Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2022)25

Unclassified

English - Or. English

1 September 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Case study for the integration of in vitro data in the developmental neurotoxicity hazard identification and characterisation using flufenacet

Series on Testing and Assessment
No. 363

JT03501671

OECD Environment, Health and Safety Publications

SERIES ON TESTING AND ASSESSMENT

NO. 363

Case study for the integration of in vitro data in the developmental neurotoxicity hazard identification and characterisation using flufenacet

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

- **Also published in the Series on Testing and Assessment: link**

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

Foreword

OECD member countries have been making efforts to expand the use of alternative methods in assessing chemicals. The OECD has been developing guidance documents and tools for the use of alternative methods such as (Q)SAR, chemical categories and Adverse Outcome Pathways (AOPs) as a part of Integrated Approaches for Testing and Assessment (IATA). There is a need for the investigation of the practical applicability of these methods/tools for different aspects of regulatory decision-making, and to build upon case studies and assessment experience across jurisdictions.

The objective of the IATA Case Studies Project is to increase experience with the use of IATA by developing case studies, which constitute examples of predictions that are fit for regulatory use. The aim is to create common understanding of using novel methodologies and the generation of considerations/guidance stemming from these case studies.

This case study was developed by Iris Mangas[1)], Andrea Terron[1)], Martina Panzarea[1)], Laura Martino[2)], Federica Barrucci[2)], Elisa Aiassa[2)], Irene Munoz Guajardo[2)], Kevin Crofton[3)], Martin Wilks[3)], Susanne Hougaard Bennekou[3)], Martin Paparella[3)], Ioanna Tzoulaki[3)], Antonio Hernandez-Jerez[3)] for illustrating practical use of IATA and submitted to the 2021 review cycle of the IATA Case Studies Project.

Affiliation at time of contribution:

1) Pesticides Peer Review Unit, European Food Safety Authority (EFSA)

2) Assessment and methodological support Unit, EFSA

3) EFSA PPR Panel WG of external experts on DNT IATA, EFSA

The case study was reviewed by the project team, and endorsed at the 6th meeting of the Working Party on Hazard Assessment in June 2022.

The case study is illustrative examples, and their publication as OECD monographs does not translate into direct acceptance of the methodologies for regulatory purposes across OECD countries. In addition, the cases study should not be interpreted as official regulatory decisions made by the authoring member countries.

This document is published under the responsibility of the Chemicals and Biotechnology Committee of the OECD.

Table of contents

FIGURES

TABLES

Abbreviations and acronyms

5-HT	5-hydroxytryptamine
ADHD	Attention-deficit/hyperactivity disorder
ADME	Absorption, distribution, metabolism and excretion
ADOS	Autism Diagnostic Observation
AMPARs	alpha-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
ASD	Autism Spectrum Disorder
BDNF	Brain-Derived Neurotrophic Factor
BDNF	Brain-derived neurotrophic factor
BMCs	Benchmark Concentrations
BMRs	Benchmark Responses
BRIEF	Behaviour Rating Inventory of Executive Function
BV2	Cells derived from raf/myc-immortalised murine neonatal microglia
Bw	Body weight
Camk1g	Calcium/Calmodulin Dependent Protein Kinase IG
CBCL	Child Behaviour Check List
CNN	Convolutional neuronal network
CNS	Central Nervous System
CREB	cAMP response element-binding protein
DIV	Days In Vitro
DNT	Developmental neurotoxicity
DNT-IVB	DNT in vitro battery
DQ	Development quotient
GABA	Gamma-Aminobutyric acid
GD	Gestation Day
GLP	Good Laboratory Practice
HCI	High Content imaging
hiPSC	human induced pluripotent stem cells
Hnnf	Human neuronal network formation
hNPC	Human neural progenitor cells
HOS	Human observational studies
IATA	Integrated Approaches for Testing and Assessment
IUF	Leibniz Research Institute for Environmental Medicine
IVB	In vitro battery
IVIVE	in vitro to in vivo extrapolation
KE	Key Events
KER	Key Event Relationship
LDH	lactate dehydrogenase
LOAEL	lowest-observed-adverse-effect level
LUHMES	Lund human mesencephalic cells
MAP2	Microtubule Associated Protein

