Thus, T4 attenuation was generally associated with the absence of statistically significant neurodevelopmental findings. For Aroclor 1254 and DE-71, offspring serum T4 decrements were generally more pronounced on PND 6–14 than earlier or later during lactation.

For those studies that included several timepoints of serum T4 measurement in the dams or offspring, it was evaluated if serum T4 decrements either aggravated or attenuated over the course of the exposure period (*Columns K and Q in Spreadsheet "Overview T4 T3 TSH DNT"*). Aggravations vs. attenuations in maternal serum T4 decrements were not associated with the presence or absence of neurodevelopmental alterations (*Column S*), and it was also not possible to assign such trends over time to specific MoAs (*Column A*).

As regards offspring serum T4 decrements (*Column Q*), the identification of trends was impaired by the circumstance that timepoints of measurement (*Column P*) differed considerably between studies. The further evaluation of trends in offspring serum T4 decrements only considered those 18

studies in which offspring serum T4 had been measured several times over the course of the exposure period and thus also several times during the critical period for neurodevelopment (i.e. between approximately GD 20–PND 4 and the end of lactation). This evaluation yielded the following outcome:

Offspring serum T4 decrements aggravated over the course of the exposure period for PTU (Goldey et al. 1995a; Schneider et al. 2009 [SOT poster, SI-7]; Johnstone et al. 2013; Ramhøj et al. 2020a [SOT poster]), ETU (Marty et al. 2013 in European Commission 2017a), amitrole (Ramhøj et al. 2021), perchlorate (York et al. 2005a), Aroclor 1254 (Goldey et al. 1995b; Goldey and Crofton 1998; Crofton et al. 2000b) and DE-71 as per Kodavanti et al. (2010) (Figures 2(A,B)). By comparison, in the DE-71 study by Ramhøj et al. (2020a [SOT

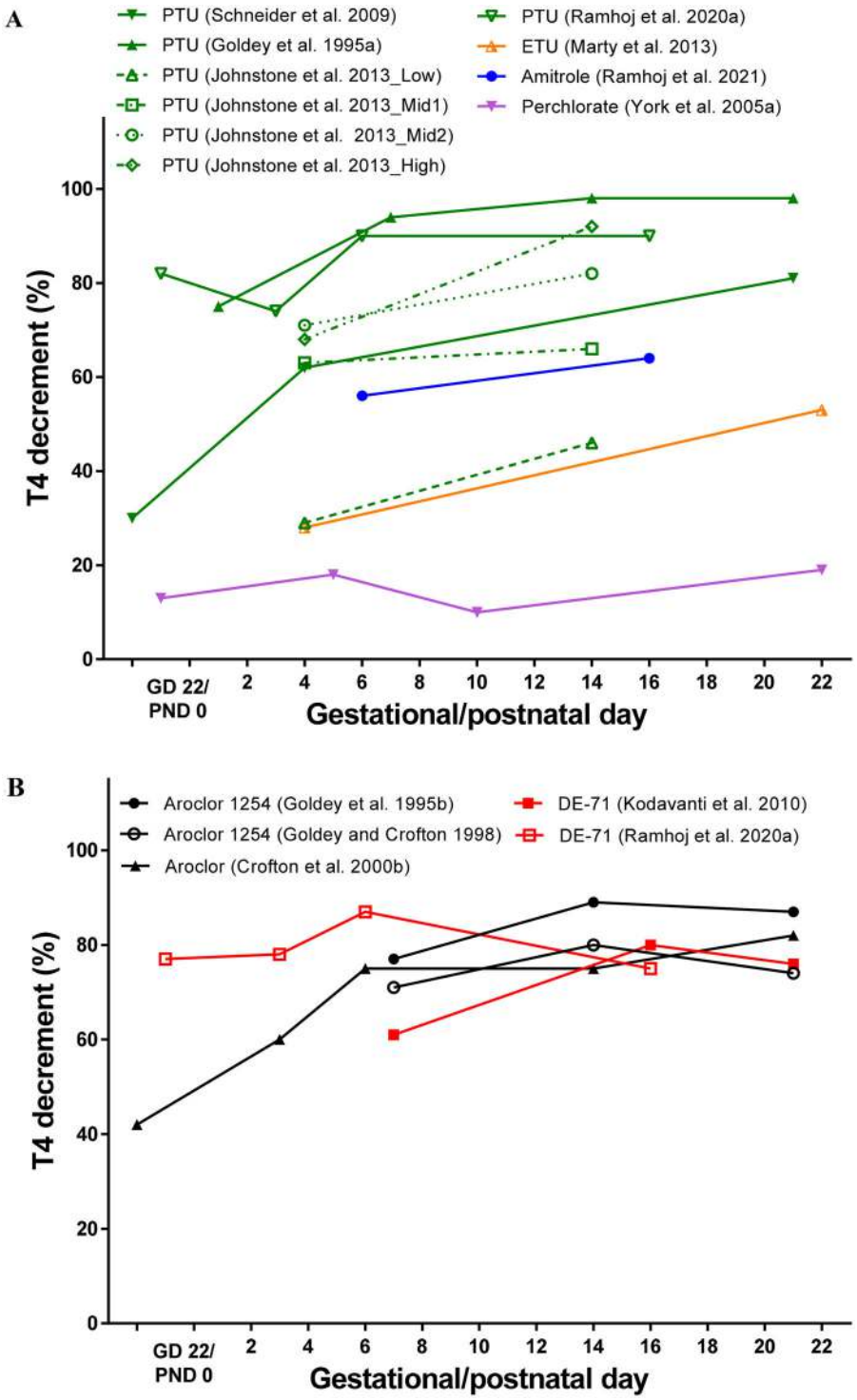


Figure 2. Trends in offspring serum T4 decrements during exposure period.
(A) Case Study 1—PTU, ETU, amitrole: aggravation of offspring serum T4 decrements during exposure period; Case Study 2—perchlorate: unclear trend for offspring serum T4 decrement.
(B) Case Study 3—Aroclor 1254, DE-71: studies in which offspring serum T4 decrements were most pronounced around mid-lactation and/or aggravated during exposure period.
(C) Case Study 1—MMI, cyanamide; Study 3—PFHxS, triclosan, one Aroclor 1254 study; Case Study 4—OMC: attenuation of offspring serum T4 decrements during exposure period.
ETU: ethylene thiourea; GD: gestational day; MMI: methimazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: Propylthiouracil; T4: Thyroxine. See Section 3.2.1.4 for explanation of these figures.

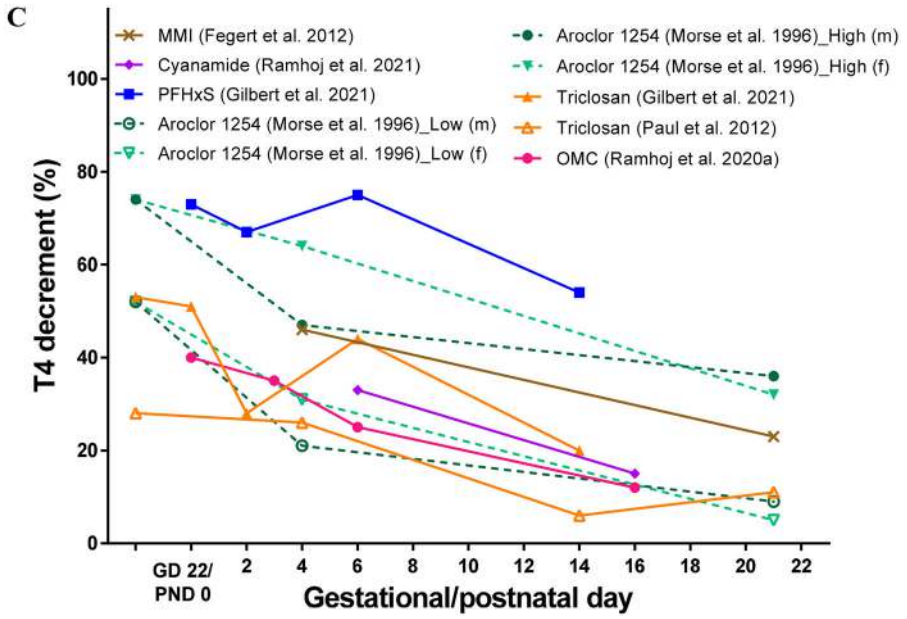


Figure 2. (Continued).

poster]), there was no clear trend for offspring serum T4 decrements over the course of the exposure period (Figure 2(B)).

Offspring serum T4 decrements attenuated over the course the exposure period for MMI in the EOGRTS-like study (Fegert et al. 2012), cyanamide (Ramhøj et al. 2021), PFHxS and triclosan (Gilbert et al. 2021; Paul et al. 2012) as well as OMC (Ramhøj et al. 2020a [SOT poster]). Hence, the attenuation of offspring serum T4 reduction between late gestation and the end of lactation (i.e. over the course of the exposure period) was generally associated with either the absence of brain-related findings or the presence of only single statistically significant neurodevelopmental findings from amongst a broad spectrum of brain-related parameters assessed (Figure 2(C)).

Further, in the studies evaluating Aroclor 1254 and DE-71, offspring serum T4 decrements were generally more pronounced during mid-lactation (PND 6–14) than earlier or later during lactation (*Column Q, cells with blue shading, in Spreadsheet "Overview T4 T3 TSH DNT"*). Specifically, from amongst the Aroclor 1254 studies, offspring serum T4 reduction was most pronounced on PND 14 in the studies by Goldey et al. (1995b) and Goldey and Crofton (1998) (Figure 2(B)). By comparison, in the Aroclor 1254 study by Crofton et al. (2000b), offspring serum T4 reduction was most pronounced on PND 21 (Figure 2(B)). However, in this study, the top dose was 6 mkd Aroclor 1254 as compared to 8 mkd in the studies by Goldey et al. and Goldey and Crofton. Therefore, the dose applied by Crofton et al. (2000b) might not have been sufficiently high to elicit a more pronounced offspring serum T4 decrement around PND 14. In the Aroclor 1254 study by Morse et al. (1996), T4 reduction was more pronounced on GD 20 than on PND 4, 21, or 90 (Figure 2(C)); however, this study included exposure only from GD 10 to 16. Therefore, the exposure duration may well have been too short to elicit a more pronounced offspring serum T4 decrement around PND 14 (on which day T4 was further not measured in this study).

From amongst the DE-71 studies, Kodavanti et al. (2010) reported that offspring serum T4 reduction was most pronounced on PND 14. Further, Ramhøj et al. (2020a [SOT poster]) reported that offspring serum T4 reduction was most pronounced on PND 6 (Figure 2(B)). The DE-71 studies by de-Miranda et al. (2016) and Bowers et al. (2015) were excluded from this evaluation since the timepoints of T4 measurement (PND 23 and PND 21, 50, 100, respectively) did not allow determining if T4 reduction was most pronounced on PND 14.

Finally, in the PFHxS and triclosan studies by Gilbert et al. (2021), offspring serum T4 reductions generally attenuated between the first and last timepoints of measurement (GD 20/PND 0 *vs.* PND 14), but they were higher on PND 6 than on PND 2 (Figure 2(C)).

3.2.2. Serum levels of T3 in the dams and offspring

Key messages: Maternal or offspring serum T3 decrements were generally less pronounced than the corresponding serum T4 decrements. The ECETOC T4 TF empirically set a threshold of $\geq$20% *and* statistically significant offspring serum T3 reduction and evaluated whether T3 levels meeting this threshold are indicative of statistically significant neurodevelopmental findings. In the present database, there is some association between extent of offspring serum T3 reduction and the occurrence of statistically significant neurodevelopmental effects in rats.

Findings related to maternal and offspring serum T3 levels are summarised in *Table SI-1, Spreadsheet "Overview T4 T3."* In 19 of the 40 studies (see introduction to Section 3.2), serum T3 levels were measured in the dams, and in 10 of these 19 studies, maternal serum T3 levels were either not or only non-statistically significantly altered. If maternal serum T3 levels were altered, they were generally reduced. Increased maternal serum T3 levels were only recorded in the cyanamide study by Ramhøj et al. (2021), and this increase was

non-statistically significant. In the mercaptobenzimidazole study by Ramhøj et al. (2021), where maternal T4 levels were non-statistically significantly increased both on GD 15 and LD 22 (Section 3.2.1.1), maternal serum T3 was statistically significantly reduced by 31% on GD 15. Generally, maternal serum T3 alterations were not more pronounced than the corresponding maternal serum T4 reductions (*Column J in Spreadsheet "Overview T4 T3"*).

Offspring serum T3 levels were measured in 27 of the 40 studies, and in nine of these 27 studies, offspring serum T3 levels were either not or only non-statistically significantly altered (at any timepoint in case of multiple timepoints of measurement). Offspring serum T3 levels were also generally reduced if they were altered. Exceptions are a statistically significant increase by 77% in the EOGRTS-like MMI study by Fegert et al. (2012) measured on PND 90 (whereas alterations were non-significant on PND 4 and 21) and a statistically significant increase by 43% in the female adult offspring in TBBPA study by Lilienthal et al. (2008) and Van der Ven et al. (2008). The reliability of the data from the TBBPA study has been questioned (EFSA 2011b), see Section 4.2.3.1 for discussion. Increased offspring serum T3 was not observed in any other TBBPA study.

In four of the 27 studies, offspring serum T3 reduction was more pronounced than offspring serum T4 reduction at least at one timepoint of measurement (*Column O in Spreadsheet "Overview T4 T3"*), i.e. in two perchlorate studies (York et al. 2004, 2005a, 2005b) one triclosan study (Gilbert et al. 2021) and one OMC study (Ramhøj et al. 2020a [SOT poster]). In these studies, offspring serum T4 and T3 reductions were generally mild to moderate. An exception is the perchlorate study by York et al. (2004), in which offspring serum T3 was significantly reduced by 57% on PND 5 whereas offspring serum T4 was only reduced by 26%.

The ECETOC T4 TF further evaluated the present database to empirically set an offspring serum T3 reduction threshold that was indicative of statistically significant neurodevelopmental findings (following the same considerations as described in Section 3.2.1.3 for the setting of the offspring serum T4 reduction thresholds). In empirically trying out different offspring serum T3 thresholds (see Table 13 for decision matrix), "≥20% and statistically significant" *vs.* "≥18% and statistically significant" offspring serum T3 decrements yielded similar outcomes (Supplementary Information SI-4.2; *Table SI-1, Spreadsheet "Overview T4 T3," Column M, P, Q, R*). The ECETOC T4 TF selected the higher threshold, i.e. ≥20% and statistically significant offspring serum T3 threshold, because T3/T4 decreases generally need to attain ~25% as compared to the concurrent controls to be considered "reasonably detectable" in a common regulatory setting (Li et al. 2019; Marty et al. 2021; Section 4.4). Nonetheless, in some (investigational) studies included in this review, serum T3 reductions by 10–25% were also recorded as statistically significant.

An assessment whether the ≥20% and statistically significant offspring serum T3 reduction threshold is indicative of statistically significant neurodevelopmental effects yielded seven studies as true positive, five studies as false positive, eight studies as true negative, and four studies as false negative (Table 14 and *Table SI-1, Spreadsheet "Overview T4 T3," Column R*).

The false positive studies are the perchlorate study by York et al. (2004), the PFHxS and triclosan studies by Gilbert et al. (2021) and the DE-71 and OMC studies by Ramhøj et al. (2020a [SOT poster]). Of these, the perchlorate, triclosan and OMC studies were true negative following the offspring serum T4 reduction thresholds (Section 3.2.1.3). In the OMC study, offspring serum T3 was reduced by 20% and thus ranged at the borderline to true negative. The PFHxS and DE-71 studies were also false positive in the evaluation of the offspring serum T4 reduction thresholds.

In two of the four false negative studies, the only statistically significant neurodevelopmental findings were electrophysiological alterations (Gilbert and Sui 2008; Gilbert et al. 2013). The two other false negative studies were two Aroclor 1254 studies (Goldey and Crofton 1998; Crofton et al. 2000b), which yielded auditory deficits, and which were true positive following the offspring serum T4 reduction thresholds (Section 3.2.1.3). In the Aroclor 1254 study by Crofton et al. (2000b) offspring serum T3 was significantly reduced by 19% and thus ranged at the borderline to true positive.

3.2.3. Serum levels of fT4 and fT3 in the dams and offspring

Key messages: The database for maternal and offspring serum fT4 is incomplete, and only one study evaluated serum fT3 (and only in the offspring). In few Aroclor 1254 and triclosan studies, offspring serum fT4 reduction was (slightly) more pronounced than serum T4 reduction.

Findings related to maternal and offspring serum fT4/fT3 levels are summarised in *Table SI-1, Spreadsheet "Overview T4 fT4 T3 fT3."* Maternal serum fT4 levels *(Column H)* were measured in four studies, i.e. the dietary iodine deficiency study by Zhang et al. (2012), the PFHxS study by Gilbert et al. (2021), the Aroclor 1254 study by Morse et al. (1996) and the

Table 13. Decision-matrix applied for the further evaluation of offspring serum T3 reduction.

		Offspring serum T3 reduction ≥20% and statistically significant?	
		Yes	No
Statistically significant neurodevelopmental findings?	Yes	True positive	False negative
	No	False positive	True negative

The ECETOC T4 TF set the ≥20% and statistically significant offspring serum T3 reduction threshold empirically to yield the fewest numbers of false negative and false positive findings in the database collated for the present review. The assessment schemes were then used for an evidence-based and pragmatic analysis to further evaluate whether the mentioned thresholds are indicative of statistically significant neurodevelopmental findings in this database (Sections 3.2.2 and 4.4.3).

Table 14. Application of offspring serum T4 and T3 reductions thresholds to predict statistically significant neurodevelopmental findings.

	Substance	≥ 60% / ≥ 50% T4 ↓ in TDG / LDGs?	≥ 20% and statist. significant T3 ↓?	Statistically significant neurodevelopmental findings	References
Case Study 1	PTU	True positive	NA	Altered motor activity (PND 14, 23, 7 months); 7 months: hearing deficits and effects on cognitive function (some tests)	Axelstad et al. (2008)
		True positive	True positive	Heterotopia; altered cortex gene transcription	Ramhøj et al. (2020a) [a]
		True positive	True positive	Delayed eye opening, altered motor activity and acoustic startle response, hearing deficits	Goldey et al. (1995a)
		True positive	True positive	Heterotopia; altered learning and memory (PND 15)	Johnstone et al. (2013), Gilbert et al. (2014, 2016), O'Shaughnessy et al. (2018b)
	MMI	True negative	True negative	PND 21/22, PND 70: NE motor activity; PND 70: NE FOB, brain weight or histopathology	Fegert et al. (2012)
		True positive	NA	Impaired motor activity (PND 21, 40, 60) and cognitive function (PND 80)	Darbra et al. (2003)
	ETU	True negative	NA	PND 24: NE acoustic startle response. Adult: NE motor activity, FOB (grip strength) or detailed clinical observation	Marty et al. (2013) in European Commission (2017a)
	Mancozeb	True negative	NA	NE spatial learning (radial arm maze), motor activity (activity boxes) or acoustic startle response	Axelstad et al. (2011a)
	Amitrole	True positive	True positive	Heterotopia	Ramhøj et al. (2021)
	MBI	True negative	True negative	No heterotopia	Ramhøj et al. (2021)
	Cyanamide	True negative	True negative	No heterotopia	Ramhøj et al. (2021)
Case Study 2	Perchlorate	True negative	False positive	NE motor activity, acoustic startle response, cognitive function (passive avoidance, water maze); some changes in brain morphometry	York et al. (2004)
		True negative	True negative	NE motor activity; m: trend for changes in brain morphometry, f: trend reversed	York et al. (2005a, b)
		False negative [b]	False negative [b]	NE motor activity or cognitive function (Morris water maze, fear conditioning); electrophysiological findings	Gilbert and Sui (2008)
	Dietary iodine deficiency	False negative	NA	↓ integral optical density and altered gene expression in hippocampal CA1 region; effects on learning and memory	Zhang et al. (2012)
		False negative [b]	False negative [b]	NE acoustic startle response or cognitive function (fear conditioning); electrophysiological findings	Gilbert et al. (2013)
	Substance	**≥ 60% / ≥ 50% T4 ↓ in TDG / LDGs?**	**≥ 20% and statist. significant T3 ↓?**	**Statistically significant neurodevelopmental findings**	**References**
Case Study 3	TBBPA	True negative	True negative	NE motor activity, acoustic startle response, cognitive function (passive avoidance test, water M-maze performance), detailed clinical observation or brain histopathology	Cope et al. (2015)
		True negative	True negative	No morphological alterations in brain (as indicators of DNT)	Saegusa et al. (2009, 2012)
	PFHxS	True negative	True negative	NE motor activity or learning and memory (radial arm maze), no clear alteration of gene expression	Ramhøj et al. (2018, 2020b)
		False positive	False positive	NE acoustic startle response, trace fear conditioning or gene expression	Gilbert et al. (2021)
	Aroclor 1254	True positive	True positive	Permanent auditory deficits; altered motor activity on PND 15	Goldey et al. (1995b)
		True positive	False negative	PND 13, 15: transient reduction in motor activity; hearing deficits	Goldey and Crofton (1998)
		True positive	False negative [c1]	Hearing deficits in adults	Crofton et al. (2000b)
	DE-71	False positive	True negative	NE (FOB, motor activity; PND 24, 58-60, 273)	Kodavanti et al. (2010)
		False positive	False positive	No heterotopia or gene expression change	Ramhøj et al. (2020a) [a]
		True positive	True positive	PND 90: altered acoustic startle response; PND 100: reduced rearing as only motor activity-related finding at any timepoint	Bowers et al. (2015), Gill et al. (2016)
	Triclosan	True negative	False positive	NE acoustic startle response or trace fear conditioning, no heterotopia	Gilbert et al. (2021)
CS4	OMC	False negative	NA	f: reduced motor activity at 9 and 17 weeks	Axelstad et al. (2011b)
		True negative	False positive [c2]	No heterotopia, some gene expression changes	Ramhøj et al. (2020a) [a]

CS: case study; DE-71: a mix of polybrominated diphenyl ethers; DNT: developmental neurotoxicity; ETU: ethylene thiourea; f: female offspring FOB: functional observational batter; LDGs: lower-dose groups; m: male offspring; MBI: mercaptobenzimidazole; MMI: methimazole; NA: not available; NE: no effect; NS: not statistically significant; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PTU: propylthiouracil; SOT: Society of Toxicology; T4: thyroxine; T3: triiodothyronine; TBBPA: tetrabromobisphenol A; TDG: top-dose group.

Light grey/dark grey shading: True positive/true negative following the thresholds for ≥60%/≥50% offspring serum T4 reduction in the TDG/LDGs or ≥20% and statistically significant offspring serum T3 reduction. Note: Offspring serum TSH changes are not considered in this table since the ECETOC T4 TF does not propose a TSH threshold to predict the likelihood of neurodevelopmental findings.

[a]SOT poster, see bibliography for weblink.

[b]False negative on account of electrophysiological findings, but not on account of motor activity or cognitive function, which were not altered.

[c1]Borderline positive.

[c2]Borderline negative.

triclosan study by Gilbert et al. (2021). Offspring serum fT4 levels (*Column N*) were measured in these same studies and additionally in the Aroclor 1254 study by Goldey et al. (1995b). Goldey et al. (1995b) further measured serum fT3 levels in the offspring (*Column P*), a parameter that was not recorded in the dams in any of the 40 studies.

Serum fT4 levels in both the dams and offspring, as well as serum fT3 levels in the offspring, were always reduced (either statistically significantly or non-significantly). In the dams, serum fT4 reductions were generally not more pronounced than the corresponding serum T4 reductions. An exception is the Aroclor 1254 study by Morse et al. (1996), where maternal serum T4 and fT4 levels were significantly reduced by 50% and 58%, respectively, on GD 20, whereas they were non-significantly reduced on LD 21 (*Column K*).

In the offspring, serum fT4 reductions were (slightly) more pronounced than the corresponding T4 reductions in three studies (*Column R*):

- Aroclor 1254: Goldey et al. (1995b) at all timepoints; Morse et al. (1996) on GD 20 but not during lactation
- Triclosan: Gilbert et al. (2021) on PND 14 as only timepoint of serum fT4 measurement

3.2.4. Serum levels of TSH in the dams and offspring

Key messages: Maternal and offspring serum TSH levels were only dose-dependently and statistically significantly increased upon exposure to substances with a direct thyroid-related MoA (exception: one DE-71 study). Further, the dose-dependent and statistically significant offspring serum TSH increases were most pronounced for the TPO inhibitors PTU and amitrole.

Findings related to maternal and offspring serum TSH levels are summarised in *Table SI-1, Spreadsheet "Overview T4 T3 TSH."* In 20 of the 40 studies (see introduction to Section 3.2), serum TSH levels were measured in the dams (*Column I*). If maternal TSH was statistically significantly altered, it was increased; TSH reductions were always minor and always non-statistically significant (PFHxS: Ramhøj et al. 2018, 2020b; Gilbert et al. 2021; triclosan: Gilbert et al. 2021). Dose-dependent and statistically significant increases in maternal serum TSH levels (*Column K*) were only recorded upon exposure to substances with a direct thyroid-related MoA, i.e. the TPO inhibitors PTU (Schneider et al. 2009 [SI-7]; Johnstone et al. 2013), MMI (Fegert et al. 2012), mercaptobenzimidazole (Ramhøj et al. 2021), and the NIS inhibitor perchlorate (York et al. 2005a, 2005b; Gilbert and Sui 2008) as well as dietary iodine deficiency (Zhang et al. 2012; Gilbert et al. 2013).

In the PTU study by Johnstone et al. (2013), a dose-dependent and significant increase in maternal serum TSH was recorded at 1–3 ppm PTU on LD 21, i.e. right after the final exposure. In the dams exposed to 10 ppm PTU, serum TSH was non-significantly increased by 6%. This can be explained by the circumstance that, while exposure in these animals also ended on LD 21, they were only euthanised with concomitant blood sampling on LD 27/28 to allow for an extended lactation period at the very high dose level of 10 ppm PTU.

The dietary iodine deficiency study by Zhang et al. included the most pronounced serum TSH increase measured in either the dams or offspring, i.e. an increase by 1616% in the female rats around the time of mating after 12 weeks of feeding with iodine deficient diet. Hence, this datapoint was measured after a much longer exposure period than applied in most studies considered here. Between the two dietary iodine deficiency studies, maternal and offspring serum TSH increases were less pronounced in the study by Gilbert et al. (dams: 125%; offspring: 140%, both on PND 21) than in the study by Zhang et al. (dams: 1616% around the time of mating; offspring: 425% on PND 45) despite identical exposure periods (from 12 weeks premating until LD 21) and even though iodine deficiency was more pronounced in the Gilbert et al. study. Differences in the timepoints of blood sampling may account for these differences in serum TSH levels.

Further, in the DE-71 study by Kodavanti et al. (2010), maternal serum TSH was statistically significantly increased by 127% on LD 22 in the high-dose group. However, this increase was not dose-dependent since maternal serum TSH increase was more pronounced in the low-dose group (100%, non-significant) than in the mid-dose group (50%, non-significant). Dose-dependent but non-statistically significant increases in maternal TSH levels were recorded in the amitrole study by Ramhøj et al. (2021) and the PFHxS study by Gilbert et al. (2021). The ECETOC T4 TF suggests that the greater variability in TSH measurements may impact the ability to detect significant changes (discussed in Section 4.4).

Offspring serum TSH levels were measured in 26 of the 40 studies (*Column O in Spreadsheet "Overview T4 T3 TSH"*). Again, if TSH levels were statistically significantly altered, they were increased. Dose-dependent and statistically significant increases in offspring serum TSH levels were generally only measured in studies evaluating substances with direct thyroid-related MoA (*Column Q*), i.e. the TPO inhibitors PTU (Schneider et al. 2009 [SOT poster; SI-7]; Ramhøj et al. 2020a [SOT poster]), MMI (Fegert et al. 2012), ETU (Marty et al. 2013, as cited in European Commission 2017a), and amitrole (Ramhøj et al. 2021), the NIS inhibitor perchlorate (York et al. (2004, 2005a, 2005b) as well as dietary iodine deficiency (Zhang et al. 2012; Gilbert et al. 2013). In the PTU study by Johnstone et al. (2013), offspring serum TSH was dose-dependently and statistically significantly increased on PND 14 in 1–3 ppm dose groups, but not at 10 ppm.

The dose-dependent and statistically significant increase in offspring serum TSH was most pronounced for those TPO inhibitors, which also elicited neurodevelopmental effects, i.e. PTU and amitrole. In the PTU and amitrole studies, offspring serum TSH increases generally ranged above 400% in the high-dose groups and even attained more than 1000% in PND 21 males following PTU exposure from GD 6 to PND 10 (Schneider et al. 2009 [SI-7]). In the dietary iodine deficiency study by Zhang et al. (2012), offspring serum TSH was increased by 425% on PND 45 (but by 99% on PND 7). By comparison, in the dietary iodine deficiency study by Gilbert et al. (2013) offspring serum TSH was increased by 140% on PND 21. For those TPO inhibitors, which did not elicit neurodevelopmental effects in the respective studies, the dose-dependent and statistically significant offspring serum TSH increase did not attain 400% (MMI in the EOGRTS-like study: 314%; ETU: 52%; mercaptobenzimidazole: 88% non-significant; cyanamide 42% non-significant).

Only one study investigating a substance with an indirect thyroid-related MoA also showed dose-dependently and

statistically significantly increased offspring serum TSH levels. Bowers et al. (2015) applied 0.3, 3, and 30 mkd DE-71 and recorded non-significantly increased offspring serum TSH in the two lower-dose groups (each: 21%) and a statistically significant 42% increase in the top dose (on PND 21). Hence, this dose-dependent and statistically significant offspring serum TSH increase was much less pronounced than observed for PTU or amitrole. In the other studies investigating substances with indirect thyroid-related MoAs, offspring serum TSH increases were always not statistically significant.

3.2.5. Brain levels of T4 and T3 in the offspring

Key messages: Brain T4/T3 data are only available for five of the 14 case study substances and for dietary iodine deficiency. Brain T4 reductions were much more pronounced in the PTU studies, and in a recent perchlorate study, than in the dietary iodine deficiency or Aroclor 1254 studies. In the PFHxS and triclosan studies, brain T4 reductions were transient (statistically significant only on PND 0 and GD 20–PND 2, respectively). Brain T3 levels were only significantly altered in the PTU studies and in the recent perchlorate study.

Findings related to thyroid hormone levels in the offspring brain are summarised in *Table SI-1, Spreadsheet "Overview Serum + Brain T4 DNT."* Brain T4 levels (*Column N*) were measured in 8 studies, i.e. the PTU studies by Sharlin et al. (2010); Bastian et al. (2014)/Spring et al. (2016); Johnstone et al. (2013)/Gilbert et al. (2014, 2016)/O'Shaughnessy et al. (2019), the dietary iodine deficiency study by Gilbert et al. (2013), the PFHxS study by Gilbert et al. (2021), the Aroclor 1254 study by Morse et al. (1996) and the triclosan study by Gilbert et al. (2021). Apart from the dietary iodine deficiency study by Gilbert et al. (2013), these same studies also included measurement of brain T3 levels (*Column O*). Further, the perchlorate study by Gilbert et al. (2022) that was published after data evaluation for the present review was completed includes brain T4 and T3 levels (*Table SI-1, Spreadsheet "CS2 Perchlorate"*). At 44 and 140 mkd perchlorate, brain T4 levels were reduced by 73% and 91%, respectively, and brain T3 levels by 64% and 85%, respectively, in the GD 20 foetus. These decrements were associated with offspring serum T4 decrements by 80% and 96%, respectively, and the differential expression of selected genes in the thyroid gland and brain (no data for offspring serum T3 or any other brain-related parameters).

Brain T3 levels were only significantly altered in the PTU studies and in the recent perchlorate study but not in the PFHxS, Aroclor 1254 or triclosan studies.

As regards brain T4 levels, between-study comparability is not only impaired by differences in timepoints of measurement, but also by differences in the part of the brain that was submitted to the hormone analysis, and potentially also by differences in analytical methods.

Generally, offspring from dams exposed to PTU exhibited pronounced T4 reductions in the cortex (up to 59% on PND 14/15; Sharlin et al. 2010), the entire brain (to below the lower limit of quantification on PND 14; Gilbert et al. 2016) and forebrain (up to 92% on PND 2; O'Shaughnessy et al. 2019). Consistently, these pronounced brain T4 decrements coincided with pronounced offspring serum T4 decrements by $\geq$60% (*Column K in Spreadsheet "Overview serum + brain T4 DNT"*). The recent perchlorate study by Gilbert et al. (2022) yielded a similar pattern of findings.

Offspring from dams fed with iodine deficient diets beginning premating and lasting all through gestation and lactation exhibited moderately reduced brain T4 on GD 20 (30%) that coincided with moderately reduced offspring serum T4 (43%). On PND 21, brain T4 levels were not significantly altered, whereas the serum T4 reduction had aggravated (62%). For these offspring, reduced synaptic transmission in the dentate gyrus (lower amplitudes and reduced excitatory post-synaptic potential slope) were recorded on PND 60–180 but no effects in the acoustic startle response, water maze or fear conditioning tests (Gilbert et al. 2013).

In the PFHxS study (Gilbert et al. 2021), which yielded no periventricular heterotopia or neurobehavioural findings, brain T4 levels were statistically significantly reduced by 42% and 18% on PND 0 and 14, and non-significantly reduced on PND 2 and 6 (analysed by liquid chromatography/mass spectrometry/mass spectrometry; lower limit of quantification for T4 and T3: 100 pg/g tissue). Changes in brain T3 levels were not statistically significant at these timepoints. Offspring serum T4 decrements generally exceeded the $\geq$60% offspring serum T4 reduction threshold, just as the offspring serum T3 decrements exceeded the $\geq$20% and statistically significant offspring serum T3 reduction threshold (therefore, this PFHxS study was assessed as "false positive"). Hence, from amongst all thyroid-related parameters included in this PFHxS study, only brain T3 correctly predicted the absence of brain-related effects.

In the Aroclor 1254 study (Morse et al. 1996), T4 levels in the forebrain were reduced by up to 87% on GD 20 (analysed by radioimmunoassay). While this value was marked as being non-significant, the ECETOC T4 TF assumes that low sample sizes of 2–3 animals/group in addition to high data variability accounts for this estimation. In the cerebellum, T4 levels were reduced to below the lower limit of quantification (10 and 6 pg/g tissue for T4 and T3, respectively) on GD 20. On PND 21, T4 levels in the forebrain were reduced by up to 44% in the female offspring, whereas they were not significantly altered in the males. These alterations of forebrain T4 levels coincided with moderate serum T4 decrements (36% and 32% in the male and female offspring, respectively). The only non-thyroid brain-related parameter included in this study was forebrain DIO 2 activity, and it was increased on GD 20, reduced on PND 4 and not altered on PND 90–196.

In the triclosan study (Gilbert et al. 2021), which yielded no periventricular heterotopia or neurobehavioural effects, brain T4 levels were statistically significantly reduced by 26%, 46%, and 43% on GD 20, PND 0 and PND 2, respectively. These findings coincided with non-significant offspring serum T4 decrements on GD 20 and PND 0, and a moderate but statistically significant offspring serum T4 decrement on PND 2 (28%). On PND 6 and 14, brain T4 levels were not significantly altered, whereas offspring serum T4 decrements were again moderate and statistically significant (44% and 20%, respectively). Brain T3 decrements were always non-significant.

3.2.6. Weight and histopathology of the thyroid gland of the dams and offspring

Key messages: Findings related to maternal and offspring thyroid weight/histopathology always coincided with increased serum TSH levels.

Findings related to maternal and offspring thyroid weight and histopathology are summarised in *Table SI-1, Spreadsheet "Overview T4 Thyroid WT HP."*

Ten of the 40 studies (see introduction to Section 3.2) included both maternal thyroid weight and maternal thyroid histopathology, six further studies maternal thyroid weight alone and two further studies maternal thyroid histopathology alone (*Columns I-M*). From amongst these 18 studies, it was only in the EOGRTS-like MMI study by Fegert et al. (2012) that thyroid weight and histopathology in the dams were more sensitive (in terms of dose level at which effects were observed) than maternal serum T4 measurements (*Column M*, further considering *Column F*). In the other studies, maternal thyroid gland weight and/or histopathology were either as sensitive as, or less sensitive than, maternal serum T4 reduction, or even not affected (Cope et al. 2015).

In those studies that included both maternal thyroid weight/histopathology and maternal serum TSH levels, increased thyroid weight/histopathological findings always coincided with increased serum TSH levels (but without clear association between statistically significant *vs.* non-significant TSH increases). By comparison, absence of altered maternal thyroid weight/histopathology always coincided with non-significantly *decreased* maternal serum TSH.

Ten of the 40 studies included both offspring thyroid weight and histopathology and 11 further studies offspring thyroid weight alone (*Columns Q–U*). From amongst these 21 studies, histopathology of the offspring thyroid gland was more sensitive than offspring serum T4 in the ETU study by Marty et al. (2013) as cited in European Commission (2017a) and in the perchlorate study by York et al. (2004). By comparison, thyroid histopathology was not more sensitive in the perchlorate study by York et al. (2005a, 2005b) that included higher doses of perchlorate and a longer exposure period than York et al. (2004), i.e. dosing up until LD 30 instead of up until LD 10. In the other studies, histopathology of the offspring thyroid gland was either as sensitive as, or less sensitive than, serum T4 in the offspring (*Column N*), provided that effects were observed. Offspring thyroid weight was never more sensitive than offspring serum T4 levels.

In those studies that included both offspring thyroid weight/histopathology and offspring serum TSH levels, statistically significant offspring thyroid weight increases/histopathological findings always coincided with statistically significantly increased offspring serum TSH levels (when considering comparable timepoints of measurement).

3.3. Body weight of the dams and offspring

Key messages: Maternal body weight reductions were generally $<$10% showing the maximum tolerated dose was generally not exceeded. Significant and/or $\geq$10% offspring body weight reductions were recorded across all case studies. Offspring body weight was only unaffected in studies yielding no neurodevelopmental findings (apart from electrophysiological alterations). If offspring body weight was reduced, this did not necessarily coincide with neurodevelopmental findings.

In the preceding sections, the information on body weight reduction in the dams has been used on a case-by-case basis as one parameter to determine if the maximum tolerated dose had likely been reached, especially in those studies which did not yield neurodevelopmental findings. In this section, the findings from the 40 studies (see introduction to Section 3.2) related to body weight reduction in the dams or offspring are jointly presented.

