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 IV: the ECETOC and CLE Proposal for a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS)

Stephanie Melching-Kollmuss, Kathrin Bothe, Alex Charlton, Babunilayam Gangadharan, Rashin Ghaffari, Sylvia Jacobi, Sue Marty, Heike-Antje Marxfeld, Elizabeth F. McInnes, Ursula G. Sauer, Larry P. Sheets, Christian Strupp, Helen Tinwell, Christiane Wiemann, Philip A. Botham & Bennard van Ravenzwaay

To cite this article: Stephanie Melching-Kollmuss, Kathrin Bothe, Alex Charlton, Babunilayam Gangadharan, Rashin Ghaffari, Sylvia Jacobi, Sue Marty, Heike-Antje Marxfeld, Elizabeth F. McInnes, Ursula G. Sauer, Larry P. Sheets, Christian Strupp, Helen Tinwell, Christiane Wiemann, Philip A. Botham & Bennard van Ravenzwaay (2023): Towards a science-based testing strategy to identify maternal thyroid hormone imbalance and neurodevelopmental effects in the progeny – Part IV: the ECETOC and CLE Proposal for a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS), Critical Reviews in Toxicology, DOI: 10.1080/10408444.2023.2231033

To link to this article: https://doi.org/10.1080/10408444.2023.2231033

Published online: 09 Aug 2023.

CRITICAL REVIEWS IN TOXICOLOGY
https://doi.org/10.1080/10408444.2023.2231033

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 IV: the ECETOC and CLE Proposal for a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS)

Stephanie Melching-Kollmuss[a], Kathrin Bothe[b], Alex Charlton[c], Babunilayam Gangadharan[d], Rashin Ghaffari[e], Sylvia Jacobi[f], Sue Marty[g], Heike-Antje Marxfeld[h], Elizabeth F. McInnes[i], Ursula G. Sauer[j], Larry P. Sheets[k], Christian Strupp[l], Helen Tinwell[d], Christiane Wiemann[m], Philip A. Botham[i] and Bennard van Ravenzwaay[n]

[a]BASF SE, Limburgerhof, Germany; [b]Bayer AG, CropScience, Monheim, Germany; [c]Syngenta Crop Protection, Greensboro, NC, USA; [d]Bayer SAS, Lyon, France; [e]Corteva Agriscience, Newark, DE, USA; [f]SJ-Consult, Wiesbaden, Germany; [g]Dow, Inc, Midland, MI, USA; [h]BASF SE, Ludwigshafen, Germany; [i]Syngenta, Jealott's Hill, Bracknell, United Kingdom; [j]Scientific Consultancy – Animal Welfare, Neubiberg, Germany; [k]Bayer CropScience, Chesterfield, MO, USA; [l]Gowan Crop Protection Ltd, Reading, United Kingdom; [m]BASF Oesterreich GmbH, Vienna, Austria; [n]Environmental Sciences Consulting, Altrip, Germany

ABSTRACT

Following the European Commission Endocrine Disruptor Criteria, substances shall be considered as having endocrine disrupting properties if they (a) elicit adverse effects, (b) have endocrine activity, and (c) the two are linked by an endocrine mode-of-action (MoA) unless the MoA is not relevant for humans. A comprehensive, structured approach to assess whether substances meet the Endocrine Disruptor Criteria for the thyroid modality (EDC-T) is currently unavailable. Here, the European Centre for Ecotoxicology and Toxicology of Chemicals Thyroxine Task Force and CropLife Europe propose a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS). In Tier 0, before entering the Thyroid-NDT-TAS, all available *in vivo*, *in vitro* and *in silico* data are submitted to weight-of-evidence (WoE) evaluations to determine whether the substance of interest poses a concern for thyroid disruption. If so, Tier 1 of the Thyroid-NDT-TAS includes an initial MoA and human relevance assessment (structured by the key events of possibly relevant adverse outcome pathways) and the generation of supportive *in vitro*/*in silico* data, if relevant. Only if Tier 1 is inconclusive, Tier 2 involves higher-tier testing to generate further thyroid- and/or neurodevelopment-related data. Tier 3 includes the final MoA and human relevance assessment and an overarching WoE evaluation to draw a conclusion on whether, or not, the substance meets the EDC-T. The Thyroid-NDT-TAS is based on the state-of-the-science, and it has been developed to minimise animal testing. To make human safety assessments more accurate, it is recommended to apply the Thyroid-NDT-TAS during future regulatory assessments.

ARTICLE HISTORY

Received 14 March 2023
Revised 22 June 2023
Accepted 22 June 2023

KEYWORDS

Endocrine disruptor; thyroid modality; adverse outcome pathway (AOP); mode-of-action (MoA); thyroid peroxidase (TPO); sodium-iodide symporter (NIS); deiodinase (DIO); uridine diphosphate glucuronyl-transferase (UGT); liver enzyme induction; weight-of-evidence (WoE); extended one-generation reproductive toxicity study (EOGRTS; OECD TG 443); developmental neurotoxicity study (DNT study; OECD TG 426); motor activity; cognitive function (learning and memory)

Abbreviations: ADME: absorption, distribution, metabolism, elimination; AE: adverse effect (Table Appendix 1); AhR: aryl hydrocarbon receptor; AO(P): adverse outcome (pathway); BROD: benzoxyresorufin (Figure 6); CAR: constitutive androstane receptor; Cefic LRI: European Chemical Industry Council Long-Range Research Initiative; CHO: Chinese hamster ovary (cells) (Table 3); CLE: CropLife Europe; CTA: comparative thyroid assay (Figure 3); Cyp: cytochrome p-450 (Figure 6); DIO: deiodinase; DNT: developmental neurotoxicity; EA: endocrine activity (Table Appendix 1); ECETOC: European Centre for Ecotoxicology and Toxicology of Chemicals; ECHA: European Chemicals Agency; EDC(-T): Endocrine Disruptor Criteria (for the thyroid modality); EFSA: European Food Safety Authority; EOGRTS: extended one-generation reproductive toxicity study; EP: European Parliament; EPA: Environmental Protection Agency; EU: European Union; EU NETVAL: EU Network of Laboratories for the Validation of Alternative Methods; EURL ECVAM: European Union Reference Laboratory for alternatives to animal testing; FRTL: Fischer rat thyroid follicular (cell line) (Table 3); fT3: free triiodothyronine; fT4: free thyroxine; GABA: gamma amino-butyric acid (Table Appendix 3); GD: gestational day (Table 1, Table 2); H.R.: human relevance (Figures 1–6); HP: histopathology (Figure 2); HPT: hypothalamic–pituitary–thyroid (Table Appendix 2); HTS: high-throughput screening; ILSI/HESI: International Life Sciences Institute / Health and Environmental Sciences Institute; KE(R): key event (relationship) (Figure 6, Table Appendix 3); LC: liquid chromatography (Table 3); LD: lactational day (Table 1); LDG: lower-dose groups (Figure 6); LEI: liver enzyme induction (Figure 5); MCT: monocarboxylate transporter (Table 3); MDCK: Madin Darby Canine Kidney (Table 3); MIE: molecular initiating event; MoA: mode-of-action; MS: Mass spectrometry

CONTACT Stephanie Melching-Kollmuss stephanie.melching-kollmuss@basf.com Certified Toxicologist (DGPT, ERT) – Regulation Agrochemicals, BASF SE, APD/ET - Li 444, Speyerer Strasse 2, 67117 Limburgerhof, Germany

(Table 3); MTD: maximum tolerated dose (Figure 2); NA: not addressed (Table 3); NDT: neurodevelopmental toxicity; NIS: sodium – iodide symporter; OCSPP: Office of Chemical Safety and Pollution Prevention; OECD: Organisation for Economic Co-operation and Development; PBK: physiologically based kinetic; PFHxS: perfluorohexane sulphonates (Table Appendix 2); PND: postnatal day; PPAR: peroxisome proliferator-activated receptor; PROD: pentoxyresorufin (Figure 6); PXR: pregnane X receptor; QSAR: quantitative structure activity relationship (Table 3); REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals; SBP: serum binding protein (Figure 5, Figure 6); SC: Scientific Committee; T-modality: thyroid modality for endocrine disruption; T3: triiodothyronine; T4: thyroxine; TBG: thyroid binding globulin (Table 3); TDG: top-dose group (Figure 6); TF: Task Force; TG: Test Guideline; TH: thyroid hormone; ThyM: thyroid method (Table 3); Thyroid-NDT-TAS: Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme; TK: toxicokinetics (Figure 3); TPO: thyroid peroxidase; TR: thyroid receptor (nuclear) (Table 3, Figure 5); TRH: thyrotropin-releasing hormone (Table 3); TSAR: Tracking System for Alternative Methods towards Regulatory Acceptance; TSH: thyroid stimulating hormone; TTR: transthyretin (Table 3, Figure 6, Table Appendix 3); UGT: uridine diphosphate glucuronyltransferase; Val: validation (Table 3); WHO / IPCS: World Health Organisation / International Programme on Chemical Safety; WoE: weight-of-evidence

Table of contents

1. Introduction

1.1. Background and scope

Thyroid perturbations during pregnancy and lactation can result in a variety of developmental alterations including neurodevelopmental impairment (Zoeller et al. 2007; Gilbert et al. 2012, 2020). Therefore, 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) that was developed for substances regulated under the Plant Protection

Products and Biocidal Products Regulations (EP and Council 2009, 2012).

Appendix A of the EFSA and ECHA (2018) Endocrine Disruptor Guidance presents *Additional considerations for how to assess the potential for thyroid disruption for human health*. Specifically, Appendix A describes patterns of thyroid-related effects in experimental animals that are considered to represent a concern for thyroid perturbation-mediated neurodevelopmental impairment in humans so further investigations are required. A general testing scheme is then proposed that is applicable to liver enzyme inducers only, i.e. substances that mediate enhanced thyroid hormone clearance (Curran and DeGroot 1991). The general testing scheme presented in Appendix A includes serum thyroid hormone measurements in the most sensitive populations, comparative *in vitro* studies of liver enzyme activities induced by the test substance in animal and human test systems, and the exclusion of other possible thyroid-related modes-of-action (MoAs).

However, neither Appendix A nor the further clarifications that have since been published in the EFSA (2020) *Technical report on the outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology* provide broader guidance covering different thyroid-related MoAs or indicate how the data should be evaluated in a weight-of-evidence (WoE) approach to reach a conclusion on whether, or not, a substance meets the European Commission (2017, 2018) Endocrine Disruptor Criteria (EDC). Appendix A recognises that the identification of thyroid-related hazards is currently hampered by a lack of internationally validated test methods. Overall, it is currently unclear how specific thyroid-related MoAs should be identified and how the (non-)human relevance of thyroid effects and/or neurodevelopmental effects observed in rats should be established (see also reviews by Gilbert et al. 2012, 2020; Kortenkamp et al. 2020).

To address these uncertainties, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) convened the Thyroxine (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 thyroid hormone imbalance and potentially also neurodevelopmental effects in the progeny; (2) if effects observed in rodents are *not* relevant for humans in line with the EDC (i.e. in accordance with the EDC, effects observed in rodents are by default considered relevant for humans, and it is the non-human relevance that would need to be established); and (3) if a threshold for thyroid hormone decrements can be identified below which neurodevelopmental effects are not to be expected. Building upon the evidence collated in the three previous reviews by the ECETOC T4 TF (Sauer et al. 2020; Marty et al. 2021, 2022), the present review fulfils the overarching goal of the TF and proposes a science-based tiered testing strategy. To enhance the range of underlying expertise, the ECETOC T4 TF has collaborated with CropLife Europe (CLE) in developing the ECETOC and CLE Proposal for a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS).

Note, throughout this article, the term neurodevelopmental toxicity (NDT) is used to describe adverse neurodevelopmental effects which are specifically related to thyroid hormone perturbations. This is in contrast to developmental neurotoxicity (DNT), which is used throughout as term to describe all types of substance-mediated neurodevelopmental adverse effects regardless of the underlying cause, i.e. also those that are unrelated to thyroid hormone perturbations.

The Thyroid-NDT-TAS, as it is presented here, only considers NDT as the adverse outcome of thyroid-related MoAs but not, e.g. thyroid carcinoma in rats. The human relevance of these tumours has generally been questioned (Bartsch et al. 2018; Foster et al. 2021). However, the Thyroid-NDT-TAS is not prescriptive. It presents a generic concept for the assessment of thyroid hormone imbalance and possibly resulting adverse effects.

The Thyroid-NDT-TAS is proposed to help overcome the impasse resulting from the fact that the current Appendix A of the EFSA and ECHA (2018) Endocrine Disruptor Guidance does not provide clear guidance on how to establish whether, or not, a substance meets the European Commission (2017, 2018) EDC for the thyroid modality (EDC-T; see Section 1.2 below for details on the EDC). The Thyroid-NDT-TAS is based on the state-of-the-science; it has been developed to make human safety assessments more accurate while at the same time minimising animal testing in line with the 3Rs principle to replace, reduce and refine animal testing (Russell and Burch 1959) that has been implemented in the Plant Protection Products and Biocidal Products Regulations (EP and Council 2009, 2012) as well as in *Directive 63/2010/EU on the protection of animals used for scientific purposes* (EP and Council 2010).

Further, the Thyroid-NDT-TAS is conceived to comply with current European Union (EU) legislation on the determination of endocrine disrupting properties. This legislation is outlined in Section 1.2 below as the starting point for the rationale underlying the Thyroid-NDT-TAS. Based thereupon, Section 2 describes the elements of the Thyroid-NDT-TAS, and Section 3 draws conclusions on the applicability of the Thyroid-NDT-TAS to determine whether, or not, a substance meets the European Commission (2017, 2018) EDC-T, including the determination whether effects observed in rat studies are not relevant for humans.

1.2. EU legislation of relevance for the Thyroid-NDT-TAS

The Commission Delegated Regulation (EU) 2017/2100 (European Commission 2017) and the Commission Regulation (EU) 2018/605 (European Commission 2018) provide the legal framework setting out scientific criteria for the determination of endocrine disrupting properties in the context of the Biocidal Products Regulation (EP and Council 2012) and the Plant Protection Products Regulation (EP and Council 2009), respectively.

A substance shall be considered as having endocrine disrupting properties if it (a) elicits adverse effects (unless it can be shown that these are not relevant to humans), (b) has an endocrine activity, and (c) the adverse effect is the consequence of the endocrine activity, i.e. the two are linked by an endocrine MoA (Box 1). Further, this link between the endocrine activity and the adverse effect should be biologically plausible (discussed in Section 2.1.1 and Section 2.3.1.2).

Point 1(a) of the EDC (Box 1) clearly states that the observed effects must be adverse, i.e. they must result in an impairment of functional capacity, of the capacity to compensate for additional stressors or of an increase in

susceptibility to other influences. This definition for adversity follows the definition for adversity by the World Health Organisation/International Programme on Chemical Safety (WHO/IPCS 2009). By comparison, if a substance that has endocrine activity only causes adaptive, non-adverse effects, it does not fulfil the EDC and shall not be considered as having endocrine disrupting properties. Adaptive, non-adverse effects have been defined as biological effects that do not cause biochemical, behavioural, morphological or physiological changes that affect the general well-being, growth development or life span of an animal (Lewis et al. 2002).

Box 1. Endocrine disruptor criteria (European Commission 2017, 2018).

For active substances in biocidal products, Point 1 of Section A (Endocrine-disrupting properties with respect to humans) of the Annex to European Commission (2017) states:

1. *A substance shall be considered as having endocrine-disrupting properties that may cause adverse effect in humans if, based on points (a) to (d) of point (2)*[#]*, it is a substance that meets all of the following criteria, unless there is evidence demonstrating that the adverse effects identified are not relevant to humans:*
   a. *it shows an adverse effect in an intact organism or its progeny, which is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences;*
   b. *it has an endocrine MoA, i.e. it alters the function(s) of the endocrine system;*
   c. *the adverse effect is a consequence of the endocrine MoA.*

[#] See Box 2 in Section 2.1.1 for details of Point 2(a-d).
A widely identical wording has been implemented for active substances in plant protection products in the Annex to European Commission (2018).

Point 1(a) of the EDC also clearly states that a substance shall not be considered as having endocrine disrupting properties if any adverse effects observed in laboratory animals are not relevant to humans. Hence, in the absence of evidence demonstrating irrelevance, effects on thyroid function and neurodevelopment observed in animal studies are considered relevant for humans.

Both Point 1(b) and 1(c) of the EDC refer to endocrine MoA (Box 1). Clarification regarding the use of this term is provided in the EFSA and ECHA (2018) Endocrine Disruptor Guidance (p. 7):

> It should be highlighted that the "endocrine MoA" as stated in point (b) should be interpreted as "endocrine activity" while the term "endocrine MoA" in point (c) covers the link between the adverse effect and the endocrine activity identified in points a) and b), respectively.

In this context, endocrine activity is defined as follows (EFSA and ECHA 2018, p. 99):

> Interaction with the endocrine system that can potentially result in a response of the endocrine system, target organs and tissues. A substance that has an endocrine activity it has the potential to alter the function(s) of the endocrine system.

The EDC implemented in European Commission (2017, 2018) for biocidal products and plant protection products are also cited in *Regulation (EC) No. 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals* (REACH; EP and Council 2006). Specifically, in Section 2.3 (Other hazards) of Annex II (Requirements for the compilation of safety data sheets) of the REACH Regulation, it is stated: "*Information shall be provided on . . . . whether the substance is a substance identified as having endocrine disrupting properties in accordance with the criteria set out in Commission Delegated Regulation (EU) 2017/2100 [. . .] or Commission Regulation (EU) 2018/605 [. . .].*" Hence, the EDC are also generally applicable to substances under the REACH Regulation.

2. The ECETOC and CLE Proposal for a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS)

The ECETOC and CLE Proposal for a Thyroid-NDT-TAS provides a structural concept that consists of these elements (Figure 1):

- Tier 0: collection and WoE evaluation of all available data to determine whether the substance of interest poses a

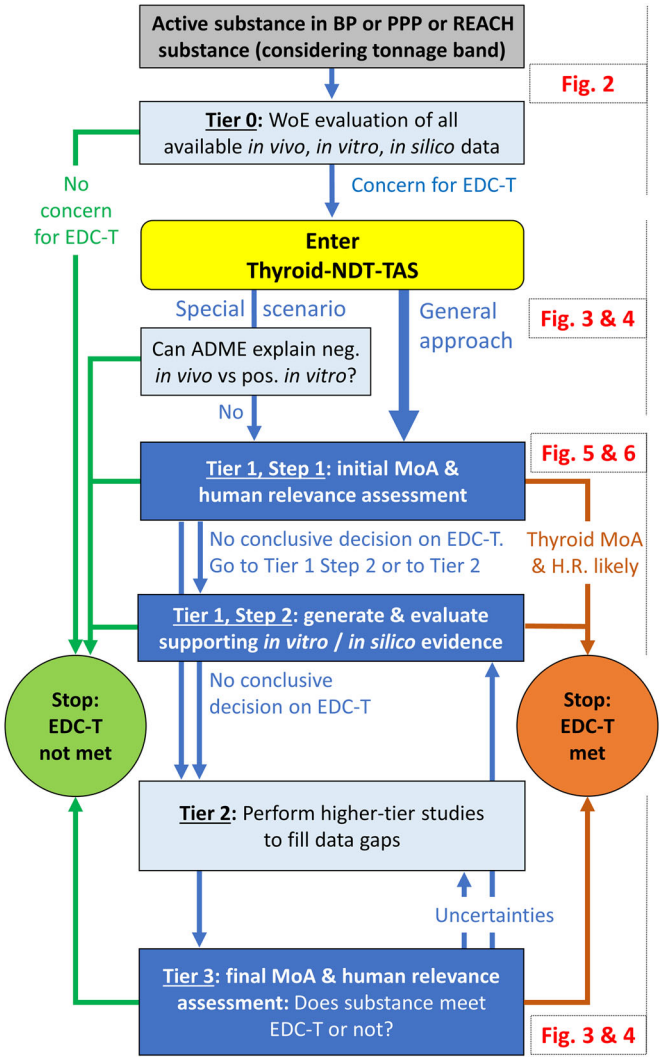

Figure 1. Overview of the ECETOC and CLE Thyroid-NDT-TAS (see Figures 2–6 for details). BP: biocidal product; EDC-T: endocrine disruptor criteria for the thyroid modality; MoA: mode-of-action; PPP: plant protection product; REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals; WoE: weight-of-evidence.

concern for endocrine disruption via the thyroid modality (T-modality) (see Section 2.1 for details)

- Tier 1–3: follow up on concerns identified in Tier 0 (see Section 2.2 for outline and Sections 2.3 and 2.4 for details)
  - Tier 1, Step 1: initial MoA and human relevance assessment of thyroid-related effects observed in animal studies
  - Tier 1, Step 2: generation and evaluation of supporting *in vitro* / *in silico* data, if relevant
  - Tier 2: identification of higher-tier testing needs to generate further thyroid- and/or neurodevelopment-related data, if relevant
  - Tier 3: final MoA and human relevance assessment and final WoE evaluations of all available data to draw a conclusion on whether, or not, the substance of interest meets the EDC-T

2.1. Tier 0: collection and WoE evaluation of all available data

2.1.1. Introduction to Tier 0 and general criteria for WoE evaluations

Figure 2 presents the Tier 0 evaluation scheme to collect all available thyroid-related data (and neurodevelopment-related data, if available) and to conduct a WoE evaluation of the thyroid-related data to determine whether the substance of interest poses a concern for endocrine disruption *via* the T-modality so that the Thyroid-NDT-TAS should be entered. Hence, Tier 0, as it is described throughout Section 2.1, focusses on the identification of thyroid-related effects. The reasons for this focus are (1) that thyroid-related data will generally be available for the Tier 0 evaluation (Section 2.1.2.1), and (2) that neurodevelopmental findings in the absence of thyroid-related effects (i.e. in the absence of endocrine activity) do not indicate a concern for endocrine disruption *via* the T-modality (but only DNT) in which case the Thyroid-NDT-TAS is not entered. By comparison, the scenario that both thyroid- and neurodevelopment-related effects are recorded in Tier 0 is described in Section 2.2.2 (i.e. the Thyroid-NDT-TAS is entered to determine whether the endocrine activity and the adverse effect are linked by a thyroid-related MoA and whether this MoA is (not) relevant to humans). See Table Appendix 1, which is included in this article after the bibliography, for possible scenarios for the Tier 0 *in vivo* database.