MBR	Mean burst rate
MDI	Cognitive impairment
MEA	Microelectrode arrays
MFR	Mean Firing Rate
MI	Mental development index
MIE	Molecular Initiating Events
MISIB	Mean Interspike Interval In Burst
MoA	Mode of Action
MRI	Magnetic resonance imaging
MSEL	Mullen Scales of Early Learning
MWM	Morris Water Maze behavioural test
NAM	New approach methodologies
NCC	Neural crest cells
NDDs	Neurodevelopmental disorders
NMDA	N-Methyl-d-Aspartic acid
NMDARs	N-Methyl-d-Aspartic acid receptors
NNF	Neuronal network formation
NPC	Neural Progenitor Cells
NPC1 test	Primary hNPC Proliferation Assay
NPC2 test	Primary hNPC Migration Assay
NPC3 test	Primary hNPC Neuronal Differentiation Assay
NPC4 test	Neuronal morphology (neurite length and area) of young neurons differentiated from hNPC
NPC5 test	Oligodendrocyte Differentiation
OHAT/NTP	The Office of Health Assessment and Translation/National Toxicology Program
OL	Oligodendrocyte
OPCs	OL progenitor cells
PBPK	Physiologically-based pharmacokinetic
pCREB	cAMP response element-binding protein phosphorylated
PDI	Impaired psychomotor development
PND	Postnatal day
PoD	Point-of-departure
pre-OLs	Pre-myelinating oligodendrocytes
PTrkB	Tropomyosin receptor kinase B phosphorylated
r/HNNf	Rat/human neuronal network formation
RoB	Risk of bias
ROS	Reactive oxygen species
RyR	Ryanodine receptors
SCOs	Synchronous Calcium oscillations
SDQ	Strengths and Difficulties Questionnaire
SNAP25	Synaptosomal-Associated Protein, 25kDa
TCPy	3,5,6-trichloro -2-pyridinol
TG426	OECD Test Guideline 426
TK	Toxicokinetics
TNF	Tumour Necrosis Factor alpha
TrkB	Tropomyosin receptor kinase B
TUBB3	Tubulin Beta 3 Class III
UKN2 test	The cMINC Neural Crest Cell Migration Assay
UKN4 test	The NeuriTox Neurite Outgrowth of CNS Neurons Test (UKN4 Test)
UKN5 test	The PeriTox Neurite Outgrowth of PNS Neurons Test
VGCCs	Voltage-gated Ca2+ channels
WISC-IV	Wechsler Intelligence Scale for Children
WoE	Weight of evidence

Executive Summary

This case study was developed to show the applicability of an *in vitro battery (IVB) for* developmental neurotoxicity (DNT) in the context of the European pesticide Regulation (EU) 283/2013 and 1107/2009. This case study uses flufenacet (an herbicide) as a prototypical compound tested in the IVB for which a developmental neurotoxicity (DNT) concern was not anticipated, this conclusion has been challenge in this IATA as a follow up of the commenting period . Flufenacet was chosen since its selective herbicide mode of action, based on the inhibition of the biosynthesis of very long chain fatty acids resulting in inhibition of cell division and cell growth, combined with one negative DNT guideline study, suggested a low concern for DNT. In this case, the DNT guideline study was triggered by a secondary endocrine concern rather than primary neurotoxic concern.

This integrated approaches for testing and assessment (IATA) case study for flufenacet is comprised of several data streams and methodologies in line with the same approach used in the deltamethrin case study. A systematic literature review of *in vivo* and *in vitro* experimental evidence as well as human observational studies (HOS) was performed. Evidence from a DNT-IVB that covers basic neurodevelopmental key events, e.g. functional and/or morphological alterations in brain cells and tissue, was also considered. The studies retrieved were submitted to a critical appraisal and quantitative uncertainty analysis. The Adverse Outcome Pathway (AOP) concept has been applied as a framework to develop the IATA in order to structure all the evidence and characterise the biological and toxicological relevance of the IVB in predicting an adverse effect.

No HOS with flufenacet were retrieved from the literature. Only one *in vivo* study was available, an OECD TG 426 DNT study. Evidence appraised in Tier 1 also included information on the adverse outcome assessment in zebrafish behavioural studies that included exposure during development.