The extent of maternal and offspring body weight reduction (and/or information on signs of systemic toxicity, which are further discussed in Section 4.5) are summarised in *Columns F and I of Table SI-1, Spreadsheet "Overview DNT details."* Information on maternal body weight or body weight gain or signs of systemic toxicity in the high-dose dams was provided in 32 of the 40 studies. If maternal body weight was altered, it was generally reduced, and such body weight reductions in the dams were generally below 10%. Exceptions are the mancozeb dose range finding study by Beck (2008a) as reported in European Commission (2017a) who recorded a 10.5% body weight reduction in the high-dose (60 mkd) dams during gestation, which resulted in a lowering of the top dose to 30 mkd in the subsequent DNT study (Beck 2008b) as reported in European Commission (2017a). Further, in the TBBPA study by Lilienthal et al. (2008) and Van der Ven et al. (2008), 10% and 15.2% body weight reductions were recorded in the high-dose dams (3000 mkd) on GD 7 and GD 21, respectively.

Information on offspring body weight or body weight gain was provided in 36 of the 40 studies. Two studies neither provided information on the body weight of the dams nor on that of the offspring. These are the dietary iodine deficiency study by Zhang et al. (2012) and the Aroclor 1254 study by Morse et al. (1996). The Supplementary Information SI-5.1 shows for which substances significant and/or $\geq$10% body weight reduction in the offspring was reported upon *in utero*/lactational exposure, and the Supplementary Information SI-5.2 shows for which substances no such effects were observed. Interestingly, absence of offspring body weight reduction coincided with absence of neurodevelopmental findings.

3.4. Liver enzyme activities

Key messages: The Case Study 3 substances generally induced UGT, ethoxy resorufin-O-deethylase (EROD), pentoxy resorufin-O-dealkylase (PROD) and/or benzyloxy resorufin-O-dealkylase (BROD) or the expression of gene transcripts mediating liver enzyme induction.

Findings related to substance-mediated liver enzyme induction, or the increased expression of gene transcripts mediating liver enzyme induction, are summarised in *Table SI-1, Spreadsheet "Overview liver enzymes."* The case study-specific Table SI-1 spreadsheets also include information on

liver weight, as available. However, the quantitative correlation between liver weight and thyroid weight was not further evaluated in the present review, which focuses on thyroid- and brain-related parameters.

Data on liver enzyme activities are available for Aroclor 1254 (Morse et al. 1996), DE-71 (Zhou et al. 2002; Bowers et al. 2015; de-Miranda et al. 2016), and triclosan (Paul et al. 2012). Enzymes considered were UGT, EROD, BROD, and PROD (see Appendix I, Section App-I.3.1.2). Further, for PFHxS and triclosan, data on liver enzyme-related gene transcripts are available (Gilbert et al. 2021).

In the dams, the measured liver enzymes were generally statistically significantly increased both at the end of gestation (GD 20) and at the end of lactation (LD 21–22). As an exception, maternal PROD and T4-UGT were only statistically significantly altered on LD 22 but not on GD 20 in the triclosan study by Paul et al. (2012).

In the offspring from dams exposed to Aroclor 1254 and DE-71, all measured liver enzymes were generally statistically significantly increased during lactation (Morse et al. 1996; Zhou et al. 2002; Bowers et al. 2015; de-Miranda et al. 2016). These studies also showed pronounced offspring serum T4 decrements $\geq$60% and neurodevelopmental findings (see Section 3.2.1.2 for T4 data recorded by de-Miranda et al.). By comparison, in the offspring from dams exposed to triclosan, only PROD was increased on PND 4, but not T4-UGT, and neither PROD nor T4-UGT were increased on PND 14 or 21 (Paul et al. 2012). In this triclosan study, offspring serum T4 was only reduced on GD 20 and PND 4 (by up to 28% and 26%, respectively) but not on PND 14 or 21. Hence, on PND 4, the moderate but statistically significant T4 reduction did not coincide with increased T4-UGT (Paul et al. 2012). While brain-related parameters were not included in the triclosan study by Paul et al. (2012), the triclosan study by Gilbert et al. (2021) showed no effects on neurobehaviour, heterotopia or brain gene expression. In this study, exposure to 300 mkd triclosan did not alter the expression of T3 responsive genes in the liver, but only liver transcripts related to phase I and II metabolism (Gilbert et al. 2021). Gilbert et al. (2021) reported similar findings for PFHxS.

3.5. Gene expression and/or transporter levels in the offspring brain

Key messages: In studies evaluating PTU or dietary iodine deficiency, the expression of a subset of investigated genes was altered, and this generally coincided with brain-related effects. No (or only one or two) gene transcripts were altered in the PFHxS and triclosan studies, and these studies also yielded no neurodevelopmental findings. Different marker genes were used between studies.

Findings related to substance-mediated alterations in the expression of genes in brain tissues or in the levels of brain-specific proteins are summarised in *Table SI-1, Spreadsheet "Overview gene expression."* Gene expression in brain tissues and/or the levels of brain-specific proteins were measured in 18 studies. This includes two PTU studies and an Aroclor 1254 study that were not considered previously in this review since their focus was on gene expression (PTU: Royland et al. 2008; Kobayashi et al. 2009; Aroclor 1254: Royland and Kodavanti 2008). Generally, the selection of genes and proteins assessed in the different studies differed considerably between research teams as well as between different publications from the same research teams. The ECETOC T4 TF only evaluated if genes were differentially expressed upon substance exposure (as well as the extent of change, if relevant) but did not evaluate how specific genes were differentially expressed upon substance exposure or how the selection of genes compared between studies. The ECETOC T4 TF also did not evaluate the relevance of sets of genes selected by the respective authors. Such evaluations would have exceeded the scope of this review.

Eight PTU studies included measurements of brain gene expression (Kobayashi et al. 2005, 2009; Royland et al. 2008; Bastian et al. 2014; Gilbert et al. 2017; O'Shaughnessy et al. 2018b, 2019; Ramhøj et al. 2020a [SOT poster]). Generally, at least some genes were differentially expressed already at the respective lowest dose tested. Genes associated with, e.g. cell differentiation, myelin development, synaptic function and neuronal signalling and migration were frequently affected and downregulated. Myelin defects are a known effect of thyroid perturbations (Rodriguez-Pena 1999; Lucia et al. 2018; Pagnin et al. 2021). Concordantly, those studies that included neurodevelopmental parameters (generally: periventricular heterotopia) also showed neurodevelopmental effects.

One dietary iodine deficiency study (Zhang et al. 2012) addressed the levels of two brain-specific proteins [i.e. brain-derived neurotrophic factor (BDNF) and neuroendocrine-specific protein A (NSP-A)] and the expression of *c-Fos* and *c-Jun*. BDNF and NSP-A are important mediators of thyroid hormones and have essential roles in brain development, and *c-Fos* and *c-Jun* are essential for spatial learning and memory consolidation in rats (further discussed by Zhang et al. 2012). As compared to the offspring from dams fed with normal iodine diet, the offspring from the dams fed with iodine-deficient diet showed reduced BDNF levels but increased NSP-A levels both on PND 7 and PND 45. Also, the expression of *c-Fos* and *c-Jun* was significantly reduced at both time points. As regards neurobehavioural effects, Zhang et al. (2012) recorded longer escape latency in the Morris water maze in the offspring of the iodine deficient group.

Two PFHxS studies (Ramhøj et al. 2018, 2020b; Gilbert et al. 2021) included the measurement of brain gene transcripts that had previously been associated with periventricular heterotopia upon exposure to PTU and on gene markers of disrupted thyroid hormone action. Generally, the selected gene transcripts and gene markers were not altered, except that *Itih3* was increased and *Klf9* was slightly decreased (Ramhøj et al. 2018, 2020b). Ramhøj et al. noted that the increased expression of *Itih3* was contrary to the reduction caused by PTU. Neither PFHxS study showed neurobehavioural effects or periventricular heterotopia.

One Aroclor 1254 study included measurements of gene expression in the offspring cerebellum and hippocampus (Royland and Kodavanti 2008). Generally, a broad number of gene transcripts was altered, and more so in the hippocampus than in the cerebellum, as well as more so on PND 7

than on PND 14 (exposure period from GD 6 to LD 21). This study did not include neurodevelopmental parameters, but other Aroclor 1254 studies showed effects on motor activity, acoustic startle response and hearing function.

Two DE-71 studies included measurements of gene expression in offspring brain tissues (Gill et al. 2016; Ramhøj et al. 2020a [SOT poster]). In the study by Gill et al. (2016), the PND 21 high-dose group female offspring (30 mkd; exposure period GD 1–PND 21) showed upregulation in markers related to most pathways considered (cytokines, tyrosine kinase C, calmodulin-dependent protein kinase II, muscarinic and nicotinic receptors) whereas few markers were downregulated. In the PND 21 male offspring, some markers were upregulated in the low-dose group (0.3 mkd), and some were downregulated in the mid- and high-dose groups (3, 30 mkd). On PND 250, some markers were downregulated in the males and females from the mid- and high-dose groups (Gill et al. 2016). In this study, neurodevelopmental assessments showed no effects on motor activity apart from reduced rearing on PND 110 in the mid- and high-dose groups, no effect on emergence latency or prepulse inhibition but altered acoustic startle response in the high-dose group on PND 90 (Bowers et al. 2015). By comparison, in the DE-71 study by Ramhøj et al. (2020a [SOT poster]) different gene markers reported to reflect disrupted thyroid hormone action were not significantly changed on PND 16 (40 mkd DE-71 applied from GD 7 to PND 22), and brain histopathology showed no signs of periventricular heterotopia.

One triclosan study included measurement of nine gene transcripts that were previously associated with periventricular heterotopia upon exposure to PTU and on eleven gene markers of disrupted thyroid hormone action (Gilbert et al. 2021). None of the transcripts or markers were significantly changed, just as there were no neurobehavioural alterations or periventricular heterotopia.

One OMC study included measurement of six gene markers of disrupted thyroid hormone action (Ramhøj et al. 2020a [SOT poster]). At 375 mkd, only *Pvalb* was significantly reduced, and at 500 mkd, *Pvalb*, *Coll11a2*, *Gjb6*, *Hopx* and *Klf9* were significantly reduced. Brain histopathology showed no signs of periventricular heterotopia.

Taken together, in different PTU studies (Kobayashi et al. 2005, 2009; Royland et al. 2008; Bastian et al. 2014; Gilbert et al. 2017; O'Shaughnessy et al. 2018b, 2019; Ramhøj et al. 2020a [SOT poster]), the dietary iodine deficiency study by Zhang et al. (2012) and the DE-71 study by Bowers et al. (2015) and Gill et al. (2016), the differential expression of larger numbers of gene markers and gene transcripts that were selected as being associated with altered neurodevelopment and/or periventricular heterotopia by the respective authors generally coincided with the corresponding brain-related effects. By comparison, no, or only one or two, gene transcripts were altered in the PFHxS studies (Ramhøj et al. 2018, 2020b; Gilbert et al. 2021), in the DE-71 study by Ramhøj et al. (2020a [SOT poster]) and in the triclosan study by Gilbert et al. (2021), and these studies also yielded no neurodevelopmental findings. An exception is the OMC study by Ramhøj et al. (2020a [SOT poster]), in which four genes were altered in the high-dose group on PND 16, but brain histopathology showed no periventricular heterotopia (*Columns H and J in Spreadsheet "Overview Gene Expression"*).

4. Discussion

4.1. Definition of case studies, selection of substances, and selection of studies

Key messages: The four case studies cover the most important thyroid-related AOPs/MoAs with neurodevelopmental outcomes in mammals. While the substances were assigned to the case studies based upon their presumed primary MIE(s), some substances may additionally trigger further MIEs (e.g. possibly, the TPO inhibitor PTU also inhibits DIOs). The selection of the case study substances was driven by the availability of rat studies that included *in utero*/lactational substance exposure and the measurement of thyroid- and brain-related parameters.

In the present review, data from rat studies have been evaluated to determine which extent of maternal and/or offspring thyroid hormone imbalance caused by *in utero*/lactational exposure to thyroid-active substances leads to neurodevelopmental or other brain-related effects in the rat offspring. Four case studies have been presented that cover the MIEs for the major thyroid-related AOPs/MoAs with neurodevelopmental outcomes in mammals as identified by Marty et al. (2021), i.e. Case Study 1: TPO inhibition; Case Study 2: NIS inhibition; and Case Study 3: liver enzyme induction and/or displacement of T4 from serum binding proteins. Case Study 4 includes OMC, which has been observed to inhibit DIO1 activity in the liver (further discussed below).

The assignment of fourteen substances to the four case studies was based upon the available evidence that indicated that they trigger the respective MIEs. However, the evidence that a specific substance triggers a particular MIE, and/or has a specific MoA, is sometimes not conclusive (see Appendix I, Sections App-I.1.2, App-I.2.2, App-I.3.2, App-I.4.2 for Case Study 1, 2, 3, 4, respectively). Often, the evidence has been derived from *in vitro* assays (e.g. enzyme inhibition, displacement of serum binding proteins), and it is not yet fully understood how the *in vitro* events correlate to *in vivo* effects. Also, it is often not known whether a substance might trigger further MIEs beyond the one identified in the given investigation.

For the two MIEs that may lead to enhanced thyroid hormone clearance, i.e. displacement of thyroid hormone from serum binding proteins and induction of liver enzymes that metabolise thyroid hormone, the available evidence indicates that all five selected substances (TBBPA, PFHxS, Aroclor 1254, DE-71, triclosan) may trigger both MIEs. For this reason, Case Study 3 covers both of these MIEs. Notably, searches in the database ToxCast (https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data [accessed 2022 August]) indicated that the Case Study 3 substance TBBPA may also inhibit DIOs or TPO *in vitro* and that the Case Study 3 substance PFHxS (CAS no. 3871-99-6) may also inhibit NIS *in vitro* (data from ToxCast evaluation not shown). However, a very high proportion of substances test positive in these

in vitro assays (see e.g. Paul Friedman et al. 2016), and there is little contextual information that the effects observed *in vitro* are biologically relevant (for definition, see EFSA 2011a, 2017). Based upon these considerations, the ECETOC T4 TF assessed the evidence that TBBPA and PFHxS enhance thyroid hormone clearance as overriding and assigned these substances to Case Study 3.

Despite such uncertainties, the pattern of thyroid-related effects observed between case studies reveals specificities for some substances that can be related to their presumed MoA as well as differences between substances with direct *vs.* indirect thyroid-related MoAs (e.g. related to offspring serum TSH; Section 3.2.4). Thereby, the findings from this review generally confirm the assignment of the substances to the four MoA-related case studies. By comparison, it was not possible to establish patterns of brain-related effects—either for specific substances, or for specific MoAs or across case studies. However, even if such patterns of effects were present (e.g. common patterns of changes in motor activity, auditory startle response and/or cognitive function for specific substances), it would most likely not have been possible to (fully) identify them due to the heterogenicity of the study designs, especially with respect to types and timepoints of neurodevelopmental assessments.

The selection of the case study substances was driven by the availability of rat studies that included *in utero*/lactational substance exposure and the measurement of both thyroid-related and brain-related parameters. Some substances with relevant MoAs were not considered due to the inability to find suitable studies (i.e. carbamazepine, PFOS, iopanoic acid; see *Table SI-1, Spreadsheet "Examples Excluded Studies Substances"*).

The UGT-inducer phenobarbital was not selected for Case Study 3 since its primary MoA in rats leads to thyroid carcinogenesis as an adverse outcome (see e.g. Plummer et al. 2021). Furthermore, phenobarbital may have other effects on neurodevelopment that are mediated by non-thyroidal MoAs, e.g. gamma aminobutyric acid-A effects (Kaindl et al. 2008) or hippocampal protein kinase C effects (Wu and Wang 2002). Also, the ECETOC T4 TF is unaware of rat studies addressing the impact of *in utero*/lactational exposure to phenobarbital on neurodevelopment.

The selection of a suitable substance was especially difficult for Case Study 4. There are hardly any DIO inhibitors that do not also exhibit other effects on thyroid function, e.g. PTU inhibits both TPO and DIO1 (Oppenheimer et al. 1972). While the oral cholecystographic agent iopanoic acid is known to inhibit DIO1 and DIO2 (Braga and Cooper 2001), the ECETOC T4 TF was unable to find a study investigating iopanoic acid that met all study inclusion criteria. For example, Tuca et al. (1994) administered iopanoic acid to pregnant dams, but this study did not include serum thyroid hormone levels in either the dams or offspring, and it also did not include neurobehavioural parameters. As regards further potential substances for Case Study 4, Shimizu et al. (2013) reported that structure-activity relationships indicated that Rose Bengal, erythrosine B, phloxine B, benzbromarone and different biphenyl ethers were potent DIO inhibitors. The ECETOC T4 TF was unable to find studies addressing the effects of *in utero*/lactational exposure to any of these substances on maternal or offspring thyroid function and neurodevelopment. Eventually, the UV filter OMC was selected as Case Study 4 substance because it has been observed to reduce DIO1 activity in the liver and to reduce T4 levels in the blood (Schmutzler et al. 2004; Klammer et al. 2007) and since two rat studies fulfilling the study inclusion criteria were available for OMC (Axelstad et al. 2011b; Ramhøj et al. 2020a [SOT poster]). However, since DIO1 activity in the liver is dependent on serum thyroid hormone levels, the observed serum T4 decrement cannot necessarily be attributed to a DIO1 inhibiting MoA. Further, it is being discussed whether OMC may affect the hormone system *via* MoAs unrelated to the hypothalamic–pituitary–thyroid (HPT) axis (Axelstad et al. 2011b). For these reasons, Case Study 4 was named "other MoAs including DIO inhibition."

While OMC inhibits DIO1 activity in the liver, it is DIO2, which is active in the brain (Marty et al. 2021). The Aroclor 1254 study by Morse et al. (1996) included measurement of brain DIO2 activity, but it did not include any other brain-related parameter. The evaluation of how substance mediated DIO2 inhibition may affect neurodevelopment would have been highly relevant for the present review. It is regrettable that this evaluation was impeded by the inability to identify a substance that (only or predominantly) triggers DIO2 inhibition and for which a study fulfilling the study inclusion criteria is available. Research work is recommended to establish under which conditions the inhibition of DIOs in either the liver or brain leads to a reduction or an increase in serum and/or brain T4/T3 levels.

4.2. Overarching evaluation of the database considered in this review

Key messages: The present review is founded on a large database. Despite large variability in the design of the studies, it provides important insight on how thyroid-related parameters in the dams and offspring and brain-related parameters in the offspring may be altered upon gestational/lactational exposure to substances mediating thyroid hormone imbalance by different MoAs.

The four case studies include 14 thyroid-active substances (and dietary iodine deficiency) that trigger the most important thyroid-related MIEs for AOPs with neurodevelopmental outcomes in mammals as identified by Marty et al. (2021). Both standard rat toxicity studies and non-standard investigational studies using rats were considered. The case studies cover 51 studies including 13 investigational studies from research work conducted by the US EPA on the effects of PTU. These 51 studies were published in altogether more than 60 publications, and all thyroid- and brain-related data presented in these publications have been considered. Thereby, the present review is founded on a large database showing how thyroid-related parameters in pregnant/lactating rats and their offspring (Tables 9 and 14) as well as brain-related parameters in the offspring (Tables 7 and 8) may be altered upon exposure to substances that may cause thyroid hormone imbalance by different MoAs.

In most studies, the exposure period covered the major parts of both gestation and lactation, generally beginning at GD 6/7 when implantation is complete in the rat to prevent substance interference with implantation. Hence, in most studies, the exposure period covered the critical periods of thyroid hormone imbalance that have been suggested for the development of periventricular heterotopia, cochlear damage, altered motor activity or impaired cognitive function (Table 5). Nonetheless, there is generally large variability in the design of the studies, and this impairs between-study comparability. (1) Most studies did not include all potentially relevant thyroid-related parameters. (2) The timepoints of thyroid hormone measurements, especially in the offspring serum, differed considerably between studies; also, timepoints did not always correspond to critical periods of neurodevelopment for the brain-related endpoints assessed. (3) The spectrum of brain-related assessments, as well as the timepoints at which these tests were performed, differed considerably between studies. (4) In a few studies, there was no report of either the presence or absence of systemic toxicity (e.g. effects on body weight or body weight gain) so that it was not possible to determine whether the observed effects were a primary response or secondary to systemic toxicity in the dams. Further, prevailing knowledge gaps, not only with respect to the substances' MoA(s) but also to critical periods of development (which may further be protracted if the substance causes developmental delays; Section 4.5) impair data evaluation.

Therefore, the findings from this review can by no means be considered to provide a final answer. Nonetheless, they do provide insight on how *in utero*/lactational exposure to substances causing thyroid hormone imbalance in rats may affect the maternal and offspring thyroid hormone systems and offspring neurodevelopment.

In Sections 4.2.1–4.2.4 below, important findings from the four case studies are further discussed. It is considered whether the findings are coherent (1) between studies for the same substance and (2) between the substances assigned to a given case study. Possible reasons for inconsistencies are discussed. Considering all findings together, it is sought to establish patterns of effects for specific case study substances or, if possible, patterns of thyroid-related effects that are related to specific case studies (and hence MoAs). It is discussed whether the findings likely indicate that the offspring thyroid hormone system showed adaptive reactions, so that neurodevelopment was not impaired, or that such adaptive mechanisms were overwhelmed, so that hypothyroid-mediated effects developed in the offspring brain. Focus is on thyroid-related findings observed in the offspring, and not the dams, since it is the extent of perturbation of the offspring thyroid system that appears imminently relevant for hypothyroid-mediated neurodevelopmental impairment. Indeed, the maternal thyroid-related data support this hypothesis, as it is also briefly addressed in the below discussions.

The discussion of the findings from the four case studies is followed by an overarching discussion of the scientific relevance of the brain-related (Section 4.3) and thyroid-related parameters (Section 4.4). Finally, Section 4.5 briefly discusses other manifestations of developmental delays, though such effects are not the primary focus of this review.

It is important to reiterate that this review does not evaluate the human relevance of findings from the rat studies. This topic will be addressed in the planned fourth paper of the ECETOC T4 TF review series. Also, the findings presented here exclusively relate to rat studies covering *in utero*/lactational substance exposure. It is not necessarily possible to extrapolate these findings to thyroid-related effects that might be caused by subchronic or chronic exposure to the case study substances (or other substances).

Finally, this review did not consider whether the case study substances might also elicit direct effects on neurodevelopment by MoAs that are unrelated to thyroid function. However, such considerations should form part of comprehensive evaluations which should include all observed effects (e.g. also in other target organs).

4.2.1. Case Study 1: Impairment of thyroid hormone synthesis via TPO inhibition

4.2.1.1. Two patterns of thyroid- and brain-related effects for TPO inhibitors.

Key messages: For PTU, amitrole and MMI at very high doses, neurodevelopmental effects coincided with pronounced alterations in all thyroid-related parameters in the offspring serum and brain, likely indicating that the adaptive mechanisms of the offspring thyroid system were overwhelmed. For MMI at the maximum tolerated dose in an EOGRTS-like study, ETU, mancozeb, mercaptobenzimidazole and cyanamide, the much less pronounced thyroid effects in the offspring together with the absence of neurodevelopmental findings likely indicate that the adaptive mechanisms of the offspring thyroid system were effective, or the tests were not sufficiently sensitive to detect an effect.

The seven TPO inhibitors considered in Case Study 1 generally showed two different patterns of brain- and thyroid-related effects. Those TPO inhibitors that elicited neurobehavioural effects or periventricular heterotopia caused a $\geq$60% offspring serum T4 decrement (i.e. PTU, amitrole and MMI at $\sim$36 mkd; but not ETU, mancozeb, mercaptobenzimidazole, cyanamide or MMI at the maximum tolerated dose in an EOGRTS-like study, i.e. 3 mkd). A similar distinction was not possible based upon maternal serum T4 decrements since these were generally low to moderate ($<$60%) for all TPO inhibitors (and coincided with unchanged maternal serum T3 and statistically significantly increased maternal serum TSH).

In the offspring, pronounced serum T4 decrements coincided with statistically significant serum T3 decrements in those studies which yielded neurodevelopmental effects (the PTU studies, the amitrole study and the recent MMI study by Ramhøj et al. (2022a), in which MMI was dosed up to 16 mkd from GD 7 to LD 22). The combination of pronounced offspring serum T4 decrement and statistically significant offspring serum T3 decrement is likely indicative of severe thyroid hormone perturbation. By comparison, statistically significantly increases in offspring serum T3, and only moderately reduced offspring serum T4, were observed on PND 90 as only timepoint of measurement in the EOGRTS-like MMI study, which did not yield any brain-related findings (Fegert

et al. 2012). Regrettably, offspring serum T3 data are unavailable for ETU or mancozeb, i.e. two TPO inhibitors that did not elicit brain-related effects. For mercaptobenzimidazole and cyanamide, that also did not elicit brain-related effects, offspring serum T4 and T3 were non-significantly altered, except that offspring serum T3 was statistically significantly increased on PND 6 in the mercaptobenzimidazole study. The ECETOC T4 TF considers that non-significant alterations in offspring serum T4 and T3 indicate a mild to moderate thyroid hormone imbalance that was compensated for by adaptive responses.

In addition to the amitrole study by Ramhøj et al. (2021), which was included in Case Study 1, Ramhøj et al. (2022a) published an additional amitrole study after data evaluation was completed for the present review. While the dosing regime was the same between the two studies (exposure to 25 and 50 mkd from GD 7 to LD 22), the slightly different hormone data (*Table SI-1, Spreadsheet "CS1 Other"*) suggest that the two publications are not founded on the same experimental study. Overall, the thyroid-related findings are consistent: On PND 16/17, both amitrole studies showed moderate offspring serum T4 reduction (<50%) at 25 mkd and pronounced offspring serum T4 reduction (≥60%) at 50 mkd. Ramhøj et al. (2021) additionally measured offspring serum T4 on PND 6 where it was reduced by up to 56%. Ramhøj et al. (2021) found that periventricular heterotopia was induced at 50 mkd (25 mkd not assessed) as sole brain-related assessment. By comparison, Ramhøj et al. (2022a) assessed motor activity on PND 21 and the differential expression of 10 genes on PND 16 and found motor activity in the male offspring to be statistically significantly increased at both dose levels as well as dose-dependent decreases in the expression of different brain genes. Hence, for the low-dose male offspring, the ≥50% threshold for serum T4 reduction was not met (44% on PND 16) so that this parameter failed to predict the increased motor activity. Finally, it is interesting to note that, between the two studies, motor activity was (at least) as sensitive as induction of periventricular heterotopia.

Regrettably, serum fT4 data are not available for any TPO inhibitors. Therefore, it is not possible to further evaluate how this parameter may have complemented the pattern of effects observed for total T4 and T3 in the offspring serum.

Offspring serum TSH was generally dose-dependently and statistically significantly increased upon exposure to TPO inhibitors, and this increase was more pronounced for PTU and amitrole (TSH increase above 400%) than it was for MMI in the EOGRTS-like study, ETU, mancozeb, mercaptobenzimidazole and cyanamide, for which brain-related effects were not observed (TSH increase below 400% and non-significant for mercaptobenzimidazole and cyanamide). As the offspring HPT axis was clearly also activated by TPO inhibitors for which brain-related effects were not observed, the extent of offspring TSH increase might provide an indication for the effectiveness of the HPT axis to address the serum thyroid hormone decrements. The very high offspring TSH increases recorded for PTU and amitrole might indicate that the TSH increase triggered by the HPT axis was no longer effective. Further research is recommended to establish whether prolonged offspring serum TSH increases exceeding a certain level indicate that the HPT axis is "overwhelmed" so that adequate thyroid activation to address substance-mediated thyroid hormone imbalance no longer occurs. Also, it remains to be established if there is an added value in measuring serum fT4/fT3 levels if total thyroid hormone levels are only moderately reduced and the pattern of neurodevelopmental or other brain-related findings is not conclusive.

4.2.1.2. No data on brain T4/T3 levels or brain gene expression changes for TPO inhibitors which did not elicit neurodevelopmental effects.

Key messages: PTU caused pronounced brain T4 and T3 decrements and induced the expression of most brain genes considered in the respective studies. The further evaluation of these parameters is impaired by the absence of data on brain T4/T3 or brain gene expression for those TPO inhibitors, which did not elicit neurodevelopmental effects.

From amongst the Case Study 1 substances, brain thyroid hormone levels are only available for PTU. In the PTU studies, pronounced decreases in T4 and/or T3 levels in different parts of the brain (by up to 92% and 97%, respectively) were observed (Section 3.2.5). Interestingly, brain cortex T3 levels were increased in the low- and mid-dose groups (by 22% and 78%, respectively) on PND 14/15 in the PTU study by Sharlin et al. (2010) whereas they were decreased by 44% in the high-dose group. Possibly, this is an indication for adaptive responses in the low- and mid-dose groups that were no longer effective in the high-dose group. Regrettably, brain T4 and T3 levels are unavailable for those TPO inhibitors for which neurobehavioural findings or periventricular heterotopia were *not* recorded. Therefore, it is not possible to determine whether these substances elicited less pronounced reductions in brain thyroid hormone levels. Further investigations of how *in utero*/lactational exposure to different TPO inhibitors, that either cause or do not cause DNT, affect offspring brain T4 and T3 levels appear as a pertinent research need.

Brain gene expression data are available for PTU, MMI (at 8 and 16 mkd) and amitrole (Section 3.5), i.e. from studies, in which neurodevelopmental effects were observed. Different brain genes were evaluated between studies. Generally, exposure to PTU, MMI (at 8 and 16 mkd) and amitrole resulted in a dose-dependently altered expression of a variety of brain genes; however, consistency across studies and an understanding of the relevance of specific gene expression changes in predicting specific neurodevelopmental alterations (e.g. altered motor activity, altered acoustic startle response) are lacking. Research work is recommended to align on a relevant brain gene expression pattern, as such information could also complete some of the elements included in thyroid-related AOPs with neurodevelopmental outcomes in mammals (further discussed in Section 4.3.6).

4.2.1.3. Possible explanations for differences in patterns of effects for different TPO inhibitors.

Key messages: To date, there is no conclusive evidence for why some TPO inhibitors elicit pronounced offspring serum T4

decrements and neurodevelopmental effects in rats whereas others do not. Likely, differences in toxicokinetics play a role.

Differences in TPO binding affinities are likely not solely decisive for the TPO inhibitors' potential to elicit pronounced offspring serum T4 decrements and neurodevelopmental effects in rats. While there is some evidence that PTU and MMI can either irreversibly or reversibly bind to TPO *in vitro*, the reversible binding appears more relevant *in vivo* (Davidson et al. 1978); see Appendix I, Section App-I.1.2, for further discussion. It has also been suggested that PTU additionally inhibits DIO1 and thus has multiple thyroid-related MoAs (Rijntjes et al. 2013). This is also one of the possible explanations for why PTU, but not ETU or mancozeb, induces neurodevelopmental effects that is presented in the European Commission (2017a) Thyroid Disruption Workshop Report (for discussion, see Appendix II, which is included in this paper after the bibliography). Since DIO1 metabolises T4 to more active T3 or inactive reverse T3, DIO1 can modulate local T3 levels in the liver and hence also systemic T3 levels. However, this reaction takes place in the periphery. Therefore, Ramhøj et al. (2022a) questioned that DIO1 inhibition contributes to the neurodevelopmental effects caused by PTU. Just like PTU, the TPO inhibitors amitrole and MMI (at 8 and 16 mkd), which do not inhibit DIO1, also cause pronounced offspring T4 and T3 decrements as well as neurodevelopmental effects (Ramhøj et al. 2022a). Other factors that may contribute to the circumstance that only some TPO inhibitors cause neurodevelopmental effects may be differences in toxicokinetics, including the passing of the blood-brain barrier, placental/lactational transfer, and/or foetal/pup blood concentrations of the active substances. In this regard, Marchant et al. (1977) suggested that the rat placenta appeared more permeable to ^{35}S-MMI than to ^{35}S-PTU. Further toxicokinetics-related research is recommended to better understand why some TPO inhibitors elicit neurodevelopmental effects in rats whereas others do not.

4.2.1.4. Inconsistencies between MMI studies.

Key messages: MMI did not cause neurodevelopmental effects in an EOGRTS-like study when tested up to the maximum tolerated dose in that study (3 mkd). When applying up to 16 mkd MMI from GD 7 to LD 22, altered motor activity and increased brain gene expression were observed. For two investigational studies using extremely high doses of ~36 mkd MMI in the drinking water, it is unclear if the maximum tolerated dose was exceeded.

The findings from the different MMI studies are not fully consistent. In the EOGRTS-like study (Fegert et al. 2012), moderate T4 reductions (by up to 35% on LD 22 in the dams and by up to 46% on PND 4 in the offspring) and no brain-related findings were observed when testing up to 3 mkd MMI, the maximum tolerated dose in that study based on body weight decrements. In the investigational study by Darbra et al. (2003), the much more pronounced offspring T4 decrement (by 78% on PND 21 as only thyroid hormone value reported) is consistent with the data by Fegert et al. since Darbra et al. applied a much higher dose (~36 mkd from GD 9 to LD 21). Darbra et al. recorded neurodevelopmental effects so that this study was assessed as "true positive" in the evaluation of the ≥60% offspring serum T4 reduction threshold. From a full hazard characterisation perspective, however, it is noteworthy that Darbra et al. (2003) did not provide information on the dams' body weight and further did not report the magnitude of body weight reduction in the offspring. Therefore, the neurodevelopmental findings cannot be related to the overall health status of the dams or offspring, and it is not possible to determine if the maximum tolerated dose may have been exceeded. Moreover, in the MMI study by Ausó et al. (2004), the same, very high dose of ~36 mkd was applied (again in the drinking water), but only for four days during gestation (GD 12–15). Ausó et al. recorded that T4 was reduced in the dams on GD 15 by ~30% (i.e. in a similar range as the maternal T4 reduction on LD 22 in the EOGRTS-like study). Ausó et al. also reported a 22% maternal serum T3 decrement on GD 15 (a parameter that was unchanged on LD 22 in the EOGRTS-like study) and that offspring serum T4 was not reduced on PND 40 as only timepoint of measurement. The significant decrease in maternal serum T3 rather suggests that thyroid hormone perturbation was pronounced in the dams, and thus likely led to thyroid hormone imbalance in the offspring. The ECETOC T4 TF suggests that PND 40 is too late to establish offspring thyroid hormone effects caused after gestational substance administration from GD 12 to 15. Indeed, Ausó et al. is the only study considered in the present review in which offspring T4 was not altered at any timepoint. It is also the only study in which the offspring brain was investigated (on PND 40) after having exposed the dams only while the foetus was still fully dependent on maternal T4 (GD 12–GD 15). As regards brain-related effects, Ausó reported altered proportions of cells in different layers of the primary somatosensory cortex and the hippocampal CA1 area and specified these as heterotopic cells. The ECETOC T4 TF notes that the exposure period applied by Ausó et al. does not overlap with the sensitive period for the development of periventricular heterotopia upon exposure to PTU identified by O'Shaughnessy et al. (2019), i.e. GD 19–PND 2. O'Shaughnessy et al. (2019) did not detect heterotopia when PTU was administered from GD 9 to GD 14. The reasons for this discrepancy between the exposure period applied by Ausó et al. and the window of susceptibility identified by O'Shaughnessy et al. are unclear. Ausó et al. assessed single ectopic cells as being heterotopic, unlike the clusters of cells recorded, e.g. by Gilbert et al. (2014), O'Shaughnessy et al. (2019), or Ramhøj et al. (2021). To the best of the ECETOC T4 TF's knowledge, the relevance of single ectopic cells as being indicative for heterotopia has not been described by other authors. By commonly accepted definition, heterotopia is a cluster of cells in an abnormal location and not the presence of single ectopic cells (see e.g. Guerrini and Parrini 2010). It is also noteworthy that the locations of the single ectopic cells described by Ausó et al. (see above) do not match the location of heterotopia in the *corpus callosum* described in the other studies.

Hasebe et al. (2008) also investigated the effects of MMI on maternal and offspring serum T4 and T3 levels and brain development (see *SI-1 Spreadsheet "Examples Excluded Studies*

Substances" for rationale to not consider this study in Case Study 1). Hasebe et al. treated dams orally with 20 mkd MMI from GD 17 until the end of lactation. Motor function and coordination were examined in the PND 30 offspring using the rota-rod test. Further, the cerebellum was evaluated both *via* morphometry and histopathology on PND 6, 9, 12, 15, 25, and 30. The major findings were impaired motor function and an increased (and in some cases decreased) sub-lobulation of the cerebellum with preservation of the normal cerebellar lobulation pattern. As stated by Hasebe et al. (2008): "*In the cerebellum of hypothyroid pups, the external granular cells were retarded both in cell migration and proliferation but present a longer time than in normal, resulting in acquisition of a greater number of the internal granular cells*." Regrettably, Hasebe et al. did not provide any morphometrical information on other parts of the brain, and especially not on the presence or absence of heterotopia in the region of the *corpus callosum* as described by O'Shaughnessy et al. (2018b, 2019), Ramhøj et al. (2020a, 2021), and Gilbert et al. (2021).

In the recent MMI study by Ramhøj et al. (2022a), 8 and 16 mkd MMI were applied from GD 7 to LD 22 (*Table SI-1, Spreadsheet "CS1 Other"*). Hence, the exposure duration (GD 7–LD 22) matches that applied in most studies included in this review, and the dose level ranges between the top dose of 3 mkd MMI applied by Fegert et al. and the dose of ~36 mkd MMI applied by Darbra et al. and Ausó et al. Ramhøj et al. (2022a) recorded significantly altered motor activity in the PND 21 male offspring of the high-dose group, which coincided with pronounced offspring serum T4 reduction and moderate offspring serum T3 reduction (T4: 92% and 86% T3: 25% and 29% in the PND 16 males and PND 17 females, respectively) as well as dose-dependent and statistically significant increases in offspring serum TSH. In the low-dose group, motor activity was not statistically significantly altered, and this coincided with only moderately reduced offspring serum T4 reduction (<60%) and non-statistically significant offspring serum T3 reduction. Taken together, the recent MMI study by Ramhøj et al. (2022a) supports the notion that MMI causes neurodevelopmental effects at high doses, and it further supports the suggested thresholds of ≥60%/≥50% offspring serum T4 reduction for the top-dose group/lower-dose groups.

4.2.2. Case Study 2: Impairment of thyroid hormone synthesis via iodine deficiency

4.2.2.1. Perchlorate: moderate thyroid hormone imbalance, no neurodevelopmental effects, but electrophysiological alterations.