Potentially relevant thyroid-related data that should be gathered for the Tier 0 evaluation include available data from *in vivo* studies (Section 2.1.2), and these are often

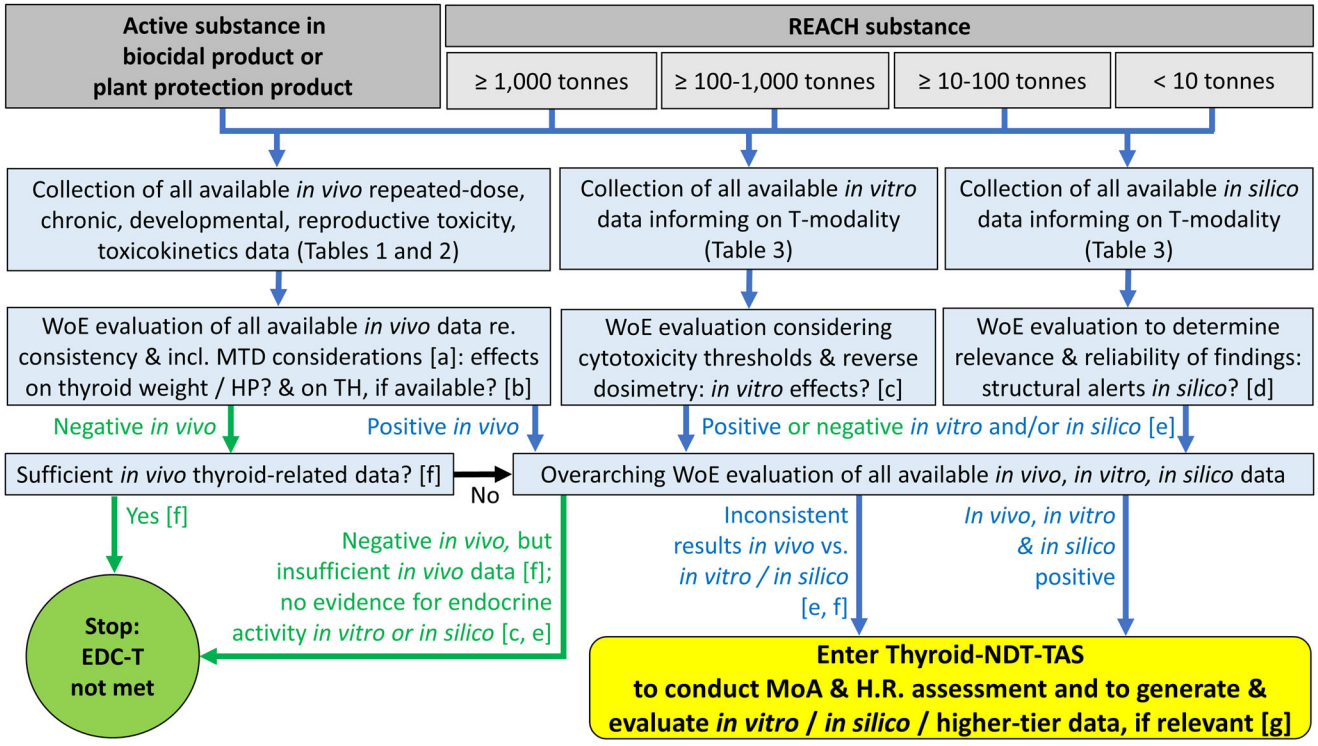

Figure 2. Tier 0: Evaluation of all available data to decide on the need to enter the ECETOC and CLE Thyroid-NDT-TAS.
EDC-T: endocrine disruptor criteria for thyroid modality; HP: histopathology; H.R.: human relevance; MoA: mode-of-action; MTD: maximum tolerated dose; T-modality: thyroid modality for endocrine disruption; TH: thyroid hormone; WoE: weight-of-evidence.
Colour legend: dark grey boxes: types of substances; light grey boxes, from right to left: production volumes (tonnage ranges) as per REACH Annexes VII-X, respectively; light blue boxes: elements of the assessment; blue arrows: continuation of evaluation; green arrows and text: negative findings; green circle: conclusion from Tier 0 evaluation that the EDC-T are not met. Yellow shape: conclusion from Tier 0 to enter the Thyroid-NDT-TAS.
[a] See Section 2.1.2.2 for elements to consider when applying expert judgement to determine whether the maximum tolerated dose was reached or exceeded.
[b] See Section 2.1.2.3 for elements to consider during the WoE evaluation of the *in vivo* thyroid-related findings.
[c] See Section 2.1.3 for elements to consider during the WoE evaluation of *in vitro* mechanistic data and to conclude that there is no evidence for *in vitro* activity.
[d] See Section 2.1.4 for elements to consider during the WoE evaluation of *in silico* data and to conclude that there is no evidence for *in silico* structural alerts.
[e] *In vitro* negative includes *in vitro* effects that only occurred at dose levels exceeding the *in vivo* top doses (as determined via *in vitro*-to-*in vivo* extrapolations).
[f] See Section 2.1.5 for aspects to consider in determining whether the *in vivo* database is sufficient. Inconsistent results *in vivo* vs. *in vitro/in silico* includes the scenarios "*in vivo* negative (*in vivo* database insufficient) combined with *in vitro/in silico* positive" and "*in vivo* positive combined with *in vitro/in silico* negative."
[g] Respect information requirements for REACH substances depending on tonnage band and applicability of the European Commission (2017, 2018) Endocrine Disruptor Criteria and EFSA and ECHA (2018) Endocrine Disruptor Guidance.

complemented by available data from *in vitro* mechanistic assays (Section 2.1.3) and possibly also by non-testing information from *in silico* modelling (Section 2.1.4). Point 2(a) of the European Commission (2017, 2018) EDC also requests that the assessment is based on all available *in vivo*, *in vitro* and *in silico* data (Box 2). This may include information from read-across, i.e. "*the use of relevant information from analogous substance(s) (the "source" information) to predict properties for the 'target' substance(s) under consideration*" (ECHA 2017). However, read-across is not further discussed in this manuscript since it focusses on the collection, generation and evaluation of substance-specific data.

Box 2. Collection of all available data and WoE evaluation to identify whether a substance has endocrine disrupting properties (European Commission 2017, 2018).

For active substances in biocidal products, Point 2 of Section A (Endocrine-disrupting properties with respect to humans) of the Annex to European Commission (2017) states:

(2) The identification of a substance as having endocrine-disrupting properties that may cause adverse effect in humans in accordance with point (1) [see Box 1] *shall be based on all of the following points:*

a. *all available relevant scientific data (in vivo studies or adequately validated alternative test systems predictive of adverse effects in humans or animals; as well as in vivo, in vitro, or, if applicable, in silico studies informing about endocrine MoAs):*
   i. *scientific data generated in accordance with internationally agreed study protocols, in particular those referred to in Annexes II and III of Regulation (EU) No 528/2012;*
   ii. *other scientific data selected applying a systematic review methodology;*
b. *an assessment of the available relevant scientific data based on a WoE approach in order to establish whether the criteria set out in point (1)* [see Box 1] *are fulfilled; in applying the WoE determination, the assessment of the scientific evidence shall, in particular, consider all of the following factors:*
   i. *both positive and negative results;*
   ii. *the relevance of the study designs for the assessment of adverse effects and of the endocrine MoA;*
   iii. *the quality and consistency of the data, considering the pattern and coherence of the results within and between studies of a similar design and across different species;*
   iv. *the route of exposure, toxicokinetic and metabolism studies;*
   v. *the concept of the limit dose, and international guidelines on maximum recommended doses and for assessing confounding effects of excessive toxicity;*
c. *using a WoE approach, the link between the adverse effect(s) and the endocrine MoA shall be established based on biological plausibility, which shall be determined in the light of current scientific knowledge and under consideration of internationally agreed guidelines;*
d. *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.*

A widely identical wording has been implemented for active substances in plant protection products in the Annex to European Commission (2018).

Following Tier 0 of the Thyroid-NDT-TAS, these different types of data (i.e. data from *in vivo* studies, *in vitro* assays and/or *in silico* modelling) are first submitted to separate WoE evaluations and then to an overarching WoE evaluation. Based thereupon, it is determined whether the substance of interest poses a human health concern for endocrine disruption *via* the T-modality. All WoE evaluations included in Tier 0 (just as those which form part of Tier 1–3) should be conducted in line with Point 2(b–d) of the EDC (Box 2). Accordingly, the WoE evaluation should consider both positive and negative results, the relevance of study designs, the quality and consistency of the data, the route of exposure, toxicokinetic and metabolism studies as well as the concept of the limit dose and "*international guidelines on maximum recommended doses and for assessing confounding effects of excessive toxicity*" (Section 2.1.2.2). Based upon the WoE evaluation, "*the link between the adverse effect(s) and the endocrine MoA shall be established based on biological plausibility*" (Point 2(c) of the EDC), and "*adverse effects that are non-specific secondary consequences of other toxic effects*" (Point 2(d) of the EDC) shall be ruled out.

The ECETOC T4 TF and CLE recommend conducting all WoE evaluations following pre-defined approaches thereby enhancing their transparency, objectivity and consistency. Such pre-defined approaches may be based on the ECHA templates for WoE and uncertainty evaluation in risk assessment (https://www.echa.europa.eu/web/guest/support/guidance-on-reach-and-clp-implementation/formats [accessed 2023 May]) and/or the EFSA Scientific Committee *Scientific opinion on the guidance on the use of the WoE approach in scientific assessments* (EFSA SC 2017).

2.1.2. Tier 0: collection and WoE evaluation of all available in vivo data

2.1.2.1. In vivo database that is generally available for Tier 0 WoE evaluation. For all substances that may need to be assessed for endocrine disrupting properties (i.e. active substances in biocidal products and plant protection products as well as REACH substances), mandatory information requirements have been implemented in the EU (EP and Council 2006, 2012; European Commission 2013). In accordance with the applicable legislation (Table 1), data from repeated dose toxicity studies and at least screening-level developmental and reproductive toxicity studies in rats are generally available for the Tier 0 evaluation (i.e. except for the very low tonnage REACH substances, which are generally assumed to have very low exposure potential). Further, a multi-generation reproductive toxicity study is generally available for active substances in biocidal products and plant protection products. All corresponding Organisation for Economic Co-operation and Development (OECD) Test Guidelines (TGs) include mandatory assessments of the thyroid gland, i.e. gross inspection, measurements of absolute and relative organ weight and histopathological investigations (for most recent versions of all OECD TGs, see https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm [accessed 2023 May]). Measurements of serum T4 are mandatory in the 90-day repeated dose toxicity study (OECD TG 408, version of 2018), in the pre-natal developmental toxicity study (OECD TG 414, version of 2018), in the reproduction and developmental toxicity screening tests (OECD TG 421/422, versions of 2016) and in the extended one-generation reproductive toxicity study (EOGRTS; OECD TG 443, version of 2018). Measurements of serum triiodothyronine (T3) are only mandatory in the 90-day repeated dose

Table 1. Overview of *in vivo* database that will generally be available for different types of substances during the Tier 0 evaluation.

Test method	Active substances in biocidal products [a]	Active substances in plant protection products [b]	Substances regulated under the REACH Regulation [c]			
			Annex VII: ≥ 1 tonne	Annex VIII: ≥ 10 tonnes	Annex IX: ≥ 100 tonnes	Annex X: ≥ 1,000 tonnes
28-day short-term repeated dose toxicity study (OECD TG 407)	✓, preferred species rat	✓, where available	REACH Annex VII for substances manufactured or imported in quantities of 1 tonne or more does not include standard information requirements for repeated dose, developmental or reproductive toxicity, or toxicokinetics. Nonetheless, existing data concerning these endpoints have to be submitted.	✓, unless reliable 90-day study available	✓	✓
90-day subchronic repeated dose toxicity study (OECD TG 408)	✓, preferred species rat	✓, rodents (usually rats) and non-rodent species		✓, only if indicated by specific concerns	✓	✓
Long-term (≥ 12 months) repeated dose toxicity study (e.g. OECD TG 452)	✓, preferred species rat	✓, rat (and mouse, unless shown that not necessary)		Generally, not available	Generally, not available	✓, only under certain conditions
Reproduction/ developmental screening study (OECD TG 421 or 422) [d]	May be available as range finding study for 90-day or developmental toxicity study, but not mandatory	Generally, not available		✓, preferred species rat	✓, preferred species rat	✓, preferred species rat
(Prenatal) developmental toxicity study (e.g. OECD TG 414) [d]	✓, initially one species, "*preferred species rabbit*," OECD TG 414 rat study will generally be available	✓, rabbit and rat; OECD TG 414 rat study will generally be available		Generally, not available, unless "*serious concerns about potential adverse effects on sexual function, fertility or development*"	✓, one species, generally rat, second species as relevant	✓, second species if first species negative
Two-generation reproduction toxicity study (OECD TG 416) or EOGRTS (OECD TG 443)	✓, rat preferred species, EOGRTS (with Cohort 1 A & 1B including 2nd generation) preferred test method for new studies	✓, rat (majority of compounds: OECD TG 416; EOGRTS only if conducted recently)		Generally, not available, unless "*serious concerns about potential adverse effects on sexual function, fertility or development*"	✓, EOGRTS (Cohorts 1 A and 1B without F_2 generation, one species), if concerns [e]	✓, EOGRTS preferred test method
Developmental neurotoxicity study (OECD TG 426)	✓, or any relevant study (set) providing equivalent information, or Cohorts 2 A and 2B of an EOGRTS with additional investigation for cognitive functions	Generally, not available		Generally, not available	Generally, not available	Generally, not available
Toxicokinetics study (OECD TG 417)	"*The toxicokinetics and metabolism studies should provide basic data about the rate and extent of absorption, the tissue distribution and the relevant metabolic pathway including the degree of metabolism, the routes and rate of excretion and the relevant metabolites.* "	"*Information on blood and tissues concentration of the active substance and relevant metabolites … shall be generated in short and long-term studies on relevant species to enhance the value of the toxicological data generated in terms of understanding the toxicity studies.*"		"*Assessment of the toxicokinetic behaviour of the substance to the extent that can be derived from the relevant available information*"	No additional standard information requirements [f]	No additional standard information requirements [f]

EOGRTS: extended one-generation reproductive toxicity study; GD: gestational day; LD: lactational day; OECD: Organisation for Economic Co-operation and Development; REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals; TG: Test Guideline.

[a] Database available for active substances under the Biocidal Products Regulation (EP and Council 2012, consolidated version of 15 April 2022), as per "Core Data Set" presented in its Annex II (Information requirements for active substances), Title 1, considering only information requirements related to Section 8.8 (toxicokinetics and metabolism data in mammals), Section 8.9 (repeated dose toxicity) and 8.10 (reproduction toxicity). Waiving of reproduction toxicity testing is possible as per Section 8.10 if "*the substance is of low toxicological activity (no evidence of toxicity seen in any of the tests available provided that the dataset is sufficiently comprehensive and informative), it can be proven from toxicokinetic data that no systemic absorption occurs via relevant routes of exposure (e.g. plasma or blood concentrations below detection limit using a sensitive method and absence of the substance and of metabolites of the substance in urine, bile or exhaled air) and the pattern of use indicates that there is no or negligible human or animal exposure.*"

[b] Database for active substances under the Plant Protection Products Regulation (EP and Council 2009) as per data requirements set out in the Annex to Commission Regulation (EU) No 283/2013 (European Commission 2013, consolidated version 17 November 2014), considering only data requirements related to Annex, Section 5.1 (studies on absorption, distribution, metabolism and excretion in mammals), Section 5.3 (short-term toxicity), Section 5.5 (long-term toxicity) and Section 5.6 (reproductive toxicity).

[c] Database for substances under the REACH Regulation (EP and Council 2006, consolidated version of 14 October 2022), as per standard information requirements set out in its Annexes VII, VIII, IX and X for substances manufactured or imported in quantities of ≥ 1 tonne, ≥ 10 tonnes, ≥ 100 tonnes and ≥ 1,000 tonnes, respectively, considering only information requirements related to the corresponding Sections 8.6 (repeated dose toxicity), 8.7 (reproduction toxicity) and 8.8 (toxicokinetics). For reproduction toxicity, these Annexes include the following rules for adaptation from the standard information requirements:

- Annex VIII, Section 8.7: The OECD TG 421/422 does not need to be conducted when there is no relevant human exposure.
- Annex IX and Annex X, Section 8.7: The reproductive toxicity studies do not need to be conducted if: *"the substance is of low toxicological activity (no evidence of toxicity seen in any of the tests available), it can be proven from toxicokinetic data that no systemic absorption occurs via relevant routes of exposure (e.g. plasma/blood concentrations below detection limit using a sensitive method and absence of the substance and of metabolites of the substance in urine, bile or exhaled air) and there is no or no significant human exposure."* (See also Footnote [e] below for conditions under which the EOGRTS is required as per Annex IX.).

[d] The (combined repeated dose toxicity with the) reproduction/developmental toxicity screening test (OECD TG 421/422, version of 29 July 2016) includes exposure from at least 14 days pre-mating until LD 13 and mandatory measurements of thyroid weight, histopathology and serum thyroid hormones. The prenatal developmental toxicity study (OECD TG 414, version of 25 June 2018) includes substance exposure from GD 6 – GD 19 in the rat and mandatory measurements of thyroid weight and histopathology as well as serum thyroid hormone measurements.

[e] Conditions under which EOGRTS is required as per REACH Annex IX standard information requirements: *"the available repeated dose toxicity studies indicate adverse effects on reproductive organs or tissues or reveal other concerns in relation with reproductive toxicity."* Further, as per rules for adaptation of the standard information requirements:

- the extension of Cohort 1B to include the F_2 generation may be necessary if "(a) *the substance has uses leading to significant exposure of consumers or professionals…, and (b) any of the following conditions are met: the substance displays genotoxic effects in somatic cell mutagenicity tests in vivo …, or there are indications that the internal dose for the substance and/or any of its metabolites will reach a steady state in the test animals only after an extended exposure, or there are indications of one or more relevant MoAs related to endocrine disruption from available in vivo studies or non-animal approaches."*
- the inclusion of the DNT Cohort 2 A may be necessary if *"existing information on the substance itself derived from relevant available in vivo or non-animal approaches (e.g. abnormalities of the CNS, evidence of adverse effects on the nervous or immune system in studies on adult animals or animals exposed prenatally), or specific mechanisms/MoAs of the substance with an association to (developmental) neurotoxicity and/or (developmental) immunotoxicity (e.g. cholinesterase inhibition or relevant changes in thyroidal hormone levels associated to adverse effects), or existing information on effects caused by substances structurally analogous to the substance being studied, suggesting such effects or mechanisms/MoAs."*

[f] As per REACH Annex I (General provisions for assessing substances and preparing chemical safety reports), Section 11 (Toxicological information), the safety data sheet shall include *"where appropriate information on toxicokinetics, metabolism and distribution."* Also, as per specific rules for adaptation from the standard information requirements implemented in REACH Annexes VII-X, toxicokinetic data can be used to support, e.g. the waiving of repeated dose and/or reproduction toxicity studies, e.g. when *"no systemic absorption occurs via relevant routes of exposure."*

toxicity study and in the prenatal developmental toxicity study, and measurements of thyroid stimulating hormone (TSH) are mandatory in these two studies as well as in the EOGRTS (Table 2). Please see Beekhuijzen et al. (2019), Li AA et al. (2019) and Marty et al. (2021, 2022) for in-depth discussions of issues to be considered during thyroid hormone measurements.

In accordance with the OECD (2012) *Conceptual Framework for testing and assessment of endocrine disruptors* and the OECD (2018) *Guidance Document No. 150 on standardised test guidelines for evaluating chemicals for endocrine disruption*, these OECD TGs are assigned to different levels that reflect the extent of information they can provide for the assessment of endocrine disruption:

Level 5 studies are defined as "*in vivo assays providing more comprehensive data on adverse effects on endocrine relevant endpoints over more extensive parts of the life cycle of the organism*" (OECD 2012). Two OECD TGs have been identified as Level 5 studies, the EOGRTS (OECD TG 443) and the two-generation reproductive toxicity study (OECD TG 416; version of 2001).

Level 4 studies are defined as "*in vivo assays providing data on adverse effects on endocrine relevant endpoints*" (OECD 2012). Level 4 studies include, amongst others, the 28- and 90-day repeated dose toxicity studies (OECD TG 407 and 408), the prenatal developmental toxicity study (OECD TG 414), the (combined repeated dose toxicity study with the) reproduction and developmental toxicity screening studies (OECD TG 421/422), if optional thyroid endpoints are included, the DNT study (OECD TG 426, version of 2007) as well as the pubertal development and thyroid function assays in peripubertal male and female rats (Environmental Protection Agency (EPA) Office of Chemical Safety and Pollution Prevention (OCSPP) Guidelines 890.1450 and 890.1500; https://www.epa.gov/sites/default/files/2019-10/documents/ocspp-testguidelines_masterlist-2019-09-24.pdf [accessed 2023 May]).

Level 3 studies include the uterotrophic assay and the Hershberger assay, which are generally not relevant for the T-modality.

Level 2 assays cover *in vitro* assays (Section 2.1.3).

Level 1 includes physico-chemical information, *in silico* modelling (Section 2.1.4) and other non-testing approaches.

The EFSA and ECHA (2018) Endocrine Disruptor Guidance has adopted the OECD (2012, 2018) scheme to assign different types of studies and information to these same levels.

2.1.2.2. Consideration of maximum tolerated dose.

According to the Tier 0 scheme to decide on the need to enter the Thyroid-NDT-TAS, the WoE evaluation of the available *in vivo* data considers whether thyroid-related effects, if present, were only observed above the maximum tolerated dose. If this is the case, the effects are considered "*adverse effects that are non-specific secondary consequences of other toxic effects*" (Point 2(d) of the EDC) and thus do not necessarily indicate a concern for endocrine disruption *via* the T-modality (further discussed below). The EFSA and ECHA (2018) Endocrine Disruptor Guidance lists elements to consider in evaluating whether the maximum tolerated dose was

Table 2. Measurements of serum T3, T4 and TSH in OECD test guidelines.

Test method	OECD TG, date version was adopted	T4	T3	TSH
28-day short-term repeated dose toxicity study	OECD TG 407 3 October 2008	*" … it may be helpful to retain plasma or serum samples to measure T3, T4 and TSH (optional) if there is an indication for an effect on the pituitary-thyroid axis"*		
90-day subchronic repeated dose toxicity study	OECD TG 408 27 June 2018	Mandatory at study termination		
(Prenatal) developmental toxicity study	OECD TG 414 25 June 2018	Dams: mandatory at study termination (GD 20/21)		
Two-generation reproduction toxicity study	OECD TG 416 22 January 2001	No mandatory measurements of T3, T4 or TSH		
(Combined repeated dose toxicity with the) reproduction/ developmental screening test	OECD TG 421/422 29 July 2016	Mandatory: PND 13 pups, adult males Optional: Dams, PND 4 pups	Optional	Optional
Developmental neurotoxicity (DNT) study	OECD TG 426 16 October 2007	No mandatory measurements of T3, T4 or TSH		
Extended one-generation reproductive toxicity study (EOGRTS)	OECD TG 443 25 June 2018	Mandatory: PND 22 pups, P_0/F_1 adults at study termination Optional: PND 4 pups	Not mentioned	Mandatory: PND 22 pups, P_0/F_1 adults at study termination
Long-term (≥12 months) repeated dose toxicity study	OECD TG 451–453 27 June 2018	No mandatory measurements of T3, T4 or TSH		

GD: gestational day; PND: postnatal day; T3: triiodothyronine; T4: thyroxine; TG: Test Guideline; TSH: thyroid stimulating hormone.

reached in a study and emphasises that expert judgement is required for this evaluation (p. 21):

> Elements to consider are alterations in physiological function, including: no more than 10% decrease in body weight gain relative to control, target organ toxicity and alterations in clinical pathological parameters. Although these parameters can only be considered indicative and expert judgement is necessary to define the maximum tolerated dose on a case-by-case basis. Elements which indicate that the maximum tolerated dose has been exceeded are reported in the *OECD Guidance on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation* (OECD 2000).