The revised outcome of the analysis indicates that evidence is available for which a probability of at least 66% would support the identification of a DNT AO. No hazards were identified in any of the studies retrieved from literature or in the DNT IV-B (i.e., there was no endpoint category assessed as being affected with a probability higher than 66%). When integrating the data it was concluded that flufenacet is unliky to be a DNT throughout a direct mechanism. In this case, the integration of the IV-B, has been critical in the weight of evidence. Several in vitro endpoints were assessed through an iterative process providing mechanistic information relevant for the IATA addressing chemical-induced changes in DNT downstream KEs. All of them were assessed in the uncertainty analysis as not being triggered by flufenacet (with a probability higher than 66%). Therefore, the iterative process was stopped and no DNT AOP was postulated. However, as flufenacet is able to induce changes in the circulating levels of thyuroid hormones alternative AOPs should be investigated and postulated but this was out of the scope for this IATA.

Overall this IATA supports the conclusion, derived from in vitro, that there is no evidence that flufenacet is a direct developmental neurotoxicant. The case study shows also the applicability of the DNT-IVB in an specific hazard characterisation context. The integration of the in vitro data, including the DNT-IVB, exemplifies how data from the DNT IVB can be incorporated as part of the weight of evidence within the AOP-informed IATA framework to support the weight of evidence for interpretation of in vivo study. However, the potential causal relationship of metabolites or thyroid hormone disruption remains

uncertain. Thus, the case study also includes considerations associated with testing for metabolites and recommendations on how to address this in the OECD *in vitro* DNT guidance.

1 Introduction

1.1. Background

The consequences of human developmental neurotoxicity (DNT) liabilities have a relevant socio-economic impact and the current increase in prevalence of neurodevelopmental disorders cannot be solely explained by genetic factors. The developing nervous system is especially vulnerable to certain chemicals and exposure may result in altered neural development with consequences that may be different from those observed with the chemical in adult nervous system. This is due to a plethora of neurodevelopmental processes that play only marginal roles in the adult brain or distinct roles of e.g. neurotransmitter signalling in the developing compared to the adult nervous system (Nguyen et al., 2001). Moreover, certain developing brain cell types, as for e.g. oligodendrocyte precursor cells, are more sensitive to stressors than their adult counterparts, rendering white matter development as a susceptible target. Despite the higher sensitivity of the developing nervous system, DNT hazard identification remains a 'hidden hazard' with the main reasons being the limited data requirements for DNT across the different jurisdictions and the uncertainties associated with the current *in vivo* DNT models (Paparella et al., 2020).

Annex II of the **Regulation No 1107/2009** on the placing of plant protection products on the market in the European Union (EU) considers DNT as a critical effect of particular significance, and studies may be carried out when there is indication for potential developmental neurotoxicity from others studies, or when the concern arises from the systematic literature review (Regulation 283/2013). In the EU, recognised testing methods include those listed in the **Commission Communication No 2013/C 95/01** (1998 US EPA DNT Guideline, OECD TG 426 and OECD TG 443 with a DNT cohort) (US EPA, 1998; OECD 2007 and 2018b).

In vivo DNT studies are complex in both study design and data interpretation, frequently resulting in diverging conclusions (Makris et al., 2009). A variety of *in vitro* methods (DNT-IVB) are now available, which utilise human-relevant cell systems capable of testing chemicals for the potential to disrupt key neurodevelopmental processes (Frank et al., 2017; Bal-Price et al., 2018; Masjosthusmann et al., 2020). The assumption is that dysregulation of a fundamental process in brain development has the potential to lead to an adverse neurodevelopmental outcome. Therefore, by means of mechanistic understanding, this IATA case study has been developed to acknowledge the contribution of the DNT-IVB in the DNT hazard characterisation of chemical substances and to understand the applicability of the DNT-IVB in cases where the DNT concern is unlikely to be directly linked to the intrinsic neurotoxic properties of a chemical.

1.2. Problem formulation

This case study is using flufenacet as a model chemical to test the regulatory problem formulation on DNT hazard characterisation (see below). Flufenacet is an approved herbicide active substance under

Regulation No 1107/2009. In Europe Flufenacet is not classified as a DNT substance. However, a OECD 426 DNT study was submitted, based on the evidence of changes in thyroid hormones in general toxicity studies. The US-EPA provided an assessment of this study and considered that a low concern exists for increase susceptibility of the pups (qualitative and quantitative based on decreased body weight observed at all dose levels, and additional effects, including decreased motor activity, delayed developmental landmarks, and decreases in morphometric measurements at mid and high doses). Therefore, this IATA case study is intended to use the outcome of the DNT-IVB to provide mechanistic information on the direct impact of flufenacet on the developing brain. Scientific peer-reviewed open literature studies were also submitted by the applicant during the peer-review process and appraised in line with Regulation 1107/2009.