Key messages: Perchlorate caused moderate maternal and offspring thyroid hormone imbalance and no neurodevelopmental effects, apart from electrophysiological alterations. However, when restricting iodine-levels in the diet to prevent a masking of detrimental effects, offspring serum T4 decrements were pronounced and coincided with pronounced brain thyroid hormone decrements and altered brain gene expression.

Case Study 2 included rat studies investigating the *in utero*/lactational effects of the NIS inhibitor perchlorate and, for comprehensiveness, dietary iodine deficiency studies, even though the focus of this review is on substance-mediated effects. Overall, the perchlorate studies that applied doses up to 30 mkd (York et al. 2004, 2005a, 2005b; Mahle et al. 2003) yielded consistent results. Serum T4 decrements were generally moderate (<60%) in both the dams and offspring. Serum T3 decrements in the dams were generally low and mostly not statistically significant, whereas maternal serum TSH was dose-dependently and statistically significantly increased. Most likely, the pattern of maternal T4, T3 and TSH changes indicates that the HPT axis compensated for the substance-mediated thyroid hormone imbalance. In the offspring, serum T3 reduction was overall low to moderate (i.e. <60% and mostly even lower). However, upon *in utero*/lactational exposure to high dose levels of 10 and 30 mkd perchlorate, offspring serum T3 reduction was more pronounced than the corresponding offspring serum T4 reduction (York et al. 2004; 2005a). Offspring serum TSH was dose dependently and significantly increased by at most 49% in the perchlorate studies, i.e. in a similar range as for those TPO inhibitors for which no brain-related effects were observed. The ECETOC T4 TF interprets these thyroid-related findings as indicating that the offspring thyroid hormone system was perturbed but that the adaptive mechanisms of the HPT axis were not necessarily overwhelmed. In agreement with this assumption (as well as with the hypothesis that <60% offspring serum T4 reduction is indicative of "absence of statistically significant neurodevelopmental effects"), the perchlorate studies also showed no effects on motor activity, acoustic startle response or cognitive function. However, Gilbert and Sui (2008) observed altered brain electrophysiology at 4.5–140 mkd perchlorate; see Section 4.3.5 for discussion of brain electrophysiology.

Gilbert et al. (2022) published a further perchlorate study after data evaluation for the present review was finalised (*Table SI-1, Spreadsheet "CS2 Perchlorate"*). In this study, up to 140 mkd perchlorate, the same dose range as applied by Gilbert and Sui (2008), were administered to pregnant dams in the drinking water from GD 6 to GD 20. Unique to this perchlorate study, iodine levels in the diet were controlled and below 25% "*of that found in typical rodent chows*" to prevent potential masking of detrimental effects of NIS inhibitors on hormone production while ensuring sufficient iodine supply (Gilbert et al. 2022). Presumably on account of the iodine-restricted diet, the offspring serum T4 decrements were much more pronounced in this study than in the perchlorate study by Gilbert and Sui (2008), i.e. 80% and 96% at 44 and 140 mkd, respectively, on GD 20 (Gilbert et al. 2022) as compared to non-significant alterations at 44 and 140 mkd on PND 4 (Gilbert and Sui 2008) despite administration of the same dose range. The pronounced serum T4 decrements in the GD 20 foetus coincided with pronounced offspring brain T4 and T3 decrements (up to 91% and 85%, respectively) and altered gene expression in both the maternal and offspring thyroid gland and offspring forebrain. Further, levels of monoiodotyrosine, diiodotyrosine, reverse T3, T4 and T3 were significantly reduced at these dose levels in both the maternal and offspring thyroid gland (both: GD 20). Apart from forebrain gene expression, this study did not include any other brain-related assessments.

Apparently, perchlorate exposures combined with iodine-restricted diet led to pronounced offspring serum and brain T4 and T3 decrements and altered brain gene expression; regrettably, neurobehavioural endpoints are unavailable for this study.

4.2.2.2. The two dietary iodine deficiency studies: inconsistent brain-related findings.

Key messages: Dietary iodine deficiency studies do not include xenobiotics, which might elicit neurodevelopmental impairment by (additional) MoAs unrelated to thyroid hormone imbalance. Therefore, a comprehensive review of dietary iodine deficiency studies may enhance the understanding of how maternal and offspring thyroid hormone imbalance affects neurodevelopment. However, the two dietary iodine deficiency studies considered here yielded inconsistent brain-related findings.

An abundance of dietary iodine deficiency studies using rats is available. For example, the search query "*iodine deficiency rat thyroid (brain OR neuro*)*" yielded 162 results in the database PubMed on 10 February 2022. A comprehensive review of all relevant studies would have exceeded the scope of this review, which focussed on substance-mediated effects. Notwithstanding, even a preliminary comparison of the findings from the two dietary iodine deficiency studies that were selected for illustrative purposes (Zhang et al. 2012; Gilbert et al. 2013) with the findings recorded in the studies for the 14 case study substances is impaired by the circumstance that the two dietary iodine deficiency studies yielded inconsistent findings for both the thyroid- and the brain-related parameters.

The differences in offspring serum T4 decrements [moderate (<60%) on PND 7/PND 45 and on GD 20/PND 14 as per Zhang et al. and Gilbert et al., respectively, but pronounced (≥60%) on PND 14 as per Gilbert et al.; Section 3.2.1.3] can most likely be explained by differences in the levels of iodine deficiency between the two studies (1.2 μg *vs.* 0.2–0.3 and 0.7–1.1 μg iodine/day), and possibly further by the differences in timepoints of T4 measurement. Taken together, the thyroid-related findings indicate that the adaptive mechanisms of the HPT axis were (at least) partially effective in preventing the occurrence of adverse neurodevelopmental effects (acoustic startle response, water maze learning and fear conditioning not altered as per Zhang et al.). Both maternal and offspring serum T3 were only non-significantly altered, whereas maternal and offspring serum TSH levels were dose-dependently and statistically significantly increased (but less so in the study by Gilbert et al. than in the study by Zhang et al. despite identical exposure periods and even though iodine deficiency was more pronounced in the Gilbert et al. study; Section 3.2.4).

The brain-related findings, and specifically those from the Morris water maze test, are also not consistent between the two dietary iodine deficiency studies (Section 3.1.2). It can only be speculated if differences in timepoints of measurement of cognitive function and/or interlaboratory variability account for the different outcomes of the water maze tests. At PND 40, the timepoint at which Zhang et al. conducted the water maze test, male and female rats are in the peripubertal phase (US EPA 2011a, 2011b). Therefore, it cannot be excluded that stress was more prominent during this phase than when Gilbert et al. tested the animals at PND 60–180.

As dietary iodine deficiency is known to affect neurodevelopment, the reasons for why (more) pronounced brain-related findings were not recorded in the two iodine deficiency studies are unclear. An earlier iodine deficiency study by Li et al. (1986) included histopathological evaluation of the offspring brain and thyroid (on GD 16, 17, 18, 19, 20 and PND 1, 5, 10, 20, 30, 60) after dams had been fed with an iodine deficient diet for four months prior to mating. The iodine deficient diet included 45 μg iodine/kg diet (or 0.9 μg iodine/day assuming food consumption of 20 g/day), which was chosen to reflect the level of iodine deficiency that had induced endemic cretinism in a village in China. Controls were fed a diet containing 547 μg iodine/kg diet (or 11 μg iodine/day). The main findings in the iodine deficient group were goitre, higher uptake of ^{125}iodine, reduced serum T4 (measured only in the dams and only during gestation—without further specification of timepoint) and reduced brain weight. While the density of brain cells had increased, the mean neuron size was reduced. The external granular layer of the cerebellum persisted longer in the iodine deficient pups. The layering of the cerebral cortex was not recognisable in layers 2, 3, and 4. The myelination of the *corpus callosum* was reduced in iodine deficient pups (see Section 4.3.4 for discussion of brain histopathology).

4.2.3. Case Study 3: Enhancement of thyroid hormone clearance via (1) displacement of thyroid hormone from serum binding proteins and/or (2) induction of liver enzymes that metabolise thyroid hormones

Key messages: Different patterns of thyroid- and brain-related effects were observed for the five Case Study 3 substances that enhance thyroid hormone clearance.

Evidence in the scientific literature indicates that all five Case Study 3 substances can both displace thyroid hormone from serum binding protein and induce liver enzymes (Appendix I; Section App-I.3.2). The ECETOC T4 TF did not further investigate which of these MIEs was predominant for which of the five substances, and most likely, this is not yet fully understood. Differences in the predominant MIE/MoA, as well as in toxicokinetic properties, might account for differences in patterns of effects observed for the Case Study 3 substances. This review was not designed to establish the specific molecular or cellular events underlying such differences. Other research teams reported that different liver enzyme inducers may selectively induce either T3-UGT or T4-UGT thereby leading to different patterns of thyroid-related effects (Klaassen and Hood 2001; Vansell and Klaassen 2001). More recently, Vansell (2022) suggested that increased membrane transport of thyroid hormones, likely in conjunction with increased T3 glucuronidation, which was considered to be more relevant than T4 glucuronidation, provide a better indication of thyroid disrupting potential than UGT induction alone.

Based upon the database collated here, all five Case Study 3 substances induced liver enzymes related to phase I and/or II metabolism (Section 3.4), but only Aroclor 1254 and DE-71 also induced neurodevelopmental effects. Aroclor 1254 induced low frequency hearing loss (further discussed in Section 4.2.3.4) and DE-71 induced some effects on motor activity, acoustic startle response and/or learning and memory in some studies but not in others (further discussed in Section 4.2.3.5).

4.2.3.1. TBBPA and triclosan: low to moderate thyroid hormone imbalance and generally no neurodevelopmental effects.

Key messages: TBBPA and triclosan generally elicited at most moderate thyroid hormone imbalance and no neurodevelopmental effects.

TBBPA and triclosan generally elicited at most moderate (<60%) serum T4 reduction in the dams and offspring. These two substances also generally elicited only low and mostly non-significant alterations in serum T3 and only non-significant alterations in serum TSH in the dams and offspring. Lai et al. (2015) postulated four hypotheses for the absence of an adaptive increase in TSH in response to TBBPA-mediated T4 reduction. (1) TBBPA interferes with thyroid hormone binding; (2) TBBPA stimulates UGT upregulation and increased hepatic metabolism of T4; (3) TBBPA changes the sequestration of T4 in body storage sites; (4) TBBPA impairs T4 synthesis in the thyroid gland (considered unlikely since TBBPA did not elicit histopathological changes of the thyroid gland; Lai et al. 2015). Beyond these hypotheses by Lai et al., Klaassen and Hood (2001) suggested that TSH increases are more closely related to T3 decrements than to T4 decrements. However, T3 was mostly non-significantly decreased upon exposure to TBBPA so that TSH increases may not have been expected.

Triclosan further elicited a 25% reduction in offspring serum fT4 on PND 14 and an up to 46% reduction in brain T4 on GD 20–PND 2 (i.e. much lower than the brain T4 reduction caused by PTU), but it did not significantly change brain T4 on PND 6 or 14 or brain T3 at any of these timepoints (brain T4/T3 data are unavailable for TBBPA). Hence, TBBPA- and triclosan-mediated thyroid hormone imbalance was at most moderate in these studies, and changes in brain T4 levels only transient. Similarly, a comprehensive review by Witorsch (2014) concluded that available toxicity test data in a variety of mammalian species showed little evidence that triclosan adversely affects thyroid function. Consistently, the TBBPA and triclosan studies also showed no neurobehavioural or other brain-related findings. As an exception, Lilienthal et al. (2008) recorded altered BAEPs in an enhanced one-generation reproductive toxicity study testing up to 3000 mkd TBBPA. The EFSA (2011b) Panel on Contaminants in the Food Chain, in its *Scientific Opinion on TBBPA and its derivatives in food*, raised concerns about benchmark dose modelling in this study and requested independent confirmation of BAEP study findings before using these data in risk assessment. In line with the EFSA (2011b) conclusions, the ECETOC T4 TF considers the negative outcome of the studies by Cope et al. and Saegusa et al. as superseding in the overall evaluation of the TBBPA studies.

4.2.3.2. PFHxS: pronounced offspring serum T4 reduction, which attenuated during exposure period, transiently reduced brain T4, no neurodevelopmental findings.

Key messages: PFHxS elicited pronounced offspring serum T4 reduction, which however attenuated over the course of the exposure period; brain T4 was transiently reduced, and neurodevelopment remained unaffected.

PFHxS caused pronounced (≥60%) offspring serum T4 reduction during the first postnatal days (PND 0, 2, and 6), whereas offspring serum T4 was only moderately reduced (<60%) on PND 14 and 16 (Ramhøj et al. 2018, 2020b; Gilbert et al. 2021; Section 3.2.1.3). A variety of neurodevelopmental assessments were negative in both studies; similarly, the expression of most genes was not statistically significantly altered. In the evaluation of these findings, the pronounced offspring serum T4 reduction observed on PND 0, 2, and 6 (i.e. during the window of susceptibility to periventricular heterotopia) in the PFHxS study by Gilbert et al. (2021) was considered as being overriding to the moderate reduction observed on PND 14. Therefore, this study was assessed as false positive in the evaluation of the ≥60%/≥50% offspring serum T4 reduction thresholds. However, offspring serum T4 reduction attenuated over the course of the exposure period, a finding that was generally only recorded in studies yielding no brain-related findings. Further, offspring serum T3 was significantly reduced, whereas offspring serum TSH levels were only non-significantly altered (Sections 3.2.2 and 3.2.4). Offspring serum fT4 reduction (measured by Gilbert et al. 2021) was almost equal to the corresponding serum T4 reduction (52% *vs.* 54%). Just as was observed for triclosan, PFHxS-mediated brain T4 reductions (observed on PND 0, 2 and 6) were much less pronounced than those observed for the TPO inhibitor PTU (Section 3.2.5) and were superseded by an increase in brain T4 on PND 14. Further, brain T3 levels, which were reduced upon exposure to PTU, were unchanged in the PFHxS study at all four timepoints. Reflecting the amount of active thyroid hormone present in the target organ, brain T3 may be the best suitable parameter to predict the occurrence or absence of neurodevelopmental effects (Section 4.4.6). Also, brain levels of fT4/fT3 (which, however, were not measured in any study) might have remained unchanged in the PFHxS studies, resulting in an euthyroid status in the developing brain and thus unaffected neurodevelopment. Research work is recommended to determine whether the increased brain T4 levels on PND 14 reflect an adaptive reaction to PFHxS-mediated thyroid hormone imbalance and whether brain T3 may be a suitable parameter to predict neurodevelopmental impairment.

4.2.3.3. Possible MoAs for PFHxS and PFOS.

Key messages: PFHxS likely triggers MIEs for several thyroid-related MoAs, which, however, apparently do not overwhelm the adaptive mechanisms of the thyroid system of rat pups.

Gilbert et al. (2021) suggested that, while total serum T4 may have been decreased in their PFHxS study due to substance interference with serum binding proteins, fT4 transport into the brain was likely maintained so that neurodevelopment remained unaffected. As per Gilbert et al., this hypothesis is consistent with clinical presentations of patients and with findings from knockout mice possessing "loss of function" mutations in serum binding proteins. Ramhøj et al. (2020b) suggested that PFHxS reduces serum thyroid hormone levels by triggering multiple MIEs, including T4 displacement from transthyretin and albumin and liver enzyme induction, both leading to increased T4 excretion. The findings from the present review support the view that PFHxS-mediated serum T4 decrements did not trigger HPT axis activation after *in utero*/lactational substance exposure nor subsequent key events leading to measurable neurodevelopmental effects. Chang et al. (2008) suggested a similar hypothesis with respect to the MoA of PFOS: (1) PFOS competes (directly or indirectly) with T4 for serum carrier protein binding sites, and the displacement of the thyroid hormones leads to a transient elevation of fT4. (2) This transient elevation in fT4 leads to increased turnover and elimination of T4. (3) PFOS does not affect the ability of the pituitary to release TSH or to respond to hypothalamic thyrotropin releasing hormone (Chang et al. 2008; PFOS was not selected as Case Study 3 substance due to the inability to find suitable studies; *Table SI-1, Spreadsheet "Excluded Substances & Studies"*). While the hypotheses on possible MoAs for PFHxS and PFOS presented and discussed in the literature appear plausible, research work is recommended to expand the database on substance-mediated changes in brain T4 and T3 levels to enhance the understanding of the implications of different patterns of thyroid hormone imbalance on neurodevelopment (Section 4.4.6).

4.2.3.4. Aroclor 1254: low frequency hearing loss and accelerated eye opening.

Key messages: Aroclor 1254 elicited pronounced offspring serum T4 reduction, which was likely not related to an activation of the HPT axis. Aroclor 1254 also caused low frequency hearing loss and accelerated eye opening (for which a thyroid-unrelated MoA is being discussed).

The four Aroclor 1254 studies considered in Case Study 3 generally yielded consistent findings (Goldey et al. 1995b; Goldey and Crofton 1998; Crofton et al. 2000b; Morse et al. 1996). In the dams, serum T4 decrements were moderate (<60%), and serum T3 levels were either not changed or only moderately reduced (26%). None of the Aroclor 1254 studies assessed maternal serum TSH. Interestingly, maternal serum fT4 reduction was slightly more pronounced than serum T4 reduction on GD 20 (58/50%; Morse et al. 1996). In the offspring, serum T4 reduction was always pronounced (i.e. ≥60%) both at the end of gestation and during lactation up until PND 30, and serum fT4 reduction was often even (slightly) more pronounced (Goldey et al. 1995b; Morse et al. 1996). Further, offspring serum T4 reduction during the exposure period was most pronounced on PND 14 (Section 3.2.1.4). Offspring serum T3 was moderately reduced (e.g. by 29% on PND 21; Goldey et al. 1995b), whereas serum TSH was only non-significantly changed. This pattern of effects likely indicates that the Aroclor 1254-mediated hypothyroxinaemia was not related to an activation of the HPT axis (after *in utero*/lactational substance exposure). This supports the assignment of Aroclor 1254 to Case Study 3 as substance with an indirect thyroid-related MoA.

The pronounced serum T4 reduction in the newborn and juvenile offspring caused by *in utero*/lactational exposure to Aroclor 1254 coincided with hearing deficits upon adulthood. These hearing deficits further coincided with pronounced T4 reductions in the brain on GD 20 (forebrain: up to 87%, cerebellum: below the lower limit of quantification), which had attenuated by PND 21 (forebrain: males: no significant effect: females: 44% reduction in the high-dose group). The organ of Corti of offspring of Aroclor 1254-treated female rats showed a mild to moderate loss of outer hair cells in the upper-middle and apical turns; inner hair cells and ganglion cells were not affected (Crofton et al. 2000a). Brain T3 levels were not altered by *in utero*/lactational exposure to Aroclor 1254 (Morse et al. 1996).

Goldey et al. (1995b) reported accelerated eye opening in pups after *in utero*/lactational exposure to 8 mkd Aroclor 1254; they suggested that this finding could not be attributed to "*classical effects of hypothyroidism*" but provided further evidence that Aroclor 1254 may have "*a number of different mechanisms of toxic action in the developing rat*." The pups from 8 mkd group also showed 25% and 50% mortality at PND 12 and PND 21, respectively; further, there was 15% mortality by PND 21 in the 4 mkd group (Goldey et al. 1995b). In a second study, Goldey and Crofton (1998) also applied 8 mkd Aroclor 1254. These authors reported a similar effect on early eye opening but did not report any offspring mortality. The reason for the differential response in pup mortality across studies is unclear (exposure period in both studies: GD 6–LD 20).

4.2.3.5. DE-71 studies: findings not fully concordant.

Key messages: DE-71 generally caused pronounced offspring serum T4 reduction and at most moderate offspring serum T3 reduction. The findings for offspring serum TSH were not consistent between the five DE-71 studies. Also, only some DE-71 studies showed brain-related findings. The reasons for discrepancies between DE-71 studies are unclear. DE-71 likely has complex MoAs, but these are not yet fully understood.

In the DE-71 studies, maternal serum T4, just as maternal serum T3, if available, was generally moderately reduced (<60%). While maternal serum TSH was significantly increased (by 127%) in the high-dose group of the study by Kodavanti et al. (2010), this effect was not dose-dependent (Section 3.2.4) and further less pronounced than the serum TSH increases observed for the TPO inhibitors. In the recent DE-71 study by Ramhøj et al. (2022b, 2022c), maternal serum TSH was not significantly affected, but the authors noted large data variability.

In the offspring, serum T4 reductions were generally pronounced (Section 3.2.1.3) whereas serum T3 reductions were at most moderate (Section 3.2.2). Offspring fT4 data are

unavailable for DE-71. While offspring serum TSH was non-significantly increased as per Kodavanti et al. (2010) and Ramhøj et al. (2020a [SOT poster]), it was dose-dependently and significantly increased as per Bowers et al. (2015) (Section 3.2.4). Consistent with these differences, offspring serum T3 reduction was also most pronounced in the DE-71 study by Bowers et al. (*Table SI-1, Spreadsheet "Overview T4 T3"*). Offspring serum TSH was not affected on PND 16 or 27 in the DE-71 study by Ramhøj et al. (2022b; exposure period GD 7–LD 16). Thereby, the DE-71 study by Bowers et al. is the only study investigating a substance with indirect thyroid-related MoA (Case Studies 3 and 4) in which offspring serum TSH was both dose-dependently and statistically significantly increased. Nonetheless, the up to 42% TSH increase recorded by Bowers et al. is much lower than the offspring serum TSH increases caused by those TPO inhibitors that elicited neurodevelopmental effects. Differences in offspring serum TSH increases or offspring serum T3 decreases between the DE-71 studies cannot be explained by differences in dose setting since the selected top doses were widely concordant (30–40 mkd; exception: the recent DE-71 study by Ramhøj et al. (2022b), see below). However, Bowers et al. began dosing on GD 1 whereas Kodavanti et al. and Ramhøj et al. began dosing on GD 6–7. Since the duration of thyroid hormone change also seems to determine the extent of TSH increase (see e.g. findings from the dietary iodine deficiency studies presented in Section 3.2.4), the longer dosing period might contribute to the dose-dependent and statistically significant increase in offspring serum TSH recorded by Bowers et al. (2015).

Taken together, offspring serum T4 reduction was generally pronounced in the DE-71 studies, but—apart from the study by Bowers et al.—this did not coincide with pronounced dose-dependent and statistically significant serum TSH increases or more than low to moderate serum T3 reductions (<40%). Most likely, this indicates that thyroid hormone imbalance was not triggered by direct effects on the HPT axis upon *in utero*/lactational exposure to DE-71 from GD 6 to 7 until the end of lactation.

As compared to the pronounced offspring serum T4 reduction, the DE-71 studies by Kodavanti et al. (2010) and Ramhøj et al. (2020a) yielded no brain-related findings, for which reason they were assigned as false positive in the evaluation of the ≥60% offspring serum T4 reduction threshold (Section 3.2.1.3). Also, Ramhøj et al. (2022b) found no effects on motor activity and habituation (PND 21 and 79), learning in the Morris water maze (PND 132–176) or auditory function (PND 216–233) in pups following *in utero*/lactational exposure to up to 60 mkd DE-71 from GD 7 to PND 14/16. Hence, this recent DE-71 study showed no neurodevelopmental effects even though a 1.5–2-fold higher dose was administered than in the five other DE-71 studies (also, the ECETOC T4 TF suggests there was adequate systemic toxicity in this study; see note in Cell O43-48 of *Table SI-1, Spreadsheet "CS3 DE-71"*).

The DE-71 studies by de-Miranda et al. (2016); Bowers et al. (2015)/Gill et al. (2016) and Zhou et al. (2002), as reported in the poster abstract by Taylor et al. (2003), did yield brain-related findings, but always only in few tests or at few timepoints within a comprehensive testing battery. The ECETOC T4 TF is unable to establish a plausible reason for why some DE-71 studies yielded brain-related findings whereas others did not. As described above, the top dose was widely the same between DE-71 studies, and for four of the five studies, the exposure durations covered GD 6–7 up until the end of lactation. Regrettably, brain thyroid hormone data are unavailable for DE-71 so that it is not possible to correlate the presence or absence of neurobehavioural findings to brain thyroid hormone levels.

In line with the assignment of DE-71 to Case Study 3, Ramhøj et al. (2022b) concluded that "*the DE-71 mixture is a microsomal enzyme inducer which does not cause an activation of the HPT-axis in rat offspring*." Possibly, DE-71 has further MoAs in addition to mechanisms enhancing thyroid hormone clearance (see also Ramhøj et al. 2022c). Kronborg et al. (2017) found that DE-71 inhibits the expression of TPO-related genes in human thyroid cells. However, if DE-71 were also a TPO inhibitor, this would still not explain why there were brain-related findings in some DE-71 studies but not in others. As per Bansal et al. (2014), the effect profile of DE-71 differs from that of the TPO inhibitor PTU. Bansal et al. (2014) concluded that PBDEs, such as DE-71, "*can have multiple effects on hormone clearance, tissue uptake, and even receptor activation, producing complex effects on development and adult physiology*."

4.2.3.6. Aroclor 1254 and DE-71: offspring serum T4 reduction more pronounced during mid-lactation than in early or late lactation.

Key messages: In the Aroclor 1254 and DE-71 studies, offspring serum T4 decrements were generally more pronounced on PND 6–14 than in early or late lactation.

Offspring serum T4 decrements elicited by Aroclor 1254 and DE-71 were more pronounced during mid-lactation (PND 6–14) than in early or late lactation, i.e. they were most pronounced in the middle of the exposure period. This finding (that also applies to the recent DE-71 study by Ramhøj et al. 2022b, 2022c) was not observed for any other substance included in the case studies (Section 4.4.5), and it is especially noteworthy considering that serum T4 levels in healthy pups are generally highest at approximately PND 14 (Zoeller and Tan 2007). Therefore, this timepoint has the greatest dynamic range to detect changes, which may be coincidental. Aroclor 1254 and DE-71 are lipophilic so that the maximum offspring serum T4 reduction around PND 6–14 might be related to lactational transfer of these substances into the milk and direct substance exposure in the offspring.

4.2.4. Case Study 4: OMC studies: maternal serum T4 eliminated but hardly any effects in the offspring?

Key messages: One study reported that exposure to up to 1000 mkd OMC caused nearly complete maternal serum T4 depletion, which apparently lasted for three weeks. This extreme and continued effect would need to be reproduced to allow establishing its biological relevance. Offspring serum T4 levels remained almost unchanged just as brain-related findings were rather scant.

The findings from the two OMC studies are not straightforward to explain. In the study by Axelstad et al. (2011b),

exposure to up to 1000 mkd OMC caused nearly complete maternal serum T4 depletion, which apparently lasted for three weeks (maternal serum T4 reduction: up to 96% and 100% on GD 15 and LD 15, respectively). These were the most pronounced maternal serum T4 reductions observed in any of the studies included in this review. Regrettably, Axelstad et al. did not measure maternal serum T3 or TSH, and Ramhøj et al. (2020a), the second OMC study, did not measure any maternal serum hormone levels. As compared to these tremendous maternal serum T4 decrements, offspring serum T4 levels remained almost unchanged (PND 16 offspring: no significant effects in the females, up to 35% T4 reduction in the males; adult offspring: no significant changes in either the females or males). These offspring serum T4 data match those from Ramhøj et al. (2020a) who reported up to 40% and 35% offspring T4 reduction on GD 21 and PND 3, respectively, and no significant change in offspring serum T4 at either PND 6 or PND 16.

Axelstad et al. (2011b) did not report any signs of overt toxicity in the dams or developmental delays in the offspring. The ECETOC T4 TF considers it unlikely that maternal serum T4 could have been (widely) eliminated for three weeks during gestation and lactation without causing overt toxicity in the dams and further without considerably affecting offspring serum T4 levels or offspring development. Also, the brain-related findings observed in the two OMC studies were rather scant: Following a broad spectrum of neurobehavioural assessments, most findings were either non-significant or not dose-dependent [exceptions: reduced motor activity at 9 and 17 weeks in the high-dose female offspring (Axelstad et al. 2011b) and brain gene expression changes (Ramhøj et al. 2020a)]. Axelstad et al. (2008) noted that the "*severely decreased T4 levels … did … not cause the expected behavioural effects in the offspring and no correlations with T4 levels were seen. Our results indicate that in rats, only severe postnatal T4 decreases are determining for adverse brain development.*"

The extreme and continued maternal serum T4 reduction observed by Axelstad et al. (2011b) would need to be reproduced to allow establishing its biological relevance (for definition see EFSA 2011a, 2017) and to preclude sampling or measuring errors. By comparison, Schmutzler et al. (2004) treated ovariectomised female Sprague-Dawley rats for 12 weeks with 54 and 285 mg/animal/day OMC. Assuming that the rats (14 weeks of age at the onset of the experiment) had a body weight of 280 g (https://www.criver.com/products-services/find-model/cd-sd-igs-rat?region=23 [accessed 2022 September]), the applied doses correspond to 193 and 1018 mkd. Hence, the top doses applied by Schmutzler et al. and Axelstad et al. lie in the same range. Schmutzler et al. reported that serum T4 levels were reduced by ~40% (significant) and 16% (non-significant), in the low- and high-dose groups, respectively, at the end of the treatment period. Serum T3 levels were non-significantly increased in the low-dose group but slightly reduced in the high-dose group, whereas serum TSH levels were not affected in either dose group (Schmutzler et al. 2004). Further, Klammer et al. (2007) treated ovariectomised Sprague-Dawley rats for five days with 10, 33, 100, 333, or 1000 mkd OMC and recorded significantly reduced serum T4 levels (by 25% and 41%) at the two highest doses only. Thus, T4 reductions by OMC in these studies were considerably less than reported by Axelstad et al. (2011b).

Data provided in the ECHA disseminated dossier for OMC stand in agreement with these findings (https://echa.europa.eu/registration-dossier/-/registered-dossier/15876 [accessed 2022 September]). Four-week dietary administration of 1000 mkd OMC to 11- to 13-week-old female Wistar rats resulted in slight but statistically significantly increased serum T4 concentrations (19%) but no effects on T3 or TSH levels. Furthermore, in a Good Laboratory Practice-compliant 90-day toxicity study (performed according to OECD TG 408), dietary administration of up to 1000 mkd OMC to Füllinsdorf Albino Specific Pathogen Free rats did not result in histopathological changes of the thyroid gland.

The studies by Schmutzler et al. and Klammer et al. as well as those provided on the ECHA disseminated dossier did not address *in utero* or lactational exposure to OMC, and pregnancy and lactation can undoubtedly pose additional stress to the maternal thyroid system. Nonetheless, the low to moderate serum T4 reductions observed in these studies upon up to 90-day exposure to OMC do not provide evidence to support the continued maternal serum T4 depletion reported by Axelstad et al. (2011b).

4.3. Neurodevelopmental and other brain-related findings

Key messages: Section 4.3 discusses the biological relevance of the most important brain-related parameters evaluated in the reviewed studies (for definition of biological relevance, see EFSA 2011a, 2017). Windows of susceptibility are considered, if known. Inter-study and inter-substance comparability is impaired since different brain-related parameters were evaluated between studies, at different timepoints and by different methodologies.

Brain-related parameters considered in the present review include neurobehavioural effects on the organism level (altered motor activity, cognitive function, acoustic startle response), functional changes in late-stage key events (altered electrophysiology or auditory signalling), structural changes (periventricular heterotopia, decreased brain volume or thickness of specific brain layers, altered glial cell labelling/cell density) and changes in the expression of brain genes and brain-related proteins. These brain-related parameters were selected for analysis based on sensitivity to thyroid hormone imbalance and the availability of a sufficient number of studies to allow a reasonable assessment (Section 2.2). Other brain-related parameters, that are not considered here since only single studies (testing MMI) were found, include juvenile play behaviour (Northcutt et al. 2021), social interaction (Sala-Roca et al. 2002) as well as isolation-induced ultrasonic vocalisations and homing behaviour tests as attempts to reflect characteristics of autistic-like symptoms in humans (Melancia et al. 2017). Further, the parameters included in the functional observational battery, e.g. grip strength (https://www.ecfr.gov/current/title-40/chapter-I/

subchapter-R/part-798/subpart-G/section-798.6050 [accessed 2022 August]), were not assessed in detail since they did not prove sensitive in predicting neurodevelopmental effects.

Finally, only a few studies were found that addressed magnetic resonance imaging and volumetric analysis of the offspring brain upon *in utero*/lactational exposure to PTU (Powell et al. 2012) or MMI (Hasegawa et al. 2010; Lucia et al. 2018; Salas-Lucia et al. 2020). However, Sauer et al. (2020), referring to work by Korevaar et al. (2016), identified magnetic resonance imaging as a potentially useful parameter to inform on morphological alterations of the child brain and recommended research work to determine how this technique "*can be combined with functional assessments of the brain to enable comprehensive evaluations of child neurodevelopmental outcomes*" and to establish its applicability for toxicological assessments using rodents (Sauer et al. 2020).

In the evaluation of the brain-related parameters considered here, between-study and between-substance comparability is impaired by the circumstance that different parameters (e.g. motor activity *vs.* auditory startle response) were addressed in the different studies, at different timepoints and by different methodologies to measure both the brain-related parameters as well as thyroid hormone changes. For example, periventricular heterotopia was *not* assessed for MMI, ETU/mancozeb, perchlorate or TBBPA, whereas periventricular heterotopia was the *only* brain-related parameter considered for mercaptobenzimidazole and cyanamide.

Due to this study heterogeneity, it was not possible to establish whether specific substances or specific MoAs are associated with patterns of brain-related effects. It was also not possible to compare the sensitivity of different methodologies to assess a given brain-related parameter (e.g. radial arm maze vs. Morris water maze for spatial learning).

The evaluation of the findings from the present review needs to consider that variability may influence the sensitivity of the different brain-related parameters to detect neurodevelopmental effects. For neurobehavioural endpoints, mean coefficients of variation (CVs) are reported in the DNT Guidance Document (NAFTA 2016). In a comparison of motor activity assessments across 14 laboratories, mean CVs ranged from 20% to 60% at PND 21 and 18–30% at PND 60 (Raffaele et al. 2008). Other repeated-dose toxicity studies using adult rats reported mean CVs of 20–25% (Crofton et al. 1991) or 20–53% (Moser et al. 1997) for motor activity. For the auditory startle response, mean CVs were variable across laboratories (20–110%; Raffaele et al. 2008); CVs for auditory brainstem response were not included in this document. Mean CVs for learning and memory tests are more difficult to define as methodologies (e.g. Morris water maze or passive avoidance) and variables assessed (e.g. latency, distance, speed, number of errors) differ across laboratories. For heterotopia, studies generally indicate background levels of heterotopia in control animals; however, absence of heterotopia also has been reported in study groups (e.g. Gilbert et al. 2021). Also, heterotopia has only been measured in a few laboratories and has been reported differently across studies (incidence, volume in mm^3, number of sections with heterotopia, size bins), making it more challenging to identify mean CV values. Possibly, inconsistencies between, e.g. the two dietary iodine deficiency studies (Section 4.2.2.2) and some DE-71 studies (Section 4.2.3.5) can also be explained by such variability.

Finally, it is important to note that the present review did not consider toxicokinetic properties of the case study substances including their potential to cross the placenta or the blood-brain barrier. Therefore, it is not possible to establish whether any brain-related effects might also have been mediated by direct substance effects on the offspring brain.

The sections below further discuss the most important brain-related parameters. Critical periods of neurodevelopment are addressed, if known, as they can help understand the biological relevance of thyroid hormone alterations during development and the impact of timing for thyroid hormone assessments. Critical periods have been identified for hearing function and periventricular heterotopia and, more broadly, for motor activity and cognitive function. The available data suggest that many of the critical periods occur predominantly in the postnatal period. Nonetheless, important stages of brain development in the rat take place prenatally. Indeed, all the mentioned neurodevelopmental endpoints could also be affected with sufficient prenatal alterations in thyroid hormone levels in the developing brain by affecting neuronal proliferation, differentiation, migration, etc. Biologically relevant and sustained thyroid hormone decrements during gestation might be anticipated to produce broader systemic effects rather than targeted effects on specific neurodevelopmental endpoints (see Section 4.4.5 for alignment of timepoints of offspring serum thyroid hormone measurements to cover critical periods of neurodevelopment).

4.3.1. Motor activity

Key messages: Motor activity proved to be a (relatively) sensitive parameter. There seems to be a propensity for offspring serum T4 or T3 decrements of $\geq$60% or $\geq$20%, respectively, recorded around PND 14–21, to be associated with altered motor activity.

Motor activity was frequently addressed in the reviewed studies (20 of 47 studies; Section 3.1.5), and it proved to be a (relatively) sensitive parameter (by yielding findings at lower dose levels than other neurodevelopmental parameters). Motor activity was the most sensitive neurodevelopmental parameter in the PTU study by Axelstad et al. (2008; more sensitive at 16 weeks than on PND 14, 17, or 23), in the Aroclor 1254 study by Goldey et al. (1995b; findings on PND 15 but not at earlier or later timepoints) and in the DE-71 study by Bowers et al. (2015; statistically significant and dose-dependent effects on PND 100 (110) but not at earlier timepoints). Further, from amongst the studies that included only one dose level, effects on motor activity, but no other brain-related findings, were observed in the MMI study by Darbra et al. (2003; on PND 21, 40, 60) and in the Aroclor 1254 study by Goldey and Crofton (1998; on PND 13 and 15 but not at later timepoints). Hence, in the PTU studies, effects on motor activity were *not* observed at PND 16 or 17 but at earlier and later timepoints (Kobayashi et al. 2005; Axelstad et al. 2008), whereas effects on motor activity were *only*

observed on PND 13–15 in the Aroclor 1254 studies but not at earlier or later timepoints (Goldey et al. 1995b; Goldey and Crofton 1998). The most common pattern for hypothyroid-mediated developmental alterations in motor activity includes hypoactivity prior to weaning followed by hyperactivity in weanling-to-adult rats (Axelstad et al. 2008). The present review did not consider whether the thyroid hormone imbalance might also have mediated altered motor activity in the dams; also, there would be very little—if no—information on this in the studies considered.