Guidance related to the maximum tolerated dose has been published in several OECD documents. In the OECD (2002) *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies*, the maximum tolerated dose is defined as the highest dose to produce toxic effects without causing death or significant morbidity *and* a no more than 10% decrease in body weight relative to controls. OECD TGs generally recommend using either a limit dose of 1,000 mg/kg body weight/day or the maximum tolerated dose as top dose. Further, the OECD (2018) *Guidance Document No. 150 on standard test guidelines for evaluating chemicals for endocrine disruption* highlights that endocrine effects observed in the presence of clear systemic toxicity are unlikely to be due to endocrine activity:

> 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.

More recently, Sewell et al. (2022), building on concepts developed in the ECETOC (2021) *Technical Report No. 138 on Guidance on dose selection*, recommended that the highest dose in repeated dose toxicity studies should cause "*minimal but evident toxicity to the test animals without significantly compromising their well-being*." Sewell and co-authors highlighted the importance of understanding the human relevance of kinetics to inform study design:

> For animal data, a translational understanding of human relevance of kinetics is important to inform study design in relation to several scopes, including hazard identification, risk assessment […] Typically, the risk of non-relevance may be higher at higher dose levels, where: A) high systemic exposures may disrupt physiological detoxification processes or other homeostatic processes leading to overt toxicity, potentially confounding appropriate evaluation of the toxicological results, and B) high systemic exposures may be quantitatively and qualitatively different from potential human systemic exposure.

The ECETOC (2021) Technical Report includes further details on "*pragmatic approaches to selecting dose levels that allow accurate risk assessment and also enable hazard-based classification based on identification of relevant hazards*."

Beyer et al. (2011) provided recommendations from International Life Sciences Institute/Health and Environmental Sciences Institute (ILSI/HESI) workshops on how to establish maternal toxicity in developmental and reproductive toxicity studies. Regarding the maximum tolerated dose, Beyer and colleagues concluded that "*a decrease in body weight gain of 20*% [in pregnant animals] *was considered excessive for most test articles/test materials*."

A European Society of Toxicologic Pathology expert workshop provided advice regarding the types of liver effects that indicate that the maximum tolerated dose was likely exceeded in studies investigating liver enzyme inducers (Hall et al. 2012):

> A dose level of a xenobiotic that in short-term tests induced either structural or biochemical evidence of hepatocellular damage, or produced increases in liver weight of approx. ≥150% would be considered adverse in the context of dose setting and exceed the maximum tolerated dose.

Hall et al. (2012) also indicated which types of liver effects should be considered non-adverse adaptive reactions: "*Hepatomegaly as a consequence of hepatocellular hypertrophy*

without histologic or clinical pathology alterations indicative of liver toxicity was considered an adaptive and a non-adverse reaction."

Taken together, the ECETOC T4 TF and CLE conclude that expert judgement is required to determine on a case-by-case basis whether the maximum tolerated dose was reached or exceeded in a study. Use of the kinetically derived maximum dose, i.e. the highest dose at kinetically linear doses or slightly above the point-of-departure from linear kinetics, has been suggested as an alternative to the maximum tolerated dose to address potential discrepancies between external and internal doses (Terry et al. 2016; Sewell et al. 2020; Felter et al. 2021). While a further discussion of the concepts of the maximum tolerated and kinetically derived maximum doses would exceed the scope of the present article, the ECETOC T4 TF and CLE recommend considering toxicokinetic data and/or physiologically based kinetic (PBK) modelling in establishing meaningful dose ranges for subchronic or longer-term studies.

2.1.2.3. Recommendations for the weighting of in vivo findings. During Tier 0 to decide on the need to enter the Thyroid-NDT-TAS, the available *in vivo* thyroid-related data are evaluated to determine whether there is a concern for endocrine disruption *via* the T-modality. Hence, the available *in vivo* data are evaluated to determine whether the substance of interest elicits effects on the thyroid gland and/or on serum thyroid hormone levels. Thereby, the *in vivo* data may provide information related to endocrine activity (Point 1(b) of the EDC) and a potential thyroid-related MoA (Point 1(c) of the EDC). If neurodevelopmental data from the EOGRTS (OECD TG 443) or DNT study (OECD TG 426) are available during Tier 0, they provide information on the substance's potential to elicit an adverse effect (Point 1(a) of the EDC). In this case, the Thyroid-NDT-TAS is entered to determine whether the endocrine activity and the adverse effect are linked by a thyroid-related MoA (Point 1(c) of the EDC) and whether this MoA is not relevant to humans (Section 2.2.2).

If effects on the thyroid gland and/or altered serum T4 (TSH and T3) levels are recorded in Tier 0, they are jointly evaluated in a WoE approach following the general criteria for WoE evaluations outlined in Section 2.1.1. Further, the ECETOC T4 TF and CLE recommend considering the following elements in the WoE evaluation:

- Absence of effects on thyroid weight and histopathology observed in 28-day (or longer-term) repeated dose toxicity studies is generally sufficient to conclude that the EDC-T are not met even if serum thyroid hormone data are unavailable. Justification: Prolonged thyroid perturbations lead to compensatory reactions of the thyroid gland regardless of the substance's thyroid-related MoA (further discussed in Section 2.3.2). In this regard, the EFSA (2020) Technical Report states (p. 7): "*… the dataset for thyroid can be considered complete on a case-by-case basis, pending whether the duration and doses selection allow a proper assessment of the thyroid histopathology (thyroid histopathology is generally considered more sensitive and informative than thyroid weight).*" This premise stands in line with findings from the EFSA (2019a) Scientific Report *Establishment of cumulative assessment groups of pesticides for their effects on the thyroid*. Of 127 active substances causing T4/T3 decrements and/or TSH increases, only two did not also alter relative thyroid weight or thyroid gland histopathology, and EFSA concluded that these two active substances were unlikely to cause hypothyroidism (EFSA 2019a).
- Thyroid-related effects are weighted lower if only one of several thyroid-related parameters is altered.
- Altered thyroid weight/histopathological findings are weighted lower if only observed in one (or few) out of several studies.
- Altered serum T4, T3 and/or TSH levels are weighted lower if observed in studies including small group sizes (e.g., $n < 10$). Justification: Due to the versatility of the hormone system, hormone data that are based on such small group sizes generally have weak statistical power. Assuming group sizes of 10 animals, T3/T4 decreases by approx. 25% and a TSH increase of approx. 40% as compared to the concurrent controls can be detected as statistically significant (for details, see Li AA et al. 2019; Marty et al. 2021). Appendix B of the EFSA and ECHA (2018) Endocrine Disruptor Guidance provides *Recommendations for design, conduction and technical evaluation of hormonal studies*.
- If thyroid effects are observed in more than one species (e.g. rats, rabbits and/or dogs) in studies with similar design, such concordances may add weight to the findings. However, if thyroid effects are only observed in one of several species, the aim should be to explain these using, e.g. information on absorption, distribution, metabolism, elimination (ADME) and/or data from *in vitro* mechanistic assays. Similarly, data from the amphibian metamorphosis assay (OECD TG 231, version of 2009) or the *Xenopus* eleutheroembryonic thyroid assay (OECD TG 248, version of 2019), as two ecotoxicological Level 3 assays (OECD 2012, 2019), may support WoE evaluations as the hypothalamic-pituitary-thyroid axis is highly conserved across vertebrates (see also EFSA and ECHA 2018, p. 9 and 102).

The WoE evaluation should also consider possible differences in sensitivity to substance-mediated thyroid perturbation (1) between adult males and/or non-pregnant/non-lactating females vs pregnant/lactating females; and (2) between pregnant/lactating females vs foetuses and/or new-born pups.

The ECETOC T4 TF and CLE also recommend considering the following elements to support the WoE evaluation of the available *in vivo* data. These elements reflect observational findings from Marty et al. (2022) who evaluated all thyroid- and brain-related data from 51 rat studies that included *in utero*/lactational exposure to substances causing thyroid hormone imbalance:

- Consider weighting offspring serum T4 decrements higher if they are $\geq$60% and $\geq$50% in the top- and lower-dose groups, respectively. Justification: Marty et al.

(2022) found that offspring serum T4 decrements exceeding these thresholds indicated an increased likelihood for statistically significant adverse neurodevelopmental effects.

- Consider weighting offspring serum T3 decrements higher if they are ≥20% *and* statistically significant. Justification: Marty et al. (2022) found that offspring serum T3 decrements exceeding this threshold indicated an increased likelihood for statistically significant adverse neurodevelopmental effects.
- Consider weighting offspring serum TSH increases higher if they are ≥400%. Justification: Marty et al. (2022) found that thyroid peroxidase (TPO) and sodium – iodide symporter (NIS) inhibitors (i.e. substances with a direct thyroid-related MoA; Section 2.1.3) that elicited statistically significant adverse neurodevelopmental effects in the respective studies (that included *in utero*/lactational substance exposure) also elicited dose-dependent and statistically significant offspring serum TSH increases ≥ 400%.
- Consider weighting thyroid-related effects lower if offspring serum T4 decrements attenuate over the course of an *in utero*/lactational exposure period. Justification: Marty et al. (2022) found that attenuations of offspring serum T4 decrements over the course of *in utero*/lactational exposure periods were associated with absence of statistically significant adverse neurodevelopmental effects.
- Consider weighting thyroid-related effects lower if they are not accompanied by offspring body weight changes. Justification: Marty et al. (2022) found that absence of offspring body weight changes was associated with absence of statistically significant adverse neurodevelopmental effects.

The ECETOC T4 TF and CLE recommend making use of future assessments of developmental and reproductive toxicity addressing the T-modality to further investigate the relevance and reliability of these empirical parameters and thresholds observed by Marty et al. (2022) and hence their possible applicability and robustness for regulatory assessments.

The database considered by Marty et al. (2022) was insufficient to conclude on the suitability of maternal or offspring serum free T4 (fT4)/free T3 (fT3) data to predict the likelihood for statistically significant neurodevelopmental findings, let alone to suggest thresholds for fT4/fT3 alterations indicating adverse outcomes. Serum fT4 is the most frequent thyroid-related parameter measured in human studies addressing maternal thyroid function and child neurodevelopment (Sauer et al. 2020). The ECETOC T4 TF and CLE recommend considering the generation of offspring serum fT4/fT3 data in future rodent developmental and reproductive toxicity studies to support the further investigation of these parameters.

Finally, a comprehensive evaluation of a substance's potential to elicit thyroid perturbations should also consider its toxicokinetic properties (Marty et al. 2022). The Tier 0 WoE evaluation considers all available toxicokinetics data to ensure that any observed effects reflect toxicological properties of the substance at doses that are relevant to the potentially exposed human population (Tan et al. 2021), with due consideration of the toxicokinetics-related part of the uncertainty factors that may be applied on a case-by-case basis. This stands in line with Point 2(b)(iv) of the EDC (Section 1.2).

2.1.3. Tier 0: collection and WoE evaluation of available in vitro data

During Tier 0 to decide on the need to enter the Thyroid-NDT-TAS, all available *in vitro* mechanistic data are collected and evaluated to record possible endocrine activity *via* the T-modality. A spectrum of *in vitro* assays has been developed to investigate whether substances may trigger the molecular initiating events (MIEs) or subsequent key events of the most important thyroid-related MoAs, including (1) inhibition of NIS, which mediates uptake of iodide into the thyroid gland as first step in thyroid hormone synthesis; (2) inhibition of TPO as enzyme mediating thyroid hormone synthesis; (3) induction of liver enzymes mediating thyroid hormone metabolism (predominantly: uridine diphosphate glucuronyltransferase (UGT)); and (4) displacement of thyroid hormones from serum binding proteins. The latter two key events may lead to enhanced thyroid hormone clearance and thus form part of indirect thyroid-related MoAs. By comparison, TPO inhibition and NIS inhibition, which both affect thyroid hormone synthesis, are MIEs for two direct thyroid-related MoAs. Less frequent MIEs of thyroid-related MoAs that can be investigated *in vitro* include the inhibition of different deiodinases (DIOs), as enzymes mediating thyroid hormone metabolism, and substance interaction with thyroid hormone receptors; see Table 3 for an overview of potentially relevant *in vitro* assays and Noyes et al. (2019) and Marty et al. (2021) for comprehensive reviews of the scientific evidence on the MIEs and key events of thyroid-related MoAs.

The *in vitro* assays are Level 2 assays in accordance with the OECD (2012) Conceptual Framework, i.e. "*in vitro assays providing data about selected endocrine mechanism(s)/pathways(s).*" Hence, the *in vitro* data may provide information that is relevant to determine (1) whether the substance of interest has endocrine activity (Point 1(b) of the EDC) and (2) to support the determination of its likely MoA (related to Point 1(c) of the EDC; Box 1).

Regarding regulatory applicability, the validation of *in vitro* assays to investigate thyroid activity is a matter of extensive research work worldwide. The EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) and the EU Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL) are engaged in the validation of *in vitro* assays targeting the MIEs and/or key events of all major thyroid-related MoAs (Zuang et al. 2022; JRC 2023). Details on the validation status of the *in vitro* thyroid methods are provided in the EURL ECVAM *Tracking System for Alternative Methods towards Regulatory Acceptance* (TSAR; https://tsar.jrc.ec.europa.eu/). As per 24 May 2023, the experimental parts of the validation have been finalised for *in vitro* assays addressing TPO inhibition, thyroid hormone glucuronidation, substance binding to transthyretin or thyroxin binding globulin, DIO1 activity and thyroid receptor

Table 3. Overview of *in vitro* assays, chemical methodologies and *in silico* models that allow investigating MIEs or key events of thyroid-related MoAs in mammals.

***In vitro* assay, chemical assay, *in silico* model, PBK model**	**MIE or key event addressed**	**Test system (from rat and human = facilitates H.R. assessment)**	**EU TSAR status [a]**	**EPA: HTS readiness [b]**	**References**
Radioactive-iodide uptake assay using human hNIS-HEK293T-EPA cells	NIS inhibition (MIE of AOP 54, 134)	Rat and human cell lines	NA	Existing	Hallinger et al. (2017) Wang et al. (2018)
Radioactive-iodide uptake inhibition screening assay using the FRTL-5 cell line			NA	Existing	Buckalew et al. (2020)
Activation of NIS based on Sandell–Kolthoff reaction		Human liver microsomes as per Zuang et al. (2022)	ThyM 2d: Val. ongoing	NA	TSAR [a]; Zuang et al. (2022); JRC (2023)
Amplex UltraRed-TPO assay	TPO inhibition (MIE of AOP 42)	Paul Friedman et al.: rat microsomes ThyM 2a: FTC-238 cells transfected with human recombinant TPO	ThyM 2a: Val. finalised, evaluation ongoing [c]	Existing	Paul Friedman et al. (2016); TSAR [a]; JRC (2023)
Guaiacol oxidation assay		Paul et al.: rat and pig microsomes	NA	NA	Paul et al. (2013)
Characterisation of distinct multiple catalytic activities of TPO (LC-MS/MS)		Tater et al.: rat microsomes			Tater et al. (2021)
Tyrosine iodination using LC	TPO-mediated tyrosine iodination (MIE of AOP 42)	FTC-238 cells transfected with human recombinant TPO	ThyM 2c: Val. finalised, evaluation ongoing [c]	NA	TSAR [a]; JRC (2023)
QSAR	TPO inhibition (MIE of AOP 42)	Limited opportunity for human relevance assessment	NA	NA	Gadaleta et al. (2021)
QSAR	TPO inhibition (MIE of AOP 42) NIS inhibition (MIE of AOP 54, 134) TR inhibition; DIO1/2/3 inhibition (MIEs in AOP network [d])		NA	NA	Garcia de Lomana et al. (2021) Dracheva et al. (2022)
PBK modelling for TPO inhibition	Modelling of AOs in response to potencies of MIEs for AOP 42, 54	Modelling of effects in rats	NA	NA	Handa et al. (2021) Hassan et al. (2017)
In vitro assays to address a substance's potential to displace TH from TTR, TBG (and/or albumin)	Interaction with serum binding proteins (MIE of AOP 152)	ThyM 3a: human TBG, TTR ThyM 3b: human TTR	ThyM 3a/3b: Val. finalised, evaluation ongoing [c]	Existing	Collet et al. (2020) Zuang et al. (2022) JRC (2023)
In vitro HTS assays to assess chemical binding and activation of specific nuclear receptors, which mediate liver enzyme induction	Liver nuclear receptor activation (MIE of AOP 8)	PXR (Rosenfeld et al.): liver samples from knock out mice; CAR (Maglich et al.): primary human hepatocytes; AhR (He et al.): plasmids transfected into mouse hepatoma Hepa1c1c7 cells	NA	Existing	Noyes et al. (2019) citing: He et al. (2011) Maglich et al. (2003) Romanov et al. (2008) Rosenfeld et al. (2003)
Comparative liver enzyme induction (Phase 2)	UGT induction (key event of AOP 8)	Primary rat and human hepatocytes	NA	Promising	Bomann et al. (2021) Parmentier et al. (2022) Tinwell and Bars (2022) Wiemann et al. (2023)
LC/MS assay addressing inhibition of TH glucuronidation	TH metabolism (relates to key events of AOP 8)	Human liver microsomes	ThyM 4b: Val. finalised, evaluation ongoing [c]	NA	Zuang et al. (2022) JRC (2023)

(*continued*)

Table 3. Continued.

In vitro assay, chemical assay, *in silico* model, PBK model	MIE or key event addressed	Test system (from rat and human = facilitates H.R. assessment)	EU TSAR status [a]	EPA: HTS readiness [b]	References
Thyroid organ-on-a-chip: spheroids of primary human (and/or rodent) thyroid microtissue	Functional assay covering TH synthesis (relates to key events of AOP 42, 54, 134)	Rat and human microtissues	NA	NA	Deisenroth et al. (2020) Moroni et al. (2020)
Liver-thyroid multi-organ-on-a-chip, rat and human	Functional assay covering TH synthesis and transport (relevant for all thyroid-related MoAs)		NA	NA	Karwelat et al. (2023) Kühnlenz et al. (2023)
PBK modelling to estimate serum/brain TH concentrations in rats vs humans	Modelling of functional assay for homeostasis of TH after exposure to TH disruptors (relevant for all thyroid-related MoAs)	Model developed to enable species differences assessments	NA	NA	Leonard et al. (2016)
In vitro inhibition of human DIO1	DIO inhibition (MIE in AOP network [d])	Rat or human liver microsomes	NA	Existing	Renko et al. (2015) Olker et al. (2019) Weber et al. (2022)
DIO1 activity based on Sandell–Kolthoff reaction	DIO activity (MIE in AOP network [d])	Human liver microsomes	ThyM 4a: Val. finalised, evaluation ongoing [c]	NA	TSAR [a]; JRC (2023)
In vitro inhibition of human DIO2 and DIO3	DIO inhibition (MIE in AOP network [d])	Lysate from recombinant cells	NA	Existing	Hornung et al. (2018) Olker et al. (2019)
Inhibition of MCT8 based on Sandell–Kolthoff reaction	TH transport through cell membranes (MIEs in AOP network [d])	MDCK cells stably transfected with human MCT8	ThyM 5a: Val. ongoing	Early	TSAR [a]; JRC (2023)
TR transactivation assays (ThyM 6a: human TRα and TRβ reporter gene transactivation – agonist activities ThyM 6b: CALUX human TRβ reporter gene transactivation – agonist and antagonist activities)	TR agonism/antagonism (MIE in AOP network [d])	ThyM 6a: engineered proprietary human HEK293 cells ThyM 6b: proprietary TRβ CALUX® cells (origin: human osteoblastic osteosarcoma U2OS line)	ThyM 6a: Val. finalised, evaluation ongoing [c] ThyM 6b: Val. ongoing	Existing	Paul-Friedman et al. (2019) TSAR [a]; JRC (2023)
TSH receptor activation based on cAMP measurement	Integrated tissue responses to alterations in TH levels (MIE in AOP network [d])	CHO-K1 cells transfected with human TSH receptor	ThyM 1b: Val. ongoing	Promising	TSAR [a]; JRC (2023)
Measurement of proliferation of rat pituitary-derived cell line GH3	Interference of test item with TH on cellular level (pituitary weight change)	Rat pituitary-derived cell line GH3	ThyM 8a: Val. ongoing	NA	TSAR [a]; JRC (2023) Jomaa et al. (2013)

AhR: aryl hydrocarbon receptor; AO(P): adverse outcome (pathway); CAR: constitutive androstane receptor; CHO: Chinese hamster ovary (cells); EURL ECVAM: European Union Reference Laboratory for alternatives to animal testing; FRTL: Fischer rat thyroid follicular (cell line); DIO: deiodinase; HTS: high-throughput screening; LC: liquid chromatography; MCT: monocarboxylate transporter; MDCK: Madin Darby Canine Kidney; MIE: molecular initiating event; MoA: mode-of-action; MS: Mass spectrometry; NA: not addressed; NIS: sodium – iodide symporter; PBK: physiologically based kinetic; QSAR: quantitative structure activity relationship; TBG: thyroid binding globulin; TH: thyroid hormone(s); ThyM: thyroid method (key word in TSAR, followed by method number as assigned in TSAR [a]); TPO: thyroid peroxidase; TR: thyroid receptor; TRH: thyrotropin-releasing hormone; TTR: transthyretin; UGT: uridine diphosphate glucuronyltransferase; Val.: validation.

[a] EURL ECVAM TSAR – Tracking System for Alternative methods towards Regulatory acceptance; https://tsar.jrc.ec.europa.eu/ [accessed 24 May 2023].

[b] Adapted from Table 2 in Noyes et al. (2019): "*In vitro HTS assay development denoted as "promising" indicates MIEs in the thyroid axis for which there is interest and/or activity in developing in vitro HTS approaches, typically with supportive in vivo and slow- or medium-throughput in vitro toxicity studies indicating chemical interactions. In vitro HTS readiness denoted as "early" indicates putative MIEs but with limited toxicity evidence and with little current activity to develop high-throughput alternatives.*"

[c] Validation Study Part 1 (Assessment of method robustness, reliability and within laboratory reproducibility using few test items) and Part 2 (Assessment of the method's mechanistic relevance using selected validation set chemicals covering the underlying mechanism of this method) finalised. Assessment of the data and study reports and decision if further validation work is needed in planning or ongoing; see JRC (2023) for details on the Thyroid Validation Study.

[d] See Figure 2 in Noyes et al. (2019): "*AOP network for chemically induced thyroid activity showing the integration of multiple individual AOPs under development and proposed.*"

transactivation (see Column "EU TSAR status" in Table 3). All finalised validation study reports and associated data are provided to an *ad hoc* OECD Expert Group for Thyroid Disruption Methods for data analysis and the decision if further validation work is needed (JRC 2023).

In the USA, the EPA has been using a broad spectrum of *in vitro* assays that address MIEs or key events for all major thyroid-related MoAs for high-throughput screening (HTS). Noyes and colleagues from the EPA have marked the "HTS readiness" as "existing" for *in vitro* assays addressing NIS, TPO and DIO inhibition, interaction with serum binding proteins, interaction with hepatic nuclear receptors and transactivation of nuclear thyroid hormone receptors (Noyes et al. 2019; Table 3; Column "EPA: HTS readiness"). These assays have been used within the EPA Endocrine Disruptor Screening Program (https://www.epa.gov/endocrine-disruption) and the EPA research effort ToxCast (https://www.epa.gov/chemical-research/toxicity-forecasting) [both websites accessed 2023 May].

The ECETOC T4 TF and CLE recommend considering all available relevant *in vitro* data for the Tier 0 evaluation, giving due consideration to the scientific validation status and applicability domains of the underlying methods. The evaluation of *in vitro* mechanistic data should consider the physico-chemical properties and analytical chemistry data of the test substance to ensure that it was amenable to screening and likely to be present in the test system. Importantly, *in vitro* effects should only be considered relevant if observed at non-cytotoxic doses. Further, if permitted by the available *in vitro* and *in vivo* database, *in vitro*-to-*in vivo* extrapolations and/or reverse dosimetry based on PBK modelling should be conducted to determine whether *in vitro* effects were observed at dose levels reflecting relevant *in vivo* plasma concentrations (see, e.g. Li H et al. 2017; Louisse et al. 2017). If *in vitro* effects only occurred at dose levels exceeding the internal/plasma concentrations recorded in the top dose groups from animal studies, the outcome is assessed as negative. If PBK modelling and/or *in vivo* toxicokinetics data are not available for the substance of interest, the ECETOC T4 TF and CLE recommend only using *in vitro* effects for the assessments if they were observed at or below the respective assay-specific top concentrations (e.g. $\leq$ 100 µM as used for ToxCast assays; Judson et al. 2016; Filer et al. 2017; Whalley et al. 2017; Franzosa et al. 2021) as threshold values to optimise the relevance of *in vitro* findings for human safety assessments.

Finally, when jointly evaluating *in vitro* data from different assays, it is important to note that these assays generally have high false positive rates (see, e.g. Paul Friedman et al. 2016; Noyes et al. 2019). This suggests high biological relevance of an overall negative outcome of a battery of *in vitro* assays.

Taken together, the ECETOC T4 TF and CLE recommend that all available and relevant *in vitro* data on thyroid activity should be considered within the Tier 0 WoE evaluations. Further, it is recommended that "sufficient *in vitro* database and no evidence for *in vitro* effects" may be concluded if negative *in vitro* data are available for MIEs or key events covering all major thyroid-related MoAs that are potentially relevant for the substance of interest. These may be TPO inhibition, NIS inhibition, UGT induction and interference with serum binding proteins. In addition, *in vitro* effects on DIO interaction and interference with thyroid receptors may need to be considered if deemed relevant based on *in silico* structural alerts (Section 2.1.4). As described above, HTS assays are available for all these events (Noyes et al. 2019). Even though such HTS assays have generally not yet been formally adopted (e.g. as OECD TG), the ECETOC T4 TF and CLE recommend seeking opportunities to facilitate their applicability within the Thyroid-NDT-TAS. For example, Ball et al. (2022) presented a framework for chemical safety assessment as pragmatic approach to integrate different types of data from TG-conforming and non-TG-conforming studies. This framework allows new *in silico* models and *in vitro* assays to be incorporated in the assessment scheme "*as they develop through continuous selective evolution rather than periodic revolution.*" Within this framework, a simple categorisation scheme for exposure estimation is suggested, which considers both the level and duration of exposure, and a tiered approach for hazard assessment. Importantly, the framework by Ball et al. includes derivation of regulatory decision-making outputs (i.e. classification/categorisation, limit values and exposure estimates) after each tier. Thereby, the level of precision required to make an appropriate safety assessment can be adapted to the given information needs (Ball et al. 2022). The ECETOC T4 TF and CLE suggest following the tiered chemical safety assessment framework described by Ball et al. (2022) when considering whether data from non-TG-conforming studies or assays may be useful for the WoE evaluations embedded in the Thyroid-NDT-TAS.

2.1.4. Tier 0: collection and WoE evaluation of available in silico data

The Tier 0 evaluation of the available data to decide on the need to enter the Thyroid-NDT-TAS should preferably also consider non-testing information from *in silico* modelling (see, e.g. Garcia de Lomana et al. 2021; Dracheva et al. 2022). Generally, *in silico* models are Level 1 assays in accordance with the OECD (2012) Conceptual Framework, i.e. "*existing data and non-test information.*"

In silico models to predict a substance's potential to trigger the MIEs of thyroid-related MoAs have not yet been formally accepted for regulatory purposes. Nonetheless, it is recommended to seek opportunities to integrate such approaches into regulatory assessments as this may serve the goal to replace, reduce and refine animal testing (Russell and Burch 1959; EP and Council 2010), for example by applying the Ball et al. (2022) framework to incorporate new *in silico* models and *in vitro* assays into assessment schemes (Section 2.1.3).

2.1.5. Tier 0: overarching WoE evaluation of all available data and conclusion

After separate WoE evaluations of the available *in vivo*, *in vitro* and *in silico* data, the final step of Tier 0 includes (1) an evaluation of the sufficiency of the *in vivo* thyroid-related

database, and (2) an overarching WoE evaluation of all available *in vivo*, *in vitro* and *in silico* data. Regarding the sufficiency of the *in vivo* database, the EFSA and ECHA (2018) Endocrine Disruptor Guidance states (p. 31):

> To have the T-mediated adversity with regard to humans and mammals (as non-target organisms) sufficiently investigated, the thyroid parameters foreseen to be investigated in the following studies OECD test guidelines 407, 408, 409 (and/or the one-year dog study, if available), 416 (or 443 if available) and 451-3 should have been measured and the results included in the dossier. If there is no indication of effects in these studies, the T modality is considered to be sufficiently covered.

The further clarifications that have since been published in the EFSA (2020) *Technical report on the outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology* have detailed that if no effects are observed in comprehensive histopathological assessments of the thyroid gland, this is generally sufficient to conclude that the EDC-T are not met even if serum thyroid hormone data are unavailable. As described in Section 2.1.2.3, this scenario may apply to substances whose database includes older OECD TG versions that did not yet include serum hormone measurements. In addition, the ECETOC T4 TF and CLE suggest that absence of thyroid-related findings in an enhanced OECD TG 421/422 reproduction and developmental toxicity screening test, together with absence of thyroid-related findings upon subchronic or longer-term substance exposure, may be sufficient to conclude that there is no indication for thyroid-related effects (even if *in vitro* assays were positive). In this context, enhanced implies that the reproduction and developmental toxicity study includes both serum thyroid hormones and thyroid histopathology.

If the *in vivo* thyroid-related database is sufficient and consistently indicates no substance-mediated effects on thyroid function, there is no evidence for endocrine activity *via* the T-modality (Point 1(b) of the EDC), and it is concluded that the EDC-T are not met. Accordingly, there is no need to enter the Thyroid-NDT-TAS.

If the *in vivo* thyroid-related database is not sufficient, or if *in vivo* thyroid-related effects were observed, all available *in vivo*, *in vitro* and *in silico* data are submitted to an overarching WoE evaluation, which should again follow the general principles for WoE evaluations (Section 2.1.1). This overarching WoE evaluation can yield four different outcomes:

- *In vivo*, *in vitro* and *in silico* thyroid-related data are consistently negative, but insufficient *in vivo* thyroid-related database: If this scenario applies and the *in vitro* database is complete and consistently provides negative outcomes for all potentially relevant thyroid-related MoAs (Section 2.1.3), there is no evidence for endocrine activity via the T-modality (Point 1(b) of the EDC), and it is concluded that the EDC-T are not met. This conclusion considers the high false positive rate of the *in vitro* assays, which enhances the biological relevance of an overall negative *in vitro* outcome.
- *In vivo* thyroid-related data are negative (insufficient *in vivo* thyroid-related database), *in vitro* and/or *in silico* data are positive: If this scenario applies, the further evaluation should consider which *in vitro* assays were positive and whether the discrepancy is explainable (e.g. by *in vitro*-to-*in vivo* extrapolations). If the *in vitro* response was observed at relevant concentrations, the WoE evaluation should aim to establish the level of confidence in the negative *in vivo* results. It may be considered, for example, whether serum thyroid hormone data are already available, and, if so, whether they were recorded at sensitive stages (e.g. pregnancy) and/or in the offspring. Further, the WoE evaluation may consider whether the compound is rapidly metabolised *in vivo*.
  - If the WoE evaluation conclusively shows that the *in vivo* negative findings are biologically more relevant than the *in vitro* positive findings (e.g. because discrepancies in the *in vitro*-to-*in-vivo* extrapolations can be shown), there is no evidence for endocrine activity via the T-modality (Point 1(b) of the EDC), and it is concluded that the EDC-T are not met.
  - If the WoE evaluation indicates that the *in vitro* positive findings may be biologically more relevant than the *in vivo* negative findings (e.g. because the *in vitro* assays were conducted using human test systems and species differences in the metabolism of the substance are suspected), the Thyroid-NDT-TAS should be entered, and the generation of additional data should preferably first focus on the conduct of comparative metabolism assays to further investigate such potential species differences.
  - If the WoE evaluation is inconclusive, the Thyroid-NDT-TAS should be entered. As relevant, the generation of additional data may first focus on the generation of ADME data to help explain the discrepancy between the negative *in vivo* studies and positive *in vitro* assays (e.g. when the compound is rapidly metabolised *in vivo*) and/or on further investigating the MoA that likely underlies the positive *in vitro* response(s). Preferably, orthogonal assays, i.e. assays including different endpoint detection methods, should be applied for this purpose (Section 2.3.4).
- *In vivo* thyroid-related data are positive, but *in vitro/in silico* data are negative: If this scenario applies, the Thyroid-NDT-TAS should be entered to evaluate the sufficiency of the *in vitro* database to assess all potentially relevant thyroid-related MoAs. Further, the WoE evaluation may consider whether *in vivo* findings were likely due to an active metabolite that was not formed *in vitro*. Accordingly, the added value of comparative *in vitro* metabolism assays, *in silico* modelling and/or read-across should be considered to investigate whether active metabolites recorded in animal studies may also be formed in humans.
- *In vivo*, *in vitro* and *in silico* thyroid-related data are consistently positive, i.e. they indicate a concern for endocrine disruption via the T-modality: If this scenario applies, the Thyroid-NDT-TAS should be entered to conduct a MoA and human relevance assessment and, possibly, to determine higher-tier testing needs.

Taken together, if the available thyroid-related data consistently indicate no concerns for endocrine activity *via* the T-modality, the EDC-T are not met, and the Thyroid-NDT-TAS is not entered. *Vice versa*, if the available thyroid-related data consistently indicate concerns for endocrine activity *via* the T-modality, the Thyroid-NDT-TAS is entered. In case of inconsistent findings, it should be the goal to explain these inconsistencies and to identify the need to generate further data when passing through the Thyroid-NDT-TAS. If the database is already complete (in accordance with the state-of-the-art) and the overall WoE evaluation is inconclusive (e.g. because findings are equivocal), an overall consideration of the entire hazard profile of the compound is recommended on a case-by-case basis before risk management is conducted.

2.2. Tier 1–3 of the Thyroid-NDT-TAS

Section 2.2.1 presents the scenario that the *in vivo* thyroid- and/or neurodevelopment-related database is insufficient at the onset of Tier 1 of the Thyroid-NDT-TAS and outlines the principal elements of Tier 1–3 for this scenario. By comparison, Section 2.2.2 presents the scenario that the *in vivo* thyroid- and neurodevelopment-related database is sufficient at the onset of Tier 1, i.e. that data from all potentially relevant OECD TGs are already available. These two scenarios were selected as illustrative examples to demonstrate how the Thyroid-NDT-TAS, that is not prescriptive, can be adapted to different information needs. In practice, the available database, just as the spectrum of thyroid- and neurodevelopment-related findings recorded at the onset of Tier 1 will vary on a case-by-case basis (Table Appendix 1). Expert judgement is required to determine how the Thyroid-NDT-TAS should be passed through, i.e. how it needs to be adapted to the given information needs.

2.2.1. Tier 1–3 of the Thyroid-NDT-TAS; example: in vivo thyroid- and/or neurodevelopment-related database insufficient at onset of Tier 1

Figure 3 presents the Thyroid-NDT-TAS illustrating the scenario that the *in vivo* thyroid- and/or neurodevelopment-related database is insufficient when entering Tier 1 of the Thyroid-NDT-TAS. To avoid, as far as possible, new animal testing, the Thyroid-NDT-TAS begins by conducting, in Tier 1, Step 1, a MoA and human relevance assessment and by completing in Tier 1, Step 2, the *in vitro*/*in silico* database, as relevant (Section 2.3). If the outcome of Tier 1 permits the conclusion that the EDC-T are either met or not met, the Thyroid-NDT-TAS is terminated. Only if the MoA is unclear and/or the human relevance of thyroid effects observed in the rat studies (and/or in studies using further species) cannot be excluded, the Thyroid-NDT-TAS is continued to Tier 2 to determine higher-tier testing needs to complete the *in vivo* thyroid- and/or neurodevelopment-related database. Further, the Tier 2 evaluations should consider, on a case-by-case basis, whether toxicokinetic data are relevant to establish, e.g. test substance concentrations and half-lives in the maternal, foetal and/or pup blood, placental/lactational substance transfer and/or the substance's potential to pass the blood-brain barrier.

As the first part of Tier 2, the *in vivo* thyroid-related database is completed. The measurement of serum levels of T4 in the offspring is considered pivotal for this purpose, possibly supplemented by offspring serum TSH and T3 data, as offspring serum hormone levels, at the end of gestation or during lactation, are more closely associated with neurodevelopmental outcomes than the corresponding maternal serum levels (Marty et al. 2022). *In vivo* studies to measure offspring serum thyroid hormone levels include the comparative thyroid assay (US EPA 2005) and the OECD TG 421/422 developmental toxicity tests if enhanced to include serum thyroid hormone measurements at suitable timepoints. The ECETOC T4 TF and CLE recommend considering such shorter-term studies to enhance the understanding of thyroid function before considering the EOGRTS (OECD TG 443) or the enhanced two-generation reproductive toxicity study (OECD TG 416) supplemented with thyroid and offspring brain parameters, as Level 5 assays (Section 2.1.2.1). Alternatively, a DNT study (OECD TG 426), enhanced with measurements of thyroid-related parameters, or a male/female pubertal assay, as Level 4 assays, may be considered useful. Generally, expert judgement is required to determine which test methods are best suited for the substance of interest while adhering to the 3Rs principle (EP and Council 2010).

If the selected studies do not show any effects on offspring serum thyroid hormone levels, endocrine activity for the T-modality (Point 1(b) of the EDC; Box 1) is not present in the offspring, and it is concluded that the EDC-T are not met. If, however, offspring serum thyroid hormones are statistically significantly altered, endocrine activity for the T-modality is likely present, and the next part of Tier 2 aims to complete the neurodevelopment-related database to determine whether the substance also causes an adverse effect (Point 1(a) of the EDC). Formally adopted TGs to assess neurodevelopmental effects include the Level 4 DNT study (OECD TG 426), and the Level 5 EOGRTS (OECD TG 443), which is the only OECD TG that mandates measurements of offspring serum thyroid hormone, thyroid histopathology and neurodevelopmental effects in one study. Following the parameters addressed in OECD TG 426 or 443, adverse neurodevelopmental effects that can possibly be observed in rats include, e.g. alterations in motor activity, acoustic startle response, learning and memory and/or brain morphology/histopathology (Section 2.2.3).

If the selected assessments provide no evidence for adverse neurodevelopmental effects, it is concluded that the EDC-T are not met since there is no adverse outcome. Of note, the Thyroid-NDT-TAS, as it is described in this article, only considers NDT as the adverse outcome of thyroid-related MoAs (Section 1.1). Further developmental manifestations of substance-mediated thyroid hormone imbalance (e.g. altered offspring growth, developmental delays, disrupted thermoregulation, increased mortality and early eye opening) are not further considered here but were discussed in Section 4.5 of Marty et al. (2022).

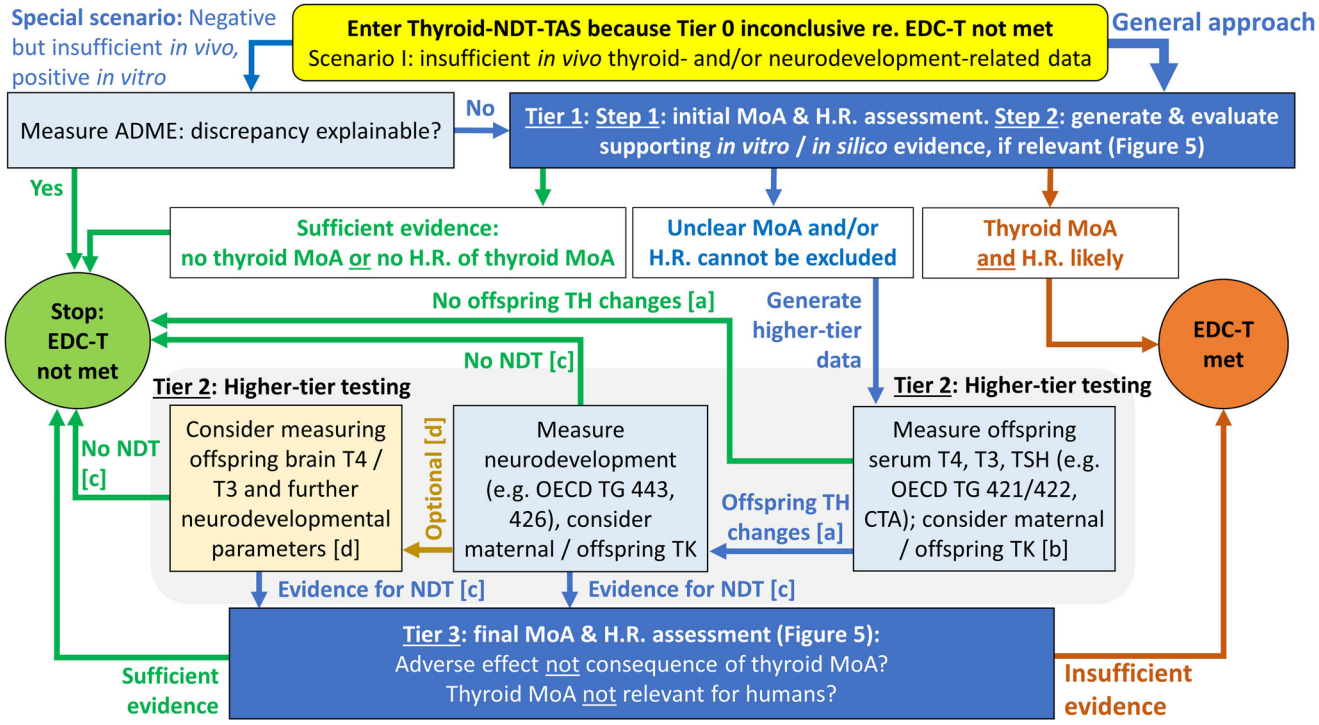

Figure 3. The ECETOC and CLE Thyroid-NDT-TAS. Scenario I: insufficient *in vivo* thyroid- and/or neurodevelopment-related data at onset of Tier 1.
ADME: absorption, distribution, metabolism, elimination; CTA: comparative thyroid assay; DNT: developmental neurotoxicity; EDC-T: endocrine disruptor criteria for thyroid modality; EOGRTS: extended one-generation reproductive toxicity study; H.R.: human relevance; MoA: mode-of-action; NDT: neurodevelopmental toxicity; OECD: Organisation for Economic Co-operation and Development; T3: triiodothyronine; T4: thyroxine; TG: test guideline; TK: toxicokinetics; TSH: thyroid stimulating hormone.
Colour legend: yellow shape: conclusion from Tier 0 evaluation to enter the Thyroid-NDT-TAS; dark blue boxes: MoA and human relevance assessment (Figure 5); light blue boxes: elements of the assessment; blue arrows: continuation of evaluation; ochreous box and arrows: optional elements of the assessment (as the respective parameters have not yet been formally validated or adopted for regulatory use); grey shading: elements of the higher-tier testing; red-brown vs green arrows and text: findings leading to conclusion that the EDC-T are met (red-brown circle)/are not met (green circle).
[a] Consider offspring serum T4, T3, and TSH thresholds observed by Marty et al. (2022) to support the determination of the biological relevance of findings (Section 2.1.2.3).
[b] Following expert judgement, further serum thyroid hormone data may not be necessary. Measurements of maternal and offspring plasma concentrations of the test material may be used to calculate placental transfer.
[c] See Section 2.2.3 for details on neurodevelopmental assessments.
[d] Following expert judgement, consider additional investigations using culled pups from the EOGRTS (or DNT study) and/or the performance of *in vitro* mechanistic assays and/or (not TG-conforming) perinatal studies to measure, e.g. brain thyroid hormones and/or receptor occupancies using immunohistochemistry. As relevant, consider measuring brain thyroid hormone already during the performance of the EOGRTS (or DNT study). Measurements of maternal and offspring plasma concentrations of the test material may be used to calculate placental transfer.

If, however, there is evidence for adverse neurodevelopmental effects, Tier 3 of the Thyroid-NDT-TAS addresses the final MoA and human relevance assessment (Section 2.4) to establish whether the adverse effect (Point 1(a) of the EDC) and the endocrine activity (Point 1(b) of the EDC) are linked by an endocrine MoA (Point 1(c) of the EDC).

2.2.2. Tier 1–3 of the Thyroid-NDT-TAS; example: in vivo thyroid- and neurodevelopment-related database sufficient at onset of Tier 1

Figure 4 presents the Thyroid-NDT-TAS illustrating the scenario that the *in vivo* thyroid- and neurodevelopment-related databases are sufficient at the onset of the Thyroid-NDT-TAS (and that *in vivo* thyroid effects were recorded in Tier 0). If this scenario applies, the Thyroid-NDT-TAS is generally passed through in the same manner as described in Section 2.2.1 for the scenario "*in vivo* database insufficient." However, since neurodevelopmental data are already available, these are assessed first (Section 2.2.3), and Tier 1 is only entered if adverse neurodevelopmental effects were observed or could not be excluded. If the Tier 1 MoA and human relevance assessment is inconclusive, the supportive higher-tier testing will generally be restricted to the measurement of non-guideline parameters since all potentially relevant TG-conforming data are already available.