The precise developmental window during which thyroid hormone imbalance affects motor activity is still unknown. In the developing motor system of the rat, firing of motor neurons begins on GD 15 before the foetal thyroid is functional (Clarac et al. 2004). Nonetheless, the critical window for hypothyroid-mediated effects on motor activity in rats may lie predominantly in the postnatal period (Clarac et al. 2004). This is consistent with OECD TG 426 DNT testing wherein motor activity is first assessed around PND 13, then on PND 17 and around weaning (PND 21), when adult patterns of motor activity are fully developed, and finally around PND 60–70, i.e. in the adult offspring. Similar provisions are included in the OECD TG 443 EOGRTS but with testing in weanlings and adults only (see also NAFTA 2016). In the present review, the exposure period in studies, in which altered motor activity was observed, covered at least parts of gestation and lactation and thus the (broad) critical window for thyroid-related effects on motor activity.

Against this background, the below discussion of the relationship between thyroid hormone levels and motor activity focusses on offspring T4 and T3 levels around PND 14 and PND 21 during the ontogeny of motor activity response. Effects on motor activity upon exposure to PTU, to a very high dose of MMI or to Aroclor 1254 always followed pronounced offspring serum T4 decrements ($\geq$60%) and sometimes also offspring serum T3 decrements ($\geq$60% for PTU, but no change for Aroclor 1254) on PND 16 or 20–22. Similarly, in the DE-71 study by Bowers et al. (2015), pronounced offspring serum T4 and T3 reduction on PND 21 was associated with reduced rearing in the PND 100 (110) offspring. By comparison, in the DE-71 studies by Kodavanti et al (2010) and Zhou et al. (2002)/Taylor et al. (2003), pronounced offspring serum T4 reductions ($\geq$60%) were observed on PND 14–21 but motor activity remained unaffected (on PND 24, 58–60, or 273; Kodavanti et al. 2010). Differences in findings may reflect variability in methods/timing for motor activity assessment, duration of thyroid hormone alteration, failure to evaluate the most critical developmental period, and/or other unaccounted for variables that can affect motor activity. Finally, in the perchlorate study by York et al. (2005a, 2005b), mild offspring serum T4 reductions ($<$20%) on GD 21, PND 5, 10 and 22 were associated with unchanged motor activity on PND 14, 18, 22, and 59.

As regards offspring serum T3 levels, in the PTU studies, altered motor activity that persisted until adulthood was associated with at least moderate offspring serum T3 reduction (e.g. 59% on PND 13, 45% on PND 21; Kobayashi et al. 2005; Johnstone et al. 2013). By comparison, in the Aroclor 1254 studies, transiently altered motor activity (around PND 13–15) was associated with no or only non-significant offspring serum T3 reduction on PND 14 (Goldey et al. 1995b; Goldey and Crofton 1998). Further research is recommended to establish if different levels of thyroid hormone imbalance during lactation determine whether altered motor activity is likely transient or persistent.

Taken together, there appears to be a propensity for offspring serum T4 or T3 decrements of $\geq$60% or $\geq$20%, respectively, recorded around PND 14–21, to be associated with altered motor activity, with less pronounced T4/T3 reductions having a lower likelihood of being associated with altered motor activity (Figures 3(A–D) and 4(A–C)). However, across case studies, there is some overlap between magnitude of change in offspring serum T4 and T3 with and without altered motor activity (Supplementary Information SI-6). An association between magnitude of offspring serum T4 or T3 decrement is also not observable for specific MoAs (see Figures 3(A–D) for the offspring serum T4 and motor activity data available for each of the four case studies; and Figures 4(A–C) for the offspring serum T3 and motor activity data available for Case Studies 1–3 (the Case Study 4 OMC studies did not include serum T3 measurements). These findings provide further evidence that offspring serum T4 and T3 levels should preferably be considered together with brain T4 and T3 levels to establish whether the offspring thyroid hormone imbalance is so pronounced that neurodevelopmental effects, including altered motor activity, will occur (further discussed in Section 4.4.7).

4.3.2. Cognitive function

Key messages: Parameters related to cognitive function were generally not the most sensitive brain-related parameters. Opportunities to test higher cognitive function in rodents are limited.

Cognitive function was the most frequently addressed neurodevelopmental parameter in the studies considered here (assessed in 22 of 47 studies; Section 3.1.5). Nonetheless, parameters related to cognitive function were generally not the most sensitive brain-related parameters (as regards dose level at which neurodevelopmental findings were observed; Table 7). Exceptions where learning and memory had equal or greater sensitivity to other brain-related endpoints included two studies that only used one dose level and where a test for cognitive function was positive whereas at least one other test (for any brain-related parameter) was not. These are the MMI study by Darbra et al. (2003), in which *in utero* and lactational exposure to ~36 mkd MMI resulted in altered passive avoidance tasks on PND 80 (and effects on motor activity on PND 21, 40, and 60), whereas novelty-directed exploratory behaviour and anxiety behaviour remained unchanged, and the DE-71 study by de-Miranda et al. (2016), in which female offspring exposed to 30 mkd DE-71 during lactation made more errors in the radial arm maze, whereas motor activity remained unchanged on PND 40–42, 70, and 100.

While deficits in cognitive function, i.e. learning and memory as well as attention, are amongst the most

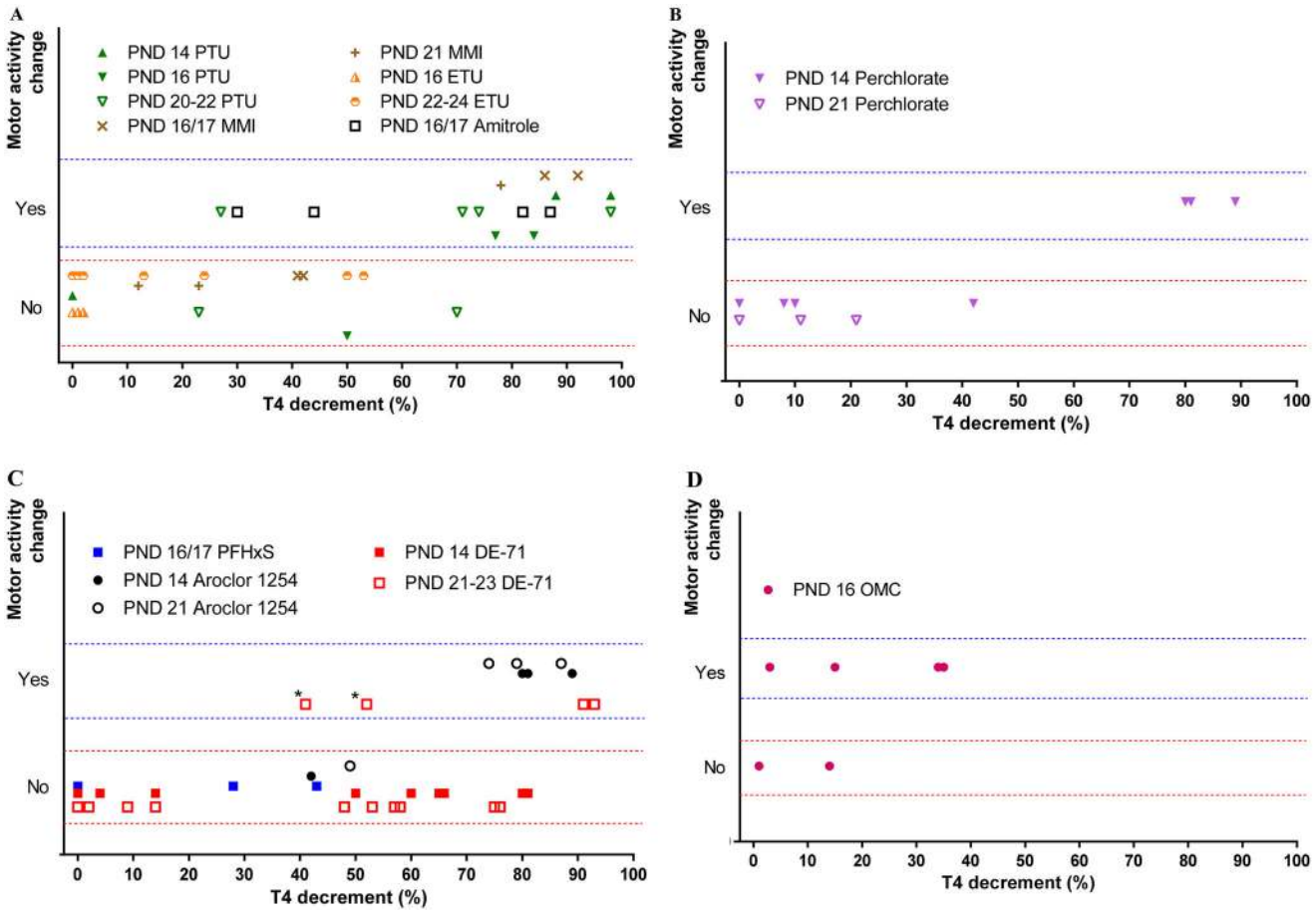


Figure 3. Pup serum T4 decrements (between PND 14 and end of lactation) associated with motor activity changes, sorted by case study.
(A) Case Study 1—PTU, MMI, ETU, amitrole.
(B) Case Study 2—perchlorate.
(C) Case Study 3—PFHxS, Aroclor 1254, DE-71.
(D) Case Study 4—OMC.
ETU: ethylene thiourea, MMI: methimazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T4: thyroxine.
These graphs include all (statistically significant or non-significant) data recorded at all dose levels from the studies that tested for motor activity and measured offspring serum T4 between PND 14 and the end of lactation (i.e. the window of susceptibility to motor activity changes). All data expressed as percent of concurrent control group for the respective study.
Y-axis: Binary representation of presence of change (upper range) or absence of change (lower range). All instances of altered motor activity recorded as statistically significant. Data points at 0% T4 decrement may represent either unaffected or increased T4 levels.
*Figure 3C: Decrease in rearing recorded by Bowers et al. (2015) at 3 mkd DE-71 but no effect on ambulation or immobility time (i.e. no effect on motor activity *per se*).
Timepoints as designated in the respective studies. The birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ± 1 day may represent the same stage of development.

profound neurodevelopmental effects associated with maternal thyroid perturbation in humans, opportunities to test higher cognitive function in rodents are limited (Makris et al. 2009). There is simply no test that is equivalent to assessing "intelligence" in rodents. In standard toxicity test methods, assessments of cognitive function are only included in the OECD TG 426 DNT study but not in the OECD TG 443 EOGRTS. Tests for cognitive function mentioned in OECD TG 426 (therein Section 37) include passive avoidance tests, Morris water maze, radial arm maze and T-maze. Cognitive function is often assessed both at weaning and during adulthood to determine whether effects appeared later in development/maturation or whether early effects on learning and memory reflect permanent alterations in neuronal function.

As regards critical windows of neurodevelopment, the evaluation of the data on learning and memory yields similar findings as have been discussed above for motor activity, although gestational development likely plays a more critical role. Learning and memory tests require locomotor/swimming, vision and audition, which are developed by weaning. This is consistent with DNT testing (OECD TG 426) wherein evaluation of learning and memory is conducted in weanlings beginning approximately PND 23. In the present review, most studies that addressed effects on learning and memory dosed the animals from GD 6 to LD 21 and thus included exposures over the period of the critical windows of development. As regards association between offspring serum T4 changes and altered cognitive function, the data generally support that T4 decrements of <60% (Figure 5A and Figure 5B for measurements at GD 20–PND 10 and PND 14–PND 30, respectively) and T3 decrements of <20% (Figure 6) have a lower likelihood of altering learning and memory than larger decrements.

Consistent with the findings from this review, an analysis of the results from OECD TG 426 DNT studies investigating a

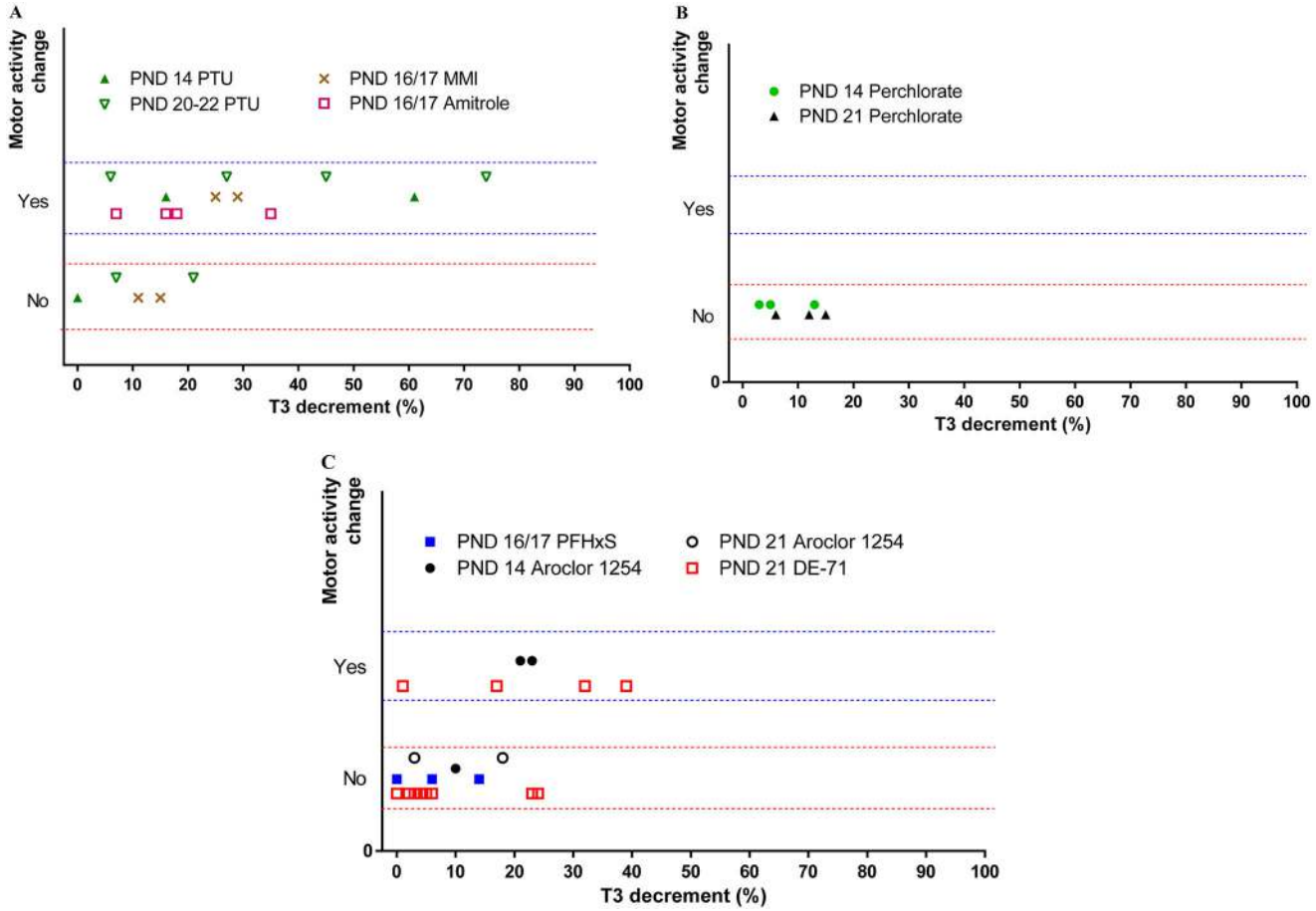


Figure 4. Pup serum T3 decrements (between PND 14 and end of lactation) associated with motor activity changes.
(A) Case Study 1—PTU, MMI, amitrole.
(B) Case Study 2—perchlorate.
(C) Case Study 3—PFHxS, Aroclor 1254, DE-71.
MMI: methimazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T3: triiodothyronine.
These graphs include all (statistically significant or non-significant) data recorded at all dose levels from the studies that tested for motor activity and measured offspring serum T3 between PND 14 and the end of lactation (i.e. the window of susceptibility to motor activity changes). All data expressed as percent of concurrent control group for the respective study. Note: There is no panel for Case Study 4 since the OMC study that addressed motor activity did not also address serum T3.
Y-axis: Binary representation of presence of change (upper range) or absence of change (lower range). All instances of altered motor activity recorded as statistically significant. Data points at 0% T3 decrement may represent either unaffected or increased T3 levels.
Timepoints as designated in the respective studies. The birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ± 1 day may represent the same stage of development.

total of 69 pesticides showed that the standard tests for cognitive function were less sensitive for detecting learning and memory effects than measures of motor activity and acoustic startle habituation (Raffaele et al. 2010), see also review by Vorhees and Makris (2015). As per EFSA (2020b), 14 of these 69 pesticides exhibit strong evidence for having a MoA directly related to hypothyroidism.

4.3.3. Acoustic responses

Key messages: Tests for acoustic startle response proved useful. For Aroclor 1254 (and PTU), pronounced offspring serum T4 decreases beginning around PND 4–6 and lasting until PND 15–18 were associated with altered cochlear development.

Parameters related to acoustic responses include the acoustic startle response (Crofton and Sheets 1989), which is commonly included in DNT studies, and hearing function. The acoustic startle response assesses a primitive form of learning *via* startle habituation. Since testing for the acoustic startle response typically includes a strong startle stimulus, these assessments are not sensitive to more subtle changes in hearing threshold. By comparison, altered hearing function can be measured as a change in auditory brainstem response or hearing thresholds across frequencies.

The acoustic startle response was included in 17 of the 47 studies that included brain-related parameters (Section 3.1.5). Even though it is a rather crude measurement, it was the most sensitive brain-related parameter in the PTU study by Kobayashi et al. (2005). Regrettably, the acoustic startle response was not assessed in the other PTU studies.

Tests for hearing function including auditory brainstem responses were included in seven studies. Hearing deficits were recorded in all Aroclor 1254 studies investigating this endpoint, and hearing function was also (amongst) the most sensitive endpoint in these studies (possibly together with transiently altered motor activity on PND 15).

Thyroid hormone is required for cochlear development. The thyroid hormone-sensitive period for ototoxicity is slightly better defined that those for effects on motor activity

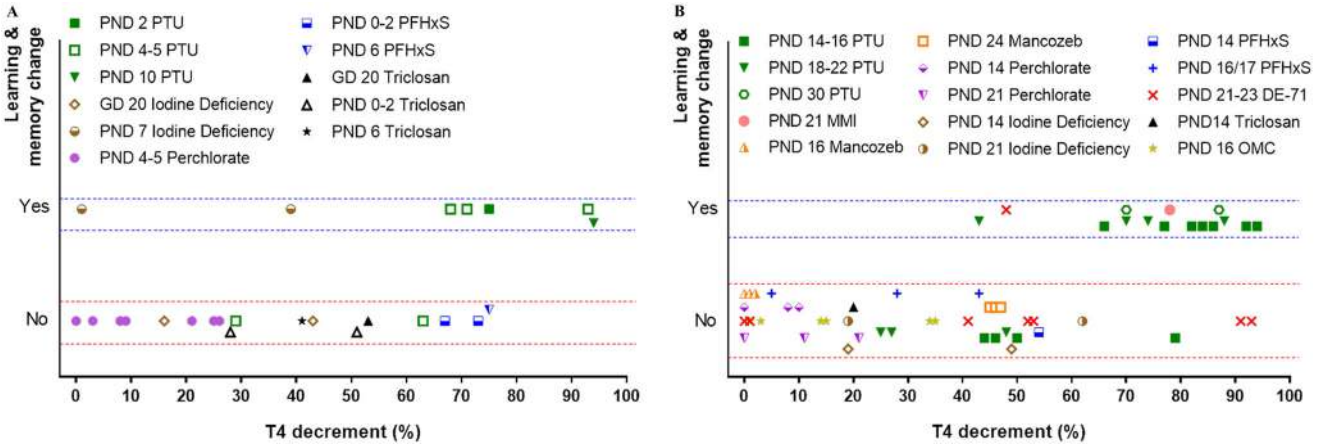


Figure 5. Foetal and pup serum T4 decrements associated with learning and memory changes.
(A) Foetal and pup serum T4 decrements (between GD 20 and PND 10) associated with learning and memory changes.
(B) Pup serum T4 decrements (between PND 14 and PND 30) associated with learning and memory changes.
GD: gestational day; MMI: methimazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T4: thyroxine.
These graphs include all (statistically significant or non-significant) data recorded at all dose levels from the studies that tested for learning and memory and measured offspring serum T4 either, for Figure 5A, on GD 20–PND 10 (i.e. the first part of the window of susceptibility to learning and memory changes—no datapoints for PND 11–13) or, for Figure 5B, on PND 14–PND 30 (i.e. the second part of the window of susceptibility to learning and memory changes). All data expressed as percent of concurrent control group for the respective study.
Y-axis: Binary representation of presence of change (upper range) or absence of change (lower range). All instances of altered learning and memory recorded as statistically significant. Data points at 0% T4 decrement may represent either unaffected or increased T4 levels.
Timepoints as designated in the respective studies. The day the dams were "sperm positive" was either defined as GD 1 or GD 0, and the birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ±1 day may represent the same stage of development.

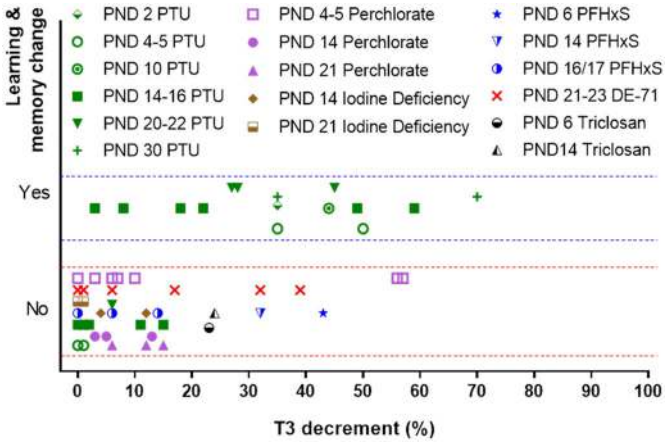


Figure 6. Pup serum T3 decrements (between PND 2 and PND 30) associated with learning and memory changes. PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T3: triiodothyronine. This graph includes all (statistically significant or non-significant) data recorded at all dose levels from the studies that tested for learning and memory and measured offspring serum T3 from PND 2 to PND 30 (i.e. the window of susceptibility to learning and memory changes). All data expressed as percent of concurrent control group for the respective study. Y-axis: Binary representation of presence of change (upper range) or absence of change (lower range). All instances of altered learning and memory recorded as statistically significant. Data points at 0% T3 decrement may represent either unaffected or increased T3 levels. Timepoints as designated in the respective studies. The birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ±1 day may represent the same stage of development.

or cognitive function. The sensitive period for effects on the organ of Corti (i.e. hair cell loss in the cochlea affecting low frequency hearing) is postnatal prior to PND 21 with some evidence for PND 4 or 6 to PND 15 or 18. The NAFTA (2016) guidance document, referring to research work by, e.g. Sheets et al. (1988), indicates that cochlear development in the rat begins during the second postnatal week and reaches adult auditory patterns by PND 21. Following a cross-fostering study, Crofton et al. (2000b) determined that postnatal lactation exposure to Aroclor 1254 has a major impact on ototoxicity in rats. This is consistent with the period of cochlear development. Further, Crofton (2004) established that 50–60% decreases in pup serum T4 levels on PND 14–21 induced by, e.g. Aroclor 1254, other PCBs and PTU were associated with altered cochlear development. In these studies, thyroid hormone measurements were collected at a time that largely coincided with this developmental event. Crofton (2004) suggested that the 50–60% decrease in pup serum T4 levels may represent a threshold beyond which serum T4 decrements in pups alter cochlear development affecting low frequency hearing. This threshold is consistent with the ≥60%/≥50% offspring serum T4 reduction threshold suggested here as being predictive of any statistically significant neurodevelopmental effect.

4.3.4. *Brain histopathology including the evaluation of periventricular heterotopia*

Key messages: Histopathological evaluations of the brain require expert judgement. Periventricular heterotopia was observed in some studies even though offspring serum T4 reduction was not pronounced. In other studies, periventricular heterotopia was not observed even though offspring serum T4 reduction was pronounced.

In the studies considered here, general brain histopathology (as well as brain weight and gross observations) was generally only included in the standard toxicity studies, e.g. the OECD TG 426 DNT study and the OECD TG 443 EOGRTS.

These test guidelines include brain weight and morphometry as well as gross observations and histopathology in the perfusion-fixed animals at two time points after birth (optional at PND 22 and mandatory at PND 90). Histopathological and morphometric examinations can inform on substance-mediated structural abnormalities. However, in the case of brain development, it is unlikely that such examinations directly point to the pathogenesis of a given lesion due to the plethora of factors influencing this process (Barkovich et al. 2012).

Regarding the sensitivity of the morphometric endpoints, there is general agreement amongst neuropathologists that effects on linear brain measures indicating a selective effect on brain development should be seen in both juveniles (PND 11 and 22) and young adults (PND 70) (Li et al. 2019). By comparison, altered linear measures only seen in juveniles (i.e. without brain-related findings at PND 70) and in conjunction with decreased body weight in the dams or pups and/or with maternal care issues suggest that the difference in brain measurements in the juveniles, relative to controls, might rather be related to generalised delays or adverse effects on growth and development than a specific neurotoxic effect (Li et al. 2019). Further advice on the interpretation of neuropathological DNT data is provided in Kaufmann and Gröters (2006) and Garman et al. (2016).

The formation of periventricular heterotopia is not included as a standard parameter in OECD TG 426 or OECD TG 443; however, provisions on the types and timepoints of histopathological investigations of the brain would facilitate the identification of heterotopia in these studies. The EU-funded ATHENA project (*Assays for the identification of Thyroid Hormone axis-disrupting chemicals: Elaborating Novel Assessment strategies*) is exploring the usefulness and relevance of the formation of heterotopia as an endpoint for *in vivo* assays to evaluate effects on brain development in mammals (Kortenkamp et al. 2020).

The potential of thyroid-active substances to elicit periventricular heterotopia has been a matter of extensive research work that is being led and coordinated by the US EPA. Specifically, knowledge on the window of susceptibility to periventricular heterotopia is founded on work by O'Shaughnessy et al. (2019) who administered 10 ppm PTU (∼1–1.3 mkd) at varying intervals to pregnant and/or lactating dams. Since periventricular heterotopia was observed in all offspring exposed to PTU from GD 19 to PND 2, O'Shaughnessy et al. (2019) identified this period as a critical window for neuronal migration in the *corpus collosum* and, thus, as a period that is sensitive to offspring thyroid perturbations. However, thyroid hormone decrements may persist due to the toxicokinetic properties of PTU (Leonard et al. 2016; O'Shaughnessy et al. 2018a). Indeed, O'Shaughnessy et al. (2019) reported offspring serum T4 decrements by 98%, 99%, and 74% on PND 0, 2 and 6, respectively, which were associated with brain T4 decrements by 78%, 92%, and 68%, respectively, despite discontinuing dosing on PND 2. Accordingly, the window of susceptibility to periventricular heterotopia upon exposure to PTU may extend beyond PND 2. Possibly, it is for this reason that the periventricular heterotopia studies by Ramhøj et al. (2020a, 2021) measured offspring thyroid hormones up to PND 6.

In an empirically derived quantitative computer model using data from PTU studies, brain heterotopia has been linearly associated with T4 decrements in the maternal and offspring serum and in the offspring brain (Hassan et al. 2017). Nonetheless, the relationship between foetal/pup serum T4 levels and the development of periventricular heterotopia is likely complex. In some studies, periventricular heterotopia was observed even though offspring serum T4 reduction was not pronounced, and in other studies, periventricular heterotopia was not observed even though offspring serum T4 reduction was pronounced. This observation applies in both foetuses (Figure 7(A)) and PND 0–6 pups (Figure 7(B)) with T4 measurements during the sensitive window for periventricular heterotopia. In the PTU study by Gilbert et al. (2014) and Johnstone et al. (2013), statistically significant periventricular heterotopia was observed at a PTU dose level of only 1 ppm (∼0.1–0.13 mkd; exposure period: GD 6–LD 21). Thereby, periventricular heterotopia has proven a sensitive endpoint since it was reported at lower PTU dose levels than those at which other brain-related effects occurred. The periventricular heterotopia observed at 1 ppm PTU was associated with non-significant offspring serum T4 reduction (29% and 46% on PND 4 and 16, respectively; Johnstone et al. 2013). However, this experiment did not include prenatal T4 levels, which can contribute to periventricular heterotopia (see above). At the same dose level (1 ppm PTU), Hassan et al. (2017) recorded that foetal serum T4 was reduced by 61% on GD 20, and this study followed the same study design as that by Johnstone et al. (2013). Further, in the PTU study by Spring et al. (2016; exposure period: GD 6–LD 22), only small, non-statistically significant periventricular heterotopia was observed at 1 ppm PTU, and this corresponded to moderate offspring serum T4 decrements (∼52% on PND 2). Also, in the PTU study by Ramhøj et al. (2020a), maternal exposure to an ∼10-fold higher dose of 1 mkd PTU resulted in non-statistically significant periventricular heterotopia and moderate offspring serum T4 decreases (51%, 37%, and 59% on GD 21, PND 3 and PND 6, respectively; exposure period: GD 7–LD 22). Hence, neither the heterotopia findings nor the offspring serum T4 data are fully concordant between these PTU studies, which have comparable study designs. Regrettably, brain T4 levels were not reported during the window of susceptibility to periventricular heterotopia in these studies. However, brain T4 was reduced by 73% at 1 ppm on PND 14 as per Experiment II in O'Shaughnessy et al. (2018b), which presents brain thyroid hormone data for the study by Johnstone et al. (2013) and Gilbert et al. (2014). Finally, statistically significant periventricular heterotopia was also recorded upon exposure to 50 mkd amitrole, and this finding was associated with an offspring serum T4 decrement of 56% on PND 6 (but by 64% on PND 16) (Ramhøj et al. 2021).

By comparison, periventricular heterotopia was not recorded upon exposure to mercaptobenzimidazole, cyanamide (Ramhøj et al. 2021), PFHxS, triclosan (Gilbert et al. 2021), DE-71 or OMC (Ramhøj et al. 2020a [SOT 2020 poster]). In the mercaptobenzimidazole, cyanamide, triclosan and OMC studies, absence of periventricular heterotopia coincided with only low to moderate offspring serum T4

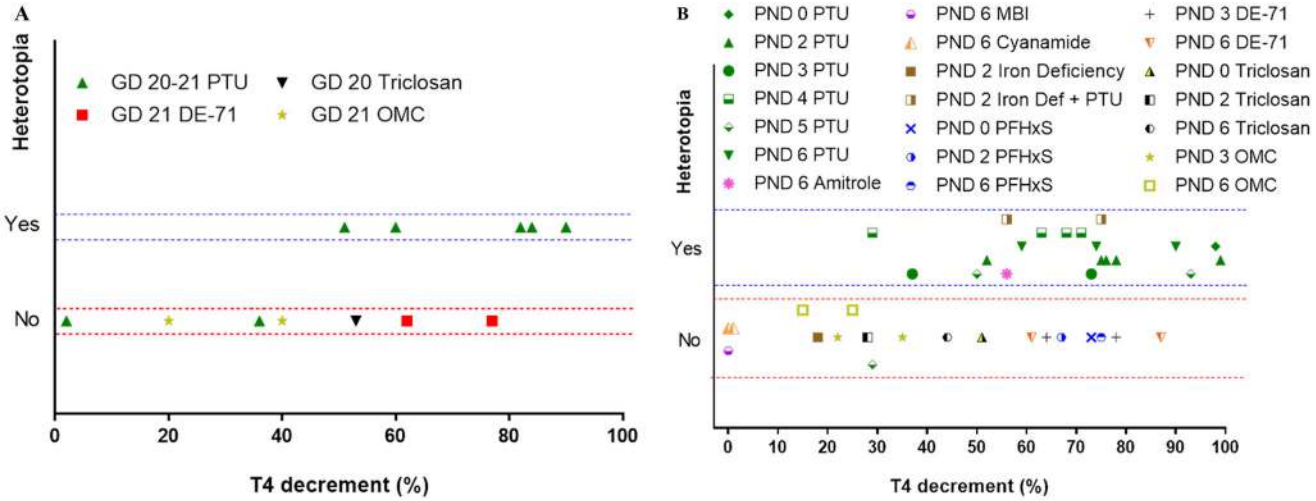


Figure 7. Foetal and pup serum T4 decrements associated with periventricular heterotopia.
(A) Foetal serum T4 decrements (GD 20–GD 21) associated with periventricular heterotopia.
(B) Pup serum T4 decrements (PND 0–PND 6) associated with periventricular heterotopia.
GD: gestational day; MBI: mercaptobenzimidazole; OMC: octyl methoxycinnamate; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T4: thyroxine.
These graphs include all (statistically significant or non-significant) data recorded at all dose levels from the studies that tested for heterotopia and measured offspring serum T4, for Figure 7A, on GD 20–GD 21 (i.e. the first part of the window of susceptibility to the development of heterotopia), or for Figure 7B, on PND 0–PND 6 (i.e. the second part of the window of susceptibility to the development of heterotopia). All data expressed as percent of concurrent control group for the respective study. Data points at 0% T4 decrement may represent either unaffected or increased T4 levels.
Y-axis: Binary representation of presence of change (upper range) or absence of change (lower range). All instances of heterotopia recorded as statistically significant. Timepoints as designated in the respective studies. The day the dams were "sperm positive" was either defined as GD 1 or GD 0, and the birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ±1 day may represent the same stage of development.

decrements (<60%). However, in the PFHxS and DE-71 studies, offspring serum T4 decrements exceeded 60% on GD 20/21 and PND 6, respectively.

It is unclear whether potency, toxicokinetics or a specific MoA related to PTU make this compound more effective at inducing periventricular heterotopia than PFHxS, DE-71, triclosan or OMC. However, the available data support the conclusion that data on foetal/pup serum T4 levels alone are insufficient to accurately predict periventricular heterotopia (or any other neurodevelopmental effect), even when these changes occur during the critical period. This may not be surprising given the number of key events that follow serum thyroid hormone decrements in the different thyroid-related AOPs/MoAs (Section 1.2). Presumably, brain T4 and T3 levels are more predictive: Even though offspring serum T4 reduction was pronounced in the PFHxS study by Gilbert et al. (2021) but only moderate in the triclosan study by these same authors (<60%), brain T4 decrements were widely concordant (Section 3.2.5), and neither substance altered brain T3 levels (which the TPO inhibitor PTU did).

4.3.5. Brain electrophysiology

Key messages: While electrophysiological investigations proved to be sensitive methodologies to evaluate substance-mediated brain-related effects, knowledge gaps and limited proficiency across laboratories currently prevent their wider application for toxicity testing.

Electrophysiological investigations were included in the perchlorate study by Gilbert and Sui (2008) and in the iodine deficiency study by Gilbert et al. (2013). In both studies, brain electrophysiology proved a sensitive parameter yielding findings at dose levels at which other brain-related parameters were not affected. The electrophysiological investigations were conducted under anaesthesia, electrodes were surgically implanted, and for the recordings, the animals were mounted in a stereotaxic frame. A similar approach to electrophysiological assessments was applied in the EPA PTU studies by Gilbert and Sui (2006), Gilbert et al. (2007), and Gilbert (2011). Further, the EPA PTU studies by Sui and Gilbert (2003) and Sui et al. (2005) included electrophysiological investigations in hippocampus slices from the treated rats. Hence, while the 51 studies included in this review were conducted by a broad spectrum of research teams, all studies that included electrophysiological investigations originated from a single research team at the National Health and Environmental Effects Research Laboratory of the US EPA. Highly specialised expertise and refined technical equipment are required to conduct electrophysiological investigations and to collect reliable and reproducible electrophysiological data. Indeed, electrophysiology is not included as a parameter in any test guideline. The investigations are not standardised, and it is unclear how it should be established that any electrophysiological findings were indeed substance-mediated and not, e.g. caused or influenced by the handling of the animals.

In conclusion, while electrophysiological investigations are apparently sensitive methodologies to evaluate substance-mediated brain-related effects—and thus potentially useful for developmental toxicity studies, knowledge gaps and limited proficiency across laboratories currently prevent their wider application for toxicity testing. It is especially on account of the expertise, time for data collection and

technical equipment required to conduct electrophysiological investigations that the ECETOC T4 TF considers it unlikely that electrophysiological investigations will be amenable to the routine testing laboratory setting in the foreseeable future.

4.3.6. Gene expression and/or transporter levels in the offspring brain

Key messages: Hypothyroidism may affect the expression of a broad variety of genes in the developing brain. Therefore, assessments of gene expression changes could be sensitive methodologies to evaluate hypothyroid-mediated neurodevelopmental effects. However, knowledge gaps currently prevent their wider application for toxicity testing. Consensus on a set of relevant marker genes has not yet been established or validated, and the available methodologies to measure gene expression changes lack standardisation for regulatory application. Nonetheless, this field is under active development, and advancements may become available in the not-too-distant future.

In the studies included in the present review, the differential expression of larger numbers of gene markers and gene transcripts generally coincided with altered neurobehaviour and/or periventricular heterotopia. If no, or only one or two gene transcripts were altered, the studies generally did not show altered neurobehaviour or periventricular heterotopia. Thus, brain gene expression appears as a potentially useful parameter to support evaluations. However, the available molecular assays are not yet standardised, and it is often unclear how altered gene expression should be linked to phenotypic alterations (Buesen et al. 2017). To date, gene expression is not included as parameter in any test guideline, even though activities are under way to facilitate its applicability in a regulatory setting (see e.g. Buesen et al. 2017).

This review showed that, so far, consistent genetic or protein markers have not been identified to support the determination of hypothyroid-mediated neurodevelopmental effects. Different research teams selected different sets of gene transcripts for their assessments. The selected candidate genes belonged to different functional classes from myelin regulating genes to genes regulating synaptic function and growth factors. Similarly, the areas of the brain in which gene expression was assessed, as well as the timepoints of measurement, differed between studies. All such differences currently impair between-study comparability. Furthermore, gene expression changes are being measured in the developing brain, in which numerous inherent changes in gene expression are ongoing to support neurodevelopment. Therefore, factors like developmental delays may be difficult to discern from more permanent (adverse) effects on gene expression.

O'Shaughnessy et al. (2019) identified a set of nine genes whose expression was altered on PND 6 (mostly decreased) and which was associated with periventricular heterotopia upon exposure to PTU. Functionally, these genes related to thyroid signalling, neurogenesis and neuronal differentiation, cell migration and growth factors as well as apoptosis. Gilbert et al. (2021) then selected this set of nine genes, and further a set of eleven "*gene markers of disrupted thyroid hormone action*" (identified in PTU studies by Royland et al. 2008; Shiraki et al. 2014; O'Shaughnessy et al. 2018b), to assess whether PFHxS and triclosan cause brain-related effects—and found gene expression unchanged (and no periventricular heterotopia). Ramhøj et al. (2022a) selected a very similar (but not identical) set of 10 genes for their MMI and amitrole studies and found the expression of most genes to be dose-dependently and statistically significantly reduced with concordant alterations in motor activity (see *Table SI-1 Spreadsheet "Overview Gene Expression"* for details on the selected genes).