2.2.3. Neurodevelopmental assessments within the thyroid-NDT-TAS

The ECETOC T4 TF and CLE recommend performing all neurodevelopmental assessments within the Thyroid-NDT-TAS in line with the provisions from the OECD TG 426 DNT study (the OECD TG 443 EOGRTS refers to the OECD TG 426 in this regard). The OECD TG 426 (version of 2007) includes substance administration during gestation and lactation and observations of the offspring up until adulthood, and it requests addressing these landmarks and endpoints (see Table 1 in OECD TG 426):

- Physical and developmental landmarks:
  - Body weight and clinical observations: weekly pre-weaning; thereafter, at least every two weeks
  - Brain weight and neuropathology: on postnatal day (PND) 22 and at study termination
  - Sexual maturation and other developmental landmarks, such as eye opening: as appropriate
- Functional/behavioural endpoints:

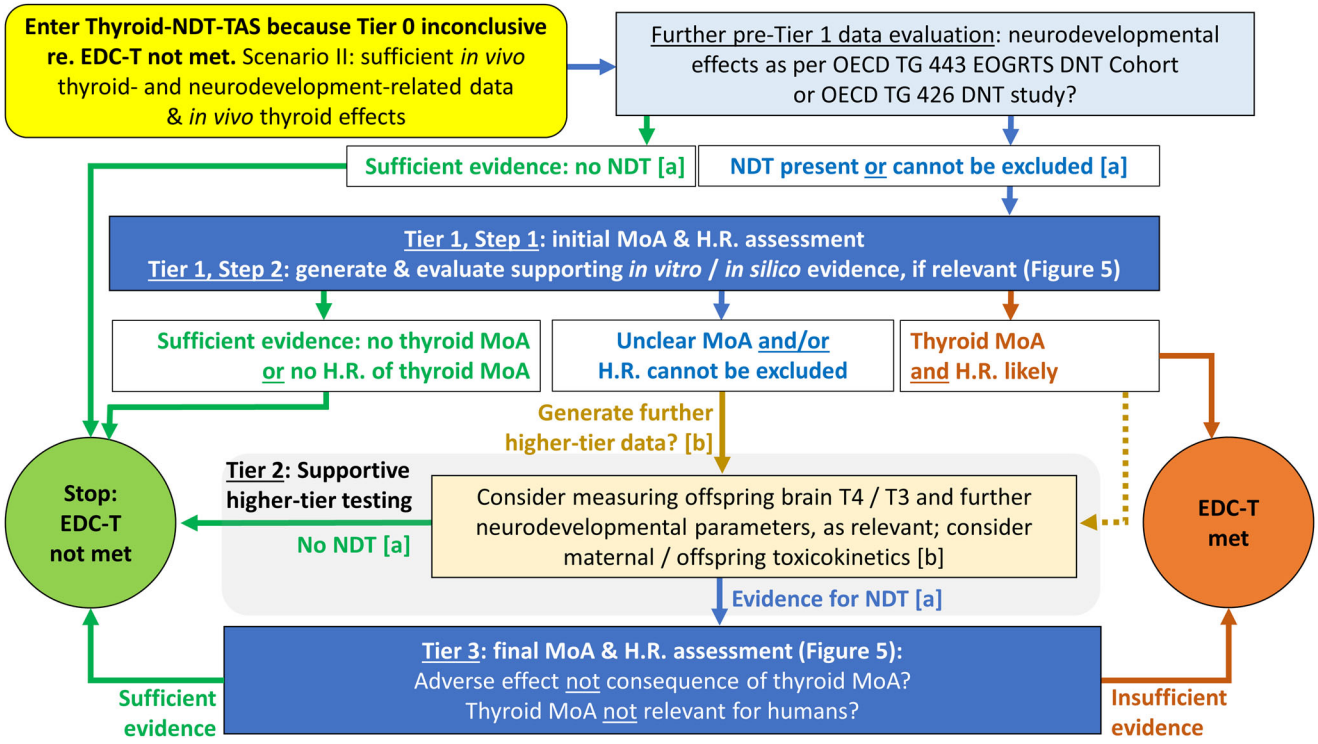

Figure 4. The ECETOC and CLE Thyroid-NDT-TAS: Scenario II: sufficient *in vivo* thyroid- and neurodevelopment-related data at onset of Tier 1.
DNT: developmental neurotoxicity; EDC-T: endocrine disruptor criteria for thyroid modality; EOGRTS: extended one-generation reproductive toxicity study; H.R.: human relevance; MoA: mode-of-action; NDT: neurodevelopmental toxicity; T3: triiodothyronine; T4: thyroxine; TG: test guideline.
Colour legend: yellow shape: conclusion from Tier 0 evaluation to enter the Thyroid-NDT-TAS; light blue boxes: elements of the assessment; blue arrows: continuation of evaluation; ochreous box and arrows: optional elements of the assessment (as the respective parameters have not yet been formally validated or adopted for regulatory use); dotted ochreous arrow: expert judgement that offspring brain thyroid hormones and/or further neurodevelopmental parameters are relevant to substantiate or rule out NDT; dark blue box: MoA and human relevance assessment (Figure 5); red-brown vs green arrows and text: findings leading to conclusion that the EDC-T are met (red-brown circle)/are not met (green circle).
[a] See Section 2.2.3 for details on neurodevelopmental assessments.
[b] Following expert judgement, consider additional investigations using culled pups from the EOGRTS and/or the performance of *in vitro* mechanistic assays and/or (not TG-conforming) perinatal studies to measure e.g. brain thyroid hormones and/or receptor occupancies using immunohistochemistry. Measurements of maternal and offspring plasma concentrations of the test material may be used to calculate placental transfer.

- Behavioural ontogeny: at least two measures pre-weaning
- Motor activity including habituation: 1–3 times pre-weaning, once in the young adults (PND 60–70)
- Motor and sensory function as well as learning and memory: once each in the adolescent offspring (PND 25 ± 2) and young adults (PND 60–70)

The ECETOC T4 TF and CLE consider all these landmarks and endpoints generally suitable to assess NDT. However, effects observed in the offspring should be put into context with, e.g. available data for adult animals across studies to determine whether such effects are due to systemic toxicity (including acute neurotoxicity) and not signs of developmental toxicity. Further, for substances with a known direct neurotoxic MoA, the Thyroid-NDT-TAS should be only entered if there is evidence that maternal hypothyroxinaemia has contributed to the neurotoxic effects in the offspring. For example, if thyroid function is only affected at higher dose levels than offspring neurotoxicity, the neurodevelopmental effect is unlikely to be caused by the thyroid perturbation (see Section 2.3.1.2 for evolved Bradford Hill consideration dose-response concordance).

Opportunities to test learning and memory (i.e. cognitive function) in rodents are limited. In the comprehensive review of 51 rat studies that included *in utero*/lactational exposure to substances causing thyroid hormone imbalance, Marty et al. (2022) found that:

> … 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.

This outcome is consistent with an analysis of the results from OECD TG 426 DNT studies investigating a total of 69 pesticides, which showed that the standard learning and memory tests were less sensitive than measures of motor activity and acoustic startle habituation (Raffaele et al. 2010). If testing for learning and memory is to be performed, expert judgement is required to select the most appropriate test that the given laboratory is equipped to run.

Heterotopias (mentioned in the above quote) are clusters of ectopic neurons in the brain indicative of altered neuronal migration (Goodman and Gilbert 2007; Gilbert et al. 2014; O'Shaughnessy et al. 2018, 2019). The potential of thyroid-active substances to elicit periventricular heterotopia in rats has been a matter of extensive research work that is being

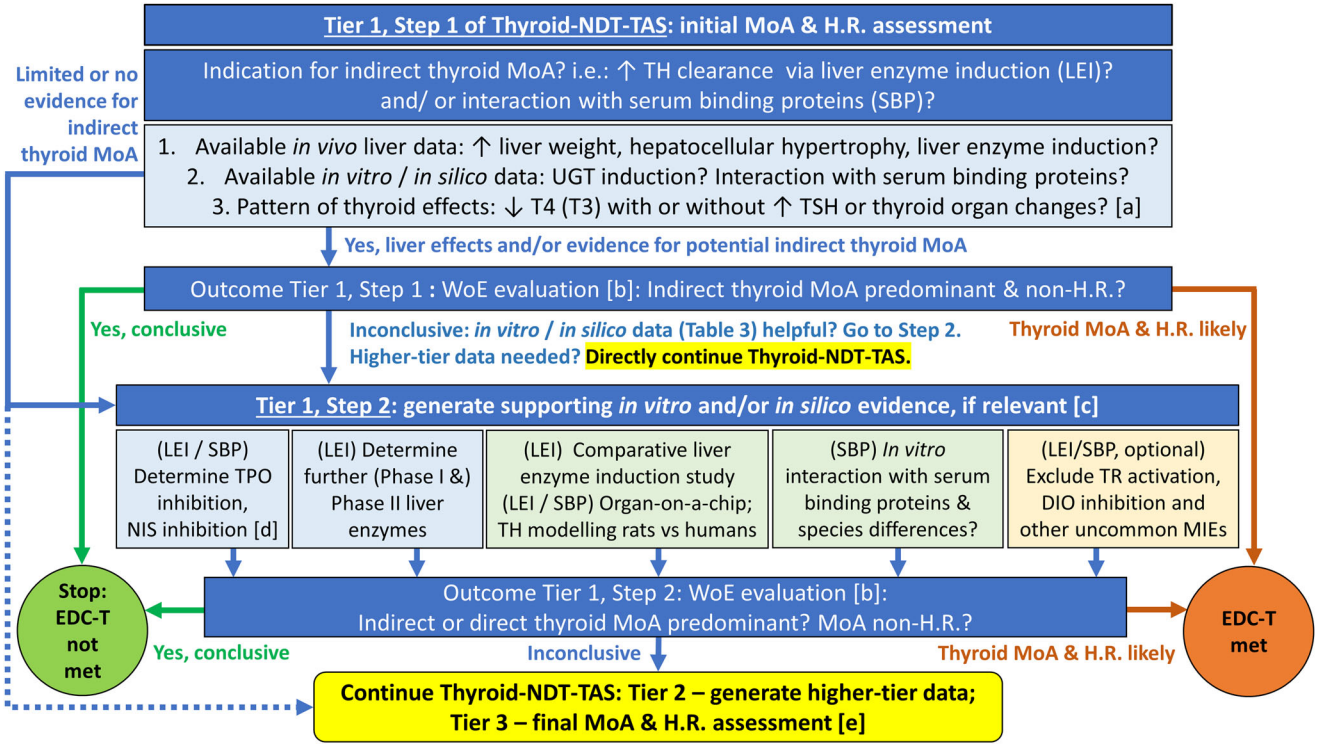

Figure 5. Decision-tree for MoA and human relevance assessment embedded in Tier 1 and Tier 3 of the ECETOC and CLE Thyroid-NDT-TAS.
DIO: deiodinase; EDC-T: endocrine disruptor criteria for thyroid modality; H.R.: human relevance; LEI: liver enzyme induction; MoA: mode-of-action; NIS: sodium – iodide symporter; SBP: serum binding protein; T3: triiodothyronine; T4: thyroxine; TH: thyroid hormone; TPO: thyroid peroxidase; TR: thyroid receptor (nuclear); TSH: thyroid stimulating hormone; UGT: uridine diphosphate glucuronyltransferase.
Colour legend: dark blue boxes: Step 1 and Step 2 of the MoA and human relevance assessment; light blue boxes: elements of the assessment; blue arrows: continuation of evaluation; dotted blue arrow: expert judgement that Step 2 of Tier 1 should be skipped to directly continue to Tier 2 to generate higher-tier data; light green boxes: optional elements of the assessment as the respective parameters have not yet been formally adopted for regulatory use; ochreous box: optional assessment as the corresponding MIEs seem to be less frequent; red-brown vs green arrows and text: findings leading to conclusion that the EDC-T are met (red-brown circle)/are not met (green circle). Yellow shape: continuation of Thyroid-NDT-TAS.
[a] TSH likely not increased, and no thyroid organ changes, if only in utero/developmental exposure to substances enhancing thyroid hormone clearance (Marty et al. 2022).
[b] See Section 2.3.1 for elements to consider during the WoE evaluation.
[c] Apply expert judgement to determine which type of supporting *in vitro* and/or *in silico* evidence may be relevant for the substance of interest.
[d] Primarily TPO and NIS inhibition need to be excluded, as thyroid-related parameters are similarly affected by substances acting via a direct thyroid-related MoA. Also, depending on the thyroid hormone effect pattern (e.g. increased serum T4), substance interaction with DIOs needs to be excluded.
[e] The final MoA and human relevance assessment shall serve to answer the questions: Is the adverse effect *not* a consequence of thyroid MoA? If the substance has a thyroid MoA, is it *not* relevant for humans?

led and coordinated by the EPA (see, e.g. Goodman and Gilbert 2007; Gilbert et al. 2014; O'Shaughnessy et al. 2018, 2019). While the assessment of periventricular heterotopia has not yet been standardised for regulatory use, brain histopathology performed around PND 16, or later up until adulthood in accordance with the respective TG, will inform on the presence of periventricular heterotopia provided that the histopathological sections include the respective brain location (Marty et al. 2022). To the best of the knowledge of the ECETOC T4 TF and CLE, the development of hypothyroid-mediated heterotopia has not yet been described in humans. Whereas cortical heterotopia has been associated with neuronal disorders and certain genetic mutations in afflicted patients, it is "*unknown what, if any, environmental influences are also linked to this malformation*" (O'Shaughnessy et al. 2018). Further, a causal relationship between heterotopias and serum thyroid hormone decrements has not been established, and control rat brains may also have heterotopias. It is also not clear whether neurodevelopmental disorders can be present without brain heterotopia. For all these reasons, the ECETOC T4 TF and CLE suggest caution in the interpretation of hypothyroid-mediated heterotopia in rats.

Generally, neuropathological evaluations (gross inspections, morphometry and histopathology of the brain) are important elements of neurodevelopmental assessments to identify substance-mediated structural abnormalities; however, the evaluation of neuropathological findings requires expert judgement (further discussed in Section 4.3.4 of Marty et al. 2022). Details on the interpretation of neuropathological evaluations are provided in Section 44 of the OECD TG 426 and references therein. The ECETOC T4 TF and CLE are aware of evidence relating the following neurodevelopmental insults (listed in Section 44 of the OECD TG 426) to maternal or offspring thyroid perturbation in rats. Hence, these histopathological findings appear relevant for the evaluation of NDT:

- Alterations in the relative size of various brain regions (e.g. external germinal layer of cerebellum, *corpus callosum*): rat studies by Li J et al. (1986) and Shibutani et al. (2009), see also reviews by Chen and Hetzel (2010) and Valdés Hernández et al. (2013); note that Shibutani et al. (2009) further recorded reduced numbers of oligodendroglial cells (reflecting impaired oligodendroglial

development) in the cerebral deep cortex of rats exposed to propyl thiouracil or methimazole.
- Alterations in neuronal migration and differentiation: rat studies by, e.g. Goodman and Gilbert (2007), Gilbert et al. (2014) and O'Shaughnessy et al. (2018, 2019).
- Alterations in patterns of myelination: rat studies by Salas-Lucia et al. (2020).

If assessments addressing the standard neurodevelopmental landmarks and endpoints that are included in the OECD TG 426 (or OECD TG 443) do not provide conclusive evidence on the presence or absence of adverse neurodevelopmental effects, the ECETOC T4 TF and CLE recommend considering additional neurodevelopmental assessments. For example, it might be considered to measure brain T4 and T3 in the culled pups from the DNT study or EOGRTS. It is important to note, however, that brain T4/T3 measurements have not yet been formally standardised and are currently not included in any formal TG. In the review by Marty et al. (2022), extensive data gaps compromised the establishment of associations between offspring brain T4/T3 decrements and the occurrence or absence of specific neurodevelopmental effects. Marty et al. (2022) concluded:

> 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 [i.e. studies that include a developmental component] evaluating thyroid-active substances [especially in prenatal/preweaning offspring]. 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.

The measurement of offspring brain T4/T3 levels is a matter of active research (O'Shaughnessy and Gilbert 2020; Ford et al. 2023; see also research project EMSG59 funded by the European Chemical Industry Council Long-Range Research Initiative (Cefic LRI) *Developing a quantitative AOP for liver-mediated thyroid modulation after prenatal exposure to a xenobiotic compound in the rat*; http://cefic-lri.org/projects/emsg59-developing-a-quantitative-aop-for-liver-mediated-thyroid-modulation-after-prenatal-exposure-to-a-xenobiotic-compound-in-the-rat/ [accessed 2023 May]). As measurements of brain T4/T3 levels become more established (and a historical control database is developed), this may enhance the understanding of how these parameters may contribute to regulatory assessments of NDT (see Table Appendix 2, which is included in this article after the bibliography, for relevant research needs). Further potentially relevant new approaches, that may complement serum T4 measures when evaluating NDT (but that have also not yet been standardised and validated for formal use), include immunohistochemical and/or gene expression assays to identify biomarkers for adverse brain-related effects (O'Shaughnessy and Gilbert 2020).

Taken together, expert judgement is required (1) to select the spectrum of standard neurodevelopmental parameters that appear relevant for the substance of interest; (2) to determine the feasibility and usefulness of non-standardised neurodevelopmental measurements and investigations; and (3) to conduct a WoE evaluation of all relevant data to determine whether, or not, NDT is present. Generally, the ECETOC T4 TF and CLE suggest that there is "sufficient" evidence to conclude on the absence of NDT if the spectrum of neurodevelopmental landmarks and endpoints listed in the OECD TG 426 DNT study or OECD TG 443 EOGRTS have been covered and the WoE evaluation provides no indication for statistically significant and biologically relevant adverse neurodevelopmental effects. Further, the ECETOC T4 TF and CLE recommend generally assessing NDT observed in rats as relevant for humans, unless it can be shown, e.g. that the underlying MoA is not relevant for humans (Section 2.3). Human observational studies addressing neurodevelopmental impairment secondary to maternal thyroid hormone imbalance include a broad spectrum of parameters including psychomotor and mental development, cognitive function (intelligence quotient), expressive vocabulary or educational attainment, and, in single studies, clinical diagnoses of autism or attention deficit hyperactivity disorder or brain morphology assessed by magnetic resonance imaging (see comprehensive review by Sauer et al. 2020). Of note, the outcomes of some of the human observational studies considered by Sauer and colleagues may have been influenced by non-consideration of confounding lifestyle and/or medical parameters when selecting mothers for the respective cohorts. While the neurodevelopmental assessments that can be included in rat studies are *per se* relatively insensitive (by addressing rather crude endpoints, including limited animal numbers per group and yielding high data variability), they are invariably conducted at high dose levels relative to possible human exposure.

2.3. Tier 1: initial MoA and human relevance assessment and generation and evaluation of supporting in vitro/in silico evidence

Figure 5 presents the decision-tree for the initial and final MoA and human relevance assessments that are embedded in Tier 1 and Tier 3 of the Thyroid-NDT-TAS, respectively. Tier 1 consists of two steps:

- Step 1: The initial MoA and human relevance assessment (Section 2.3.2). At this step, focus is on the identification of *indirect* thyroid-related MoA(s). Thereby, best possible use is made of the available *in vivo* database, which will generally include liver-related data. The outcome of Step 1 of Tier 1 may indicate:
  - That the EDC-T are met since a human-relevant thyroid-related MoA is likely, or that the EDC-T are not met since there is conclusive evidence that the substance has an indirect thyroid-related MoA that is not relevant for humans (Section 2.3.1.3 and Section 2.3.3). In these cases, the Thyroid-NDT-TAS is terminated after Tier 1, Step 1.
  - That the available evidence is inconclusive. In this case, expert judgement is required to determine, which specific additional data are needed to

determine the substance's MoA and to establish whether the MoA in rats is (not) relevant for humans.
    - If the expert judgement indicates that the need for *in vitro* and/or *in silico* data is predominant, Step 2 of Tier 1 is entered.
    - If the expert judgement indicates that the need for higher-tier data is predominant, Step 2 of Tier 1 is skipped and Tier 2 is entered.
- Step 2: The generation and evaluation of supportive *in vitro* and/or *in silico* evidence, as relevant (Section 2.3.4). Step 2 aims to complete the *in vitro*/*in silico* database by generating additional information that appears relevant to identify *both indirect and direct* thyroid-related MoAs. Just as the outcome of Step 1, the outcome of Step 2 may indicate:
    - That the EDC-T are either met or not met. In these cases, the Thyroid-NDT-TAS is terminated, and higher-tier testing is not necessary.
    - That the available evidence is still inconclusive. In this case, expert judgement is required to determine which higher-tier data may be relevant (Section 2.3.5).

If Tier 2 (higher-tier testing) is entered as an outcome of either Step 1 or Step 2 of Tier 1, the relevant *in vivo* data are identified, generated and evaluated to then proceed to Tier 3, which includes the final MoA and human relevance assessment and the final WoE evaluation (Section 2.4).

As a starting point for the detailed description of the MoA and human relevance assessments, important underlying elements are presented below, i.e. (Section 2.3.1.1) the adverse outcome pathway (AOP) concept; (Section 2.3.1.2) the evolved Bradford Hill considerations; and (Section 2.3.1.3) scientific evidence on the non-human relevance of indirect thyroid-related MoAs observed in rats.

2.3.1. Background information to MoA and human relevance assessment

2.3.1.1. Adverse outcome pathway (AOP) concept. The ECETOC T4 TF and CLE recommend applying information on AOPs to organise the MoA and human relevance assessments as the AOP concept has proven useful to address the biological plausibility of substance-specific MoAs and to enhance an understanding of toxicological effects (Marty et al. 2021). AOPs are linear sequences of events beginning with an MIE that may lead to early cellular events, followed by events in organs and organ systems and, ultimately, an observable adverse outcome in the organism, or population in case of ecologically relevant AOPs (Ankley et al. 2010; OECD 2017; Vinken et al. 2017). As compared to MoAs, AOPs are not substance-specific and thus do not include exposure or metabolism considerations (Marty et al. 2021). AOPs provide a structured approach to investigate the sequence of MIEs, key events and adverse outcomes that a substance may trigger. In biological reality, however, AOPs are hardly ever truly linear but embedded in e.g. converging events and feedback loops that aim at restoring balance (Knapen et al. 2018; Villeneuve et al. 2018).

In the central AOP repository "OECD AOP Wiki" (https://aopwiki.org [accessed 2023 May]), six AOPs relate to thyroid hormone imbalance and adverse neurodevelopmental outcomes in mammals:

- AOP 8: Upregulation of thyroid hormone catabolism via activation of hepatic nuclear receptors (leading to UGT induction), and subsequent adverse neurodevelopmental outcomes in mammals
- AOP 42: Inhibition of TPO and subsequent adverse neurodevelopmental outcomes in mammals (Crofton et al. 2019)
- AOP 54: Inhibition of NIS leading to learning and memory impairment (Rolaki et al. 2019)
- AOP 134: NIS inhibition and subsequent adverse neurodevelopmental outcomes in mammals
- AOP 152: Interference with thyroid serum binding protein transthyretin and subsequent adverse human neurodevelopmental toxicity
- AOP 300: Thyroid receptor antagonism and subsequent neurodevelopmental adverse outcomes in mammals

Hence, AOP 8 and 152 are relevant for the evaluation of indirect thyroid-related MoAs (liver enzyme induction and interaction with serum binding proteins); AOPs 42, 54 and 134 are relevant for the further evaluation of direct thyroid-related MoAs (TPO and NIS inhibition); and AOP 300 reflects events in the brain cells. Table Appendix 3, which is included in this article after the bibliography, presents the MIEs, key events and adverse outcomes of these AOPs (adapted from Table 1 in Marty et al. 2021). Table Appendix 3 also illustrates the (estimated) strength of evidence for key event relationships and their quantitative understanding. While AOP 42 and AOP 54 have been endorsed (Crofton et al. 2019; Rolaki et al. 2019), the other AOPs are marked in the OECD AOP Wiki as being "under development." Often, the evidence for key event relationships and their quantitative understanding has been tentatively indicated as "moderate" or "weak," especially for the later key events that lie closer to the adverse outcome (Table Appendix 3).

Further, Noyes et al. (2019) presented and discussed an "*AOP network for chemically induced thyroid activity showing the integration of multiple individual AOPs under development and proposed*" (Figure 2 in Noyes et al. 2019). This AOP network includes additional MIEs, such as:

- Activation of the hepatic nuclear receptors constitutive androstane receptor (CAR), aryl hydrocarbon receptor (AhR) and peroxisome proliferator-activated receptor (PPAR) – in addition to pregnane X receptor (PXR) activation, the MIE for AOP 8
- Substance interaction with the serum binding proteins albumin and thyroid binding globulin – in addition to interaction with transthyretin, the MIE for AOP 152
- DIO1, DIO2, DIO3 inhibition

Despite its recognised scientific limitations, the AOP concept provides a useful, structured framework for MoA and human relevance assessments. Therefore, the ECETOC T4 TF

and CLE recommend that all MoA and human relevance assessments within the Thyroid-NDT-TAS are structured following the sequence of MIE(s), key events and adverse outcome(s) of those AOP(s) that reflect the hypothesised MoA(s) (Section 2.3.3). Such an approach also stands in line with the EFSA (2019b) *Administrative guidance on submission of dossiers and assessment reports for the peer-review of pesticide active substances* and the EFSA (2020) *Technical report on the outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology*, which request that the relevant data should be transparently tabulated to evaluate the dose- and temporal-response.

2.3.1.2 Evolved Bradford Hill considerations. In addition to consideration of the AOP concept, the ECETOC T4 TF and CLE recommend conducting all MoA and human relevance assessments following pre-defined considerations. For example, the *WHO/IPCS MoA and Species Concordance Analysis Framework* includes five "evolved Bradford Hill considerations" (Meek, Boobis, et al. 2014; Meek, Palermo, et al. 2014; see also Table 2 in Meek, Palermo, et al. 2014):

- Consistency: "*Is the pattern of effects across species/strains/-organs/test systems what would be expected?*"
- Essentiality of key events: "*Is the sequence of events reversible if dosing is stopped or a key event prevented?*"
- Temporal concordance: "*Are the key events observed in hypothesised order?*"
- Dose-response concordance: "*Are the key events observed at doses below or similar to those associated with the end (adverse) effect?*"
- Biological concordance: "*Does the hypothesised MoA conflict with broader biological knowledge? How well established is the MoA in the wider biological database?*"

The elements for WoE analysis presented in the ECHA templates for WoE and uncertainty evaluation in risk assessment (Section 2.1.1) are also based on these evolved Bradford Hill considerations. Similarly, the EFSA (2019b) *Administrative guidance on submission of dossiers and assessment reports for the peer-review of pesticide active substances* includes an Appendix I *Template for presentation of the assessment of endocrine disrupting properties*, which lists predefined criteria that should be considered during the MoA analysis. These predefined criteria are biological plausibility and empirical support for each key event relationship (MIE to key event 1, key event 1 to key event 2, etc. up until final key event to adverse outcome), essentiality of each key event as well as consistency, analogy and specificity of the MoA.

2.3.1.3 Indirect thyroid-related MoA, a rodent-specific mechanism. Evidence in the scientific literature indicates that thyroid perturbations caused by substance-mediated induction of Phase II liver enzymes (mainly: UGT) and/or interaction with serum binding proteins are rodent-specific MoAs that are generally not relevant for humans (see, e.g. Papineni et al. 2015; Plummer et al. 2021; Strupp et al. 2020; Bomann et al. 2021; Parmentier et al. 2022; Tinwell and Bars 2022). In a comprehensive review of the underlying mechanistic evidence, Marty et al. (2021) showed that rats and humans differ considerably in the major thyroid hormone metabolism pathways. Generally, UGT-mediated glucuronidation is the major metabolic pathway in rats but only a minor pathway in humans whose most prominent route for thyroid hormone metabolism is by deiodination (Cavalieri and Pitt-Rivers 1981) and to a lesser extent by sulphotransferases (Richardson et al. 2014). See Table 3 for *in vitro* assays to evaluate species differences in the activity of these enzymes. Further research may elucidate other pathways that impact thyroid hormone metabolism.

Rats and humans also differ considerably in thyroid hormone distribution across the three major serum binding proteins (thyroid binding globulin, transthyretin and albumin) and in the binding affinities of these proteins (Marty et al. 2021). Also, none of the publications reviewed by Sauer et al. (2020) provided relevant information to establish a link between substance-mediated induction of those liver enzymes that are relevant for thyroid hormone metabolism in humans and thus increased serum thyroid hormone clearance, let alone to establish a further link between maternal hypothyroxinaemia and child neurodevelopmental impairment.

The EFSA (2020) *Technical Report on the outcome of the pesticides peer review meeting on recurring issues on mammalian toxicology* also notes that liver enzyme induction-mediated thyroid effects may not be relevant for humans (EFSA 2020, p. 8):

> … endocrine mediated adverse effects that are secondary to other toxicities (including liver toxicity) should not be considered for concluding that EDC are met. In this case, it is necessary to demonstrate by means of comparative MoA analysis that thyroid toxicity is secondary to e.g. liver toxicity. In the comparative MoA analysis, a MoA for thyroid toxicity and one for liver toxicity should be postulated in a comparative manner. The applicant should transparently tabulate the data in order to evaluate the dose- and temporal-response …
>
> The assessment of human relevance is mainly applicable to those cases where the T-mediated effect is through a liver-mediated mechanism i.e. liver enzyme induction resulting in an increase of THs [thyroid hormones] clearance. In this case, three pieces of information should be provided to evaluate whether the thyroid findings are likely or not to be human relevant: 1) analysis of T3, T4 and TSH in the repeated dose studies; 2) *in vitro* comparative studies to evaluate liver enzyme induction in the tested species (i.e. rat, mouse and dog) and humans; 3) evaluation of other potential *in vitro* mechanisms involved in the thyroid disruption. Finally, all the available evidences should be weighed, including interspecies differences and lack of any concomitant molecular initiating event.

This EFSA Technical Report also states that "… *EFSA confirmed that a CAR/PXR-mediated MoA that can also be expected to be functional in humans, leading to an increased clearance of THs would be considered relevant*" (EFSA 2020, p. 8). This statement from EFSA (2020) underlines that the non-human relevance of MoAs needs to be shown on a case-by-case basis.

In line with the available scientific evidence, the human relevance assessments that are embedded in the Thyroid-NDT-TAS focus on indirect thyroid-related MoAs that are

triggered by UGT induction and/or substance interaction with serum binding proteins.

2.3.2. Tier 1, Step 1: relevant data to establish indirect thyroid-related MoA

To make best possible use of the database that is available at the onset of the Thyroid-NDT-TAS, the Tier 1, Step 1, initial MoA and human relevance assessment (Figure 5) includes the evaluation of:

1. All available *in vivo* data on liver effects, including *in vivo* UGT induction data, if available
2. All available *in vitro* and/or *in silico* data on UGT induction and/or on substance interaction with serum binding proteins
3. The observed pattern of *in vivo* thyroid effects

Accordingly, Step 1 addresses the question whether the available evidence indicates that the substance of interest enhances thyroid hormone clearance *via* liver enzyme (UGT) induction and/or interaction with serum binding proteins and thus is likely to have an indirect thyroid-related MoA. As per Marty et al. (2022), some compounds trigger both MoAs, i.e. they induce liver enzymes and interact with serum binding proteins.

The available *in vivo* data on liver effects are jointly evaluated to determine whether adverse or adaptive liver effects are present, thereby indicating a concern for liver enzyme induction even if data on UGT induction are unavailable. Useful parameters for this evaluation include absolute and relative liver weight, gross necropsy and histological assessments of the liver as well as clinical pathology addressing the activities of, e.g. alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase and γ-glutamyltransferase (see, e.g. Hall et al. 2012). Molecular methods, such as receptor binding assays and newer toxicogenomics technologies, may also provide supportive evidence for liver enzyme induction-mediated liver toxicity (Hall et al. 2012). Further, Yamada et al. (2013) reported that substances with direct vs indirect thyroid-related MoAs could be discerned by the ratio of relative thyroid weight increase vs relative liver weight increase ($\geq$1.7 but generally $\geq$3.2 for substances with direct thyroid-related MoA and $\leq$1.2 for substances with indirect thyroid-related MoA; Yamada et al. 2013).

The available *in vitro* and/or *in silico* data on UGT induction and/or substance interaction with serum binding proteins are used as supporting evidence to determine whether an indirect thyroid-related MoA is likely. Further, if *in vitro* data from both rat and human test systems (cells, tissues, etc.) are already available during Step 1 of Tier 1, a comparative assessment of the findings can help identify species differences in effects.

The pattern of thyroid effects is considered to determine whether the liver enzyme induction coincides with serum T4 decrements and, possibly, increased TSH and follicular cell hypertrophy/hyperplasia of the thyroid gland. After at least 28-day exposure of adult rats, liver enzyme inducers generally not only cause serum T4 decrements, but these decrements also trigger a response of the hypothalamus-pituitary-thyroid axis, which is recognisable by serum TSH increases and altered thyroid weight/histopathology (see, e.g. Smith et al. 1991; Strupp et al. 2020; Parmentier et al. 2022).

However, for perfluorooctane sulphonates, Chang et al. (2007, 2008) hypothesised that these liver enzyme inducers do not trigger TSH increases since they also compete with T4 for serum binding proteins in rats. The resulting increase in free T4 would increase thyroid hormone availability to peripheral tissues for utilisation, metabolic conversion and excretion without direct interference with regulatory functions of the rat hypothalamus-pituitary-thyroid axis. Thereby, the rats would maintain an euthyroid state (Chang et al. 2007, 2008). This hypothesis further supports the relevance of thyroid histopathology as parameter for sustained and physiologically relevant thyroid hormone disruption in rats. Also, this hypothesis may help explain why Marty et al. (2022) observed offspring (and maternal) serum T4 decrements, but no TSH increases, upon gestational and lactational exposure of rats to substances that induce enhanced thyroid hormone metabolism and clearance and/or that affect T4 binding to serum carrier proteins [Marty et al. 2022, citing studies assessing perfluorohexane sulphonates (Ramhøj et al. 2018, 2020; Gilbert et al. 2021), the polychlorinated biphenyl Aroclor 1254 (Morse et al. 1996; Goldey et al. 1995) or DE-71, a mixture of polybrominated diphenyl ethers (Kodavanti et al. 2010; Bowers et al. 2015)]. In these studies, either no hypothyroid-mediated effects on neurodevelopment were observed, or such effects were less clear (e.g. absence of heterotopia) than observed for the substances with a direct thyroid MoA (Marty et al. 2022).

2.3.3. Outcome of Tier 1, Step 1: does substance have indirect thyroid-related MoA? If so, is it non-relevant to humans?

All data collected for the Tier 1, Step 1, initial MoA and human relevance assessment are submitted to a WoE evaluation to determine whether the available evidence is sufficient to conclude that the substance of interest has (or does not have) an indirect thyroid-related MoA. If liver effects are observed at lower doses and earlier timepoints than the thyroid effects, enhanced thyroid hormone clearance *via* liver enzyme induction is likely to be a relevant MoA (as the modified Bradford Hill considerations temporal concordance and dose-response concordance are fulfilled; Section 2.3.1.2). The sensitivity of the investigated liver parameters vs. thyroid parameters should always be considered before a final assessment of the temporal concordance and dose-response concordance of the events can be made.

If the outcome of Tier 1, Step 1, indicates that the substance likely has an indirect thyroid-related MoA, it is assessed whether this MoA is (not) relevant for humans. As noted in Section 2.3.1.1, the ECETOC T4 TF and CLE recommend that this MoA and human relevance assessment is structured following the sequence of events of the AOP(s) that reflect the hypothesised MoA(s). For the assessment of an indirect thyroid-related MoA, Figure 6 illustrates the sequences of events that lead from (1) hepatic nuclear

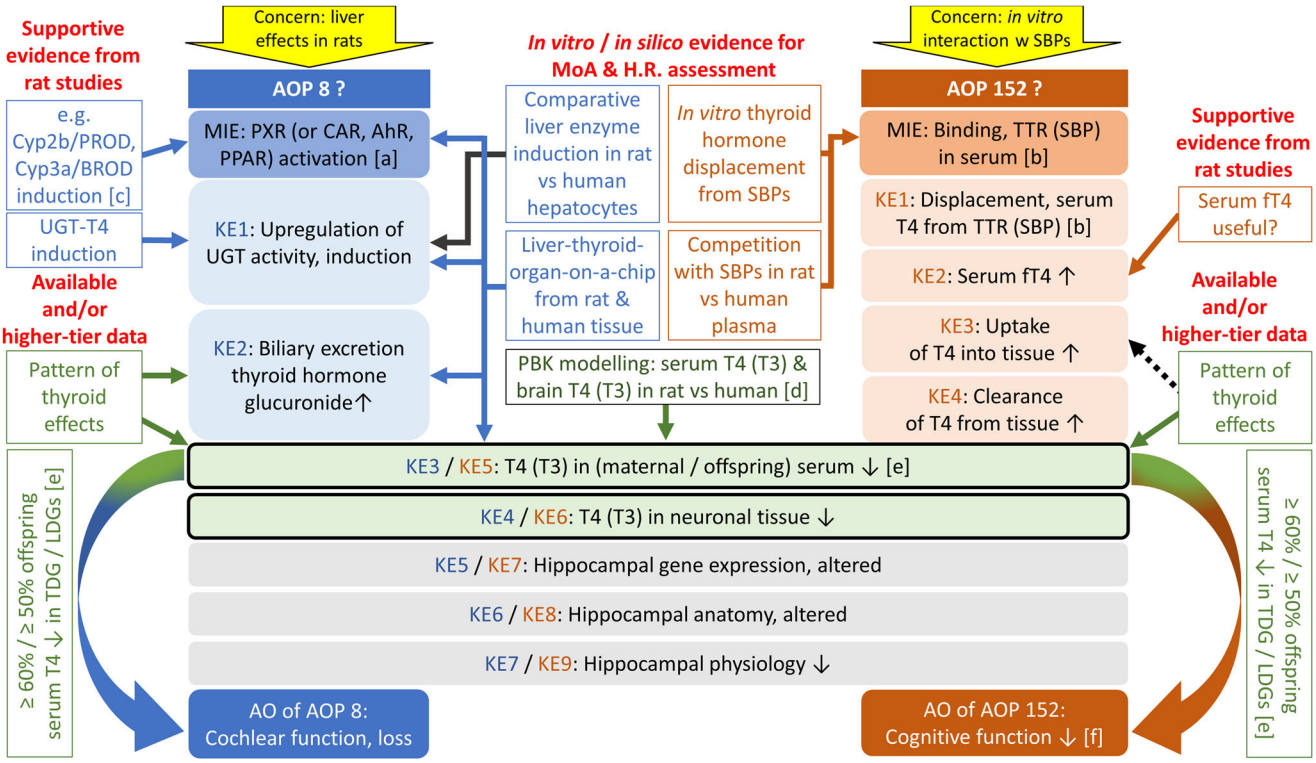

Figure 6. Sequence of events that may lead from liver enzyme induction (AOP 8) and interaction with serum binding proteins (AOP 152) to adverse neurodevelopmental outcomes in mammals and opportunities for their further investigation.

AhR: aryl hydrocarbon receptor; AO(P): adverse outcome (pathway); BROD: benzoxyresorufin; CAR: constitutive androstane receptor; Cyp: cytochrome p-450; fT4: free thyroxine; H.R.: human relevance; KE: key event; LDG: lower-dose groups; MIE: molecular initiating event; NIS: sodium – iodide symporter; PBK: physiologically based kinetic; PPAR: peroxisome proliferator-activated receptor; PROD: pentoxyresorufin; PXR: pregnane X receptor; SBP: serum binding protein; T3: triiodothyronine; T4: thyroxine; TDG: top-dose group; TSH: thyroid stimulating hormone; TTR: transthyretin; UGT: uridine diphosphate glucuronyltransferase.

Colour legend: yellow arrow: available data indicating that AOP 8/AOP 152 may be relevant for the MoA and human relevance assessment. Boxes with blue/red-brown shading: MIE, early KEs and AOs that relate to AOP 8/AOP 152. White boxes with blue/red-brown text: Supportive *in vivo* or *in vitro* evidence that may inform on the MIE or specific KEs for AOP 8/AOP 152 (linked by blue/red-brown arrows; black arrow for enhanced traceability). Boxes with green shading: KEs relating to serum/brain T4 decrements; these KEs are central to five of the six potentially relevant AOPs included in the OECD AOP Wiki (exception: AOP 300 on thyroid receptor antagonism; Table Appendix 3). White box with green text: *In vivo* data or PBK modelling to inform on thyroid-related events. Boxes with grey shadings: KEs relating to the hippocampus; these KEs are central to five of the six potentially relevant AOPs (exception: AOP 54 on NIS inhibition leading to impaired learning and memory). Dotted black arrow: T4 measurements in relevant tissues, if available.

[a] In the OECD AOP Wiki, the MIE of AOP 8 is recorded as PXR activation. Noyes et al. (2019) indicated CAR, AhR and PPAR activation as further MIEs leading to liver enzyme induction. The ECETOC T4 TF and CLE contend that all these MIEs are not indispensable to trigger UGT upregulation. Therefore, assessments addressing hepatic nuclear receptor activation may not be needed for the MoA assessment of the substance of interest.

[b] While AOP 152 only refers to TTR, the available *in vitro* assays generally allow measuring all three major serum binding proteins, i.e. TTR, albumin and thyroid binding globulin. Since thyroid hormone distribution across these three serum binding proteins and their binding affinities differ considerably between rats and humans, the ECETOC T4 TF and CLE recommend considering *in vitro* substance interaction with all three serum binding proteins, as relevant.

[c] See Tinwell and Bars (2022) for details on the indirect assessment of CAR/PXR activation in rat studies via induction of transcript level and corresponding enzyme activity associated with each receptor (Cyp2b/PROD and Cyp3a/BROD for CAR and PXR, respectively).

[d] PBK modelling: Estimate serum/brain T4 levels in rat vs human considering relevant parameters, such as binding constants, potencies of MIEs and/or liver enzyme inductions in rat vs human tissue.

[e] The AOPs in the OECD AOP Wiki (Table Appendix 3) only generally refer to "T4 in serum, decrease" without distinction between maternal and offspring serum T4 levels; also, none of the AOPs considers serum T3 (or TSH). Following the observations by Marty et al. (2022), maternal serum T4 levels do not appear predictive of neurodevelopmental effects. However, there seems to be some association between ≥ 60%/≥ 50% offspring serum T4 decrements in the TDG/LDGs (and ≥ 20% *and* statistically significant offspring serum T3 decrements) and the occurrence of statistically significant neurodevelopmental effects. Therefore, the ECETOC T4 TF and CLE recommend considering offspring serum T4 as predominant parameter related to serum thyroid hormone levels. In addition, information on maternal serum T4, maternal and/or offspring serum T3 and offspring brain T4/T3 should be considered, if available (see Section 2.1.2.3 for further discussion).

[f] For AOP 152 (as well as AOP 42 and AOP 134), "cochlear function, decreased/loss" was indicated as adverse outcome in the OECD AOP Wiki as per 13 September 2019, whereas it was indicated as "cognitive function, decreased" as per 15 October 2019.

receptor activation and liver enzyme induction or (2) substance interaction with serum binding proteins to adverse neurodevelopmental outcomes, i.e. the sequences of events that reflect AOP 8 and AOP 152, respectively (Section 2.3.1.1). Figure 6 also shows which *in vivo* or *in vitro* data may be useful to determine whether, and at which dose level, the respective event was triggered by the substance of interest. When applying test systems using cells or tissues from both rat and human origin or when conducting PBK modelling to compare estimated thyroid hormone levels in rat vs human (Leonard et al. 2016), the *in vitro* assessments/PBK models can also be used to determine species concordances or differences of effects (see Table 3; Column "Test system, from rat & human = facilitates H.R. assessment").

For each hypothesised MoA, the (rat and/or human) evidence indicating that the MIE and subsequent key events were triggered (and at which dose level) should be recorded. Tinwell and Bars (2022) and Parmentier et al. (2022) have shown how relevant information can be sorted by MIE and key event to support the MoA and human relevance assessment for substances that exhibit an indirect thyroid-related MoA in rats.

2.3.4. Tier 1, Step 2: generation of supportive in vitro/in silico evidence

Tier 1, Step 2, of the Thyroid-NDT-TAS serves to complete the *in vitro*/*in silico* database, as relevant for the substance of interest (see Section 2.1.3 for definition of "sufficient" *in vitro* database). Such supportive *in vitro* and/or *in silico* evidence may be relevant (1) to inform on the MoA(s) of the substance of interest, and (2) to establish whether thyroid-related MoAs observed in rats are (not) relevant for humans. Generally, expert judgement is required to select suitable *in vitro* assays, *in silico* models and/or PBK models (Table 3; Section 2.1.3) and to evaluate the corresponding findings.

If (additional) evidence is required to establish or rule out a direct thyroid-related MoA, *in vitro*/*in silico* data on TPO and NIS inhibition are helpful; however, the high false positive rate of the corresponding assays needs to be considered during the evaluation of any positive findings (Section 2.1.3). As relevant, orthogonal assays using different endpoint detection methods can be used to support preliminary evidence that the substance may trigger a specific MIE. For example, if an Amplex UltraRed TPO assay indicates that the substance may inhibit TPO, this concern may be followed up by assessing tyrosine iodination *via* liquid chromatography. Similarly, if a radioactive-iodide uptake assay indicates that the substance may inhibit the NIS, this concern may be followed up by assessing NIS activation based on the Sandell–Kolthoff reaction (see Table 3 for assay references).

If (additional) evidence is required to establish or rule out an indirect thyroid-related MoA (e.g. to follow up on adverse liver effects observed in the available rat studies), comparative liver enzyme induction assays and/or liquid chromatography (coupled with mass spectrometry) assays measuring inhibition of thyroid hormone glucuronidation and/or suitable PBK modelling may be relevant (see Table 3 for assay references).

2.3.5. Outcome of Tier 1, Step 2: does substance have direct or indirect thyroid-related MoA? If so, is it non-relevant to humans?

As an outcome of Tier 1, Step 2, a further WoE evaluation is conducted that generally follows the same outline as the WoE evaluation conducted as an outcome of Tier 1, Step 1, with the exception that the *in vitro*/*in silico* database is now complete so that the overall evaluation can address both indirect and direct thyroid-related MoAs. If the findings indicate that more than one (thyroid-related or non-thyroid-related) MoA may be triggered, the WoE evaluation should consider which link between thyroid activity and adverse outcome appears most plausible – and thus which MoA is likely predominant. For example, if the substance of interest causes *in vitro* NIS inhibition at very high concentrations, but thyroid effects are not observed at comparable *in vivo* dose levels, the modified Bradford Hill criterion of dose-response concordance is not met, and the MIE NIS inhibition is assessed as not biologically relevant. Similarly, *in vitro* effects that are recorded at concentrations ranging close to the cytotoxic concentrations are weighted lower than *in vitro* effects recorded at lower concentrations.