Hence, in all these studies, the selected set of genes was generally based upon prior studies that investigated PTU. O'Shaughnessy and Gilbert (2020) reviewed evidence obtained, e.g. in knockout mice to inform on the structural and/or functional phenotypic consequences of deficiencies in specific genes. Nonetheless, a comprehensive picture of how alterations in selected sets of genes affect neurodevelopment in rats, or humans, and how this may lead to specific neurobehavioural deficits, is currently unavailable. Also, the ECETOC T4 TF questions whether PTU-specific changes are a suitable reference standard to investigate thyroid-related neurodevelopmental effects of substances with different thyroid-related MoAs. Preferably, the selection of genes as well as the areas of the brain under investigation and the timepoints of measurement should be founded on a clear link between the pattern of observed gene expression changes and a specific functional phenotype. In selecting relevant genes, consideration of brain gene expression changes caused by dietary iodine deficiency may also prove relevant since such changes will not be confounded by possible direct substance effects on the brain.

As discussed by O'Shaughnessy and Gilbert (2020), further research is necessary to facilitate the applicability of molecular assays in the context of toxicity testing:

> Published genetic experiments in mice provide some evidence on how interference with thyroid signaling molecules may or may not affect neurodevelopment, but fully translating these experiments to toxicology can be challenging. Many knockout experiments utilize global deletions, convoluting translation to a chemical exposure. For example, some chemicals may never cross the brain's blood and cerebrospinal fluid barriers *in vivo*, whereas a conventional knockout mouse is missing a functional copy of a particular gene in all cell types across the body. While conditional knockouts do alleviate some of these issues, mechanistic *in vivo* toxicology is ultimately required … Ideally, highly specific "model" chemicals could be utilized to interrogate various MIEs, and then serum biochemistry and neurodevelopmental outcomes reported using the AOP framework. This would provide a fingerprint regarding what key events are expected given various scenarios of thyroid disruption that may, or may not, resemble the effects induced by PTU and MMI (O'Shaughnessy and Gilbert 2020).

For neurodevelopmental studies, a set of reference genes has been proposed that appear relatively unchanged upon exposure to PTU, DE-71, or OMC (Ramhøj et al. 2019). Such reference genes (also known as housekeeping genes) are relevant to normalise relative gene expression data obtained from quantitative real-time polymerase chain reactions (Ramhøj et al. 2019). Therefore, such activities contribute to

the standardisation of molecular assays, thereby facilitating their regulatory applicability (successful formal validation preconditioned).

In conclusion, while this review shows that the assessment of gene expression changes could be a promising and sensitive methodology to evaluate a substance's potential to elicit neurodevelopmental effects, knowledge gaps and formal validation currently prevent their wider application for toxicity testing. However, this field is currently under active development, and progress may be made in the not-too-distant future. Any testing strategy to assess hypothyroid-mediated neurodevelopmental impairment should only include specific brain-related genes if it is understood that they provide added value for the determination of adverse effects.

4.4. Thyroid-related findings

Key messages: All thyroid-related data (or any other findings) have been presented relative to the concurrent study controls. The ECETOC T4 TF estimated many hormone (and body weight) data from graphs. While this compromised data precision, self-monitoring indicated that the variation was likely negligible. The comprehensiveness of the database underlying this review is unprecedented. Nonetheless it has gaps especially related to serum fT4/fT3 and brain T4/T3, but also related to serum T3 and TSH.

In the present database, thyroid-related parameters include T4, T3, fT4 and fT3 in the maternal and/or offspring serum and in different parts of the offspring brain. Further, they include maternal and/or offspring serum TSH, as additional hormone of the HPT axis, as well as maternal and/or offspring thyroid gland weight and/or histopathology.

Generally, the thyroid hormone system is highly versatile, and this impairs the determination whether an altered serum thyroid hormone level is an adaptive or an adverse effect. There are no agreed thresholds for adversely altered serum T4, T3, or TSH levels in rodents (Marty et al. 2021). Following the common study design of regulatory rodent toxicity studies ($n = 10–15$ rats/group), T3/T4 decreases generally need to attain ~25% and TSH increases ~40% as compared to the concurrent controls to be detected as statistically significant (Li et al. 2019; Marty et al. 2021; referring to OECD 2008; US EPA 2009).

Further, to establish whether statistically significant findings from rodent toxicity studies are biologically relevant, they should be compared to historical control data. Historical control data reflect the normal distribution range for the respective sub-population, which includes all control animals that were used in a given laboratory for a specific test method and may cover up to several hundred individuals (Marty et al. 2021). Similarly, internationally accepted generic ranges for "normal," "low," or "high" TSH, T4, or T3 levels in humans do not exist; therefore, population-based reference ranges are generally applied in human observational studies, with the relevant population reflecting the individuals included in the respective cohorts (Sauer et al. 2020). In both human and rodent studies, population-based ranges for "normal" T4, T3, and TSH levels are not only dependent on different properties of the given population, but they are also assay- and laboratory-dependent (Sauer et al. 2020).

The population-, assay- and laboratory-dependence of "normal" T4, T3, and TSH levels generally impairs between-study comparability. To address this limitation, all thyroid-related data (or any other findings) have been presented in this review relative to the concurrent study controls. For some studies, this implied converting the published absolute data to relative data (expressing the relative data as the fraction *by which* the respective value differed from the concurrent controls). Further, hormone data (or data on thyroid or body weight) were often only provided in column or point graphs, and the ECETOC T4 TF estimated the respective relative data from these graphs. In doing so, great care was taken to derive the respective values with the best possible accuracy. Additionally, for quality control, all estimations were re-estimated by a further co-author. Nonetheless, such derivations of hormone data (or of data on thyroid or body weight) from graphs undoubtedly compromised data precision for the present database. However, self-monitoring using publications in which hormone data were presented both as graphs and as numerical data showed that the derivations from the graphs generally only differed from the numerical data by ~1–2% (data not shown). The ECETOC T4 TF considers such variation negligible as compared to the variability of hormone measurements. Overall, the approach taken appears suitable to draw up a comprehensive database on substance-mediated thyroid-related effects in pregnant/lactating rats and their offspring making use of a broad spectrum of differently designed investigational and standard toxicity studies. To the best of the ECETOC T4 TF's knowledge, the comprehensiveness of the database presented in this review is unprecedented. Therefore, it is made available in SI-1 in the form of Excel spreadsheets to facilitate further research work using this database.

Despite its comprehensiveness, the database has gaps. Whereas maternal and/or offspring serum T4 was included in most studies, fewer studies also measured serum T3, serum TSH and/or thyroid weight/histopathology in the dams and/or offspring, and even fewer studies addressed maternal and/or offspring serum fT4 or fT3 or offspring brain thyroid hormone levels. Also, the total number of 51 studies considered, published in altogether more than 60 publications, is clearly not sufficient to establish conclusive evidence on patterns of thyroid-related effects that are causative for neurodevelopmental impairment in rats, or on thresholds for such effects. Nonetheless, this review yields important observational findings regarding regularities for, e.g. different studies assessing the same substances, or substances having a common MoA, or substances for which brain-related effects were, or were not, observed. Further research work is recommended to substantiate and refine the insight from this review to support the biological relevance of the observational findings.

4.4.1. Maternal T4, T3, and TSH levels

Key messages: Most likely, both maternal and offspring thyroid hormone data have a value for hazard assessment. However, in the present database, none of the maternal

thyroid-related parameters were predictive of neurodevelopmental effects in the offspring.

Maternal *vs.* offspring serum T4 decrements were compared separately for samples taken between GD 20 and PND 0, while lactational substance transfer is not yet relevant, and for samples taken during lactation (Section 3.2.1.2; Figure 1). Based on available gestational and neonatal timepoints, maternal serum T4 decrements tended to be associated with offspring serum T4 decrements, however, not to the same magnitude. During or at the end of lactation, serum T4 decrements were often more pronounced in the offspring than in the dams. Presumably, this indicates that maternal blood sampling for T4, T3, and TSH measurements are most meaningful during gestation.

The difference between maternal and offspring serum T4 decrements is unsurprising since maternal thyroid hormone changes are further upstream in the AOP than offspring serum thyroid hormone changes and thus more distant from the adverse outcome. (While all relevant thyroid-related AOPs include "serum T4 reduction" as key event, none distinguishes between maternal *vs.* offspring serum T4 reduction; Marty et al. 2021.) The likelihood that different degrees of thyroid hormone imbalance in pregnant/lactating rats will lead to different degrees of thyroid hormone imbalance in the offspring (or ultimately neurodevelopmental impairment) also depends on the toxicokinetics of the substance in the dams (e.g. half-life, lactational transfer; Costa and Giordano 2007; Zhang et al. 2011) and the toxicokinetics of the substance in the foetal/neonatal offspring including its potential to cross the placenta and the blood-brain barrier (Schröder-van der Elst et al. 1998; Clewell et al. 2003, 2007; Ruis et al. 2019).

Even though maternal serum T4 reduction does not appear predictive of neurodevelopmental effects, this parameter, together with information on maternal body weight, is relevant to establish that the maximum tolerated dose was reached in the respective study—provided that the substance is known to cause thyroid hormone imbalance (Section 4.4.2). Further, whenever maternal serum T4, T3 and TSH levels are considered together, this information can indicate whether the maternal thyroid hormone system has likely undergone an adaptive *vs.* an adverse reaction. Statistically significant reductions in maternal serum T3 (i.e. above ~20–25%), together with significant maternal serum T4 decrements and extremely high increases in serum TSH, likely indicate that the adaptive mechanisms have been overwhelmed. Further, dose-dependent and statistically significant increases in maternal serum TSH likely indicate that the test substance has a direct thyroid-related MoA that affects the entire HPT axis.

Only few studies addressed maternal serum fT4 levels, and none for substances with direct thyroid-related MoA. Also, none of the rat studies addressed maternal serum fT3 levels. Therefore, this review does not allow concluding on the relevance of these parameters for (neuro-)developmental toxicity studies. The few available data indicate that maternal serum total T4 reasonably reflects maternal serum fT4; however, this observation should not be considered conclusive. By comparison, in human observational studies addressing hypothyroxinaemia in pregnant mothers, maternal fT4 is the most frequently measured thyroid parameter (Sauer et al. 2020). This discrepancy between the predominant thyroid parameter measured in human studies *vs.* rat studies impairs cross-species comparisons.

4.4.2. Body weight changes in the dams: appropriate dose setting for developmental toxicity studies

Key messages: Information on maternal body weight is relevant to confirm that the selected dose range was appropriate.

Appropriate dose setting is pivotal to ensure meaningful outcomes of any toxicity study. For studies addressing the different modalities of endocrine disruption, including thyroid hormone imbalance, appropriate dose setting is especially important to ensure that any observed endocrine effects are not secondary to systemic toxicity. In this regard, the European Commission (2017b, 2018) Endocrine Disruptor Criteria state that "*adverse effects that are non-specific secondary consequences of other toxic effects shall not be considered for the identification of the substance as endocrine disruptor*."

Vice versa, if the respective study does not yield any thyroid- or brain-related findings, it needs to be ensured that the top dose was sufficiently high that such effects would have been observed if the tested substance had the potential to elicit them. Generally, OECD TGs recommend using either a limit dose of 1000 mkd or the maximum tolerated dose, i.e. the highest dose to produce toxic effects without causing death or significant morbidity *and* a $<$10% decrease in body weight relative to controls (OECD 2002). The OECD (2018) *Guidance Document No. 150 on standard test guidelines for evaluating chemicals for endocrine disruption* states:

> The top dose or concentration should be sufficiently high to give clear systemic (i.e. non endocrine-specific) toxicity in order to ensure that a wide range of exposures (high to low) is tested. However, endocrine effects observed solely in the presence of clear systemic toxicity should be interpreted with caution and may be disregarded when sufficiently justified to be caused by secondary effects which are unlikely to be due to endocrine activity (OECD 2018).

4.4.3. Offspring serum T4 levels and $\geq$60%/$\geq$50% offspring serum T4 reduction thresholds

Key messages: In the present database, there is some association between $\geq$60%/$\geq$50% offspring serum T4 decrements in the top-/lower-dose groups and the occurrence of statistically significant neurodevelopmental effects (sensitivity: 83%; specificity: 82%; accuracy: 83%, when excluding electrophysiology from the evaluation). The presence of false positive and false negative outcomes indicates that offspring serum T4 reduction is not necessarily the primary parameter that is decisive for neurodevelopmental impairment.

The ECETOC T4 TF empirically set thresholds to establish levels of offspring serum T4 decrements that were associated with statistically significant neurodevelopmental effects (Section 3.2.1.3). Regarding the number of neurodevelopmental endpoints assessed, a conservative approach was applied in which any statistically significant change in any

neurodevelopmental parameter was assessed as indicative of DNT. However, altered brain gene expression, DIO2 activity and BAEPs were excluded from threshold considerations as these endpoints do not indicate neurodevelopmental impairment *per se*. Further, electrophysiology, which was very sensitive, was considered separately since these investigations are not standardised and require specialised techniques that limit widespread use (Section 4.3.5).

The offspring serum T4 reduction thresholds were set to facilitate an evidence-based pragmatic and observational analysis, i.e. they were aligned to yield the fewest possible numbers of false positive and false negative outcomes in the present database. As a result, the ECETOC T4 TF set two thresholds, i.e. ≥60% and ≥50% offspring serum T4 reduction for the top- and lower-dose groups, respectively. Evidently, these definitions mix biological effects (T4 reduction) and dose setting and further do not consider the number of neurodevelopmental endpoints assessed in each study. However, all studies generally selected the top dose to reflect the maximum tolerated dose. Therefore, the impact of dose setting on the biological outcome (extent of T4 reduction) appears subordinate, or even negligible, for the top dose and hence the ≥60% threshold. For many studies considered in this review, ≥60% offspring serum T4 reduction at (or close to) the maximum tolerated dose was associated with statistically significant neurodevelopmental effects, whereas <60% offspring serum T4 reductions at (or close to) the maximum tolerated dose were associated with the absence of statistically significant neurodevelopmental effects.

By comparison, dose setting for the lower doses, i.e. the intervals between dose levels, generally differed between studies. Therefore, the extent of offspring serum T4 reduction in the lower-dose groups could also be affected by the dose spacing and shape of the dose-response curve. For example, if offspring serum T4 reduction was just below 50% in the mid-dose group, it could not be ruled out that it would have exceeded this threshold if the mid dose would have been slightly higher.

Despite these limitations, both the ≥60% and ≥50% thresholds were necessary for the pragmatic evaluation of the present database to not only consider neurodevelopmental effects observed at the maximum tolerated dose but to additionally relate offspring serum T4 reduction to the lowest dose at which a statistically significant neurodevelopmental effect was observed. Since the thresholds were set empirically to indicate an association between the respective T4 levels and the occurrence of statistically significant neurodevelopmental effects (and not a causation), it is unsurprising that different thresholds were identified for the top-dose group than for the lower-dose groups (see Section 4.4.7 for research needs to establish which thresholds might be useful for regulatory toxicity testing).

In the evaluation of the ≥60%/≥50% offspring serum T4 reduction thresholds, the ECETOC T4 TF further considered offspring serum T4 data during the window of susceptibility to the respective brain-related findings, if known; thereby 5 studies were excluded from the evaluation (Table 12). Of the remaining studies, 10 were established as true positive, 3 as false positive, 11 as true negative and 4 as false negative (Tables 11 and 14). This yields a sensitivity of 71%, a specificity of 80% and an accuracy of 76% (Supplementary Information SI-4.3). Apart from electrophysiological findings which were often more sensitive, the ≥60% and ≥50% thresholds were applicable regardless of the type of brain-related parameter considered (i.e. motor activity, cognitive function, acoustic responses, periventricular heterotopia). When excluding electrophysiology from the evaluation, sensitivity, specificity and accuracy of the thresholds increased to 83%, 82%, and 83%, respectively (SI-4.3).

The three false positive studies (≥60%/≥50% offspring serum T4 reduction but no statistically significant neurodevelopmental findings) encompass a PFHxS study (Gilbert et al. 2021) and two DE-71 studies (Kodavanti et al. 2010; Ramhøj et al. 2020a [SOT poster]). Overall, the findings from the five DE-71 studies are not fully concordant, and the reasons for discrepancies are not clear (Section 4.2.3.5). Two of the four false negative studies (<60%/<50% offspring serum T4 reduction but statistically significant neurodevelopmental findings), i.e. a dietary iodine deficiency study (Gilbert et al. 2013) and a perchlorate study (Gilbert and Sui 2008), can be explained by the higher sensitivity of electrophysiological measurements as compared to testing for cognitive function or motor activity. The two further false negative studies are the second dietary iodine deficiency study (Zhang et al. 2012), in which cognitive function was affected despite less pronounced iodine deficiency than applied by Gilbert et al. (Section 4.2.2.2), and an OMC study (Axelstad et al. 2011b). In this OMC study, only single brain-related parameters from amongst a broad spectrum of neurodevelopmental measurements were statistically significant, and the association between the continuous almost complete maternal serum T4 depletion as compared to the moderate effects observed in the offspring is unclear (Section 4.2.4). (The level of statistical significance assigned in the different studies was generally not considered in the present review.)

The presence of false positive and false negative outcomes indicates that, while there appears to be some association between the extent of offspring serum T4 reduction and statistically significant neurodevelopmental findings, offspring serum T4 reduction is not necessarily the primary parameter that is decisive for neurodevelopmental impairment (Section 4.4.7).

4.4.4. Offspring serum T3, TSH, fT4, and fT3 levels and thyroid weight and histopathology

Key messages: Statistically significant offspring serum T3 decrements by at least 20% were often associated with statistically significant neurodevelopmental findings (sensitivity: 70%; specificity: 75%; accuracy: 67%, when excluding electrophysiology). The database on offspring fT4 and fT3 is too incomplete to conclude on the relevance of these parameters. Offspring (or maternal) TSH levels were generally only dose-dependently and statistically significantly increased for substances with direct thyroid-related MoAs. Increased thyroid weight/histopathological findings always coincided with increased serum TSH levels.

Since T3 is generally the active thyroid hormone, mechanisms are in place in the versatile thyroid system to convert inactive T4 to T3 to re-establish homeostasis in case serum T3 levels decrease. Therefore, it is unsurprising that serum T3 decrements were generally less pronounced than serum T4 decrements. Offspring serum T3 reduction was only more pronounced than offspring serum T4 reduction in four studies, and in these studies the hormone reductions were generally mild (<25%; Section 3.2.2) so that the differences may well reflect data variability.

More pronounced offspring serum T3 decrements likely indicate that the adaptive mechanisms of the thyroid system have been overwhelmed so that there is an increased likelihood for neurodevelopmental impairment. (To the best of the ECETOC T4 TF's knowledge, it is currently unclear how long T3 levels need to be perturbed before neurodevelopmental impairment will occur.) Statistically significant offspring serum T3 decrements by at least 20% were often associated with statistically significant neurodevelopmental findings. By comparison, offspring serum T3 reductions that were non-significant and/or <20% (or T3 increases or no effects on T3 levels) were often associated with the absence of statistically significant neurodevelopmental findings. The evaluation of the ≥20% and statistically significant offspring serum T3 reduction threshold yielded a sensitivity of 60%, a specificity of 62% and an accuracy of 61% (70%, 75%, and 67%, when excluding electrophysiology; Supplementary Information SI-4.3). Hence, the offspring serum T3 reduction threshold is less accurate than the offspring serum T4 reduction thresholds (Section 4.4.3), highlighting that offspring serum T3, on its own, is not sufficient to assess thyroid perturbations.

A comprehensive evaluation of substance-mediated thyroid hormone homeostasis *vs.* imbalance may also need to consider serum levels of fT4/fT3, and it is regrettable that the database is incomplete for these parameters. Importantly, offspring serum fT4/fT3 data were only found for studies addressing dietary iodine deficiency as well as different Case Study 3 substances, which enhance thyroid hormone clearance. Offspring serum fT4/fT3 data were not found for the different TPO inhibitors, the NIS inhibitor perchlorate or for OMC, which reduces DIO1 activity in the liver but for which the exact MoA is unclear. Research work is recommended to establish the relevance of fT4/fT3 when assessing substances causing maternal/offspring thyroid hormone imbalance by different MoAs and ultimately possibly also effects on neurodevelopment. Such research work should also address technical difficulties in determining serum fT4 and fT3 levels in rodents (Marty et al. 2021).

TSH is a further hormone of the HPT axis that is measured in toxicity studies to investigate thyroid hormone imbalance. In the present database, TSH levels in the offspring (or dams) were generally only dose-dependently and statistically significantly increased if the substances exhibited a direct thyroid-related MoA. *Vice versa*, if the TSH levels were either not increased or dose-dependently but not statistically significantly increased, the MoA was generally not direct. Further, the dose-dependent and statistically significant increase in offspring serum TSH was most pronounced for the TPO inhibitors PTU and amitrole, i.e. >400% (Section 3.2.4), the two TPO inhibitors that also elicited neurodevelopmental effects.

Klaassen and Hood (2001) observed that substances that induce UGT-mediated T3 elimination also increased serum TSH, whereas Vansell and Klaassen (2001) observed that substances that induce UGT-mediated T4 elimination did not increase serum TSH. These authors hypothesised that TSH increases are more closely associated with serum T3 decrements than with serum T4 decrements. The findings from the present review also indicate that different substances that enhance thyroid hormone clearance may have different patterns of effects.

Finally, thyroid-related parameters include measurements of thyroid weight and histopathology. In the present review, increased thyroid weight/histopathological findings always coincided with increased serum TSH levels. Similarly, Hood et al. (1999) noted that small increases in serum TSH can be sufficient to stimulate thyroid follicular cell proliferation. The present database does not include a single study in which thyroid weight in either the dams or offspring was more sensitive than the corresponding serum T4 alteration. Measurements of thyroid weight generally show pronounced variability, especially due to technical challenges in collecting and dissecting the thyroid gland. Such variability impairs the relevance of thyroid weight data. Thyroid histopathology was generally not more sensitive than the corresponding serum T4 measurements, regardless of the substance's MoA. Nonetheless, thyroid histopathology appears useful to complement serum hormone measurements during toxicological assessments.

4.4.5. Suitable timepoints for offspring serum T4, T3, TSH (and fT4/fT3) measurements

Key messages: The timepoints of offspring thyroid hormone and TSH measurements should preferably be aligned with windows of susceptibility to the neurodevelopmental effects, which in turn should be covered by the exposure period. Knowledge on the toxicokinetics of the substance under investigation will facilitate the determination of suitable timepoints.

The timepoints of serum hormone measurements in developmental toxicity studies conducted for experimental and/or regulatory purposes should be suitable to identify substance-mediated effects on thyroid function and their association with neurodevelopmental effects. Apart from those guideline toxicity studies that evaluated substance-mediated effects from premating of the parental animals all through one or two generations of offspring (OECD TG 443, enhanced OECD TGs 415 and 416), substance exposure in the studies considered here was generally restricted to (parts of) gestation and/or lactation. By comparison, the timepoints of offspring serum hormone measurements varied considerably between studies. In eight studies, the earliest timepoint of offspring serum T4 measurement was at the end of lactation (PND 20/21) or even later, i.e. at the end of the exposure period or only thereafter (Section 3.2.1.1).

However, many of the case study substances elicited more marked thyroid hormone decrements prior to weaning. Further, given the altricial state of the rat (i.e. they are born very immature) and the occurrence of many neurodevelopmental alterations during the first postnatal weeks, it seems reasonable that offspring thyroid hormone measurements might be particularly useful in understanding effects on neurodevelopment if conducted prior to weaning. Hormone measurements post-weaning only (i.e. after substance exposure and foetal/early postnatal neurodevelopment have ended) may underestimate thyroid perturbations during critical periods.

When assessing periventricular heterotopia, offspring serum T4 measurements appear especially suitable between GD 19 and PND 2 (possibly up to PND 6), and when measuring ototoxicity, PND 14 (or PND 14–21) may be best suitable. For other neurodevelopmental endpoints, the findings from the present review do not provide conclusive evidence on the best suitable timepoints (or numbers) of offspring serum T4, T3, and TSH measurements. Whenever the precise window of susceptibility to the respective neurodevelopmental effect is not known, the best suitable timepoint(s) and numbers of offspring serum thyroid hormone measurements (to determine the likelihood that thyroid perturbation is so pronounced that neurodevelopment will be impaired) may need to be determined on a case-by-case basis. Knowledge on the toxicokinetics of the substance under investigation will facilitate the determination of suitable timepoints. For example, if lactational transfer of the substance is unlikely, measurements at multiple postnatal timepoints might not be relevant. If, however, lactational substance transfer is possible, offspring serum thyroid hormone measurements at different timepoints during lactation might provide relevant information.

If a study included several timepoints of serum T4 measurement during the exposure period, effects generally either aggravated or attenuated during the exposure period. Offspring serum T4 decrements that attenuated over the course of the exposure period (i.e. before the end of lactation) were generally associated with the absence of brain-related findings (Section 3.2.1.4). By comparison, offspring serum T4 changes upon exposure to Aroclor 1254 and DE-71 did not attenuate or aggravate over the course of lactation, but rather reached maximum levels around PND 6–14 (Section 3.2.1.4). It is unclear whether this observation was coincidental (possibly, some of the maximum values around PND 6–14 can be explained by data variability) or whether it can be explained by specific pathophysiological or toxicokinetic processes.

Similarly, it is currently unclear whether a very pronounced offspring serum T4 decrement at one timepoint during the period of susceptibility is more predictive of neurodevelopmental effects than a continuous and aggravating decrement that might extend beyond the critical period. Research work is recommended to enhance the understanding of best suitable timepoints of serum (and brain) thyroid hormone measurements. Such research work should also investigate whether information on the attenuation *vs.* aggravation of offspring T4 changes over the course of the exposure period provides added value to predict the occurrence/absence of neurodevelopmental findings. If so, offspring serum hormone measurements paced at relevant intervals during the exposure period, e.g. PND 4, 14, and 21 to cover early, mid and late lactation, possibly further considering late gestation (e.g. GD 19–21), may prove useful.

However, while a set of different timepoints of offspring serum hormone measurements might be useful in an investigational setting, such considerations may need to be balanced against the needs for routine testing conducted for regulatory purposes. The timepoints of measurement included in the current OECD TG 443 EOGRTS (i.e. PND 4 and 21) have generally proven suitable to detect substance-mediated thyroid hormone imbalance. Data from the OECD TG 421/422 reproductive toxicity screening studies may provide additional insight (none of the studies included here followed this guideline).

4.4.6. Offspring brain T4/T3 levels

Key messages: Brain T4 and T3 levels may well be the most relevant thyroid-related parameters to predict whether neurodevelopmental impairment will occur upon *in utero*/lactational exposure to thyroid-active substances. It is recommended to include brain T4/T3 measurements in rat developmental toxicity studies evaluating thyroid-active substances. Thereby, opportunities to standardise brain T4/T3 measurements can be identified, and the understanding of how altered brain T4/T3 levels are linked to neurodevelopmental impairment will be enhanced.

Alterations in brain T4 and T3 levels are the thyroid-related events that lie closest to the adverse neurodevelopmental outcome. In the present database, brain T4 data are available from studies investigating PTU, perchlorate, dietary iodine deficiency, PFHxS, Aroclor 1254, and triclosan. Brain T3 data are available for these same five substances, but not for dietary iodine deficiency. Hence, brain T4 and T3 data are unavailable for 9 of the 14 case study substances. Further, there are no brain T4/T3 data from studies that assessed motor activity or hearing function or from studies in which altered acoustic startle response or cognitive function were observed. These extensive data gaps compromise the establishment of associations between reduced brain T4/T3 levels and the occurrence or absence of specific neurodevelopmental effects.

Nonetheless, this review has provided important insight that merits follow up research work. Brain T3 levels were only significantly altered in the PTU studies and in the recent perchlorate study (in which iodine intake was restricted) but not in the PFHxS, Aroclor 1254 or triclosan studies. Similarly, brain T4 decrements were more pronounced in the PTU studies and in the recent perchlorate study than they were in the studies investigating dietary iodine deficiency, PFHxS, Aroclor 1254 or triclosan. By comparison, offspring serum T4 reduction was not only pronounced in the PTU studies and in the recent perchlorate study but also in the PFHxS and Aroclor 1254 studies. Hence, there may be a disconnect between offspring serum T4 and brain T4 decrements, as has also been pointed out by O'Shaughnessy and Gilbert (2020). These

authors proposed that cellular and molecular assays could serve as targeted readouts of thyroid hormone action within the developing brain. Performed in parallel with serum thyroid hormone measurements, these molecular readouts could be causal to, or associated with, later structural and functional alterations (O'Shaughnessy and Gilbert 2020).

Figure 8(A) provides an overview of the relationship between foetal/pup serum T4 and brain T4 decrements as measured between GD 20 and PND 21. Some datapoints lie close to the trendline (e.g. PTU on GD 20 and, to a lesser extent, Aroclor 1254 on GD 20). However, overall, the datapoints are scattered and do not show a clear correlation between foetal/pup serum T4 and brain T4 decrements. Thus, foetal/pup brain T4 measurements likely have an added value beyond offspring serum T4 measurements alone in evaluating whether offspring thyroid hormone imbalance is so pronounced that neurodevelopmental effects are likely to occur. The largest serum and brain T4 decrements were observed for PTU and GD 20 Aroclor 1254 (which were both associated with brain-related findings). By comparison, for triclosan and PFHxS, offspring serum and brain T4 decrements were less pronounced, and they were not associated with brain related findings. For PND 21 Aroclor 1254, serum and brain T4 levels were likely measured after the critical window of susceptibility and therefore not directly related to the brain-related finding (mainly: altered hearing function).

Similar findings were seen for offspring serum and brain T3 although fewer datapoints are available (Figure 8(B)). Here, none of the specific datapoints lie close to the trendline.

Taken together, it is recommended to include foetal/pup brain T4/T3 measurements in studies evaluating thyroid- and brain-related effects upon *in utero* and/or lactational exposure to thyroid-active substances. As experience is gained with these measurements, opportunities for their standardisation can be identified thereby facilitating their consistent performance. Further research work is recommended to establish the best suitable brain tissues and the best suitable timepoints of brain T4/T3 measurements. In this regard, it remains to be determined whether brain T4/T3 levels measured during foetal development or in the early postnatal weeks are most relevant to predict (specific types of) neurodevelopmental effects. Research work is also recommended to establish thresholds of altered brain T4/T3 that reflect dose levels at which the adaptive mechanisms of the highly versatile thyroid system are overwhelmed so that adverse neurodevelopmental effects will occur. Finally, while the studies included here determined brain levels of total T4 and total T3, brain levels of fT4 and fT3 might also be relevant; however, these may be exceptionally low and difficult to quantify.

4.4.7. Conclusions on the measurement of thyroid-related parameters in the offspring

Key messages: The findings from the present review underline that there is no single thyroid-related parameter that is decisive, on its own, for the assessment of substance-mediated thyroid hormone imbalance in rats or, ultimately, neurodevelopmental impairment. Expert knowledge is required for an overarching evaluation of the pattern of thyroid-related effects to determine if the adaptive mechanisms of the maternal and offspring thyroid system have been overwhelmed so that neurodevelopmental effects are likely to occur.

It is unsurprising that no single thyroid-related parameter is decisive on its own given the complexity of the interplay between different hormones of the thyroid system both in

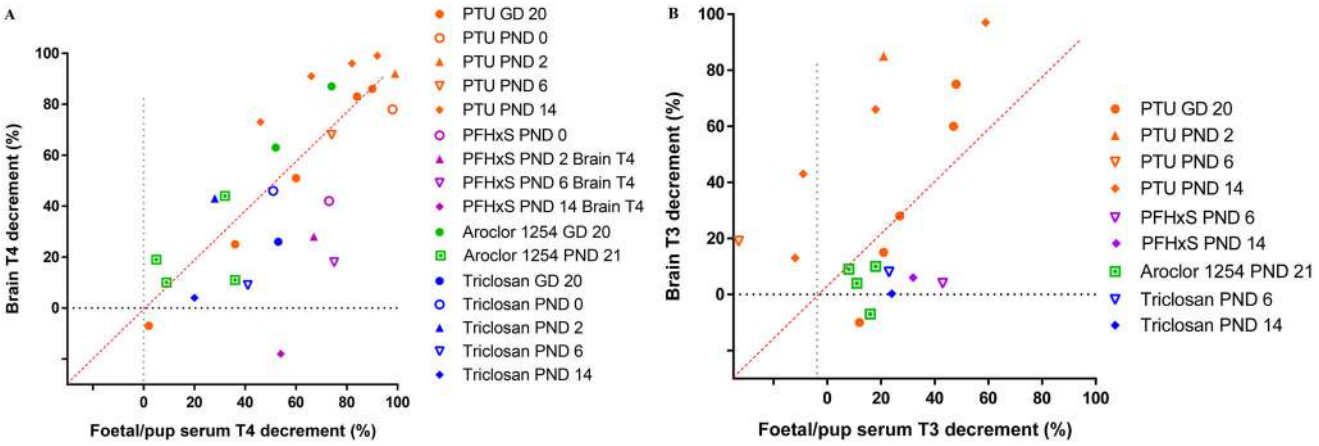


Figure 8. Relationship between offspring brain and serum thyroid hormone decrements.
(A) Relationship between offspring brain and serum T4 decrements (GD 20–PND 21).
(B) Relationship between offspring brain and serum T3 decrements (GD 20–PND 21).
GD: gestational day; PFHxS: perfluoro hexane sulphonates; PND: postnatal day; PTU: propylthiouracil; T3: triiodothyronine; T4: thyroxine.
These graphs include all (statistically significant or non-significant) data recorded at all dose levels from the studies that measured offspring serum T4 / T3 and offspring brain T4 / T3. All data expressed as percent of concurrent control group for the respective study. Values below the dotted vertical and/or horizontal lines represent increased T4 / T3 levels. The dotted red lines are arbitrary lines with a slope of 1 that was used as visual aid but not for any specific statistical analysis. Timepoints as designated in the respective studies. The day the dams were "sperm positive" was either defined as GD 1 or GD 0, and the birthdate was either defined as PND 0 or PND 1. Thus, between datapoints, timepoints ±1 day may represent the same stage of development.

the blood and in the target organs. A full assessment of maternal and offspring thyroid hormone status (i.e. serum levels of T4, T3, TSH, thyroid histopathology and brain levels of T4 and T3 during critical neurodevelopmental periods, possibly further considering serum and/or brain fT4/fT3 levels) may provide the best opportunity to understand the impact of thyroid perturbations in rat studies. Expert knowledge is required for an overarching evaluation of the pattern of thyroid-related effects (while also considering suitable timepoints of measurement) to determine if the findings rather indicate an adaptive reaction of the thyroid system or that such adaptive mechanisms have been overwhelmed so that adverse effects will develop in the target organ, which is the developing brain.

For a pragmatic evaluation of the database, the ECETOC T4 TF empirically set thresholds of ≥60%/≥50% offspring serum T4 reduction (for the top-/lower-dose groups) and of ≥20% and statistically significant offspring serum T3 reduction as possibly indicating that thyroid homeostasis was disrupted to an extent that adverse neurodevelopmental effects occurred. Interestingly, these thresholds were applicable across rat studies regardless of the substance's MoA or of the type of neurodevelopmental parameter considered. An exception is brain electrophysiology, which was generally more sensitive (see Section 4.3.5 for limitations of electrophysiological assessments). Further, if offspring serum T4 decrements attenuated during the exposure period, this was generally associated with the absence of neurodevelopmental effects.

In Table 15, the outcomes of the evaluation of both the offspring serum T4 and T3 thresholds are presented together. The joint evaluation of both thresholds yields consistent outcomes for 17 studies and inconsistent outcomes for six studies; thereof five studies are true negative or true positive following the T4 threshold but false positive or false negative following the T3 threshold. Even though this observation may indicate that offspring serum T4 levels are more closely associated with the occurrence of neurodevelopmental effects than offspring serum T3 levels, the database is too limited to suggest a data interpretation procedure for the joint evaluation of both thresholds. Similarly, this joint evaluation does not consider offspring serum TSH data. Based upon the present database, offspring serum TSH appears most useful to distinguish substances with direct thyroid-related MoA from substances with indirect thyroid-related MoA. By comparison, the observation that, from amongst the substances with direct thyroid-related MoA, the most pronounced offspring serum TSH increases (>400%) were observed for PTU and amitrole, which elicited brain-related effects, is founded on too few studies for the derivation of a threshold to predict the likelihood for neurodevelopmental effects.

Research work is recommended to follow up on the empirically set offspring serum T4 and T3 reduction thresholds to determine their biological relevance and hence usefulness for toxicity testing strategies. While there were some inconsistencies in the magnitude of thyroid hormone changes causing neurodevelopmental effects, these may be due to differences in selected methodologies, including differences in statistical power between study designs (which was not considered in the present review) and/or inherent variability in the measurement of both thyroid-related and neurodevelopmental parameters. Research work is also recommended to determine whether statistically significant and dose-dependent offspring TSH increases beyond ~400% indicate that exposure to substances with a direct thyroid-related MoA has overwhelmed the adaptive mechanisms of the thyroid system so that adverse neurodevelopmental effects will likely occur. Similarly, the usefulness of maternal/offspring serum fT4 and fT3 measurements remains to be established.

Importantly, information on toxicokinetics, which was not considered in this review, is relevant to determine the likelihood that a substance that causes thyroid hormone imbalance may have the potential to induce neurodevelopmental effects. Substance toxicokinetics may further impact the definition of sensitive windows and thus the identification of optimal timepoints of serum and brain hormone measurements. Ultimately, brain T4 and T3 levels, together with considerations on toxicokinetics, may prove most useful to predict a substance's potential to cause hypothyroid-mediated neurodevelopmental impairment.

Table 15. Joint evaluation of the offspring serum T3 and T4 thresholds.