2.4. Tier 2 and Tier 3: higher-tier testing and final MoA and human relevance assessment and final conclusion regarding EDC-T

If Tier 1 of the Thyroid-NDT-TAS is inconclusive, Tier 2 of the Thyroid-NDT-TAS can be entered to generate and evaluate relevant higher-tier data. If the higher-tier data indicate that the substance of interest does not elicit *both* offspring serum T4 (T3 and TSH) changes *and* NDT, the final conclusion is drawn that the substance of interest does not meet the EDC-T (since it does not have endocrine activity in the offspring and/or does not cause an adverse effect). If, however, both thyroid perturbations in the offspring and adverse neurodevelopmental effects are recorded, Tier 3 of the Thyroid-NDT-TAS is entered, and the final MoA and human relevance assessment is performed. The overarching WoE evaluation of all available and new data is used to answer the questions whether a (direct or indirect) thyroid-related MoA is predominant and whether the MoA is (not) relevant in humans. If this overall evaluation indicates that the substance likely has a thyroid-related MoA as predominant MoA and human relevance of this MoA cannot be ruled out, the final conclusion is drawn that the EDC-T are met.

3. Conclusions and outlook

A comprehensive and structured approach to assess whether active substances in plant protection products, biocidal products and REACH substances meet the EDC-T is currently unavailable. To address this shortcoming, the ECETOC T4 TF and CLE have now proposed a Thyroid Function-Related Neurodevelopmental Toxicity Testing and Assessment Scheme (Thyroid-NDT-TAS). The Thyroid-NDT-TAS provides a structured, tiered approach to determine (1) whether substances elicit adverse neurodevelopmental effects and (2) have thyroid activity and (3) whether the two are linked by an indirect or direct thyroid-related MoA, and thus meet the EDC-T, unless it can be shown that the MoA is not relevant for humans. The Thyroid-NDT-TAS is based on the state-of-the-science, and it has been developed to comply with the European Commission (2017, 2018) EDC and the EFSA and ECHA (2018) Endocrine Disruptor Guidance. Further, the Thyroid-NDT-TAS takes into consideration knowledge gaps that have been identified in the earlier ECETOC T4 TF reviews (Sauer et al. 2020; Marty et al. 2021, 2022); see Table Appendix 2 for details on research needs. In all tiers of the Thyroid-NDT-TAS, prevailing uncertainties may be addressed by a WoE approach if multiple data are available or, if possible, by the generation of new data. If the database is complete (in accordance with the state-of-the-art) and the overall WoE evaluation is inconclusive (e.g. because findings are equivocal), the entire hazard profile of the compound should be considered on a case-by-case basis before risk management is conducted. To make human safety assessments more accurate while at the same time minimising animal testing, it is recommended that registrants and regulators apply the Thyroid-NDT-TAS for future regulatory assessments.

Acknowledgements

We would like to thank the members of the ECETOC Scientific Committee as well as the in-house reviewers (see Declaration of competing interests) for their critical comments. Also, we would like to thank the former ECETOC T4 TF chair Nina Hallmark (Bayer AG, Germany) as well as the former TF members Davy Guignard (Evotec France, France), Bethany R. Hannas (Eli Lilly & Co., USA) and Daniel Urbisch (Syngenta, Switzerland) for valuable discussions that contributed to the basis for this review. We are indebted to Andreea Cuciureanu, Andrea Salvadori, Francesca Ugucchioni and the entire ECETOC secretariat for organisational and technical assistance to the ECETOC T4 TF. We thank the reviewers selected by the Editor that were anonymous to us. The comprehensive and helpful comments served to improve the manuscript.

Declaration of interest

The co-authors of this manuscript consist of members of the ECETOC T4 TF and/or of CLE (SMK, KB, AC, BG, RG, SJ, SM, HAM, EFM, LPS, CS, HT, ChW), the scientific writer (UGS) and Stewards from the ECETOC Scientific Committee (PAB, BvR). 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.

Stephanie Melching-Kollmuss (SMK), Heike-Antje Marxfeld (HAM) and Christiane Wiemann (ChW) are employed by BASF SE, Limburgerhof, Germany, BASF SE, Ludwigshafen, Germany, and BASF Oesterreich GmbH, Vienna, Austria, respectively. 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 few requests for amendments, which served clarification. SMK's responsibilities within BASF SE include being regulatory toxicologist for agrochemicals. SMK is the current Chair of the ECETOC T4 TF. HAM's responsibilities within BASF include pathological evaluation and assessment of a broad variety of regulatory studies conducted by BASF. ChW's responsibilities within BASF include being a regulatory toxicologist and human risk assessor for agrochemicals. Also, ChW is contributing to the CLE ED Thyroid Subgroup.

Kathrin Bothe (KB), Babunilayam Gangadharan (BG), Larry P. Sheets (LPS) and Helen Tinwell (HT) are employed by the Crop Science Division of Bayer AG (Germany/USA) or Bayer SAS (France). The Bayer portfolio includes substances that may have to be tested for their potential to cause maternal thyroid disruption and subsequent developmental neurotoxicity, including some of those mentioned in this paper. An in-house review of this manuscript yielded no requests for amendments. KB's responsibilities include being a regulatory toxicologist at the Bayer Crop Science Division (Germany). BG's responsibilities include pathological evaluation and assessment of a broad variety of regulatory studies conducted on Bayer compounds at the Bayer Crop Science Division (France). LPS's responsibilities include being Senior Fellow Regulatory Toxicology at the Bayer Crop Science Division. HT's responsibilities include being Regulatory Toxicology Team Leader at the Bayer Crop Science Division (France).

Alex Charlton (AC), Elizabeth F. McInnes (EFM) 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 amendments. AC's responsibilities within Syngenta include providing scientific support to research and development activities and to regulatory toxicology projects. EFM's responsibilities within Syngenta include providing peer review of all laboratory animal pathology data generated by Syngenta and providing expert advice on the possible adversity of all pathology findings to the global toxicology platform within Product Safety. 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. 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 amendments. RG's responsibilities within Corteva include providing scientific support to developmental and reproductive toxicology testing strategy and to research and development projects.

Sylvia Jacobi (SJ) is a freelance consultant and was employed by Albemarle Europe SRL as Corporate Toxicology Director until July 2022. SJ still provides scientific advice to this company. Albemarle is a chemical company that markets or previously marketed products containing some of the chemicals included in this paper. Its portfolio also includes substances that may have to be tested for thyroid disruption and subsequent neurotoxicity.

Sue Marty (SM) is employed by Dow, Inc. 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 no requests for amendments. SM's role at Dow is focused 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.

Christian Strupp (CS) is employed by Gowan Crop Protection Ltd., Reading, United Kingdom. Gowan is a company manufacturing and marketing plant protection products. An in-house review of this manuscript yielded few requests for amendments, which served clarification. CS's responsibilities include all aspects of human safety testing of Gowan's portfolio. CS was also the chair-elect of CLE's Human Health Expert Group (H2EG) at the time of manuscript compilation. CLE's H2EG contributes to improving the process of plant protection product safety assessment by scientific contributions and constructive commenting to consultations on regulatory processes.

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.

This manuscript was reviewed by the ECETOC Scientific Committee consisting of representatives of academia, regulatory agencies and industry (https://www.ecetoc.org/about-ecetoc/structure/scientific-committee/). This review yielded few requests for amendment, which served clarification.

Finally, all authors suggest that the proposed Thyroid-NDT-TAS, which was developed for regulatory purposes, should be used in future regulatory actions – both within the EU and in other jurisdictions.

This manuscript relates to work undertaken by members of the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) T4 TF and/or CropLife Europe (CLE). 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. CLE (https://croplifeeurope.eu/) represents company members that develop and supply pesticides and biopesticides and/or invest in precision applications and plant biotechnology traits, as well as national organisations related to these areas. CLE is financed by its membership.

ORCID

Kathrin Bothe http://orcid.org/0000-0001-6470-8882
Babunilayam Gangadharan http://orcid.org/0000-0001-6243-7660

Rashin Ghaffari http://orcid.org/0000-0002-3209-9549
Sylvia Jacobi http://orcid.org/0000-0003-0986-1467
Sue Marty http://orcid.org/0000-0002-5404-8515
Heike-Antje Marxfeld http://orcid.org/0000-0001-5390-0633
Ursula G. Sauer http://orcid.org/0000-0002-0352-2985
Larry P. Sheets http://orcid.org/0000-0001-6868-5948
Christian Strupp http://orcid.org/0000-0002-7001-1659
Helen Tinwell http://orcid.org/0000-0002-3165-4619
Christiane Wiemann http://orcid.org/0000-0002-9695-2270
Bennard van Ravenzwaay http://orcid.org/0000-0002-4264-9217

References

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem. 29(3):730–741. doi: 10.1002/etc.34.

Ball N, Bars R, Botham PA, Cuciureanu A, Cronin MTD, Doe JE, Dudzina T, Gant TW, Leist M, van Ravenzwaay B. 2022. A framework for chemical safety assessment incorporating new approach methodologies within REACH. Arch Toxicol. 96(3):743–766. doi: 10.1007/s00204-021-03215-9.

Bartsch R, Brinkmann B, Jahnke G, Laube B, Lohmann R, Michaelsen S, Neumann I, Greim H. 2018. Human relevance of follicular thyroid tumors in rodents caused by non-genotoxic substances. Regul Toxicol Pharmacol. 98:199–208. doi: 10.1016/j.yrtph.2018.07.025.

Beekhuijzen M, Rijk JCW, Meijer M, de Raaf MA, Pelgrom S. 2019. A critical evaluation of thyroid hormone measurements in OECD test guideline studies: is there any added value? Reprod Toxicol. 88:56–66. doi: 10.1016/j.reprotox.2019.07.014.

Beyer BK, Chernoff N, Danielsson BR, Davis-Bruno K, Harrouk W, Hood RD, Janer G, Liminga UW, Kim JH, Rocca M, et al. 2011. ILSI/HESI maternal toxicity workshop summary: maternal toxicity and its impact on study design and data interpretation. Birth Defects Res B Dev Reprod Toxicol. 92(1):36–51. doi: 10.1002/bdrb.20281.

Bomann W, Tinwell H, Jenkinson P, Kluxen FM. 2021. Metribuzin-induced non-adverse liver changes result in rodent-specific non-adverse thyroid effects via uridine 5′-diphospho-glucuronosyltransferase (UDPGT, UGT) modulation. Regul Toxicol Pharmacol. 122:104884. doi: 10.1016/j.yrtph.2021.104884.

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. doi: 10.1016/j.ntt.2015.08.002.

Buckalew AR, Wang J, Murr AS, Deisenroth C, Stewart WM, Stoker TE, Laws SC. 2020. Evaluation of potential sodium-iodide symporter (NIS) inhibitors using a secondary Fischer rat thyroid follicular cell (FRTL-5) radioactive iodide uptake (RAIU) assay. Arch Toxicol. 94(3):873–885. doi: 10.1007/s00204-020-02664-y.

Cavalieri RR, Pitt-Rivers R. 1981. The effects of drugs on the distribution and metabolism of thyroid hormones. Pharmacol Rev. 33(2):55–80.

Chang SC, Thibodeaux JR, Eastvold ML, Ehresman DJ, Bjork JA, Froehlich JW, Lau CS, Singh RJ, Wallace KB, Butenhoff JL. 2007. Negative bias from analog methods used in the analysis of free thyroxine in rat serum containing perfluorooctanesulfonate (PFOS). Toxicology. 234(1–2):21–33. doi: 10.1016/j.tox.2007.01.020.

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). Toxicology. 243(3):330–339. doi: 10.1016/j.tox.2007.10.014.

Chen ZP, Hetzel BS. 2010. Cretinism revisited. Best Pract Res Clin Endocrinol Metab. 24(1):39–50. doi: 10.1016/j.beem.2009.08.014.

Collet B, Simon E, van der Linden S, El Abdellaoui N, Naderman M, Man HY, Middelhof I, van der Burg B, Besselink H, Brouwer A. 2020. Evaluation of a panel of *in vitro* methods for assessing thyroid receptor β and transthyretin transporter disrupting activities. Reprod Toxicol. 96:432–444. doi: 10.1016/j.reprotox.2019.05.011.

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. OECD Series on adverse outcome pathways No. 13. Paris: OECD.

Crofton KM, Kodavanti PR, Derr-Yellin EC, Casey AC, Kehn LS. 2000. PCBs, thyroid hormones, and ototoxicity in rats: cross-fostering experiments demonstrate the impact of postnatal lactation exposure. Toxicol Sci. 57(1):131–140. doi: 10.1093/toxsci/57.1.131.

Crofton KM, Zoeller RT. 2005. Mode of action: neurotoxicity induced by thyroid hormone disruption during development - hearing loss resulting from exposure to PHAHs. Crit Rev Toxicol. 35(8-9):757–769. doi: 10.1080/10408440591007304.

Curran PG, DeGroot LJ. 1991. The effect of hepatic enzyme-inducing drugs on thyroid hormones and the thyroid gland. Endocr Rev. 12(2):135–150. doi: 10.1210/edrv-12-2-135.

Deisenroth C, Soldatow VY, Ford J, Stewart W, Brinkman C, LeCluyse EL, MacMillan DK, Thomas RS. 2020. Development of an *in vitro* human thyroid microtissue model for chemical screening. Toxicol Sci. 174(1):63–78. doi: 10.1093/toxsci/kfz238.

Dracheva E, Norinder U, Rydén P, Engelhardt J, Weiss JM, Andersson PL. 2022. *In silico* identification of potential thyroid hormone system disruptors among chemicals in human serum and chemicals with a high exposure index. Environ Sci Technol. 56(12):8363–8372. doi: 10.1021/acs.est.1c07762.

[ECETOC] European Centre for Ecotoxicology and Toxicology of Chemicals. 2021. Technical Report No. 138. ECETOC Guidance on dose selection. Brussels (Belgium): ECETOC.

[ECHA] European Chemicals Agency. 2017. Read-Across Assessment Framework (RAAF). ECHA-17-R-01-EN. March 2017.

[EFSA] European Food Safety Authority. 2019a. Scientific Report on the establishment of cumulative assessment groups of pesticides for their effects on the thyroid (Crivellente F, Hart A, Hernandez-Jerez AF, Hougaard Bennekou S, Pedersen R, Terron A, Wolterink G, Mohimont L). EFSA J. 17(9):5801.

[EFSA] European Food Safety Authority. 2019b. Administrative guidance on submission of dossiers and assessment reports for the peer-review of pesticide active substances. EFSA supporting publication 2019:EN-1612.

[EFSA] European Food Safety Authority. 2020. 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 and 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 (Andersson N, Arena M, Auteri D, Barmaz S, Grignard E, Kienzler A, Lepper P, Lostia AM, Munn S, Parra Morte JM, et al.). ECHA-18-G-01-EN; EFSA J. 16:1661–1170.

[EFSA SC] European Food Safety Authority Scientific Committee. 2017. Scientific opinion on the guidance on the use of the weight of evidence approach in scientific assessments (Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, et al.). EFSA J. 15(8):4971.

EP and Council. 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L. 396:1. 30 December 2006

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. 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ EU L. 276:33.

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. 27 June 2012.

European Commission. 2013. Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L. 93:1.

European Commission. 2017. 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 EU 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 EU L. 101:33–36.

Felter SP, Bhat VS, Botham PA, Bussard DA, Casey W, Hayes AW, Hilton GM, Magurany KA, Sauer UG, Ohanian EV. 2021. Assessing chemical carcinogenicity: hazard identification, classification, and risk assessment. Insight from a Toxicology Forum state-of-the-science workshop. Crit Rev Toxicol. 51(8):653–694. doi: 10.1080/10408444.2021.2003295.

Filer DL, Kothiya P, Setzer RW, Judson RS, Martin MT. 2017. tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics. 33(4):618–620. doi: 10.1093/bioinformatics/btw680.

Ford J, Riutta C, Kosian PA, O'Shaughessy K, Gilbert M. 2023. Reducing uncertainties in quantitative adverse outcome pathways by analysis of thyroid hormone in the neonatal rat brain. Toxicol Sci. 193(2):192–203. 2023 Apr 26: doi: 10.1093/toxsci/kfad040.

Foster JR, Tinwell H, Melching-Kollmuss S. 2021. A review of species differences in the control of, and response to, chemical-induced thyroid hormone perturbations leading to thyroid cancer. Arch Toxicol. 95(3): 807–836. doi: 10.1007/s00204-020-02961-6.

Franzosa JA, Bonzo JA, Jack J, Baker NC, Kothiya P, Witek RP, Hurban P, Siferd S, Hester S, Shah I, et al. 2021. High-throughput toxicogenomic screening of chemicals in the environment using metabolically competent hepatic cell cultures. NPJ Syst Biol Appl. 7(1):7. doi: 10.1038/s41540-020-00166-2.

Gadaleta D, d'Alessandro L, Marzo M, Benfenati E, Roncaglioni A. 2021. Quantitative structure-activity relationship modeling of the Amplex Ultrared Assay to predict thyroperoxidase inhibitory activity. Front Pharmacol. 12:713037. doi: 10.3389/fphar.2021.713037.

Garcia de Lomana M, Weber AG, Birk B, Landsiedel R, Achenbach J, Schleifer KJ, Mathea M, Kirchmair J. 2021. *In silico* models to predict the perturbation of molecular initiating events related to thyroid hormone homeostasis. Chem Res Toxicol. 34(2):396–411. doi: 10.1021/acs.chemrestox.0c00304.

Gilbert ME, Rovet J, Chen Z, Koibuchi N. 2012. Developmental thyroid hormone disruption: prevalence, environmental contaminants and neurodevelopmental consequences. Neurotoxicol. 33(4):842–852. doi: 10.1016/j.neuro.2011.11.005.

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. doi: 10.1111/jne.12169.

Gilbert ME, O'Shaughnessy KL, Axelstad M. 2020. Regulation of thyroid-disrupting chemicals to protect the developing brain. Endocrinol. 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.

Goldey ES, Kehn LS, Lau C, Rehnberg GL, Crofton KM. 1995. 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. doi: 10.1006/taap.1995.1210.

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. doi: 10.1210/en.2006-1276.

Hall AP, Elcombe CR, Foster JR, Harada T, Kaufmann W, Knippel A, Küttler K, Malarkey DE, Maronpot RR, Nishikawa A, et al. 2012. Liver hypertrophy: a review of adaptive (adverse and non-adverse) changes - conclusions from the 3rd International ESTP Expert Workshop. Toxicol Pathol. 40(7):971–994. doi: 10.1177/0192623312448935.

Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Stoker TE, Laws SC. 2017. Development of a screening approach to detect thyroid disrupting chemicals that inhibit the human sodium iodide symporter (NIS). Toxicol In Vitro. 40:66–78. doi: 10.1016/j.tiv.2016.12.006.

Handa S, Hassan I, Gilbert M, El-Masri H. 2021. Mechanistic computational model for extrapolating *in vitro* thyroid peroxidase (TPO) inhibition data to predict serum thyroid hormone levels in rats. Toxicol Sci. 183(1):36–48. doi: 10.1093/toxsci/kfab074.

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. doi: 10.1093/toxsci/kfx163.

He G, Tsutsumi T, Zhao B, Baston DS, Zhao J, Heath-Pagliuso S, Denison MS. 2011. Third-generation Ah receptor–responsive luciferase reporter plasmids: amplification of dioxin-responsive elements dramatically increases CALUX bioassay sensitivity and responsiveness. Toxicol Sci. 123(2):511–522. doi: 10.1093/toxsci/kfr189.

Hornung MW, Korte JJ, Olker JH, Denny JS, Knutsen C, Hartig PC, Cardon MC, Degitz SJ. 2018. Screening the ToxCast phase 1 chemical library for inhibition of deiodinase type 1 activity. Toxicol Sci. 162(2):570–581. doi: 10.1093/toxsci/kfx279.

[JRC] Joint Research Centre. 2023. Validation of a battery of mechanistic methods relevant for the detection of chemicals that can disrupt the thyroid hormone system. (Bernasconi C, Langezaal I, Bartnicka J, Asturiol D, Bowe G, Coecke S, Kienzler A, Liska R, Milcamps A, Munoz Pineiro A, Pistollato F, Whelan M). JRC Technical Report. JRC132532/EUR 31456 EN, Publications Office of the European Union.

Jomaa B, Aarts JM, de Haan LH, Peijnenburg AA, Bovee TF, Murk AJ, Rietjens IM. 2013. *In vitro* pituitary and thyroid cell proliferation assays and their relevance as alternatives to animal testing. ALTEX. 30(3): 293–307. doi: 10.14573/altex.2013.3.293.

Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, Little S, Wambaugh J, Setzer RW, Kothiya P, et al. 2016. Analysis of the effects of cell stress and cytotoxicity on *in vitro* assay activity across a diverse chemical and assay space. TOXICOL SCI. 153(2):409. Erratum for: TOXICOL SCI. 2016. 152(2):323–39. doi: 10.1093/toxsci/kfw148.

Karwelat D, Kühnlenz J, Steger-Hartmann T, Bars R, Tinwell H, Marx U, Bauer S, Born O, Raschke M. 2023. A rodent thyroid-liver chip to capture thyroid toxicity on organ functional level. ALTEX. 40(1):83–102. doi: 10.14573/altex.2108262.

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. doi: 10.1002/etc.4125.

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. doi: 10.1093/toxsci/kfq105.

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. doi: 10.3390/ijms21093123.

Kühnlenz J, Karwelat D, Steger-Hartmann T, Raschke M, Bauer S, Vural Ö, Marx U, Tinwell H, Bars R. 2023. A microfluidic thyroid-liver platform to assess chemical safety in humans. ALTEX. 40(1):61–82.

Leonard JA, Tan YM, 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. doi: 10.1093/toxsci/kfw022.

Lewis RW, Billington R, Debryune E, Gamer A, Lang B, Carpanini F. 2002. Recognition of adverse and nonadverse effects in toxicity studies. Toxicol Pathol. 30(1):66–74. doi: 10.1080/01926230252824725.

Li J, Wang X, Yan Y, Wang K, Qin D, Xin Z, Wei J. 1986. The effects of a severely iodine deficient diet derived from an endemic area on fetal brain development in the rat. Neuropathol Appl Neurobiol. 12(3):261–276. doi: 10.1111/j.1365-2990.1986.tb00139.x.

Li H, Zhang M, Vervoort J, Rietjens IM, van Ravenzwaay B, Louisse J. 2017. Use of physiologically based kinetic modeling-facilitated reverse dosimetry of *in vitro* toxicity data for prediction of *in vivo* developmental toxicity of tebuconazole in rats. Toxicol Lett. 266:85–93. doi: 10.1016/j.toxlet.2016.11.017.

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. doi: 10.1016/j.yrtph.2019.04.010.

Louisse J, Beekmann K, Rietjens IM. 2017. Use of physiologically based kinetic modeling-based reverse dosimetry to predict *in vivo* toxicity from *in vitro* data. Chem Res Toxicol. 30(1):114–125. doi: 10.1021/acs.chemrestox.6b00302.

Maglich JM, Parks DJ, Moore LB, Collins JL, Goodwin B, Billin AN, Stoltz CA, Kliewer SA, Lambert MH, Willson TM, et al. 2003. Identification of a novel human constitutive androstane receptor (CAR) agonist and its use in the identification of CAR target genes. J Biol Chem. 278(19):17277–17283. doi: 10.1074/jbc.M300138200.

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. doi: 10.1080/10408444.2021.1910625.

Marty S, Sauer UG, Charlton A, Ghaffari R, Guignard D, Hallmark N, Hannas BR, Jacobi S, Marxfeld H-A, Melching-Kollmuss S, et al. 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? Crit Rev Toxicol. 52(7):546–617. doi: 10.1080/10408444.2022.2130166.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J, Vickers C. 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol. 34(1):1–18. doi: 10.1002/jat.2949.

Meek ME, Palermo CM, Bachman AN, North CM, Jeffrey Lewis R. 2014. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J Appl Toxicol. 34(6):595–606. doi: 10.1002/jat.2984.

Moroni L, Barbaro F, Caiment F, Coleman O, Costagliola S, Conza GD, Elviri L, Giselbrecht S, Krause C, Mota C, et al. 2020. SCREENED: a multistage model of thyroid gland function for screening endocrine-disrupting chemicals in a biologically sex-specific manner. IJMS. 21(10):3648. doi: 10.3390/ijms21103648.

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. doi: 10.1006/taap.1996.0034.

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. doi: 10.1289/EHP5297.

O'Shaughnessy KL, Kosian PA, Ford JL, Oshiro WM, Degitz SJ, Gilbert ME. 2018. Developmental thyroid hormone insufficiency induces a cortical brain malformation and learning impairments: a cross-fostering study. Toxicol Sci. 163(1):101–115. doi: 10.1093/toxsci/kfy016.

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(1):4662. doi: 10.1038/s41598-019-40249-7.

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. doi: 10.1016/j.mce.2019.110663.

[OECD] Organisation for Economic Co-operation and Development. 2000. Series on testing and assessment No. 19. Guidance Document on the recognition, assessment and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. ENV/JM/MONO(2000)7. Paris: OECD Publishing.

[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 Publishing.

[OECD] Organisation for Economic Cooperation and Development. 2012. Conceptual Framework for testing and assessment of endocrine disruptors (as revised in 2012); https://www.oecd.org/env/ehs/testing/OECD%20Conceptual%20Framework%20for%20Testing%20and%20Assessment%20of%20Endocrine%20Disrupters%20for%20the%20public%20website.pdf [accessed 2023 May].

[OECD] Organisation for Economic Cooperation and Development. 2017. Series on testing and assessment No. 184. Revised guidance document on developing and assessing adverse outcome pathways. ENV/JM/MONO(2013)6. Paris: OECD Publishing.

[OECD] Organisation for Economic Cooperation and Development. 2018. Series on testing and assessment. Revised Guidance Document No. 150 on standardised test guidelines for evaluating chemicals for endocrine disruption. Paris: OECD Publishing.

Olker JH, Korte JJ, Denny JS, Hartig PC, Cardon MC, Knutsen CN, Kent PM, Christensen JP, Degitz SJ, Hornung MW. 2019. Screening the ToxCast Phase 1, Phase 2, and e1k Chemical Libraries for inhibitors of iodothyronine deiodinases. Toxicol Sci. 168(2):430–442. doi: 10.1093/toxsci/kfy302.

Papineni S, Marty MS, Rasoulpour RJ, LeBaron MJ, Pottenger LH, Eisenbrandt DL. 2015. Mode of action and human relevance of pronamide-induced rat thyroid tumors. Regul Toxicol Pharmacol. 71(3):541–551. doi: 10.1016/j.yrtph.2015.02.012.

Parmentier C, Baze A, Untrau M, Kampkoetter A, Lasserre D, Richert L. 2022. Evaluation of human relevance of Nicofluprole-induced rat thyroid disruption. Toxicol Appl Pharmacol. 435:115831. doi: 10.1016/j.taap.2021.115831.

Paul KB, Hedge JM, Macherla C, Filer DL, Burgess E, Simmons SO, Crofton KM, Hornung MW. 2013. Cross-species analysis of thyroperoxidase inhibition by xenobiotics demonstrates conservation of response between pig and rat. Toxicol. 312:97–107. doi: 10.1016/j.tox.2013.08.006.

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. doi: 10.1093/toxsci/kfw034.

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

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. doi: 10.1016/j.toxrep.2020.12.019.

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):563–572. doi: 10.1016/j.ntt.2010.04.053.

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. doi: 10.1093/toxsci/kfy055.

Ramhøj L, Hass U, Gilbert ME, Wood C, Svingen T, Usai D, Vinggaard AM, Mandrup K, Axelstad M. 2020. Evaluating thyroid hormone disruption: investigations of long-term neurodevelopmental effects in rats after perinatal exposure to perfluorohexane sulfonate (PFHxS). Sci Rep. 10(1):2672. doi: 10.1038/s41598-020-59354-z.

Renko K, Schäche S, Hoefig CS, Welsink T, Schwiebert C, Braun D, Becker N-P, Köhrle J, Schomburg L. 2015. An improved nonradioactive screening method identifies genistein and xanthohumol as potent inhibitors of iodothyronine deiodinases. Thyroid. 25(8):962–968. doi: 10.1089/thy.2015.0058.

Richardson VM, Ferguson SS, Sey YM, Devito MJ. 2014. *In vitro* metabolism of thyroxine by rat and human hepatocytes. Xenobiotica. 44(5):391–403. doi: 10.3109/00498254.2013.847990.

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. OECD Series on adverse outcome pathways No. 14; Paris: OECD.

Romanov S, Medvedev A, Gambarian M, Poltoratskaya N, Moeser M, Medvedeva L, Gambarian M, Diatchenko L, Makarov S. 2008. Homogeneous reporter system enables quantitative functional assessment of multiple transcription factors. Nat Methods. 5(3):253–260. doi: 10.1038/nmeth.1186.

Rosenfeld JM, Vargas R, Jr, Xie W, Evans RM. 2003. Genetic profiling defines the xenobiotic gene network controlled by the nuclear receptor pregnane X receptor. Mol Endocrinol. 17(7):1268–1282. doi: 10.1210/me.2002-0421.

Russell WMS, Burch RL. 1959. The principles of humane experimental technique. London. Methuen. Reprinted by UFAW, 1992: 8 Hamilton Close, South Mimms, Potters Bar, Herts EN6 3QD England. p. 238.

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. doi: 10.3389/fnana.2020.00033.

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. doi: 10.1080/10408444.2020.1839380.

Sewell F, Kimber I, Boobis AR. 2020. Use of the kinetically-derived maximum dose: opportunities for delivering 3Rs benefits. Regul Toxicol Pharmacol. 116:104734. doi: 10.1016/j.yrtph.2020.104734.

Sewell F, Corvaro M, Andrus A, Burke J, Daston G, Delaney B, Domoradzki J, Forlini C, Green ML, Hofmann T, et al. 2022. Recommendations on dose level selection for repeat dose toxicity studies. Arch Toxicol. 96(7):1921–1934. doi: 10.1007/s00204-022-03293-3.

Shibutani M, Woo GH, Fujimoto H, Saegusa Y, Takahashi M, Inoue K, Hirose M, Nishikawa A. 2009. Assessment of developmental effects of hypothyroidism in rats from *in utero* and lactation exposure to antithyroid agents. Reprod Toxicol. 28(3):297–307. doi: 10.1016/j.reprotox.2009.04.011.

Smith PF, Grossman SJ, Gerson RJ, Gordon LR, Deluca JG, Majka JA, Wang RW, Germershausen JI, MacDonald JS. 1991. Studies on the mechanism of simvastatin-induced thyroid hypertrophy and follicular cell adenoma in the rat. Toxicol Pathol. 19(3):197–205. doi: 10.1177/019262339101900301.

Strupp C, Quesnot N, Weber-Parmentier C, Richert L, Bomann WH, Singh P. 2020. Weight of evidence and human relevance evaluation of the Benfluralin mode of action in rats (Part II): thyroid carcinogenesis. Regul Toxicol Pharmacol. 117:104736. doi: 10.1016/j.yrtph.2020.104736.

Tan YM, Barton HA, Boobis A, Brunner R, Clewell H, Cope R, Dawson J, Domoradzki J, Egeghy P, Gulati P, et al. 2021. Opportunities and challenges related to saturation of toxicokinetic processes: implications for risk assessment. Regul Toxicol Pharmacol. 127:105070. doi: 10.1016/j.yrtph.2021.105070.

Tater A, Gupta A, Upadhyay G, Deshpande A, Date R, Tamboli IY. 2021. *In vitro* assays for characterization of distinct multiple catalytic activities of thyroid peroxidase using LC-MS/MS. Curr Res Toxicol. 2:19–29. doi: 10.1016/j.crtox.2021.01.001.

Terry C, Hays S, McCoy AT, McFadden LG, Aggarwal M, Rasoulpour RJ, Juberg DR. 2016. Implementing a framework for integrating toxicokinetics into human health risk assessment for agrochemicals. Regul Toxicol Pharmacol. 75:89–104. doi: 10.1016/j.yrtph.2015.10.003.

Tinwell H, Bars R. 2022. Isoflucypram: combining *in vivo* and NAMs data in a weight of evidence approach to demonstrate the human non-relevance of the mode of action leading to the subtle thyroid effects observed in the rat. Regul Toxicol Pharmacol. 131:105154. doi: 10.1016/j.yrtph.2022.105154.

[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; 12 pp. https://www.epa.gov/pesticide-registration/guidance-thyroid-assays-pregnant-animals-fetuses-and-postnatal-animals-and. [accessed 2023 May].

Valdés Hernández MdC, Wilson KL, Combet E, Wardlaw JM. 2013. Brain findings associated with iodine deficiency identified by magnetic resonance methods: a systematic review. Open J Rad. 03(04):180–195. doi: 10.4236/ojrad.2013.34030.

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. doi: 10.1002/etc.4124.

Vinken M, Knapen D, Vergauwen L, Hengstler JG, Angrish M, Whelan M. 2017. Adverse outcome pathways: a concise introduction for toxicologists. Arch Toxicol. 91(11):3697–3707. doi: 10.1007/s00204-017-2020-z.

Wang J, Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Laws SC, Stoker TE. 2018. High-throughput screening and quantitative chemical ranking for sodium-iodide symporter inhibitors in ToxCast phase i chemical library. Environ Sci Technol. 52(9):5417–5426. doi: 10.1021/acs.est.7b06145.

Weber AG, Birk B, Herrmann C, Huener H-A, Renko K, Coecke S, Schneider S, Funk-Weyer D, Landsiedel R. 2022. A new approach method to study thyroid hormone disruption: optimization and standardization of an assay to assess the inhibition of DIO1 enzyme in human liver microsomes. Applied In Vitro Toxicol. 8(3):67–82. doi: 10.1089/aivt.2022.0010.

Whalley PM, Bartels M, Bentley KS, Corvaro M, Funk D, Himmelstein MW, Neumann B, Strupp C, Zhang F, Mehta J. 2017. An *in vitro* approach for comparative interspecies metabolism of agrochemicals. Regul Toxicol Pharmacol. 88:322–327. doi: 10.1016/j.yrtph.2017.03.020.

Wiemann C, Melching-Kollmuss S, Hambruch N, Wiss L, Stauber F, Richert L. 2023. Boscalid shows increased thyroxin-glucuronidation in rat, but not in human hepatocytes *In vitro*. J Appl Toxicol. 43(6):828–844. doi: 10.1002/jat.4427.

[WHO/IPCS] World Health Organisation/International Programme on Chemical Safety. 2009. Principles and methods for the risk assessment of chemicals in food. Geneva (Switzerland): (Environmental Health Criteria), World Health Organisation.

Yamada T, Hasegawa R, Nishikawa S, Sakuratani Y, Yamada J, Yamashita T, Yoshinari K, Yamazoe Y, Kamata E, Ono A, et al. 2013. New parameter that supports speculation on the possible mechanism of hypothyroidism induced by chemical substances in repeated-dose toxicity studies. J Toxicol Sci. 38(2):291–299. doi: 10.2131/jts.38.291.

Zoeller RT, Tan SW, Tyl RW. 2007. General background on the hypothalamic-pituitary-thyroid (HPT) axis. Crit Rev Toxicol. 37(1–2):11–53. doi: 10.1080/10408440601123446.

Zuang V, Dura A, Ahs Lopez E, Barroso J, Batista Leite S, Berggren E, Bopp S, Campia I, Carpi D, Casati S, et al. 2022. Non-animal methods in science and regulation. EUR 30960 EN/JRC127780. Luxembourg: Publications Office of the EU.

Appendices

Table Appendix 1. Possible scenarios for the Tier 0 *in vivo* thyroid- and neurodevelopment-related database.

No.	Assessment of *in vivo* thyroid-related database			Assessment of *in vivo* neurodevelopment-related database			Overarching WoE evaluation: EDC-T met?	Conclusion: Need to enter Thyroid-NDT-TAS?
	Findings?	Database sufficient?	EDC: EA?	Findings?	Database sufficient?	EDC: AE?		
1	No	Yes	No	No	Yes	No	No (EA & AE no)	No: EDC-T not met
2	No	Yes	No	No	No	Unclear	No (EA no)	No: EDC-T not met
3	No	Yes	No	Yes	Yes	Yes	No (EA no, only DNT)	No: EDC-T not met
4	No	Yes	No	Yes	No	Unclear	No (EA no)	No: EDC-T not met
5	No	No	Unclear	No	Yes	No	No (AE no)	No: EDC-T not met
6	No	No	Unclear	No	No	Unclear	EA & AE unclear	Yes: Scenario I, Variant A
7	No	No	Unclear	Yes	Yes	Yes	EA unclear; AE yes	Yes: Scenario I, Variant B
8	No	No	Unclear	Yes	No	Unclear	EA & AE unclear	Yes: Scenario I, Variant A
9	Yes	Yes	Yes	No	Yes	No	No (AE no)	No: EDC-T not met
10	Yes	Yes	Yes	No	No	Unclear	EA yes, AE unclear	Yes: Scenario I, Variant C
11	Yes	Yes	Yes	Yes	Yes	Yes	EA & AE yes	Yes: Scenario II
12	Yes	Yes	Yes	Yes	No	Unclear	EA yes, AE unclear	Yes: Scenario I, Variant C
13	Yes	No	Unclear	No	Yes	No	No (AE no)	No: EDC-T not met
14	Yes	No	Unclear	No	No	Unclear	EA & AE unclear	Yes: Scenario I, Variant A
15	Yes	No	Unclear	Yes	Yes	Yes	EA unclear, AE yes	Yes: Scenario I, Variant B
16	Yes	No	Unclear	Yes	No	Unclear	EA & AE unclear	Yes: Scenario I, Variant A

AE: adverse effects; DNT: developmental neurotoxicity; EA: endocrine activity; EDC-T: endocrine disruptor criteria for the thyroid modality; MoA: mode-of-action; WoE: weight-of-evidence. Note: "Findings" relates to statistically significant and biologically relevant (thyroid- or neurodevelopment-related) findings.

Colour legend: dark grey shading: no effect/database not sufficient/no EA/no AE. Light grey shading: database inconclusive regarding EA and/or AE.

Scenario I (Section 2.2.1): Enter Thyroid-NDT-TAS to conduct (Tier 1, Step 1) initial MoA and human relevance assessment and to complete database, beginning with (Tier 1, Step 2) *in vitro* testing/*in silico* modelling, as relevant, followed by (Tier 2) higher-tier testing, if necessary for (Tier 3) final MoA and human relevance assessment and final WoE evaluation. The completion of the database addresses the generation of:

- Variant A: thyroid- and neurodevelopment-related data,
- Variant B: thyroid-related data,
- Variant C: neurodevelopment-related data.

Scenario II (Section 2.2.2): Enter Thyroid-NDT-TAS to conduct (Tier 1, Step 1) initial MoA and human relevance assessment and to perform (Tier 1, Step 2) *in vitro* testing/*in silico* modelling, as relevant, followed by (Tier 2) non-standard higher-tier testing, if relevant for (Tier 3) final MoA and human relevance assessment and final WoE evaluation.

Table Appendix 2. Research recommendations to enhance assessments of thyroid function and neurodevelopmental impairment (adapted from Marty et al. 2022).

Overarching topic	Recommended research topic/research question
General	Substantiate and refine the insight from Marty et al. (2022) 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
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?
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?
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
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
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?
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
Brain-related assessments (general)	Enhance knowledge base on (patterns of) brain-related effects by collecting multiple neurodevelopment-related endpoints in the same study
Motor activity	Do different levels of thyroid hormone imbalance during lactation indicate whether altered motor activity is transient or persistent?
Cognitive function	Enhance understanding of sensitivity of the different test methods use to assess cognitive function
Periventricular heterotopia	Standardise methodology, encourage cross-laboratory validation of this endpoint in view of potential uptake into formal test guidelines
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
Sensitive neurodevelopmental endpoints	Establish whether more appropriate and sensitive sets of neurodevelopmental endpoints to be measured in rats are needed
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)
Toxicokinetics	Develop framework for the integration of toxicokinetics into assessments of substance-mediated thyroid hormone imbalance

DIO: deiodinase; fT3: free triiodothyronine; fT4: free thyroxine; HPT: hypothalamic–pituitary–thyroid; MoA: mode-of-action; PFHxS: perfluorohexane sulphonates; PND: postnatal day; T3: triiodothyronine; T4: thyroxine; TPO: thyroid peroxidase.

Table Appendix 3. Thyroid-related AOPs including neurodevelopmental outcomes in mammals listed in the OECD AOP Wiki in May 2023 (adapted from Marty et al. 2021).

AOP Wiki	AOP 8	AOP 42	AOP 54	AOP 134	AOP 152	AOP 300
AOP title	Activation of hepatic nuclear receptors, & subsequent neurodevelopmental AOs in mammals	Inhibition of TPO & subsequent neurodevelopmental AOs in mammals	Inhibition of NIS leads to learning and memory impairment	NIS inhibition & subsequent neurodevelopmental AOs in mammals	Interference with TTR & subsequent human neurodevelopmental toxicity	TR antagonism & subsequent neurodevelopmental AOs in mammals
1st Author / initiator	Katie Paul Friedman	Crofton et al. (2019)	Rolaki et al. (2019)	Mary Gilbert	Erik R. Janus	Kevin Crofton
Status	Open for adoption; under development; in OECD workplan	Endorsed; in OECD Workplan	Endorsed; in OECD Workplan	Under development; in OECD workplan	Under development; in OECD workplan	Under development; in OECD workplan
MIE	PXR activation [a]	TPO inhibition	NIS inhibition	NIS inhibition	Binding, TTR in serum	Antagonism, TR
KE1	Upregulation of UGT activity; induction	Thyroid hormone synthesis, decreased	Thyroidal iodide, decreased	Thyroid hormone synthesis, decreased	Displacement, serum T4 from TTR	Hippocampal gene expression, altered
KE2	Clearance of T4 from serum, increased	T4 in serum; decreased	Thyroid hormone synthesis, decreased	Thyroidal iodide, decreased	Serum fT4, increased	Hippocampal anatomy, altered
KE3	T4 in serum; decreased	T4 in neuronal tissue; decreased	T4 in serum; decreased	T4 in serum; decreased	Uptake of T4 into tissue, increased	Hippocampal function, decreased
KE4	T4 in neuronal tissue; decreased	Hippocampal gene expression, altered	T4 in neuronal tissue; decreased	T4 in neuronal tissue; decreased	Clearance of T4 from tissue, increased	
KE5	Hippocampal gene expression, altered	Hippocampal anatomy, altered	Brain-derived neurotrophic factor, reduced	Hippocampal gene expression, altered	T4 in serum; decreased	
KE6	Hippocampal anatomy, altered	Hippocampal, physiology decreased	GABAergic interneurons, decreased	Hippocampal anatomy, altered	T4 in neuronal tissue; decreased	
KE7	Hippocampal physiology, decreased		Synaptogenesis, decreased	Hippocampal physiology, altered	Hippocampal gene expression, altered	
KE8			Neuronal network function, decreased		Hippocampal anatomy, altered	
KE9					Hippocampal physiology decreased	
AO	Cochlear function, loss	Cochlear function, decreased / loss // Cognitive function, decreased [b]	Impairment, learning and memory	Cochlear function, decreased // Cognitive function, decreased [b]	Cochlear function, decreased // Cognitive function, decreased [b]	Cochlear function, decreased / Cognitive function, decreased [b]

AO(P): adverse outcome (pathway); fT4: free thyroxine; GABA: gamma amino-butyric acid; KE(R): Key event (relationship); MIE: molecular initiating event; NIS: sodium – iodide symporter; PXR: pregnane X receptor; T4: thyroxine; TPO: thyroid peroxidase; TR: Thyroid receptor; TTR: transthyretin; UGT: uridine diphosphate glucuronyltransferase.

Colour legend:

- Text colour = strength of evidence of KER: Green = strong; brown: moderate; blue: weak; black: MIE (since 2nd KE of respective KER is marked) OR: information unavailable.
- Shading of cell = quantitative evidence for KER (always marking 2nd KE of the respective KER): Green = strong; brown: moderate; blue: weak; white: information unavailable.
- The bold boxes highlight that the KEs "T4 in serum; decrease" and "T4 in neuronal tissue; decrease" are central to five of the six AOPs.

AOP 8 refers to rodent studies addressing exposure to polychlorinated biphenyls to support the described sequence of key events (Crofton and Zoeller 2005), and postnatal lactation exposure is indicated as the critical period of exposure (Crofton et al. 2000). For AOP 42, the supporting evidence is mostly derived from *in vitro* studies and rodent studies, whereas AOP 54, AOP 152 and AOP 300 refer to human studies, rodent studies (and *in vitro* studies) as providing the supportive evidence. AOP 134 does not summarise the underlying evidence.

[a] Evidence assessment as presented in the AOP Wiki for AOP 8: "*Concordance of dose-response relationships: Multiple studies provide limited (2–3 doses) dose-response data for many of the key events. These studies demonstrate similar magnitudes of effect on circulating hormones for doses of PCBs* [polychlorinated biphenyls] *that are within an order of magnitude (3–25 mg/kg/day for Aroclor 1254)… Very limited data are available correlating any of the key events. One exception is the relationship between circulating serum T4 concentrations during development and the magnitude of hearing loss… There is a very good correlation between total serum T4 concentrations on postnatal day (PND) 14 and hearing loss assessed in adult offspring of PCB exposed dams… All of these events occur within a 2–3 fold dose range…."

[b] Adverse outcome "cochlear function, decreased/loss" indicated in AOP Wiki on 13 September 2019, and replaced by "cognitive function, decreased" by 15 October 2019.