Application of offspring serum T4 threshold	Application of offspring serum T3 threshold	Numbers of studies in which scenario was observed	Notes
True positive	True positive	6	Additionally: 2 studies in which T4 = true positive and T3 = not available
True negative	True negative	7	Additionally: 2 studies in which T4 = true negative and T3 = not available
False positive	False positive	2	This combination was observed for one PFHxS and one DE-71 study
False negative	False negative	2	These 2 studies were false negative on account of electrophysiological findings, but would have been true negative on account of traditional neurodevelopmental parameters Additionally: 2 studies in which T4 = false negative and T3 = not available
Number of studies with consistent outcomes		**17**	
True negative	False positive	3	This combination was observed for one perchlorate study (the other perchlorate study was 2x true negative), one triclosan study and one OMC study
True positive	False negative	2	This combination was observed for two Aroclor 1254 studies
False positive	True negative	1	This combination was observed for one DE-71 study
Number of studies with inconsistent outcomes		**6**	

T3: triiodothyronine; T4: thyroxine. See Table 14 for assignment of studies as true positive, false positive, true negative, false negative.
Light grey/dark grey shading: True positive/true negative following the thresholds for ≥60%/≥50% offspring serum T4 reduction in the top-/lower-dose groups or ≥20% and statistically significant offspring serum T3 reduction.

4.5. Other developmental manifestations of thyroid hormone imbalance

Key messages: The evaluation of neurodevelopmental findings should consider whether there is evidence for further developmental delays. To facilitate the interpretation of neurodevelopmental findings, offspring body weight should preferably be measured throughout lactation in the individual pups that are selected for the neurodevelopmental assessments.

This section briefly discusses further developmental manifestations of thyroid hormone imbalance (beyond neurodevelopmental impairment), which were not the focus of this review. Also, this review did not address the question whether substances might adversely affect neurodevelopment without affecting the thyroid system, a scenario that should always be considered in evaluating the overall outcome of developmental toxicity studies.

When interpreting developmental endpoints (e.g. anogenital distance, righting reflex, pinna detachment, eye opening, puberty onset), the body weight and maturity status of the offspring is considered. Pronounced maternal and offspring thyroid hormone imbalance may affect offspring growth. However, reductions in body weight and/or body weight gain may also be signs of systemic toxicity. Thus, any effects observed in the offspring, including neurodevelopmental effects, are potentially secondary to body weight effects, and it may be difficult to discern endocrine-mediated toxicity from systemic toxicity. Generally, thyroid-mediated effects on pup body weight occur with greater thyroid perturbations, but this is likely a continuum. Based upon the database considered here, absence of offspring body weight reduction always coincided with the absence of statistically significant neurodevelopmental effects (but not *vice versa*).

Important developmental manifestations of thyroid hormone imbalance beyond altered growth and neurodevelopmental impairment include developmental delays, disrupted thermoregulation, increased mortality and early eye opening.

Developmental delays are delays in the attainment of landmark structural or behavioural endpoints, and they may occur due to generalised, non-specific effects on development, e.g. altered growth. Delays in motor activity ontogeny, eye opening and incisor eruption have been reported in different PTU studies (Goldey et al. 1995a; Kobayashi et al. 2005; Axelstad et al. 2008), although these effects were non-specific and transient. When evaluating the outcomes of motor activity tests, it should be considered if any observed effects might also be secondary to developmental delays.

As regards disrupted thermoregulation, Johnstone et al. (2013) measured significant decreases in body temperature ($\leq$0.4 °C) in adult rats exposed to 1, 2, or 10 ppm PTU from GD 6 to LD 21. Sánchez-Huerta et al. (2015) also reported failure of PTU-treated adult rats to maintain body temperature when under anaesthesia, despite placement on a heating pad. Further, *in utero* and lactational exposure to 7 ppm PTU has been associated with altered arterial smooth muscle tone in two-week-old pups, presumably due to an alteration of the nitric oxide and Rho-kinase pathways (Sofronova et al. 2017), and this effect persisted in the adult offspring (Gaynullina et al. 2017).

Taken together, to facilitate the interpretation of neurodevelopmental findings, offspring body weight should preferably be measured throughout lactation and not just at the end of the study (e.g. on PND 21) or on the day animals were tested. Also, the body weight or body weight gain of the individual pups that are selected for testing should be measured and not only the average litter weight. The evaluation of neurodevelopmental findings should also consider whether there is evidence for further developmental delays, e.g. related to sexual maturation, onset of eyelid opening, pinnae detachment or the righting response.

5. Summary and conclusions revisiting the questions raised at the onset of the review

This third review prepared by the ECETOC T4 TF has served to establish which patterns of thyroid- and brain-related effects are seen in rats upon gestational and/or lactational exposure to substances causing thyroid hormone imbalance by different MoAs. Four case studies covered a total of 14 substances that (1) inhibit TPO; (2) cause iodine deficiency in the thyroid gland *via* NIS inhibition (as well as dietary iodine deficiency); (3) enhance thyroid hormone clearance; and (4) have other MoAs (and that trigger DIO1 inhibition). In collating a comprehensive database, all thyroid- and brain-related parameters included in 51 rat studies (published in ~60 papers) were considered. Most studies included substance exposure from GD 6 to 7 (i.e. after implantation is completed in the rat) up until the end of lactation. This exposure period is meaningful to determine acute gestational and/or lactational effects, and it covers the known windows of susceptibility to different neurodevelopmental effects. Based upon the findings from the four case studies, Section 5 answers the questions that were raised in Section 1.3.

Question 1: Which specific neurodevelopmental effects are observed in rat studies upon gestational and/or lactational exposure to substances that are known to cause thyroid hormone imbalance in rats via different MoAs?

Answer: Neurodevelopmental effects observed in the rat studies included altered motor activity, cognitive function, acoustic startle response and hearing function. Of these, the OECD TG 426 DNT study and the OECD TG 443 EOGRTS include motor activity and acoustic startle response as mandatory parameters, and the OECD TG 426 DNT study additionally cognitive function as well as crude assessments of hearing function. Further, the induction of periventricular heterotopia was assessed. While this parameter is currently not included in any TG, heterotopia may generally be observed during brain histopathology, which is mandatory following OECD TG 426 and 443. None of these parameters was superior to the others in identifying hypothyroid-mediated neurodevelopmental impairment, but tests for cognitive function were generally not the most sensitive. Also, it was not possible to establish patterns of brain-related effects—either for specific substances, or for specific MoAs or across case studies. However, even if such patterns of effects were present, it

would most likely not have been possible to observe them due to the heterogeneity of the study designs, especially with respect to types of neurodevelopmental parameters considered and timepoints of measurement.

For the time being, a spectrum of neurodevelopmental tests is recommended to assess potential for neurodevelopmental impairment. Brain electrophysiology appears more sensitive than the standard neurodevelopmental parameters, but it is (currently) not suitable for routine laboratory use. Similarly, measurements of gene expression changes appear sensitive, but research work is necessary to determine suitable sets of genes, which are linked to relevant phenotypic alterations, and to standardise the methodologies for regulatory use.

Overarching question to Questions 2a–2e: When are altered thyroid hormone levels indicative of potential concern for neurodevelopmental effects in rats (and hence indicative of the need for further testing)?

Question 2a: Is it sufficient to measure maternal serum thyroid hormone levels during gestation and/or lactation, or are foetal and/or pup serum thyroid hormone measurements necessary?

Answer: Maternal serum thyroid hormone levels are not sufficient to determine whether neurodevelopment will be impaired in rats. However, for substances that cause thyroid hormone imbalance, maternal serum thyroid hormone levels are relevant to determine the appropriateness of the selected dose range (see answer to Question 4). Measurement of maternal serum thyroid hormone might be most meaningful during gestation. Foetal and/or pup serum thyroid hormone measurements inform on a substance's potential to cause neurodevelopmental impairment (see answer to Question 2d).

Question 2b: Is there an added value of monitoring serum fT4, T3, fT3, and TSH in the dams and/or offspring, or is serum total T4 sufficient?

Answer: For maternal hormone measurements, see answer to Question 2a. With respect to offspring serum T3 and TSH, the present database indicates that such measurements have added value (see answer to Question 2d). Research work is recommended to enhance the reliability of serum T3 measurements. For offspring serum fT4/fT3, the database is too limited to conclude whether these parameters are useful or not.

The suitability of offspring T4, T3, and TSH data also depends on the timepoint of measurement. The findings from this review are not conclusive with respect to the best suitable timepoints (or numbers) of offspring serum T4, T3, and TSH measurements. Since offspring serum T4 decrements were often more pronounced during lactation than at its end, measurements at the end of lactation only (PND 21) may be insufficient. Additional measurements around PND 4, and possibly also on approximately PND 14–16, may be relevant. Research work is recommended to refine these observational findings from the present review.

Question 2c: Is there an added value in measuring brain thyroid hormone levels in foetuses or pups? If so, can a specific period be determined during which measurement of brain thyroid hormone is predictive of neurodevelopmental outcomes?

Answer: Based upon the present database, there may be an added value in measuring both brain T4 and brain T3. Due to substantial data gaps, it is currently not possible to determine the most relevant brain tissues or specific foetal and/or early postnatal periods during which measurement of brain thyroid hormone levels are predictive of (specific types of) neurodevelopmental outcomes. Brain T4 decrements were pronounced and brain T3 decrements significant in the PTU studies, which yielded neurodevelopmental findings, and in a perchlorate study, in which brain gene expression was the only non-thyroid parameter. By comparison, brain T4 and T3 levels were only non-significantly or transiently reduced in the studies investigating dietary iodine deficiency, PFHxS, Aroclor 1254 or triclosan, and in these studies, electrophysiological alterations or altered brain DIO2 were the only brain-related findings. Taken together, the present database is insufficient to suggest any threshold for brain T4 and T3 decrements to predict the likelihood for neurodevelopmental effects.

Question 2d: Is it possible to identify thresholds for specific thyroid-related parameters that are indicative of neurodevelopmental or other brain-related findings?

Answer: The database is most comprehensive for offspring serum T4. Based upon the present database, thresholds of $\geq$60%/$\geq$50% offspring serum T4 reduction in the top-/lower-dose groups may indicate an increased likelihood for statistically significant neurodevelopmental effects. These thresholds were applicable regardless of the substance's MoA or of the neurodevelopmental parameter addressed, except for electrophysiological alterations, which were often more sensitive. (This evaluation did not consider brain-related parameters that do not *per se* indicate neurodevelopmental impairment, such as gene expression changes.) Further, if offspring serum T4 decrements attenuate during the exposure period, this may indicate an increased likelihood for the absence of neurodevelopmental effects. Fewer data are available for offspring serum T3. A threshold of $\geq$20% and statistically significant offspring serum T3 reduction was tentatively set as possibly being predictive of neurodevelopmental effects. This threshold lies in the same range as the extent of T3 reduction that can reasonably be determined as statistically significant in standard toxicity studies. Finally, in studies addressing substances with a direct thyroid-related MoA, offspring serum TSH increases of $\geq$400% may indicate an increased likelihood for neurodevelopmental effects, but the present database is too limited to suggest this value as a threshold.

Based upon the present database, one single thyroid-related parameter is not sufficient to predict the likelihood for neurodevelopmental effects. However, if all thyroid-related parameters are considered together, this provides information on whether adaptive mechanisms of the highly versatile thyroid system have been induced or whether thyroid homeostasis has been disrupted to an extent that neurodevelopmental effects may occur. Expert knowledge is required for an overarching evaluation of the pattern of thyroid-related effects.

Question 2e: Are thyroid weight and thyroid histopathology in the dams and/or offspring sensitive parameters to indicate potential concern for neurodevelopmental effects?

Answer: The present database does not include a single study in which thyroid weight in either the dams or offspring was more sensitive than the corresponding serum T4 alteration, and it was often less sensitive. Measurements of thyroid weight generally show pronounced variability, questioning the relevance of thyroid weight data. Thyroid

histopathology was generally not more sensitive than the corresponding serum T4 measurements, regardless of the substance's MoA, and it was also often less sensitive. Nonetheless, thyroid histopathology appears useful to complement serum thyroid hormone measurements when assessing thyroid-active substances. Finally, increased thyroid weight and/or histopathological findings in the thyroid gland always coincided with increased serum TSH levels.

Question 3: Are similar changes in thyroid hormone magnitude during critical periods of development associated with the same neurodevelopmental outcomes in rats regardless of the MoA of the test compound? Is it possible to identify patterns of thyroid- and brain-related effects for substances with different MoAs?

Answer: In the present database, the empirically set offspring serum T4 and T3 reduction thresholds were generally applicable regardless of the substance's MoA or of neurodevelopmental parameter addressed. Overall, the observational findings indicate that the extent of thyroid hormone imbalance caused by the substance under investigation (whether this is an adaptive reaction to regain homeostasis or whether such adaptive processes have been overwhelmed) is more important than its MoA to predict the likelihood for

Table 16. Research recommendations resulting from the present review.

Overarching topic	Recommended research topic/research question	Section
General	Substantiate and refine the insight from this review to support the biological relevance of the observational findings and to robustly establish which thyroid-related parameters and threshold(s) is/are useful for a toxicity testing strategy to assess maternal/offspring thyroid hormone imbalance and, ultimately, the substance's potential to also cause neurodevelopmental impairment	4.4, 4.4.3
MoA: TPO inhibition	Can differences in the extent of hypothyroxinaemia caused by different TPO inhibitors (that determine whether neurodevelopmental effects will occur) be explained by, e.g. differences in toxicokinetics, the presence of further MoAs or other yet unknown reasons?	4.2.1.3
MoA: DIO inhibition	Under which conditions does DIO inhibition in the liver or brain lead to a reduction or an increase in serum and/or brain T4/T3 levels?	4.1
Substance-specific (PFHxS)	Do increased brain T4 levels on PND 14 reflect an adaptive reaction to PFHxS-mediated thyroid hormone imbalance? Is the lack of effect on brain T3 levels observed for PFHxS predictive of the absence of neurodevelopmental effects?	4.2.3.2
Offspring serum T4 and T3 measurement	Enhance understanding of best suitable timepoints for offspring serum thyroid hormone measurement; investigate whether information on attenuation *vs.* aggravation of offspring T4 decrements during exposure period provides added value to predict neurodevelopmental effects Follow up on the empirically set offspring serum T4 and T3 reduction thresholds to determine their usefulness for a toxicity testing strategy Enhance the reliability of serum T3 measurements	4.4.5, 4.4.7, 5
Offspring serum fT4/fT3 measurement	Establish the relevance of fT4/fT3 when assessing substances causing maternal/offspring thyroid hormone imbalance by different MoAs and ultimately possibly also effects on neurodevelopment Address technical difficulties in determining serum fT4 and fT3 levels in rodents	4.2.1.1, 4.4.4
Offspring serum TSH measurement	For substances with direct thyroid-related MoA: Do prolonged offspring serum TSH increases exceeding a certain level indicate that the HPT axis is "overwhelmed" so that adaptation for substance-mediated thyroid hormone imbalance no longer occurs?	4.2.1.1
Brain T4/brain T3	Expand the database on substance-mediated changes in brain T4 and T3 levels to enhance the understanding of the implications of different patterns of thyroid hormone imbalance on neurodevelopment; aim to establish thresholds of altered brain T4/T3 that reflect dose levels at which neurodevelopmental impairment will occur Establish the best suitable brain tissues and the best suitable timepoints of measurement of brain T4/T3 levels	4.2.3.3, 4.4.6
Brain-related assessments (general)	Enhance knowledge base on (patterns of) brain-related effects by collecting multiple neurodevelopment-related endpoints in the same study	4.3
Motor activity	Do different levels of thyroid hormone imbalance during lactation indicate whether altered motor activity is transient or persistent?	4.3.1
Cognitive function	Enhance understanding of sensitivity of the different test methods use to assess cognitive function	4.3.2
Periventricular heterotopia	Standardise methodology, encourage cross-laboratory validation of this endpoint in view of potential uptake into formal test guidelines	4.3.4
Brain gene expression	Determine suitable sets of genes to assess how thyroid hormone imbalance is linked to neurodevelopment impairment; link these sets of genes to relevant phenotypic alterations Standardise methodologies for regulatory use	4.2.1.2, 5
Sensitive neurodevelopmental endpoints	Establish whether more appropriate and sensitive sets of neurodevelopmental endpoints to be measured in rats are needed	Appendix II
Magnetic resonance imaging	Determine how this technique "*can be combined with functional assessments of the brain to enable comprehensive evaluations of child neurodevelopmental outcomes*" and establish its applicability for toxicological assessments using rodents (Sauer et al. 2020)	4.3
Toxicokinetics	Develop framework for the integration of toxicokinetics into assessments of substance-mediated thyroid hormone imbalance	4.4.5

DIO: deiodinase; fT3: free thyroxine; fT4: free triiodothyronine; HPT: hypothalamic–pituitary–thyroid; MoA: mode-of-action; PFHxS: perfluorohexane sulphonates; PND: postnatal day; T4: triiodothyronine; T4: thyroxine; TPO: thyroid peroxidase.

neurodevelopmental effects. For example, some TPO inhibitors (PTU, amitrole, MMI at very high doses) caused pronounced offspring serum T4 and T3 decrements and very high offspring serum TSH increases, which coincided with neurodevelopmental effects. For other TPO inhibitors (ETU, mancozeb, cyanamide, mercaptobenzimidazole, MMI in an EOGRTS-like study), offspring thyroid hormone imbalance at the maximum tolerated dose was much less pronounced and neurodevelopmental effects were not observed. Similarly, for some substances that enhance thyroid hormone clearance, neurodevelopmental effects were observed (Aroclor 1254, DE-71 in some studies), whereas none were observed for others (TBBPA, PFHxS, triclosan). Nonetheless, the present review also revealed patterns of effects for substances with different MoAs. For example, dose-dependently and statistically significantly increased offspring serum TSH was only observed for substances with a direct thyroid-related MoA.

Question 4: How should systemic toxicity (body weight effects, organ toxicity, developmental delays) be considered in rat studies in the evaluation of thyroid hormone imbalance and potential neurodevelopmental effects?

Answer: Measurements of maternal body weight (together with maternal serum T4 during gestation if the substance is known to cause thyroid hormone imbalance) inform on whether the maximum tolerated dose has been reached or exceeded in the respective study. Regarding offspring body weight, in the present database, statistically significant neurodevelopmental effects always coincided with reduced offspring body weight. Hence, absence of offspring body weight reductions might indicate a lower likelihood for neurodevelopmental impairment. Expert judgement is required to determine whether thyroid hormone imbalance or systemic toxicity are causative for offspring body weight reductions and (neuro)developmental impairment. The evaluation of neurodevelopmental findings should always consider whether there is evidence for further developmental delays, e.g. related to sexual maturation, onset of eyelid opening, pinnae detachment or the righting response.

In conclusion, the four case studies have provided important insight on how thyroid- and brain-related parameters are affected in rats upon *in utero* and/or lactational exposure to substances causing thyroid hormone imbalance in the exposed dams. However, it is important to note that this review neither addressed the human relevance of findings nor the sensitivity of the brain-related parameters included in the rat studies in predicting neurodevelopmental impairment in children.

Together with the findings from the first two reviews (Sauer et al. 2020; Marty et al. 2021), the insight from the present review will be used in the fourth and final review by the ECETOC T4 TF to draw up a tiered testing strategy to determine whether *in utero* and/or lactational substance exposure may elicit maternal and/or offspring thyroid hormone imbalance and potentially also neurodevelopmental effects. This testing strategy will also address how the human relevance of findings observed in rat studies may be established in line with the European Commission (2017b, 2018) Endocrine Disruptor Criteria and the EFSA and ECHA (2018) Endocrine Disruptor Guidance. Finally, the fourth review prepared by the ECETOC T4 TF will further explore how knowledge gaps need to be addressed to enhance the reliability and relevance of the planned testing strategy (see Table 16 for research recommendations resulting from the present review).

Acknowledgements

We would like to thank the members of the ECETOC Scientific Committee as well as the in-house reviewers (see Declaration of interest) for their critical comments. Also, we would like to thank all members of the ECETOC T4 Task Force for valuable discussions that contributed to the basis for this review. We are indebted to Maryam Bourja, Andreea Cuciureanu, Andrea Salvadori, and Virginie van der Steeg as well as the entire ECETOC secretariat for organisational and technical assistance to the ECETOC T4 Task Force. We thank the reviewers selected by the Editor that were anonymous to us. The comprehensive and helpful comments served to improve the manuscript.

Supplemental material

Supplemental data for this article is available online at https://doi.org/10.1080/08923973.2022.2078727.

Declaration of interest

This manuscript relates to work undertaken by the European Centre for Ecotoxicology and Toxicology of Chemicals T4 Task Force. ECETOC (www.ecetoc.org) is a scientific organisation which provides a collaborative space for scientists from industry, academia and governments. Its mission is to develop and promote practical, trusted and sustainable solutions to scientific challenges which are valuable to industry, as well as to the regulatory community and society in general. ECETOC is financed by its membership, which are the leading companies with interests in the manufacture and use of chemicals, biomaterials and pharmaceuticals (http://www.ecetoc.org/ecetoc-membership/member-companies/). ECETOC Task Force members work within their regular working hours but do not receive compensation by ECETOC.

The co-authors of this manuscript consist of Task Force members (SM, AC, RG, DG, NH, BRH, SJ, SMK, HAM, LPS, DU), Stewards from the ECETOC Scientific Committee (PAB, BvR,) and the scientific writer (UGS). The views expressed in this article are solely those of the co-authors and may not represent those of the sponsoring organisations. None of the co-authors received funding in cash or kind for their contribution to this manuscript, with the exception of UGS, a freelance scientific writer.

Sue Marty (SM) is employed by The Dow Chemical Company. The issues of hazard identification and risk assessment of thyroid-active compounds, and how these are assessed by regulatory/other agencies, impact substances of interest to the corporation. An in-house review of this manuscript yielded only few minor requests for changes that were either editorial or served clarification. SM's role at Dow is focussed on Dow's science strategy and testing, which includes the assessment of endocrine-active compounds.

Ursula G. Sauer (UGS), a freelance scientific writer, was hired by ECETOC to assist in the preparation of this manuscript.

Alex Charlton (AC), Daniel Urbisch (DU; since March 2022) and Philip A. Botham (PAB) are employed by Syngenta, an international agribusiness that markets crop protection chemicals and seeds. The Syngenta portfolio includes substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity. An in-house review of this manuscript yielded no requests for changes. AC's responsibilities within Syngenta include providing scientific support to research and development activities and to regulatory toxicology projects. DU is a regulatory toxicologist for agrochemicals with global responsibilities. PAB's responsibilities within Syngenta are to provide strategic scientific advice on product safety issues to the company's Product Safety, Business Sustainability and Crop Protection Development organisations.

Rashin Ghaffari (RG) is employed by Corteva Agriscience, USA, and Bethany R. Hannas (BRH) was employed by Corteva Agriscience until mid-October 2021. Corteva markets products (or previously marketed products)

containing some of the chemicals included in this paper. Further, the Corteva portfolio includes substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity. An in-house review of this manuscript yielded no requests for changes. RG's responsibilities within Corteva include providing scientific support to developmental and reproductive toxicology testing strategy and to research and development projects. BRH led the Corteva Agriscience Haskell R&D Centre DART & Endocrine Group.

BRH has been employed by Eli Lilly & Co. (USA) since November 2021. Eli Lilly & Co. develops and markets a wide range of pharmaceutical compounds. BRH's responsibilities at Eli Lilly & Co. include providing subject matter expertise in developmental and reproductive toxicology and project leadership for nonclinical safety testing.

Davy Guignard (DG), Nina Hallmark (NH) and Larry P. Sheets (LPS) are employed by the Crop Science Division of Bayer AG. The Bayer portfolio includes substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity. An in-house review of this manuscript yielded no requests for changes. NH's responsibilities include being the Bayer Crop Science Division Environment Safety Ecotoxicology Terrestrial Vertebrates Team Leader. Up until January 2020, NH was Chair of the ECETOC T4 Task Force. LPS's responsibilities include being Senior Fellow Regulatory Toxicology at the Bayer Crop Science Division. DG's responsibilities include being Research Toxicologist at the Bayer Crop Science Division.

SJ was employed by Albemarle Europe SRL until her retirement on 1 July 2022. The Albemarle portfolio includes substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity. SJ's responsibilities within Albemarle included heading the corporate toxicology department of Albemarle Corporation, worldwide management of regulatory toxicology and risk assessment of the chemicals produced by Albemarle Corporation.

Stephanie Melching-Kollmuss (SMK) and Heike-Antje Marxfeld (HAM) are employed by BASF SE, Limburgerhof, Germany, and BASF SE, Ludwigshafen, Germany, respectively. Bennard van Ravenzwaay (BvR) was Senior Vice President of the BASF SE Experimental Toxicology and Ecology Department (which is independent of any business unit and ISO 17020 certified) up until the end of 2021. DU was employee at BASF, Limburgerhof, Germany, until February 2022. BASF produces a very wide range of chemicals including some of those mentioned in this paper and/or substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity. An in-house review yielded only one minor request for change that served clarification. SMK's responsibilities within BASF SE include being regulatory toxicologist for agrochemicals. SMK is the current Chair of the ECETOC T4 Task Force. HAM's responsibilities within BASF include pathological evaluation and assessment of a broad variety of regulatory studies conducted by BASF.

BvR has been an independent consultant (Environmental Sciences Consulting) since the beginning of 2022. BvR is an Associate Professor of Reproduction Toxicity of the University of Wageningen, Netherlands, and the Chairman of the ECETOC Scientific Committee.

Finally, this manuscript was reviewed by the ECETOC Scientific Committee consisting of representatives of academia and industry (http://www.ecetoc.org/about-ecetoc/scientific-committee/). This review yielded few minor requests for amendment that served clarification.

ORCID

M. Sue Marty http://orcid.org/0000-0002-5404-8515
Rashin Ghaffari http://orcid.org/0000-0002-3209-9549
Davy Guignard http://orcid.org/0000-0002-3882-8925
Nina Hallmark http://orcid.org/0000-0002-4210-628X
Heike-Antje Marxfeld http://orcid.org/0000-0001-5390-0633
Bennard van Ravenzwaay http://orcid.org/0000-0002-4264-9217

References

Akaike M, Kato N, Ohno H, Kobayashi T. 1991. Hyperactivity and spatial maze learning impairment of adult rats with temporary neonatal hypothyroidism. Neurotoxicol Teratol. 13(3):317–322.

Ali JM, Roberts SM, Gordon DS, Stuchal LD. 2019. Derivation of a chronic reference dose for perfluorohexane sulfonate (PFHxS) for reproductive toxicity in mice. Regul Toxicol Pharmacol. 108:104452.

Andersen SL, Andersen S, Vestergaard P, Olsen J. 2018. Maternal thyroid function in early pregnancy and child neurodevelopmental disorders: a Danish nationwide case-cohort Study. Thyroid. 28(4):537–546.

Ausó E, Lavado-Autric R, Cuevas E, Rey del FD, Morreale de Escobar G, Berbel P. 2004. A moderate and transient deficiency of maternal thyroid function at the beginning of fetal neocorticogenesis alters neuronal migration. Endocrinol. 145(9):4037–4047.

Axelstad M, Hansen PR, Boberg J, Bonnichsen M, Nellemann C, Lund SP, Hougaard KS, Hass U. 2008. Developmental neurotoxicity of propylthiouracil (PTU) in rats: Relationship between transient hypothyroxinemia during development and long-lasting behavioural and functional changes. Toxicol Appl Pharmacol. 232(1):1–13.

Axelstad M, Boberg J, Nellemann C, Kiersgaard M, Jacobsen PR, Christiansen S, Hougaard KS, Hass U. 2011a. Exposure to the widely used fungicide Mancozeb causes thyroid hormone disruption in rat dams but no behavioral effects in the offspring. Toxicol Sci. 120(2):439–446.

Axelstad M, Boberg J, Hougaard KS, Christiansen S, Jacobsen PR, Mandrup KR, Nellemann C, Lund SP, Hass U. 2011b. Effects of pre- and postnatal exposure to the UV-filter octyl methoxycinnamate (OMC) on the reproductive, auditory and neurological development of rat offspring. Toxicol Appl Pharmacol. 250(3):278–290.

Bansal R, Tighe D, Danai A, Rawn DF, Gaertner DW, Arnold DL, Gilbert ME, Zoeller RT. 2014. Polybrominated diphenyl ether (DE-71) interferes with thyroid hormone action independent of effects on circulating levels of thyroid hormone in male rats. Endocrinol. 155(10):4104–4112.

Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB. 2012. A developmental and genetic classification for malformations of cortical development: update 2012. Brain. 135(Pt 5):1348–1369.

Barter RA, Klaassen CD. 1994. Reduction of thyroid hormone levels and alteration of thyroid function by four representative UDP-glucuronosyltransferase inducers in rats. Toxicol Appl Pharmacol. 128(1):9–17.

Bastian TW, Prohaska JR, Georgieff MK, Anderson GW. 2014. Fetal and neonatal iron deficiency exacerbates mild thyroid hormone insufficiency effects on male thyroid hormone levels and brain thyroid hormone-responsive gene expression. Endocrinol. 155(3):1157–1167.

Beetstra JB, van Engelen JG, Karels P, van der Hoek HJ, de Jong M, Docter R, Krenning EP, Hennemann G, Brouwer A, Visser TJ. 1991. Thyroxine and 3,3′5-triiodothyronine are glucuronidated in rat liver by different uridine diphosphate-glucuronyltransferases. Endocrinol. 128(2):741–746.

Bianco AC, Salvatore D, Gereben B, Berry MJ, Larsen PR. 2002. Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. Endocr Rev. 23(1):38–89.

Bowers WJ, Wall PM, Nakai JS, Yagminas A, Wade M, Li N. 2015. Behavioral and thyroid effects of *in utero* and lactational exposure of Sprague-Dawley rats to the polybrominated diphenyl ether mixture DE71. Neurotoxicol Teratol. 52(Pt B):127–142.

Boyes WK, Degn L, George BJ, Gilbert ME. 2018. Moderate perinatal thyroid hormone insufficiency alters visual system function in adult rats. Neurotoxicol. 67:73–83.

Braga M, Cooper DS. 2001. Clinical review 129: oral cholecystographic agents and the thyroid. J Clin Endocrinol Metab. 86(5):1853–1860.

Buesen R, Chorley BN, da Silva Lima B, Daston G, Deferme L, Ebbels T, Gant TW, Goetz A, Greally J, Gribaldo L, et al. 2017. Applying 'omics' technologies in chemicals risk assessment: Report of an ECETOC workshop. Regul Toxicol Pharmacol. 91 Suppl 1(1):S3–S13.

Chang SC, Thibodeaux JR, Eastvold ML, Ehresman DJ, Bjork JA, Froehlich JW, Lau C, Singh RJ, Wallace KB, Butenhoff JL. 2008. Thyroid hormone status and pituitary function in adult rats given oral doses of perfluorooctanesulfonate (PFOS). Toxicol. 243(3):330–339.

Chi Q, Zhang W, Wang L, Huang J, Yuan M, Xiao H, Wang X. 2020. Evaluation of structurally different brominated flame retardants interacting with the transthyretin and their toxicity on HepG2 cells. Chemosphere. 246:125749.

Choi JS, Lee YJ, Kim TH, Lim HJ, Ahn MY, Kwack SJ, Kang TS, Park KL, Lee J, Kim ND, et al. 2011. Molecular mechanism of tetrabromobisphenol

A (TBBPA)-induced target organ toxicity in Sprague-Dawley male rats. Toxicol Res. 27(2):61–70.
Clarac F, Brocard F, Vinay L. 2004. The maturation of locomotor networks. Prog Brain Res. 143:57–66.
Clewell RA, Merrill EA, Yu KO, Mahle DA, Sterner TR, Mattie DR, Robinson PJ, Fisher JW, Gearhart JM. 2003. Predicting fetal perchlorate dose and inhibition of iodide kinetics during gestation: a physiologically-based pharmacokinetic analysis of perchlorate and iodide kinetics in the rat. Toxicol Sci. 73(2):235–255.
Clewell RA, Merrill EA, Gearhart JM, Robinson PJ, Sterner TR, Mattie DR, Clewell HJ. 2007. Perchlorate and radioiodide kinetics across life stages in the human: using PBPK models to predict dosimetry and thyroid inhibition and sensitive subpopulations based on developmental stage. J Toxicol Environ Health A. 70(5):408–428.
Cope RB, Kacew S, Dourson M. 2015. A reproductive, developmental and neurobehavioral study following oral exposure of tetrabromobisphenol A on Sprague-Dawley rats. Toxicol. 329:49–59.
Costa LG, Giordano G. 2007. Developmental neurotoxicity of polybrominated diphenyl ether (PBDE) flame retardants. Neurotoxicol. 28(6): 1047–1067.
Crofton KM. 2004. Developmental disruption of thyroid hormone: correlations with hearing dysfunction in rats. Risk Anal. 24(6):1665–1671.
Crofton KM, Sheets LP. 1989. Reflex modification of the startle response. J Am Coll Toxicol. 8(1):199–211.
Crofton KM, Howard JL, Moser VC, Gill MW, Reiter LW, Tilson HA, MacPhail RC. 1991. Interlaboratory comparison of motor activity experiments: implications for neurotoxicological assessments. Neurotoxicol Teratol. 13(6):599–609.
Crofton KM, Ding D-L, Padich R, Taylor M, Henderson D. 2000a. Hearing loss following exposure during development to polychlorinated biphenyls: A cochlear site of action. Hear Res. 144(1-2):196–204.
Crofton KM, Kodavanti PR, Derr-Yellin EC, Casey AC, Kehn LS. 2000b. PCBs, thyroid hormones, and ototoxicity in rats: cross-fostering experiments demonstrate the impact of postnatal lactation exposure. Toxicol Sci. 57(1):131–140.
Crofton KM, Gilbert M, Friedman KP, Demeneix B, Marty MS, Zoeller RT. 2019. Adverse outcome pathway on inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes in mammals. In: OECD series on adverse outcome pathways no. 13. Paris: OECD.
Darbra S, Garau A, Balada F, Sala J, Martí-Carbonell MA. 2003. Perinatal hypothyroidism effects on neuromotor competence, novelty-directed exploratory and anxiety-related behavior and learning in rats. Behav Brain Res. 143(2):209–215.
Darras VM, van Herck SL. 2012. Iodothyronine deiodinase structure and function: from ascidians to humans. J Endocrinol. 215(2):189–206.
Davenport JW, Gonzalez LM, Hennies RS, Hagquist WM. 1976. Severity and timing of early thyroid deficiency as factors in the induction of learning disorders in rats. Horm Behav. 7(2):139–157.
Davidson B, Soodak M, Neary JT, Strout HV, Kieffer JD, Mover H, Maloof F. 1978. The irreversible inactivation of thyroid peroxidase by methylmercaptoimidazole, thiouracil, and propylthiouracil *in vitro* and its relationship to *in vivo* findings. Endocrinol. 103(3):871–882.
De Groef B, Decallonne BR, Van der Geyten S, Darras VM, Bouillon R. 2006. Perchlorate versus other environmental sodium/iodide symporter inhibitors: potential thyroid-related health effects. Eur J Endocrinol. 155(1):17–25.
de-Miranda AS, Kuriyama SN, da-Silva CS, do-Nascimento MS, Parente TE, Paumgartten FJ. 2016. Thyroid hormone disruption and cognitive impairment in rats exposed to PBDE during postnatal development. Reprod Toxicol. 63:114–124.
Deol MS. 1973. An experimental approach to the understanding and treatment of hereditary syndromes with congenital deafness and hypothyroidism. J Med Genet. 10(3):235–242.
Doerge DR. 1986. Mechanism-based inhibition of lactoperoxidase by thiocarbamide goitrogens. Biochemistry. 25(16):4724–4728.
Doerge DR, Takazawa RS. 1990. Mechanism of thyroid peroxidase inhibition by ethylenethiourea. Chem Res Toxicol. 3(2):98–101.
Dohán O, Portulano C, Basquin C, Reyna-Neyra A, Amzel LM, Carrasco N. 2007. The Na^+/I^- symporter (NIS) mediates electroneutral active transport of the environmental pollutant perchlorate. Proc Natl Acad Sci USA. 104(51):20250–20255.
Eayrs JT. 1961. Age as a factor determining the severity and reversibility of the effects of thyroid deprivation in the rat. J Endocrinol. 22: 409–419.
[EFSA] European Food Safety Authority. 2011a. EFSA scientific opinion. Statistical significance and biological relevance. EFSA J. 9(9):2372.
[EFSA] European Food Safety Authority. 2011b. EFSA panel on contaminants in the food chain (CONTAM). Scientific opinion on tetrabromobisphenol A (TBBPA) and its derivatives in food. EFSA J. 9(12):2477.
[EFSA] European Food Safety Authority. 2017. EFSA scientific opinion. Guidance on the assessment of biological relevance in scientific assessments. EFSA J. 15(8):e04970.
[EFSA] European Food Safety Authority. 2020a. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology. EFSA. Supporting Publication 2020: EN-1837.
[EFSA] European Food Safety Authority. 2020b. Cumulative dietary risk characterisation of pesticides that have chronic effects on the thyroid. EFSA J. 18(4):6088.
[EFSA, ECHA] European Food Safety Authority and European Chemicals Agency with the technical support of the Joint Research Centre. 2018. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA J. 16: 1661.
EP and Council. 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L. 309:1–50.
EP and Council. 2011. Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment (recast). OJ L. 174:88.
EP and Council. 2012. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. OJ L. 167:1.
European Commission. 2017a. Supporting the organisation of a workshop on thyroid disruption – final report. Framework Contract ENV.A.3/FRA/2014/0029 on implementation of the Community strategy on endocrine disruptors. Written by: Brunel University London and DTU National Food Institute Denmark; 165 pages.
European Commission. 2017b. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. OJ Eur L. 60:1–12.
European Commission. 2018. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out criteria for the determination of endocrine disrupting properties. OJ Eur L. 101:33–36.
Fegert I, Billington R, Botham P, Carney E, FitzGerald E, Hanley T, Lewis R, Marty MS, Schneider S, Sheets LP, et al. 2012. Feasibility of the extended one-generation reproductive toxicity study (OECD 443). Reprod Toxicol. 34(3):331–339.
Freeman S, Cherny L, Sohmer H. 1996. Thyroxine affects physiological and morphological development of the ear. Hear Res. 97(1–2):19–29.
Freyberger A, Ahr H-J. 2006. Studies on the goitrogenic mechanism of action of *N*,*N*,*N*′,*N*′-tetramethylthiourea. Toxicology. 217(2–3):169–175.
Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. Toxicol Pathol. 44(1):14–42.
Gaynullina DK, Sofronova SI, Selivanova EK, Shvetsova AA, Borzykh AA, Sharova AP, Kostyunina DS, Martyanov AA, Tarasova OS. 2017. NO-mediated anticontractile effect of the endothelium is abolished in coronary arteries of adult rats with antenatal/early postnatal hypothyroidism. Nitric Oxide. 63:21–28.
Gilbert ME, Sui L. 2006. Dose-dependent reductions in spatial learning and synaptic function in the dentate gyrus of adult rats following developmental thyroid hormone insufficiency. Brain Res. 1069(1): 10–22.

Gilbert ME, Sui L, Walker MJ, Anderson W, Thomas S, Smoller SN, Schon JP, Phani S, Goodman JH. 2007. Thyroid hormone insufficiency during brain development reduces parvalbumin immunoreactivity and inhibitory function in the hippocampus. Endocrinol. 148(1):92–102.

Gilbert ME, Sui L. 2008. Developmental exposure to perchlorate alters synaptic transmission in hippocampus of the adult rat. Environ Health Perspect. 116(6):752–760.

Gilbert ME. 2011. Impact of low-level thyroid hormone disruption induced by propylthiouracil on brain development and function. Toxicol Sci. 124(2):432–445.

Gilbert ME, McLanahan ED, Hedge J, Crofton KM, Fisher JW, Valentin-Blasini L, Blount BC. 2011. Marginal iodide deficiency and thyroid function: dose-response analysis for quantitative pharmacokinetic modelling. Toxicology. 283(1):41–48.

Gilbert ME, Rovet J, Chen Z, Koibuchi N. 2012. Developmental thyroid hormone disruption: prevalence, environmental contaminants and neurodevelopmental consequences. Neurotoxicology. 33(4):842–852.

Gilbert ME, Hedge JM, Valentin-Blasini L, Blount BC, Kannan K, Tietge J, Zoeller RT, Crofton KM, Jarrett JM, Fisher JW. 2013. An animal model of marginal iodine deficiency during development: the thyroid axis and neurodevelopmental outcome. Toxicol Sci. 132(1):177–195.

Gilbert ME, Ramos RL, McCloskey DP, Goodman JH. 2014. Subcortical band heterotopia in rat offspring following maternal hypothyroxinaemia: structural and functional characteristics. J Neuroendocrinol. 26(8):528–541.

Gilbert ME, Sanchez-Huerta K, Wood C. 2016. Mild thyroid hormone insufficiency during development compromises activity-dependent neuroplasticity in the hippocampus of adult male rats. Endocrinol. 157(2):774–787.

Gilbert ME, Goodman JH, Gomez J, Johnstone AFM, Ramos RL. 2017. Adult hippocampal neurogenesis is impaired by transient and moderate developmental thyroid hormone disruption. Neurotoxicol. 59:9–21.

Gilbert ME, O'Shaughnessy KL, Axelstad M. 2020. Regulation of thyroid-disrupting chemicals to protect the developing brain. Endocrinology. 161(10):bqaa106.

Gilbert ME, O'Shaughnessy KL, Thomas SE, Riutta C, Wood CR, Smith A, Oshiro WO, Ford RL, Hotchkiss M, Hassan I. 2021. Thyroid disruptors: extrathyroidal sites of chemical action and neurodevelopmental outcome – an examination using triclosan and perfluorohexane sulfonate (PFHxS). Toxicol Sci. 16:kfab080.

Gilbert ME, Hassan I, Wood C, O'Shaughnessy KL, Spring S, Thomas S, Ford J. 2022. Gestational exposure to perchlorate in the rat: thyroid hormones in fetal thyroid gland, serum, and brain. Toxicol Sci. 188(1):117–130.

Gill S, Hou Y, Li N, Pulido O, Bowers W. 2016. Developmental neurotoxicity of polybrominated diphenyl ethers mixture DE71 in Sprague-Dawley rats. J Toxicol Environ Health A. 79(12):482–493.

Goldey ES, Crofton KM. 1998. Thyroxine replacement attenuates hypothyroxinemia, hearing loss and motor deficits following developmental exposure to Aroclor 1254 in rats. Toxicol Sci. 45(1):94–105.

Goldey ES, Kehn LS, Rehnberg GL, Crofton KM. 1995a. Effects of developmental hypothyroidism on auditory and motor function in the rat. Toxicol Appl Pharmacol. 135(1):67–76.

Goldey ES, Kehn LS, Lau C, Rehnberg GL, Crofton KM. 1995b. Developmental exposure to polychlorinated biphenyls (Aroclor 1254) reduces circulating thyroid hormone concentrations and causes hearing deficits in rats. Toxicol Appl Pharmacol. 135(1):77–88.

Goodman JH, Gilbert ME. 2007. Modest thyroid hormone insufficiency during development induces a cellular malformation in the *Corpus callosum*: A model for cortical dysplasia. Endocrinol. 148(6):2593–2597.

Goswami A, Rosenberg IN. 1986. Iodothyronine 5′-deiodinase in brown adipose tissue: thiol activation and propylthiouracil inhibition. Endocrinol. 119(2):916–923.

Grijota-Martinez C, Diez D, Morreale de Escobar G, Bernal J, Morte B. 2011. Lack of action of exogenously administered T3 on the fetal rat brain despite expression of the monocarboxylate transporter 8. Endocrinology. 152(4):1713–1721.

Guerrini R, Parrini E. 2010. Neuronal migration disorders. Neurobiol Dis. 38(2):154–166.

Hallgren S, Darnerud PO. 2002. Polybrominated diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs) and chlorinated paraffins (CPs) in rats – testing interactions and mechanisms for thyroid hormone effects. Toxicology. 177(2–3):227–243.

Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Kester MH, Andersson PL, Legler J, Brouwer A. 2006. *In vitro* profiling of the endocrine-disrupting potency of brominated flame retardants. Toxicol Sci. 92(1):157–173.

Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Visser TJ, Van Velzen MJ, Brouwer A, Bergman A. 2008. Biotransformation of brominated flame retardants into potentially endocrine-disrupting metabolites, with special attention to 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47). Mol Nutr Food Res. 52(2):284–298.

Hamers T, Kortenkamp A, Scholze M, Molenaar D, Cenijn PH, Weiss JM. 2020. Transthyretin-binding activity of complex mixtures representing the composition of thyroid-hormone disrupting contaminants in house dust and human serum. Environ Health Perspect. 128(1):17015.

Hasegawa M, Kida I, Wada H. 2010. A volumetric analysis of the brain and hippocampus of rats rendered perinatal hypothyroid. Neurosci Lett. 479(3):240–244.

Hassan I, El-Masri H, Kosian PA, Ford J, Degitz SJ, Gilbert ME. 2017. Neurodevelopment and thyroid hormone synthesis inhibition in the rat: quantitative understanding within the adverse outcome pathway framework. Toxicol Sci. 160(1):57–73.

Hasebe M, Matsumoto I, Imagawa T, Uehara M. 2008. Effects of an antithyroid drug, methimazole, administration to rat dams on the cerebellar cortex development in their pups. Int J Dev Neurosci. 26(5):409–414.

Hebert R, Langlois JM, Dussault JH. 1985. Permanent defects in rat peripheral auditory function following perinatal hypothyroidism: determination of a critical period. Dev Brain Res. 23(2):161–170.

Henrichs J, Bongers-Schokking JJ, Schenk JJ, Ghassabian A, Schmidt HG, Visser TJ, Hooijkaas H, de Muinck Keizer-Schrama SMPF, Hofman A, Jaddoe VVW, et al. 2010. Maternal thyroid function during early pregnancy and cognitive functioning in early childhood: the Generation R Study. J Clin Endocrinol Metab. 95(9):4227–4234.

Hood A, Liu YP, Gattone VI, Klaassen CD. 1999. Sensitivity of thyroid gland growth to thyroid stimulating hormone (TSH) in rats treated with antithyroid drugs. Toxicol Sci. 49(2):263–271.

Huget-Penner S, Feig DS. 2020. Maternal thyroid disease and its effects on the fetus and perinatal outcomes. Prenat Diagn. 40(9):1077–1084.

Huisinga M, Bertrand L, Chamanza R, Damiani I, Engelhardt J, Francke S, Freyberger A, Harada T, Harleman J, Kaufmann W, et al. 2020. Adversity considerations for thyroid follicular cell hypertrophy and hyperplasia in nonclinical toxicity studies: results from the 6th ESTP International Expert Workshop. Toxicol Pathol. 48(8):920–938.

Johnstone AFM, Gilbert ME, Aydin C, Grace CE, Hasegawa M, Gordon CJ. 2013. Thermoregulatory deficits in adult Long Evans rat exposed perinatally to the antithyroid drug, propylthiouracil. Neurotoxicol Teratol. 39:1–8.

Julvez J, Alvarez-Pedrerol M, Rebagliato M, Murcia M, Forns J, Garcia-Esteban R, Lertxundi N, Espada M, Tardón AR, Galán I, et al. 2013. Thyroxine levels during pregnancy in healthy women and early child neurodevelopment. Epidemiology. 24(1):150–157.

Kaindl AM, Koppelstaetter A, Nebrich G, Stuwe J, SIfringer M, Zabel C, Klose J, Ikonomidou C. 2008. Brief alteration of NMDA or GABAA receptor-mediated neurotransmission has long term effects on the developing cerebral cortex. Mol Cell Proteomics. 7(12):2293–2310.

Kaufmann W, Gröters S. 2006. Developmental neuropathology in DNT-studies – a sensitive tool for the detection and characterization of developmental neurotoxicants. Reprod Toxicol. 22(2):196–213.

Kerstner-Wood C, Coward L, Gorman G. 2003. Protein binding of perfluorobutane sulfonate, perfluorohexane sulfonate, perfluorooctane sulfonate and perfluorooctanoate to plasma (human, rat and monkey), and various human-derived plasma protein fractions. US EPA docket AR-226–1354. Washington (DC): US Environmental Protection Agency.

Kitamura S, Kato T, Iida M, Jinno N, Suzuki T, Ohta S, Fujimoto N, Hanada H, Kashiwagi K, Kashiwagi A. 2005. Anti-thyroid hormonal activity of tetrabromobisphenol A, a flame retardant, and related compounds: Affinity to the mammalian thyroid hormone receptor, and effect on tadpole metamorphosis. Life Sci. 76(14):1589–1601.

Klaassen CD, Hood AM. 2001. Effects of microsomal enzyme inducers on thyroid follicular cell proliferation and thyroid hormone metabolism. Toxicol Pathol. 29(1):34–40.

Klammer H, Schlecht C, Wuttke W, Schmutzler C, Gotthardt I, Köhrle J, Jarry H. 2007. Effects of a 5-day treatment with the UV-filter octylmethoxycinnamate (OMC) on the function of the hypothalamo-pituitary-thyroid function in rats. Toxicol. 238(2–3):192–199.

Knapen D, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, Munn S, O'Brien JM, Pollesch N, Smith LC, et al. 2018. Adverse outcome pathway networks I: development and applications. Environ Toxicol Chem. 37(6):1723–1733.

Kobayashi K, Akune H, Sumida K, Saito K, Yoshioka T, Tsuji R. 2009. Perinatal exposure to PTU decreases expression of Arc, Homer 1, Egr 1 and Kcna 1 in the rat cerebral cortex and hippocampus. Brain Res. 1264:24–32.

Kobayashi K, Tsuji R, Yoshioka T, Kushida M, Yabushita S, Sasaki M, Mino T, Seki T. 2005. Effects of hypothyroidism induced by perinatal exposure to PTU on rat behaviour and synaptic gene expression. Toxicology. 212(2–3):135–147.

Kodavanti PR, Coburn CG, Moser VC, MacPhail RC, Fenton SE, Stoker TE, Rayner JL, Kannan K, Birnbaum LS. 2010. Developmental exposure to a commercial PBDE mixture, DE-71: neurobehavioral, hormonal, and reproductive effects. Toxicol Sci. 116(1):297–312.

Korevaar TIM, Muetzel R, Medici M, Chaker L, Jaddoe VWV, de Rijke YB, Steegers EAP, Visser TJ, White T, Tiemeier H, et al. 2016. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. Lancet Diabetes Endocrinol. 4(1):35–43.

Kortenkamp A, Axelstad M, Baig AH, Bergman A, Bornehag C-G, Cenijn P, Christiansen S, Demeneix B, Derakhshan A, Fini JB, et al. 2020. Removing critical gaps in chemical test methods by developing new assays for the identification of thyroid hormone system – disrupting chemicals – the ATHENA project. IJMS. 21(9):3123.

Kronborg TM, Hansen JF, Rasmussen ÅK, Vorkamp K, Nielsen CH, Frederiksen M, Hofman-Bang J, Hahn CH, Ramhøj L, Feldt-Rasmussen U. 2017. The flame retardant DE-71 (a mixture of polybrominated diphenyl ethers) inhibits human differentiated thyroid cell function *in vitro*. PLOS ONE. 12(6):e0179858.

Lai DY, Kacew S, Dekant W. 2015. Tetrabromobisphenol A (TBBPA): possible modes of action of toxicity and carcinogenicity in rodents. Food Chem Toxicol. 80:206–214.

Lain SJ, Rifas-Shiman SL, Pearce EN, Nassar N, Oken E. 2020. Neonatal thyroxine, maternal thyroid function, and cognition in mid-childhood in a US cohort. Matern Child Health J. 24(4):503–513.

Lans MC, Spiertz C, Brouwer A, Koeman JH. 1994. Different competition of thyroxine binding to transthyretin and thyroxine-binding globulin by hydroxy-PCBs, PCDDs and PCDFs. Eur J Pharmacol. 270(2–3): 129–136.

Lasley SM, Gilbert ME. 2011. Developmental thyroid hormone insufficiency reduces expression of brain-derived neurotrophic factor (BDNF) in adults but not in neonates. Neurotoxicol Teratol. 33(4):464–472.

Lazarus J. 2016. Thyroid regulation and dysfunction in the pregnant patient. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, Dungan K, Hershman JM, Hofland J, Kalra S, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; p. 2000.

Leonard JA, Tan Y-M, Gilbert M, Isaacs K, El-Masri H. 2016. Estimating margin of exposure to thyroid peroxidase inhibitors using high-throughput *in vitro* data, high-throughput exposure modeling, and physiologically based pharmacokinetic/pharmacodynamic modeling. Toxicol Sci. 151(1):57–70.

Lever C, Burton S, O'Keefe J. 2006. Rearing on hind legs, environmental novelty, and the hippocampal formation. Rev Neurosci. 17(1–2): 111–133.

Li J, Wang X, Yan Y, Wang K, Qin D, Xin Z, Wei J. 1986. The effects on fetal brain development in the rat of a severely iodine deficient diet derived from an endemic area: observations on the first generation. Neuropathol Appl Neurobiol. 12(3):261–276.

Li AA, Makris SL, Marty MS, Strauss V, Gilbert ME, Blacker A, Zorrilla LM, Coder PS, Hannas B, Lordi S, et al. 2019. Practical considerations for developmental thyroid toxicity assessments: what's working, what's not, and how can we do better? Regul Toxicol Pharmacol. 106: 111–136.

Lilienthal H, Verwer CM, van der Ven LT, Piersma AH, Vos JG. 2008. Exposure to tetrabromobisphenol A (TBBPA) in Wistar rats: neurobehavioral effects in offspring from a one-generation reproduction study. Toxicology. 246(1):45–54.

Lucia FS, Pacheco-Torres J, González-Granero S, Canals S, Obregón MJ, García-Verdugo JM, Berbel P. 2018. Transient hypothyroidism during lactation arrests myelination in the anterior commissure of rats. A magnetic resonance image and electron microscope study. Front Neuroanat. 12:31.

Mahle DA, Yu KO, Narayanan L, Mattie DR, Fisher JW. 2003. Changes in cross-fostered Sprague-Dawley rat litters exposed to perchlorate. Int J Toxicol. 22(2):87–94.

Makris SL, Raffaele K, Allen S, Bowers WJ, Hass U, Alleva E, Calamandrei G, Sheets L, Amcoff P, Delrue N, et al. 2009. A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. Environ Health Perspect. 117(1):17–25.

Marchant B, Brownlie BE, Hart DM, Horton PW, Alexander WD. 1977. The placental transfer of propylthiouracil, methimazole and carbimazole. J Clin Endocrinol Metab. 45(6):1187–1193.

Martin MT, Brennan RJ, Hu W, Ayanoglu E, Lau C, Ren H, Wood CR, Corton JC, Kavlock RJ, Dix DJ. 2007. Toxicogenomic study of triazole fungicides and perfluoroalkyl acids in rat livers predicts toxicity and categorizes chemicals based on mechanisms of toxicity. Toxicol Sci. 97(2):595–613.

Marty S, Beekhuijzen M, Charlton A, Hallmark N, Hannas BR, Jacobi S, Melching-Kollmuss S, Sauer UG, Sheets LP, Strauss V, et al. 2021. Towards a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny – part II: how can key events of relevant adverse outcome pathways be addressed in toxicological assessments? Crit Rev Toxicol. 51(4): 328–358.

Meerts IA, van Zanden JJ, Luijks EA, van Leeuwen-Bol I, Marsh G, Jakobsson E, Bergman A, Brouwer A. 2000. Potent competitive interactions of some brominated flame retardants and related compounds with human transthyretin *in vitro*. Toxicol Sci. 56(1):95–104.

Melancia F, Servadio M, Schiavi S, Campolongo P, Giusti-Paiva A, Trezza V. 2017. Testing the correlation between experimentally-induced hypothyroidism during pregnancy and autistic-like symptoms in the rat offspring. Behav Brain Res. 321:113–122.

Morse DC, Wehler EK, Wesseling W, Koeman JH, Brouwer A. 1996. Alterations in rat brain thyroid hormone status following pre- and postnatal exposure to polychlorinated biphenyls (Aroclor 1254). Toxicol Appl Pharmacol. 136(2):269–279.

Morreale de Escobar G, Calvo R, Obregon MJ, Escobar del Rey F. 1990. Contribution of maternal thyroxine to fetal thyroxine pools in normal rats near term. Endocrinology. 126(5):2765–72767.

Moser VC, Becking GC, Cuomo V, Frantik E, Kulig BM, MacPhail RC, Tilson HA, Winneke G, Brightwell WS, De-Salvia MA, et al. 1997. The IPCS collaborative study on neurobehavioral screening methods: IV. Control data. Neurotoxicology. 18:947–968.

[NAFTA] North American Free Trade Agreement. 2016. NAFTA Technical Working Group on Pesticides (TWG) – developmental neurotoxicity study guidance document; December 2016; https://www.epa.gov/sites/default/files/2017-02/documents/developmental_neurotoxicity_study_internal_guidance_document_final_0.pdf [accessed 2022 Sep].

Nagasaka A, Hidaka H. 1976. Effect of antithyroid agents 6-propyl-2-thiouracil and 1-mehtyl-2-mercaptoimidazole on human thyroid iodine peroxidase. J Clin Endocrinol Metab. 43(1):152–158.

Nakamura H, Noh JY, Itoh K, Fukata S, Miyauchi A, Hamada N. 2007. Comparison of methimazole and propylthiouracil in patients with hyperthyroidism caused by Graves' disease. J Clin Endocrinol Metab. 92(6):2157–2162.

[NAS/NRC] National Academy of Sciences/National Research Council. 2005. Health implications of perchlorate ingestion. Washington (DC): National Academy Press.

Nelson SM, Haig C, McConnachie A, Sattar N, Ring SM, Smith GD, Lawlor DA, Lindsay RS. 2018. Maternal thyroid function and child educational attainment: prospective cohort study. BMJ. 360:k452.

Nogimori T, Braverman LE, Taurog A, Fang SL, Wright G, Emerson CH. 1986. A new class of propylthiouracil analogs: comparison of 5′-deiodinase inhibition and antithyroid activity. Endocrinol. 118(4): 1598–1605.

Northcutt KV, Leal-Medina TS, Yoon YS. 2021. Early postnatal hypothyroidism reduces juvenile play behavior, but prenatal hypothyroidism compensates for these effects. Physiol Behav. 241:113594.

Noyes PD, Friedman KP, Browne P, Haselman JT, Gilbert ME, Hornung MW, Barone S, Jr, Crofton KM, Laws SC, Stoker TE, et al. 2019. Evaluating chemicals for thyroid disruption: opportunities and challenges with *in vitro* testing and adverse outcome pathway approaches. Environ Health Perspect. 127(9):95001.

O'Shaughnessy KL, Kosian PA, Ford RL, Oshiro WM, Degitz SJ, Gilbert ME. 2018a. Developmental thyroid hormone insufficiency induces a cortical brain malformation and learning impairments: a cross fostering study. Toxicol Sci. 163(1):101–115.

O'Shaughnessy KL, Wood C, Ford RL, Kosian PA, Hotchkiss MG, Degitz SJ, Gilbert ME. 2018b. Thyroid hormone disruption in the fetal and neonatal rat: predictive hormone measures and bioindicators of hormone action in the developing cortex. Toxicol Sci. 166(1):163–179.

O'Shaughnessy KL, Thomas SE Spring ST, Ford JL, Ford RL, Gilbert ME. 2019. A transient window of hypothyroidism alters neural progenitor cells and results in abnormal brain development. Sci Rep. 9:4662.

O'Shaughnessy KL, Gilbert ME. 2020. Thyroid disrupting chemicals and developmental neurotoxicity – new tools and approaches to evaluate hormone action. Mol Cell Endocrinol. 518:110663.

[OECD] Organisation for Economic Cooperation and Development. 2002. Series on testing and assessment No. 35 and series on pesticides No. 14. Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies. ENV/JM/MONO(2002)19. Paris: OECD.

[OECD] Organisation for Economic Cooperation and Development. 2004. Series on testing and assessment No. 20. Guidance document for neurotoxicity testing. ENV/JM/MONO(2004)25. Paris: OECD.

[OECD] Organisation for Economic Co-operation and Development. 2008. Test guideline No. 407: repeated dose 28-day oral toxicity study in rodents. Paris: OECD Publishing.

[OECD] Organisation for Economic Cooperation and Development. 2018. Series on testing and assessment. Revised guidance document 150 on standard test guidelines for evaluating chemicals for endocrine disruption. Paris: OECD Publishing.

Oken E, Braverman LE, Platek D, Mitchell ML, Lee SL, Pearce EN. 2009. Neonatal thyroxine, maternal thyroid function, and child cognition. J Clin Endocrinol Metab. 94(2):497–503.

Oppenheimer JH, Schwartz HL, Surks MI. 1972. Propylthiouracil inhibits the conversion of L-thyroxine to L-triiodothyronine. An explanation of the antithyroxine effect of propylthiouracil and evidence supporting the concept that triiodothyronine is the active thyroid hormone. J Clin Invest. 51(9):2493–2497.

Päkkilä F, Männistö T, Hartikainen AL, Ruokonen A, Surcel HM, Bloigu A, Vääräsmäki M, Järvelin MR, Moilanen I, Suvanto E. 2015. Maternal and child's thyroid function and child's intellect and scholastic performance. Thyroid. 25(12):1363–1374.

Pagnin M, Kondos-Devcic D, Chincarini G, Cumberland A, Richardson SJ, Tolcos M. 2021. Role of thyroid hormones in normal and abnormal central nervous system myelination in humans and rodents. Front Neuroendocrinol. 61:100901.

Paul KB, Hedge JM, Devito MJ, Crofton KM. 2010a. Developmental triclosan exposure decreases maternal and neonatal thyroxine in rats. Environ Toxicol Chem. 29(12):2840–2844.

Paul KB, Hedge JM, DeVito MJ, Crofton KM. 2010b. Short-term exposure to triclosan decreases thyroxine *in vivo* via upregulation of hepatic catabolism in young Long-Evans rats. Toxicol Sci. 113(2):367–379.

Paul KB, Hedge JM, Bansal R, Zoeller RT, Peter R, DeVito MJ, Crofton KM. 2012. Developmental triclosan exposure decreases maternal, fetal and early neonatal thyroxine: dynamics and kinetic evaluation of a putative mode-of-action. Toxicology. 300(1–2):31–45.

Paul KB, Thompson JT, Simmons SO, Vanden Heuvel JP, Crofton KM. 2013. Evidence for triclosan-induced activation of human and rodent xenobiotic receptors. Toxicol In Vitro. 27(7):2049–2060.

Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, Simmons SO. 2016. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast Phase I and II chemical libraries. Toxicol Sci. 151(1):160–180.

Paul Friedman P, Martin K, Crofton M, Hsu KM, Sakamuru C-W, Zhao S, Xia J, Huang M, Stavreva R, Soni DA, et al. 2019. Limited chemical structural diversity found to modulate thyroid hormone receptor in the Tox21 chemical library. Environ Health Perspect. 127(9):97009.

Paxinos G, Watson C. 1986. The rat brain in stereotaxic coordinates. 2nd ed. New York (NY): Academic Press.

Perez-Castillo A, Bernal J, Ferreiro B, Pans T. 1985. The early ontogenesis of thyroid hormone receptor in the rat fetus. Endocrinology. 117(6): 2457–2461.

Plummer S, Beaumont B, Elcombe M, Wallace S, Wright J, Mcinnes EF, Currie RA, Cowie D. 2021. Species differences in phenobarbital-mediated UGT gene induction in rat and human liver microtissues. Toxicol Rep. 8:155–161.

Pop VJ, Kuijpens JL, van Baar AL, Verkerk G, van Son MM, de Vijlder JJ, Vulsma T, Wiersinga WM, Drexhage HA, Vader HL. 1999. Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy. Clin Endocrinol. 50(2):149–155.

Powell MH, Nguyen HV, Gilbert M, Parekh M, Colon-Perez LM, Mareci TH, Montie E. 2012. Magnetic resonance imaging and volumetric analysis: novel tools to study the effects of thyroid hormone disruption on white matter development. Neurotoxicology. 33(5):1322–1329.

Prezioso G, Giannini C, Chiarelli F. 2018. Effect of thyroid hormones on neurons and neurodevelopment. Horm Res Paediatr. 90(2):73–81.

Raffaele KC, Fisher JE Jr, Hancock S, Hazelden K, Sobrian SK. 2008. Determining normal variability in a developmental neurotoxicity test: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol. 30(4):288–325.

Raffaele KC, Rowland J, May B, Makris SL, Schumacher K, Scarano LJ. 2010. The use of developmental neurotoxicity data in pesticide risk assessments. Neurotoxicol Teratol. 32(5):563–72572.

Ramhøj L, Hass U, Boberg J, Scholze M, Christiansen S, Nielsen F, Axelstad M. 2018. Perfluorohexane sulfonate (PFHxS) and a mixture of endocrine disrupters reduce thyroxine levels and cause antiandrogenic effects in rats. Toxicol Sci. 163(2):579–591.

Ramhøj L, Axelstad M, Svingen T. 2019. Validation of endogenous reference genes in rat cerebral cortex for RT-qPCR analyses in developmental toxicity studies. Peer J. 7:e7181.

Ramhøj L, Kortenkamp A, Scholze M, Svingen T, Gilbert ME, Ford J, Köhrle J, Axelstad M. 2020a. Perinatal exposure to propylthiouracil (PTU), but not DE-71, induces heterotopia in the brain despite similar effects on serum T4 levels. 59th Annual Meeting of the Society of Toxicology; March 15–19; Abstract #1486 in Meeting Abstract Book; p. 118; online version of poster [accessed 2022 Sep]. https://files.abstractsonline.com/CTRL/8A/D/878/D3A/542/410/2A1/51F/77F/3C6/B77/68/a3101_1.pdf?display.

Ramhøj L, Hass U, Gilbert ME, Wood C, Svingen T, Usai D, Vinggaard AM, Mandrup K, Axelstad M. 2020b. Evaluating thyroid hormone disruption: investigations of long-term neurodevelopmental effects in rats after perinatal exposure to perfluorohexane sulfonate (PFHxS). Sci Rep. 10(1):2672.

Ramhøj L, Frädrich C, Svingen T, Scholze M, Wirth EK, Rijntjes E, Köhrle J, Kortenkamp A, Axelstad M. 2021. Testing for heterotopia formation in rats after developmental exposure to selected *in vitro* inhibitors of thyroperoxidase. Environ Pollut. 283:117135.

Ramhøj L, Svingen T, Frädrich C, Rijntjes E, Wirth EK, Pedersen K, Köhrle J, Axelstad M. 2022a. Perinatal exposure to the thyroperoxidase inhibitors methimazole and amitrole perturbs thyroid hormone system signaling and alters motor activity in rat offspring. Toxicol Lett. 354: 44–55.

Ramhøj L, Svingen T, Mandrup K, Hass U, Lund SP, Vinggaard AM, Hougaard KS, Axelstad M. 2022b. Developmental exposure to the brominated flame retardant DE-71 reduces serum thyroid hormones in rats without hypothalamic-pituitary-thyroid axis activation or neurobehavioral changes in offspring. PLOS One. 17(7):e0271614.

Ramhøj L, Mandrup K, Hass U, Svingen T, Axelstad M. 2022c. Developmental exposure to the DE-71 mixture of polybrominated diphenyl ether (PBDE) flame retardants induce a complex pattern of endocrine disrupting effects in rats. PeerJ. 10:e12738.

Refetoff S, Feingold AB, Chrousos D. 2015. Thyroid hormone serum transport proteins. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, Hershman JM, Kaltsas G, Koch C, Kopp P, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.

Ren XM, Zhang YF, Guo LH, Qin ZF, Lv QY, Zhang LY. 2015. Structure-activity relations in binding of perfluoroalkyl compounds to human thyroid hormone T3 receptor. Arch Toxicol. 89(2):233–242.

Ren XM, Yao L, Xue Q, Shi J, Zhang Q, Wang P, Fu J, Zhang A, Qu G, Jiang G. 2020. Binding and activity of tetrabromobisphenol A monoether structural analogs to thyroid hormone transport proteins and receptors. Environ Health Perspect. 128(10):107008.

Rijntjes E, Scholz PM, Mugesh G, Köhrle J. 2013. Se- and S-based thiouracil and methimazole analogues exert different inhibitory mechanisms on type 1 and type 2 deiodinases. Eur Thyroid J. 2(4):252–258.

Rodriguez-Pena A. 1999. Oligodendrocyte development and thyroid hormone. J Neurobiol. 40:497–512.

Rolaki A, Pistollato F, Munn S, Price AB. 2019. Adverse outcome pathway on inhibition of Na^+/I^- symporter (NIS) leads to learning and memory impairment. In: OECD series on adverse outcome pathways no. 14. Paris: OECD.

Roques BB, Lacroix MZ, Puel S, Gayrard V, Picard-Hagen N, Jouanin I, Perdu E, Martin PG, Viguié C. 2012. CYP450-dependent biotransformation of the insecticide fipronil into fipronil sulfone can mediate fipronil-induced thyroid disruption in rats. Toxicol Sci. 127(1):29–41.

Royland JE, Parker JS, Gilbert ME. 2008. A genomic analysis of subclinical hypothyroidism in hippocampus and neocortex of the developing rat brain. J Neuroendocrinol. 20(12):1319–1338.

Royland JE, Kodavanti PRS. 2008. Gene expression profiles following exposure to a developmental neurotoxicant, Aroclor 1254: pathway analysis for possible mode(s) of action. Toxicol Appl Pharmacol. 231(2):179–196.

Ruis MT, Rock KD, Hall SM, Horman B, Patisaul HB, Stapleton HM. 2019. PBDEs concentrate in the fetal portion of the placenta: implications for thyroid hormone dysregulation. Endocrinol. 160(11):2748–2758.

Saegusa Y, Fujimoto H, Woo GH, Inoue K, Takahashi M, Mitsumori K, Hirose M, Nishikawa A, Shibutani M. 2009. Developmental toxicity of brominated flame retardants, tetrabromobisphenol A and 1,2,5,6,9,10-hexabromocyclododecane, in rat offspring after maternal exposure from mid-gestation through lactation. Reprod Toxicol. 28(4):456–467.

Saegusa Y, Fujimoto H, Woo GH, Ohishi T, Wang L, Mitsumori K, Nishikawa A, Shibutani M. 2012. Transient aberration of neuronal development in the hippocampal dentate gyrus after developmental exposure to brominated flame retardants in rats. Arch Toxicol. 86(9):1431–1442.

Saito K, Yamamoto K, Takai T, Yoshida S. 1983. Inhibition of iodide accumulation by perchlorate and thiocyanate in a model of the thyroid iodide transport system. Acta Endocrinol. 104(4):456–461.

Saito K, Kaneko H, Sato K, Yoshitake A, Yamada H. 1991. Hepatic UDP-glucuronyltransferase(s) activity toward thyroid hormones in rats: induction and effects on serum thyroid hormone levels following treatment with various enzyme inducers. Toxicol Appl Pharmacol. 111(1):99–106.

Sala-Roca J, Martí-Carbonell MA, Garau A, Darbra S, Balada F. 2002. Effects of dysthyroidism in plus maze and social interaction tests. Pharmacol Biochem Behav. 72(3):643–650.

Salas-Lucia F, Pacheco-Torres J, González-Granero S, García-Verdugo JM, Berbel P. 2020. Transient hypothyroidism during lactation alters the development of the corpus callosum in rats. An *in vivo* magnetic resonance image and electron microscopy study. Front Neuroanat. 14:33.

Sánchez-Huerta K, Pacheco-Rosado J, Gilbert ME. 2015. Adult onset-hypothyroidism: alterations in hippocampal field potentials in the dentate gyrus are largely associated with anaesthesia-induced hypothermia. J Neuroendocrinol. 27(1):8–19.

Sauer UG, Asiimwe A, Botham PA, Charlton A, Hallmark N, Jacobi S, Marty S, Melching-Kollmuss S, Palha JA, Strauss V, et al. 2020. Toward a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny – part I: which parameters from human studies are most relevant for toxicological assessments? Crit Rev Toxicol. 50(9):740–763.

Schmutzler C, Hamann I, Hofmann PJ, Kovacs G, Stemmler L, Mentrup B, Schomburg L, Ambrugger P, Grüters A, Seidlova-Wuttke D, et al. 2004. Endocrine active compounds affect thyrotropin and thyroid hormone levels in serum as well as endpoints of thyroid hormone action in liver, heart and kidney. Toxicology. 205(1–2):95–102.

Schneider S, Kaufmann W, Strauss V, Buesen R, van Ravenzwaay B, Hastings C, Hess F. 2009. Comparative developmental thyroid studies with 6-propyl-2-thiouracil (PTU) using two rat strains (Wistar *vs.* Sprague-Dawley) – results of treatment-related effects in thyroid hormones and thyroid morphology. Abstract #164. Poster presented at the 48th Annual Meeting of the Society of Toxicology, March 15–19, 2009, Baltimore MA, USA; Toxicologist 108:33. full poster available as Supplementary Information SI-7.

Schröder-van der Elst JP, van der Heide D, Rokos H, Morreale de Escobar G, Köhrle J. 1998. Synthetic flavonoids cross the placenta in the rat and are found in fetal brain. Am J Physiol. 274(2):E253–6.

Schussler GC, Schaffner F, Korn F. 1978. Increased serum thyroid hormone binding and decreased free hormone in chronic active liver disease. N Engl J Med. 299(10):510–515.

Sharlin DS, Tighe D, Gilbert ME, Zoeller RT. 2008. The balance between oligodendrocyte and astrocyte production in major white matter tracts is linearly related to serum total thyroxine. Endocrinology. 149(5):2527–2536.

Sharlin DS, Gilbert ME, Taylor MA, Ferguson DC, Zoeller RT. 2010. The nature of the compensatory response to low thyroid hormone in the developing brain. J Neuroendocrinol. 22(3):153–165.

Sheets LP, Dean KF, Reiter LW. 1988. Ontogeny of the acoustic startle response and sensitization to background noise in the rat. Behav Neurosci. 102(5):706–713.

Shimizu R, Yamaguchi M, Uramaru N, Kuroki H, Ohta S, Kitamura S, Sugihara K. 2013. Structure-activity relationships of 44 halogenated compounds for iodotyrosine deiodinase-inhibitor activity. Toxicology. 314(1):22–29.

Shiraki A, Saito F, Akane H, Takeyoshi M, Imatanaka N, Itahashi M, Yoshida T, Shibutani M. 2014. Expression alterations of genes on both neuronal and glial development in rats after developmental exposure to 6-propyl-2-thiouracil. Toxicol Lett. 228(3):225–234.

Sofronova AI, Gaynullina DK, Shvetsova AA, Borzykh AA, Selivanova EK, Kostyunina DS, Sharova AP, Martyanov AA, Tarasova OS. 2017. Antenatal/early postnatal hypothyroidism alters arterial tone regulation in 2-week-old rats. J Endocrinol. 235(2):137–151.

Spann MN, Cheslack-Postava K, Brown AS. 2020. The association of serologically documented maternal thyroid conditions during pregnancy with bipolar disorder in offspring. Bipolar Disord. 22(6):621–628.

Spring SR, Bastian TW, Wang Y, Kosian P, Anderson GW, Gilbert ME. 2016. Thyroid hormone-dependent formation of a subcortical band heterotopia (SBH) in the neonatal brain is not exacerbated under conditions of low dietary iron (FeD). Neurotoxicol Teratol. 56:41–46.

Sui L, Gilbert ME. 2003. Pre- and postnatal propylthiouracil-induced hypothyroidism impairs synaptic transmission and plasticity in area CA1 of the neonatal rat hippocampus. Endocrinol. 144(9):4195–4203.

Sui L, Anderson WL, Gilbert ME. 2005. Impairment in short-term but enhanced long-term synaptic potentiation and ERK activation in adult hippocampal area CA1 following developmental thyroid hormone insufficiency. Toxicol Sci. 85(1):647–656.

Szabo DT, Richardson VM, Ross DG, Diliberto JJ, Kodavanti PRS, Birnbaum LS. 2009. Effects of perinatal PBDE exposure on hepatic phase I, phase II, phase III, and deiodinase 1 gene expression involved in thyroid hormone metabolism in male rat pups. Toxicol Sci. 107(1):27–39.

Taylor MM, Hedge JM, Gilbert ME, DeVito MJ, Crofton K. 2003. Perinatal exposure to a polybrominated diphenyl ether mixture (DE-71): disruption of thyroid homeostasis and neurobehavioral development. Presented at Society of Toxicology 42nd Annual Meeting; March 9–13; Salt Lake City (UT) [accessed 2022 Sep]. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=61788.

Thorpe-Beeston JG, Nicolaides KH, Felto CV, Butler J, McGregor AM. 1991. Maturation of the secretion of thyroid hormone and thyroid-stimulating hormone in the fetus. N Engl J Med. 324(8):532–536.

Tuca A, Giralt M, Villarroya F, Vinas O, Mampel T, Iglesias R. 1994. Inhibition of iodothyronine 5′-deiodinase by iopanoic acid does not block nuclear T3 accumulation during rat fetal development. Pediatr Res. 35(1):91–95.

[US EPA] US Environmental Protection Agency. 1996. R.E.D. facts – amitrole. Pesticide reregistration. EPA-738-F96-002; August 1996 [accessed 2022 Sep]. https://www3.epa.gov/pesticides/chem_search/reg_actions/reregistration/fs_PC-004401_1-Sep-96.pdf.

[US EPA] US Environmental Protection Agency. 2000. Ethylene thiourea – 96-45-7 Summary created in April 1992, updated January 2000 [accessed 2022 Sep]. https://www.epa.gov/sites/production/files/2016-09/documents/ethylene-thiourea.pdf.

[US EPA] US Environmental Protection Agency. 2005. Guidance for thyroid assays in pregnant animals, fetuses and postnatal animals, and adult animals. Washington (DC): US EPA, Office of Pesticide Programs, Health Effects Division; p. 12 [accessed 2022 Sep]. https://www.epa.gov/pesticide-registration/guidance-thyroid-assays-pregnant-animals-fetuses-and-postnatal-animals-and.

[US EPA] US Environmental Protection Agency. 2008a. Toxicological review of 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47) (CAS No. 5436-43-1) in support of summary information on the Integrated Risk Information System, EPA/635/R-07/005F, June 2008.

[US EPA] US Environmental Protection Agency. 2008b. Toxicological review of 2,2′,4,4′,5-pentabromodiphenyl ether (BDE-99) (CAS No. 60348-60-9) in support of summary information on the Integrated Risk Information System, EPA/635/R-07/006F, June 2008.

[US EPA] US Environmental Protection Agency. 2008c. Toxicological review of 2,2′,4,4′,5,5′-hexabromodiphenyl ether (BDE-153) (CAS No. 68631-49-2) in support of summary information on the Integrated Risk Information System, EPA/635/R-07/007F, June 2008.

[US EPA] US Environmental Protection Agency. 2008d. Toxicological review of decabromodiphenyl ether (BDE-209) (CAS No. 1163-19-5) in support of summary information on the Integrated Risk Information System, EPA/635/R-07/008F, June 2008.

US EPA. 2009. Series 890 Endocrine disruptor screening test OCSPP 890.1500: Pubertal development and thyroid function in intact juvenile/peripubertal male rats US Environmental Protection Agency, Washington, DC [accessed 2022 Aug]. https://www.regulations.gov/document/EPA-HQ-OPPT-2009-0576-0010.

[US EPA] US Environmental Protection Agency. 2011a. Endocrine disruptor screening program. Pubertal development and thyroid function in intact juvenile/ female rats assay. OCSPP Guideline 890.1450. Standard evaluation procedure, August 2011.

[US EPA] US Environmental Protection Agency. 2011b. Endocrine disruptor screening program. Pubertal development and thyroid function in intact juvenile/peripubertal male rats assay. Guideline 890.1500. Standard evaluation procedure, August 2011.

[US EPA] US Environmental Protection Agency. 2014. Technical fact sheet. Perchlorate. January 2014 [accessed 2022 Sep]. https://www.epa.gov/sites/production/files/2014-03/documents/ffrrofactsheet_contaminant_perchlorate_january2014_final.pdf.

[US EPA] US Environmental Protection Agency. 2015. Perfluorochemicals. Biomonitoring. America's Children and the Environment, 3rd ed. [accessed 2022 Sep]. https://www.epa.gov/sites/production/files/2015-05/documents/biomonitoring-pfcs.pdf.

[US EPA] US Environmental Protection Agency. 2017. Technical Fact Sheet. Polybrominated diphenyl ethers. November 2017 [accessed 2022 Sep]. https://www.epa.gov/sites/default/files/2014-03/documents/ffrrofactsheet_contaminant_perchlorate_january2014_final_0.pdf.

Uziel A. 1986. Periods of sensitivity to thyroid hormone during the development of the organ of Corti. Acta Otolaryngol Suppl. 429(Suppl): 23–27.

Van der Ven LT, Van de Kuil T, Verhoef A, Verwer CM, Lilienthal H, Leonards PE, Schauer UM, Cantón RF, Litens S, De Jong FH, et al. 2008. Endocrine effects of tetrabromobisphenol-A (TBBPA) in Wistar rats as tested in a one-generation reproduction study and a subacute toxicity study. Toxicology. 245(1–2):76–89.

Van Wynsberghe DM, Klitgaard HM. 1973. The effects of thyroxine and triiodothyroacetic acid on neonatal development in the rat. Biol Neonate. 22(5):444–450.

Vansell NR. 2022. Mechanisms by which inducers of drug metabolizing enzymes alter thyroid hormones in rats. Drug Metab Dispos. 50(4): 508–517.

Vansell NR, Klaassen CD. 2001. Increased biliary excretion of thyroxine by microsomal enzyme inducers. Toxicol Appl Pharmacol. 176(3):187–194.

Villeneuve DL, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, Munn S, O'Brien JM, Pollesch NL, Smith LC, et al. 2018. Adverse outcome pathway networks II: Network analytics. Environ Toxicol Chem. 37(6):1734–1748.

Visser TJ. 1996. Pathways of thyroid hormone metabolism. Acta Med Austriaca. 23(1–2):10–16.

Vorhees CV, Makris SL. 2015. Assessment of learning, memory, and attention in developmental neurotoxicity regulatory studies: synthesis, commentary, and recommendations. Neurotoxicol Teratol. 52(Pt A): 109–115.

Vranckx R, Rouaze M, Savu L, Nunez EA, Beaumont C, Flink IL. 1990. The hepatic biosynthesis of rat thyroxine binding globulin (TBG): demonstration, ontogenesis, and up-regulation in experimental hypothyroidism. Biochem Biophys Res Commun. 167(1):317–322.

Weiss JM, Andersson PL, Lamoree MH, Leonards PE, van Leeuwen SP, Hamers T. 2009. Competitive binding of poly- and perfluorinated compounds to the thyroid hormone transport protein transthyretin. Toxicol Sci. 109(2):206–216.

Weiss JM, Andersson PL, Zhang J, Simon E, Leonards PEG, Hamers T, Lamoree MH. 2015. Tracing thyroid hormone-disrupting compounds: database compilation and structure-activity evaluation for an effect-directed analysis of sediment. Anal Bioanal Chem. 407(19):5625–5634.

Witorsch RJ. 2014. Critical analysis of endocrine disruptive activity of triclosan and its relevance to human exposure through the use of personal care products. Crit Rev Toxicol. 44(6):535–555.

Wolff J. 1998. Perchlorate and the thyroid gland. Pharmacol Rev. 50(1): 89–105.

Wu Y, Wang L. 2002. The effects of antiepileptic drugs on spatial learning and hippocampal protein kinase C gamma in immature rats. Brain Dev. 24(2):82–87.

York RG, Barnett J Jr., Brown WR, Garman RH, Mattie DR, Dodd D. 2004. A rat neurodevelopmental evaluation of offspring, including evaluation of adult and neonatal thyroid, from mothers treated with ammonium perchlorate in drinking water. Int J Toxicol. 23(3):191–214.

York RG, Lewis E, Brown WR, Girard MF, Mattie DR, Funk KA, Strawson JS. 2005a. Refining the effects observed in a developmental neurobehavioral study of ammonium perchlorate administered orally in drinking water to rats. I. Thyroid and reproductive effects. Int J Toxicol. 24(6): 403–418.

York RG, Barnett J Jr., Girard MF, Mattie DR, Bekkedal MVK, Garman RH, Strawson JS. 2005b. Refining the effects observed in a developmental neurobehavioral study of ammonium perchlorate administered orally in drinking water to rats. II. Behavioral and neurodevelopmental effects. Int J Toxicol. 24(6):451–467.

Yoshihara A, Luo Y, Ishido Y, Usukura K, Oda K, Sue M, Kawashima A, Hiroi N, Suzuki K. 2019. Inhibitory effects of methimazole and propylthiouracil on iodotyrosine deiodinase 1 in thyrocytes. Endocr J. 66(4):349–357.

Yu WG, Liu W, Jin YH. 2009. Effects of perfluorooctane sulfonate on rat thyroid hormone biosynthesis and metabolism. Environ Toxicol Chem. 28(5):990–996.

Zhang W, Cai Y, Sheng G, Chen D, Fu J. 2011. Tissue distribution of decabrominated diphenyl ether (BDE-209) and its metabolites in sucking rat pups after prenatal and/or postnatal exposure. Toxicology. 283(1): 49–54.

Zhang L, Teng W, Liu Y, Li J, Mao J, Fan C, Wang H, Zhang H, Shan Z. 2012. Effect of maternal excessive iodine intake on neurodevelopment and cognitive function in rat offspring. BMC Neurosci. 13:121.

Zhou T, Ross DG, DeVito MJ, Crofton KM. 2001. Effects of short-term *in vivo* exposure to polybrominated diphenyl ethers on thyroid hormones and hepatic enzyme activities in weanling rats. Toxicol Sci. 61(1):76–82.

Zhou T, Taylor MM, DeVito MJ, Crofton KM. 2002. Developmental exposure to brominated diphenyl ethers results in thyroid hormone disruption. Toxicol Sci. 66(1):105–116.
Zhou H, Yin N, Faiola F. 2020. Tetrabromobisphenol A (TBBPA): a controversial environmental pollutant. J Environ Sci. 97:54–66.
Zoeller RT, Crofton KM. 2005. Mode of action: developmental thyroid hormone insufficiency – neurological abnormalities resulting from exposure to propylthiouracil. Crit Rev Toxicol. 35(8–9):771–781.
Zoeller RT, Tan SW. 2007. Implications of research on assays to characterize thyroid toxicants. Crit Rev Toxicol. 37(1–2):195–210.

Appendix I:

Definition of case studies, selection of case study substances, and selection of representative studies

App-I.1. Case Study 1: Impairment of thyroid hormone synthesis via TPO inhibition

App-I.1.1. Case Study 1: Biological implications of TPO inhibition

TPO is an enzyme in the thyroid gland that is involved in the iodination of tyrosine residues and their coupling on thyroglobulin to form T3 and T4. TPO inhibition is associated with decreased thyroid hormone synthesis; see Marty et al. (2021) for details. Hence, TPO inhibitors directly affect thyroid hormone synthesis.

App-I.1.2. Case Study 1: Selected substances and evidence for the substances' MoA(s)

Case Study 1 includes seven substances that inhibit TPO, i.e. PTU, MMI, ETU, mancozeb (manganese ethylene bis-dithiocarbamate (polymeric) complex with zinc salt), mercaptobenzimidazole, amitrole (3-amino-1*H*-1,2,4-triazo) and cyanamide. Mancozeb is considered together with ETU since it is metabolised to ETU in rats.

The evidence that the Case Study 1 substances interfere with TPO-catalysed reactions has mainly been derived from *in vitro* assays (see e.g. Paul Friedman et al. 2016). Davidson et al. (1978) reported that TPO inactivation by PTU and MMI can occur by two different steps, i.e. (1) irreversible TPO inactivation in the presence of H_2O_2 and absence of iodide; and (2) reversible TPO inactivation in the presence of iodide, which is more likely to occur *in vivo* as iodide is present in healthy organisms. Similarly, Nagasaka and Hidaka (1976) reported that PTU caused reversible TPO inhibition since it reacted directly with the oxidised iodide and not with the enzyme, but they found that MMI irreversibly inhibited TPO *in vitro* (see also Doerge 1986). By comparison, TPO inhibition by ETU was found to be reversible once ETU was completely metabolised (Doerge and Takazawa 1990; Freyberger and Ahr 2006).

With regard to mercaptobenzimidazole, amitrole and cyanamide, Ramhøj et al. (2021) selected these three substances for their study on the basis of a systematic selection process, which showed that they were potent TPO inhibitors *in vitro*, elicited thyroid effects *in vivo* and were likely not carcinogenic or neurotoxic by other MoAs.

PTU has also been observed to inhibit DIO1 thereby potentially also affecting thyroid hormone levels in target tissues (see e.g. Goswami and Rosenberg 1986; Nogimori et al. 1986; Zoeller and Crofton 2005; Yoshihara et al. 2019).

App-I.1.3. Case Study 1: Studies considered

A total of 17 studies evaluating PTU are considered (*Table SI-1: Spreadsheets "CS1 PTU" and "CS1 EPA PTU;" SI-3: Case Study 1.1*). These are three rat toxicity studies addressing both thyroid-related and neurodevelopmental effects of PTU, i.e. Axelstad et al. (2008), which was conducted similar to an OECD TG 426 DNT study, Kobayashi et al. (2005) and Ramhøj et al. (2020a), an SOT poster; see bibliography for weblink. Further, a comparative thyroid assay by Schneider et al. (2009) was considered to provide supportive information on PTU-mediated thyroid-related effects even though it did not include neurodevelopmental assessments. Schneider et al. (2009) is also an SOT poster, and it is made available as Supplementary Information SI-7. Further, the database on PTU includes 13 investigational studies, published in altogether 20 publications, from extensive research work that is being led and coordinated by the US EPA on the potential of substances causing thyroid hormone imbalance to also elicit brain-related effects (Goldey et al. 1995a; Sui and Gilbert 2003; Sui et al. 2005; Gilbert and Sui 2006; Gilbert et al. 2007, 2014, 2016, 2017; Sharlin et al. 2008, 2010; Gilbert 2011; Lasley and Gilbert 2011; Johnstone et al. 2013; Bastian et al. 2014; Spring et al. 2016; Hassan et al. 2017; Boyes et al. 2018; O'Shaughnessy et al. 2018a, 2018b, 2019). Generally, these EPA PTU studies were reported across several publications, with some of the altogether 20 publications both referring to data from earlier publications and presenting new data. Despite such partial overlap between publications, the ECETOC T4 TF tried to identify which of the publications were most closely linked and assigned them to a total of 13 EPA PTU studies (*SI-3: Table SI-3*).

Three rat studies addressing potential neurodevelopmental effects of MMI are considered (*Table SI-1: Spreadsheet "CS1 MMI;" SI-3: Case Study 1.2*), i.e. Fegert et al. (2012), a study similar to an (OECD TG 443) EOGRTS, as well as two investigational studies by Ausó et al. (2004) and Darbra et al. (2003). Data presented for Fegert et al. (2012) also include hormone and body weight data from an unpublished study report that formed the basis for the Fegert et al. publication. A further MMI study was published by Ramhøj et al. (2022a) after data evaluation for the present review was completed (*Table SI-1, Spreadsheet "CS1 MMI"*). The findings from this recent MMI study are considered in the Section 4 discussions.

Four rat toxicity studies addressing potential neurodevelopmental effects of ETU or mancozeb are considered (*Table SI-1: Spreadsheet "CS1 ETU Mancozeb;" SI-3: Case Study 1.3*). ETU was investigated in an (OECD TG 443) EOGRTS of which data have been published in the European Commission (2017a) Thyroid Disruption Workshop Report (cited therein as "*Marty* et al. *2013 b in EU Risk Assessment Report*"). Mancozeb was investigated in an (OECD TG 426) DNT study and the corresponding range finding study. Data from these studies have also been published in European Commission (2017a), cited therein as "*Beck (2008a) in the draft risk assessment report*" (range finder) and "*Beck (2008b) in the draft risk assessment report*" (DNT study). Since Marty et al. (2013) and Beck (2008a, 2008b) are unpublished study reports and thus not available to the reader, the data from these reports are cited as having been published in European Commission (2017a), but they have not been included in the present bibliography. Finally, Axelstad et al. (2011a) investigated the effects of mancozeb in a DNT-like study that followed a similar protocol as the PTU study by Axelstad et al. (2008).

One publication addressing rat studies investigating potential thyroid-related and neurodevelopmental effects caused by mercaptobenzimidazole, amitrole or cyanamide is considered: Ramhøj et al. (2021; *Table SI-1: Spreadsheet "CS1 Other;" SI-3: Case Study 1.4*). A further amitrole study was published by Ramhøj et al. (2022a) after data evaluation for the present review was completed (*Table SI-1, Spreadsheet "CS1 Other"*). The findings from this recent amitrole study are considered in the Section 4 discussions.

App-I.2. Case Study 2: Impairment of thyroid hormone synthesis via iodine deficiency

App-I.2.1. Case Study 2: Biological implications of iodine deficiency

Since iodine is a critical component of thyroid hormones, iodine deficiency is an important cause of impaired thyroid hormone synthesis. The two predominant events that can result in decreased iodine levels in the thyroid gland are (1) inhibition of the NIS and (2) dietary iodine deficiency. The NIS is a transmembrane co-transporter located on the thyroid basolateral membrane, which transports one iodide into the thyroid gland in exchange for two sodium cations (NAS/NRC 2005; see Marty et al. (2021) for details on the function of the NIS).

App-I.2.2. Case Study 2: Selected substances and evidence for the substances' MoA(s)

The first part of Case Study 2 addresses effects caused by the NIS inhibitor perchlorate (NAS/NRC 2005; Dohán et al. 2007) and the second part dietary iodine deficiency.

Perchlorate is a negatively charged ion that is composed of one chlorine atom and four oxygen atoms (ClO_4^-). It is a poor complexing agent and forms a weak association with its counterion. Accordingly, perchlorate salts (e.g. sodium perchlorate, ammonium perchlorate) are extremely soluble in aqueous media and polar organic solvents. Due to this very high solubility, the health risks associated with the perchlorate salts are considered equivalent to those associated with perchlorate itself (NAS/NRC 2005).

Perchlorate ions interact with NIS as they are of similar size as iodide and favourably compete for the same binding site on the NIS. Indeed, perchlorate exhibits a 30-fold greater uptake over iodide (NAS/NRC 2005; Dohán et al. 2007). Thereby, exposure to perchlorate reduces the amount of iodide available for thyroid hormone synthesis in the thyroid gland. As a result, serum thyroid hormone levels decrease. Sustained decreases activate the HPT axis, resulting in increased TSH release from the pituitary and follicular cell hypertrophy/hyperplasia and decreased colloid in the thyroid gland (Wolff 1998; Saito et al. 1983). Other substances that affect thyroid hormone synthesis through a similar MoA include nitrate, perrhenate ($ReO4^-$), thiocyanate and chlorate (Wolff 1998; De Groef et al. 2006; Dohán et al. 2007).

App-I.2.3. Case Study 2: Studies considered

Five publications presenting rat studies investigating potential thyroid-related and neurodevelopmental effects caused by perchlorate are considered. These publications include two sequential studies by York et al. (2004) and York et al. (2005a,b) as well as studies by Gilbert and Sui (2008) and Mahle et al. (2003). The studies by York et al. widely followed the US EPA TG 870.6300 for DNT studies while also including thyroid-related parameters. The studies by Gilbert and Sui (2008) and Mahle et al. (2003) were investigational and did not follow standardised test protocols. All studies included thyroid-related parameters, and all studies, except for Mahle et al. (2003), included neurobehavioural or other brain-related parameters (*Table SI-1: Spreadsheet "CS2 Perchlorate;" SI-3: Case Study 2.1*). A further perchlorate study was published by Gilbert et al. (2022) after data evaluation for the present review was completed (*Table SI-1, Spreadsheet "CS2 Perchlorate"*). The findings from this recent perchlorate study are considered in Section 3 (relating, e.g. to brain T4 and T3 levels) and in the Section 4 discussions.

Numerous studies are available addressing the effect of dietary iodine deficiency on thyroid hormone synthesis and subsequent neurodevelopmental outcomes. Since this review focusses on substance-mediated effects, two studies addressing dietary iodine deficiency are included in Case Study 2 for illustrative purposes, i.e. Zhang et al. (2012) and Gilbert et al. (2013; *Table SI-1, Spreadsheet "CS2 Iodine Deficiency;" SI-3: Case Study 2.2*). Nonetheless, a comprehensive review of dietary iodine deficiency studies may enhance the understanding of how maternal and offspring thyroid hormone imbalance affect neurodevelopment since they do not include xenobiotic substances, which might elicit neurodevelopmental impairment by (further) MoAs unrelated to thyroid hormone imbalance.

App-I.3. Case Study 3: Enhancement of thyroid hormone clearance via (1) displacement of thyroid hormone from serum binding proteins and/or (2) induction of liver enzymes that metabolise thyroid hormones

App-I.3.1. Case Study 3: Biological implications (1) of displacement of thyroid hormones from serum binding proteins and (2) of induction of liver enzymes that metabolise thyroid hormone

Case Study 3 covers two MIEs: (1) displacement of thyroid hormones from serum binding proteins; and (2) induction of liver enzymes that metabolise thyroid hormones. This decision was taken since the available evidence indicated that the selected Case Study 3 substances generally exhibit both MoAs (see below, Section App-I.3.2). Both MIEs affect the availability of free thyroid hormone in the serum and facilitate enhanced thyroid hormone clearance. As such, they represent indirect thyroid-related MoAs.

App-I.3.1.1. Case Study 3: Displacement of thyroid hormone from serum binding proteins.

In the blood, the vast majority of thyroid hormone is bound to binding proteins, which are primarily thyroid binding globulin, transthyretin and albumin (Schussler et al. 1978; Refetoff 2015). Hormone distribution across the three binding proteins differs between humans and rodents, as well as between different life stages of the same species. In adult rats, the majority of thyroid hormone in the serum binds to albumin and transthyretin (Vranckx et al. 1990), rather than to the specific thyroid binding globulin. By comparison, only a small fraction of thyroid hormone in the blood is available as free thyroid hormone, but it is the free thyroid hormone that can bind to the thyroid receptors in target tissues. Therefore, homeostasis of the free hormone is pertinent to biological activity; see Marty et al. (2021) for details. Against this background, substances that bind to albumin, transthyretin and/or thyroid binding globulin can displace the bound thyroid hormone. Thereby, the ratio between free and bound thyroid hormone is modified, leading to increased thyroid hormone clearance and hence decreased serum levels of thyroid hormone.

App-I.3.1.2. Case Study 3: Induction of liver enzymes that metabolise thyroid hormone.

In rats, the most important route for thyroid hormone metabolism is by conjugation *via* phase II liver enzymes and specifically by UGTs that glucuronidate thyroid hormones for subsequent elimination (Beetstra et al. 1991); see Marty et al. (2021) for details, also on species differences in the pathways of thyroid hormone metabolism between humans and rats. Thyroid hormones are also metabolised *via* activation of pregnane X receptor-mediated phase I liver enzymes with subsequent conjugation (Visser 1996; Szabo et al. 2009; Roques et al. 2012). In rat studies, the liver microsome monooxygenases BROD, EROD and PROD are markers for the activity of phase I cytochrome P450 (Zhou et al. 2002; Paul et al. 2012, de-Miranda et al. 2016). Substances that induce liver enzymes that mediate thyroid hormone conjugation have the potential to enhance thyroid hormone clearance and hence decrease serum levels of thyroid hormone.

App-I.3.2. Case Study 3: Selected substances and evidence for the substances' MoA(s)

This case study considers TBBPA, PFHxS, Aroclor 1254 (a PCB), DE-71 (a mixture of PBDEs) and triclosan, i.e. five substances for which enhancement of thyroid hormone clearance by displacement of serum binding proteins and induction of liver enzymes have been reported as possible MIEs/MoAs (see also Section 4.1).

The evidence that the Case Study 3 substances displace thyroid hormone from serum binding proteins has generally been derived from *in vitro* studies. Specifically, Meerts et al. (2000) found TBBPA to have a more than 10-fold stronger *in vitro* binding affinity to transthyretin than ^{125}I-T4 (IC_{50} value for TBBPA: 7.7 nM). Hamers et al. (2006) reported that TBBPA bound to transthyretin with an IC_{50} value of 0.031 μM, which was 1.6-fold higher than the transthyretin binding potency of ^{125}I-T4. Similar findings were reported by Kitamura et al. (2005) and Chi et al. (2020). Hamers et al. (2020) measured the *in vitro* capacity of TBBPA to compete for transthyretin binding with a fluorescein isothiocyanate-T4 probe, yielding IC_{50} values of 0.034 μM for TBBPA and 0.1 μM for T4, i.e. an ~3-fold higher binding capacity of TBBPA as compared to T4. Ren et al. (2020) reported that TBBPA bound to human transthyretin with an IC_{50} value of ~0.2 μM (same relative binding potency as T4 in that study), but that it did not bind to human thyroid binding globulin *in vitro*. Hence, while the reported IC_{50} values vary between studies, which can likely be attributed to methodological differences, all findings generally confirm that TBBPA can displace T4 from transthyretin *in vitro*. Likewise, different perfluorinated alkyl acids including PFHxS were shown to have moderate *in vitro* binding affinities to albumin and transthyretin; generally, binding affinity was associated with the degree of fluorination of the alkyl chain, but it was never as strong as the affinity of T4 (Kerstner-Wood et al. 2003; Weiss et al. 2009). Ren et al. (2015) found that different perfluoroalkyl compounds can also bind to the ligand binding domain of the human thyroid hormone receptor, although with a much lower affinity than T3.

For Aroclor 1254 and the brominated flame-retardant DE-47, Hallgren and Darnerud (2002) presented *ex vivo* data indicating that these substances, or their metabolites, predominantly affect serum T4 levels by competing with binding to serum proteins and displacing thyroid hormone, which is then available for hepatic metabolism and clearance as free thyroid hormone; however, hepatic UGT induction is also of some importance. Similarly, Hamers et al. (2008) suggested that the biotransformation of BDE-47 yielded metabolites with considerable transthyretin binding properties. Lans et al. (1994) showed that hydroxylated PCBs, dibenzo-p-dioxins and dibenzofurans can inhibit *in vitro* T4 binding to transthyretin, but not to thyroxine-binding globulin.

Weiss et al. (2015) showed that triclosan has the potential to bind to transthyretin *in vitro*.

A broad variety of chemicals have been observed to induce phase I and/or phase II liver enzymes that mediate thyroid hormone metabolism. Specifically, UGT-mediated glucuronidation is the major metabolic pathway for thyroid hormones in the rat (further discussed in Marty et al. 2021). Evidence that the selected case study substances induce liver enzymes has been presented and discussed for TBBPA by Choi et al. (2011), for perfluorinated alkyl acids by Martin et al. (2007) and Yu et al. (2009), for PCBs by Saito et al. (1991) and Barter and Klaassen (1994), for PBDEs (including DE-71) by Zhou et al. (2001, 2002) and for triclosan by Paul et al. (2010b, 2012, 2013). DE-71 likely has complex MoAs, but these are not yet fully understood (Section 4.2.3.5).

App-I.3.3. Case Study 3: Studies considered

Three rat studies investigating TBBPA are considered, i.e. Cope et al. (2015), an OECD TG 416 two-generation study, Saegusa et al. (2009, 2012), an investigational study, and Lilienthal et al. (2008) and Van der Ven et al. (2008), an enhanced OECD TG 415 one-generation study *(Table SI-1: Spreadsheet "CS3 TBBPA;" SI-3: Case Study 3.1*). Two investigational rat studies investigating PFHxS are considered, i.e. Ramhøj et al. (2018, 2020b) and Gilbert et al. (2021; *Table SI-1: Spreadsheet "CS3 PFHxS;" SI-3: Case Study 3.2*). Four investigational rat studies evaluating Aroclor 1254 are considered: Goldey et al. (1995b), Goldey and Crofton (1998), Crofton et al. (2000b) and Morse et al. (1996). In the Aroclor 1254 study by Goldey et al. (1995b), the offspring of the high-dose group (8 mkd) exhibited 25% and 50% mortality on PND 12 and PND 21, respectively (exposure from GD 6 to LD 21). Further, Goldey et al. (1995b) reported that "*the controls, 1 and 4 mkd groups had only 3%, 5%, and 15% mortality, respectively*," by PND 21. Goldey et al. did not indicate which increases in mortality were statistically significant, and they further did not discuss whether they considered the increased mortality as treatment related. By comparison, Goldey and Crofton (1998), who also applied 8 mkd Aroclor 1254 from GD 6 to LD 21, did not report any pup mortality. Against this background, this review considers the data recorded at up to 8 mkd Aroclor 1254 in both Goldey et al. and Goldey and Crofton while bearing in mind the increased pup mortality observed by Goldey et al. (*Table SI-1: Spreadsheet "CS3 Aroclor 1254;" SI-3: Case Study 3.3*).

Five investigational rat studies evaluating DE-71 are considered: (1) Kodavanti et al. (2010); (2) Zhou et al. (2002), with further information on this study provided in a poster abstract by Taylor et al. (2003); (3) the SOT poster by Ramhøj et al. (2020a); (4) de-Miranda et al. (2016); and (5) Bowers et al. (2015) and Gill et al. (2016; *Table SI-1: Spreadsheet "CS3 DE-71;" SI-3: Case Study 3.4*). A further DE-71 study was published by Ramhøj et al. (2022b, 2022c) after data evaluation for the present review was completed (*Table SI-1, Spreadsheet "CS3 DE-71"*). The findings from this recent DE-71 study are considered in the Section 4 discussions.

Two investigational rat studies evaluating triclosan were considered: Gilbert et al. (2021) and Paul et al. (2012; *Table SI-1: Spreadsheet "CS3 Triclosan;" SI-3: Case Study 3.5*).

App-I.4. Case Study 4: Other MoAs including DIO inhibition

App-I.4.1. Case Study 4: Biological implications of DIO inhibition

Generally, all DIOs serve to metabolise thyroid hormones, but the three types (DIO1, DIO2 and DIO3) have different specific functions. These different functions also explain their distribution within the organism, which may further differ between species. In rats, DIO1, which mediates both the activation and inactivation of thyroid hormone, is expressed in the liver, kidney, central nervous system, pituitary, thyroid, intestine, and placenta (Bianco et al. 2002; Darras and van Herck 2012). Hence, DIO1 inhibition in the liver can impair thyroid hormone function. DIO2 converts T4 to active T3, and it is present in the foetal rat brain. Its inhibition during critical periods of neurodevelopment may lead to adverse neurodevelopmental outcomes; see Marty et al. (2021) for details on the function of DIO1, DIO2, and DIO3.

App-I.4.2. Case Study 4: Selected substance and evidence for the substance's MoA

OMC was selected as the Case Study 4 substance because it has been observed to reduce DIO1 activity in the liver and T4 levels in the blood (Schmutzler et al. 2004; Klammer et al. 2007); see discussion in Section 4.1 for difficulties in finding substances that only inhibit DIOs *and* for which relevant studies were available.

App-I.4.3. Case Study 4: Studies considered

Two rat studies evaluating OMC are considered: Axelstad et al. (2011b) and the SOT poster Ramhøj et al. (2020a; *Table SI-1: Spreadsheet "CS4 OMC;" SI-3: Case Study 4.1*).

Appendix II:

Conclusions on TPO inhibitors in the European Commission (2017a) Thyroid Disruption Workshop Report

In the European Commission (2017a) Thyroid Disruption Workshop Report, the mancozeb/ETU case is used as a starting point for exploring and discussing the neurodevelopmental effects of thyroid disrupting chemicals (page 34 in European Commission 2017a). The conclusion drawn from the mancozeb case study is that: "*… in contrast to what would be expected based on the clear disruptions by mancozeb of the thyroid hormone axis, no adverse effects on nervous system development have been observed…*" (page 35 in European Commission 2017a). Four possible explanations for this outcome are provided:

> "*1. … the tested doses were too low*" (European Commission 2017a; page 40)

The ECETOC T4 TF finds this explanation incomprehensible: In the DNT-like study by Axelstad et al. (2011a), the highest mancozeb dose (150 mkd) caused continued toxicity in the dams that was so severe that two dams were killed on GD 16. In the corresponding range finding study (Axelstad et al. 2011a), mancozeb dosed at 200–500 mkd had caused severe weight loss and signs of neurotoxicity (hind limb paralysis) in the dams within a few days. In a mancozeb DNT range finding study (Beck 2008a as cited in European Commission 2017a), the dams were dosed up to target doses of 60 mkd. At this dose, body weight, body weight gain and food consumption were significantly decreased during the gestation treatment period. Therefore, 30 mkd was selected as highest dose for the full DNT study (Beck 2008b as cited in European Commission 2017a; see page 44–45). The ECETOC T4 TF concludes that, if hormonal effects (and "downstream" effects) are observed in parallel with pronounced decreases in body weight gain, it cannot be distinguished whether the altered endocrine activity is a direct effect of the treatment or rather caused by the impaired nutritional state of the animal.

> "*2. The offspring were not hypothyroid themselves in the postnatal period (due to limited milk transfer) and since much brain development occurs postnatally in rats, the prenatal hypothyroxinemia was not severe enough to disrupt brain development*" (European Commission 2017a; page 40)

The ECETOC T4 TF finds that this scenario describes a physiological adaptive process of the highly versatile thyroid hormone system, but not an adverse effect.

> *"3. Offspring brain development was actually affected, but the used neurobehavioural assessment methods are not suited for the detection of subtle effects in the brain, caused by maternal hypothyroxinemia…"* (European Commission 2017a; page 40)

The findings from the present review do not provide evidence to support this scenario. Instead, they support the conclusion that, together with sensitive parameters, such as brain gene expression and electrophysiology, the currently available neurobehavioural and neurodevelopmental test methods also allow determining whether *in utero*/lactational exposure to a thyroid-active substance *does not* lead to evidence for DNT in rats. Regarding mancozeb (the matter of discussion in European Commission 2017a), the database covers a broad spectrum of neurodevelopmental parameters derived from a perinatal study (Axelstad et al. 2011a), an OECD TG 426 DNT study (Beck 2008a, 2008b as cited in European Commission 2017a) as well as an OECD TG 443 EOGRTS for the mancozeb metabolite ETU, with the EOGRTS being the preferred test method for the identification of endocrine disrupting properties following the EFSA and ECHA (2018) Endocrine Disruptor Guidance. The ECETOC T4 TF suggests that further research is required to establish whether more appropriate and sensitive sets of neurodevelopmental endpoints to be measured in rats are needed.

> *"4. The decreases in circulating T4 levels (due to TPO inhibition) lead to upregulation of peripheral deiodinase activity in the offspring brains, leading to a larger conversion of T4 to the active T3, which compensated for the low circulating T4 levels and provided enough T3 for normal brain development"* (European Commission 2017a; page 40)

Again, the ECETOC T4 TF finds that this scenario describes a physiological adaptive process, but not an adverse effect.

The European Commission (2017a) Thyroid Disruption Workshop Report also provides two explanations for why PTU, but not ETU, was found to induce neurodevelopmental effects (page 40):

> *"1. The majority of reported developmental PTU studies use very high doses of PTU and therefore cause very severe hypothyroidism. If PTU and ETU/mancozeb doses were chosen to result in an "identical" degree of hypothyroidism, an "identical" outcome regarding neurobehavioural development would also be seen."*

This explanation is repealed by the available literature: Axelstad et al. (2008) treated dams with 0.8, 1.6, and 2.4 mkd PTU from GD 7 to LD 17 (oral exposure). In the high-dose group, mean T4 levels were reduced by 59% in the GD 15 dams and by 84% and 11% in the PND 16 and PND 27 pups, respectively, and they were increased by 36% in the LD 27 dams (Table App-II). The serum T4 changes were statistically significant in the GD 15 dams and PND 16 pups ($p < 0.01$) but not in the LD 27 dams or PND 27 offspring. Statistically significant neurodevelopmental effects in the juvenile or adult offspring were observed in the mid- and high-dose groups but not in the low-dose group (Axelstad et al. 2008).

Axelstad et al. (2011a) initially dosed mancozeb up to 150 mkd in the main study (exposure from GD 7 until LD 16). This high dose was decreased to 100 mkd on GD 15 due to the severity of effects observed in the dams (i.e. in the range finding study, 200–500 mkd mancozeb caused severe weight loss and signs of neurotoxicity (hind limb paralysis) in the dams within few days). Mean T4 levels were reduced by 37% and 6% in the GD 15 and LD 24 dams, respectively, and they were increased by 2% and 31% in the PND 16 and PND 24 pups, respectively, relative to the respective concordant control group means (Table App-II). Maternal T4 reduction was significant in GD 15 ($p < 0.0001$), but not on LD 24, and T4 was not significantly altered in the pups either on PND 16 or on PND 20. Neurodevelopmental affects were not observed in any of the dose groups (Axelstad et al. 2011a).

In summary, PTU applied at the highest dose induced significant T4 reduction both in dams and pups, whereas mancozeb applied at the highest dose only induced significant T4 reduction in the dams (during gestation), but not in the pups. Notwithstanding, due to the severity of toxic effects observed in the high-dose group dams of the mancozeb study, it would not be possible to dose mancozeb beyond 100 mkd. Note, the threshold for significance applied in the Axelstad et al. (2008) PTU study differed by 100-fold from that used in the Axelstad et al. (2011a) mancozeb study ($p < 0.01$ *vs.* $p < 0.0001$).

The second explanation provided in European Commission (2017a) for why PTU, but not ETU, was found to induce neurodevelopmental effects (page 41) is:

> *"PTU has other modes of action than just TPO inhibition (also inhibition of deiodinase activity)…"*

The state-of-the-art regarding mechanisms of action of TPO inhibitors has advanced since the European Commission (2017a) Thyroid Disruption Workshop Report was published. This evidence is further discussed in Section 4.2.1.3 of the present review.

Table App-II. Comparison of maternal and offspring serum T4 decrements after gestational and lactational exposure to PTU or mancozeb.

			Dams		Pups		
Substance	Applied dose, exposure period	Serum T4	GD 15	LD 24/27[a]	PND 16	PND 24/27[a]	References
PTU	2.4 mkd, GD 7–PND 17	T4 (nM)	11.4**	24.0	4.79**	16.2	Axelstad et al. 2008
		Relative to control	↓ 59%	↑ 36% (NS)	↓ 84%	↓ 11% (NS)	
Mancozeb	150/100 mkd,[b] GD 7–PND 16	T4 (nM)	26.3***	23.2	33.1	24.3	Axelstad et al. 2011a
		Relative to control	↓ 37%	↓ 6% (NS)	↑ 2% (NS)	↑ 31% (NS)	

GD: gestational day; LD: lactational day; mkd: mg/kg body weight/day; NS: not statistically significant; PND: postnatal day; PTU: propylthiouracil; T4: thyroxine.
All values refer to group means; for standard deviation, see Axelstad et al. (2008, 2011a); relative data indicate the proportion by which the T4 level was altered as compared to the concurrent control group.
**$p < 0.01$: threshold for significance applied in PTU study; ***$p < 0.0001$: threshold for significance applied in mancozeb study.
[a]Mancozeb dose group tested at PND 24; PTU dose group tested at PND 27.
[b]150 mkd mancozeb was administered to the dams up until GD 15; thereafter the dose was reduced to 100 mkd.