The case study is therefore supporting to what extend the DNT-IVB informs the weight of evidence (WoE) approach and how it can provide in vitro fit-for-purpose information on the DNT hazard identification.

The specific problem formulation is therefore the following: How certain are we that the active substance flufenacet is a developmental neurotoxicant in humans based on the data collected, appraised, synthesised and integrated using an operational protocol in line with the IATA framework?

A structured scientific assessment approach using a predefined protocol for a systematic literature review, including screening for relevance, critical appraisal of the evidence and data extraction and analysis, was therefore applied. An uncertainty analysis was performed to conclude on evidence that should be integrated in the IATA iterative framework.

2 Purpose

2.1. Purpose of use

This IATA case study was developed to be included in the OECD Guidance under development by the DNT expert group of the OECD on the use and interpretation of DNT-IVB. Although additional case studies will be included in the OECD Guidance, this case study is intended to provide an example for the use and application of the DNT-IVB for single substance DNT hazard assessment. It should be noted that the DNT assessment as a consequence of hormonal disruption is not in the scope of the DNT IVB guidance.

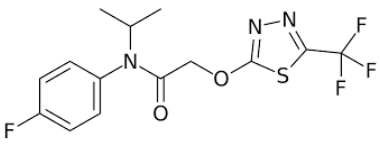
This IATA addresses DNT hazard characterisation of the herbicide active substance flufenacet in line with the drafted problem formulation: How certain are we that the active substance flufenacet is a developmental neurotoxicant in humans based on the data collected, appraised, synthesised and integrated using an operational protocol in line with the IATA framework?

A structured scientific assessment approach was employed, in line with the one used for the deltamethrin case study. This was defined by the inclusion of a predefined Systematic Review protocol for the literature (including the appraisal of the evidence with a critical appraisal tool and data extraction and analysis) and a quantitative uncertainty analysis. The AOP-informed IATA was used to integrate all the existing in vivo and in vitro data, including the DNT-IVB, in order to determine the certainty that the active substance flufenacet is or is not a developmental neurotoxicant and how this certainty is supported with the inclusion of the mechanistic information.

2.2. Target chemical(s)

The target chemical in this case was the synthetic aromatic amide oxyacetanilide herbicide flufenacet, which is the common name of *N*-(4-Fluorophenyl)-*N*-isopropyl-2-{[5-(trifluoromethyl)-1,3,4-thiadiazol-2-yl]oxy}acetamide (IUPAC name). Its chemical structure is shown in **Figure 1**.

Figure 1: Chemical structure of flufenacet



Flufenacet is a selective herbicide used in agriculture as foliar sprays in monocot and dicot crops; the mode of action is based on the inhibition of the biosynthesis of very long chain fatty acids resulting in inhibition of cell division and cell growth. It is a synthetic aromatic amide, approved under Regulation (EC) No 1107/2009 as an herbicide. It is rapidly absorbed after oral administration (around 75–94%) and excreted via urine 90%. It is widely distributed and extensively metabolised. Acute clinical signs are

non-specific and include ataxia, laboured breathing, decreased activity and, lacrimal, nasal and perianal staining. The main target organs after repeated exposure are liver, thyroid, kidney and eye and in dogs also nervous system (See EFSA-Q-2014–00810 for further details).Flufenacet was selected as a prototypical compound for which an OECD 426 study is available, but based on the known mechanism of action in plants (interference with cell wall synthesis), is not expected to alter mammalian neurodevelopment

2.3. Endpoint(s)

This IATA is performed for the assessment of the endpoint DNT following *in vitro* and *in vivo* exposure to flufenacet. Any adverse effect on the normal development of the nervous system structure and/or function, predefined by different specific endpoints measured *in vivo*, *in vitro* and in HOS, are presented in **Table 1** (details are presented in **Appendix A)**.

2.4. Exposure information

This IATA case study concerns hazard characterisation; therefore, exposure considerations are not included. Nevertheless, it is worth noting that flufenacet is an herbicide active substance and information on dietary and non-dietary exposures will be used in the risk assessment process while considering all routes of exposure.

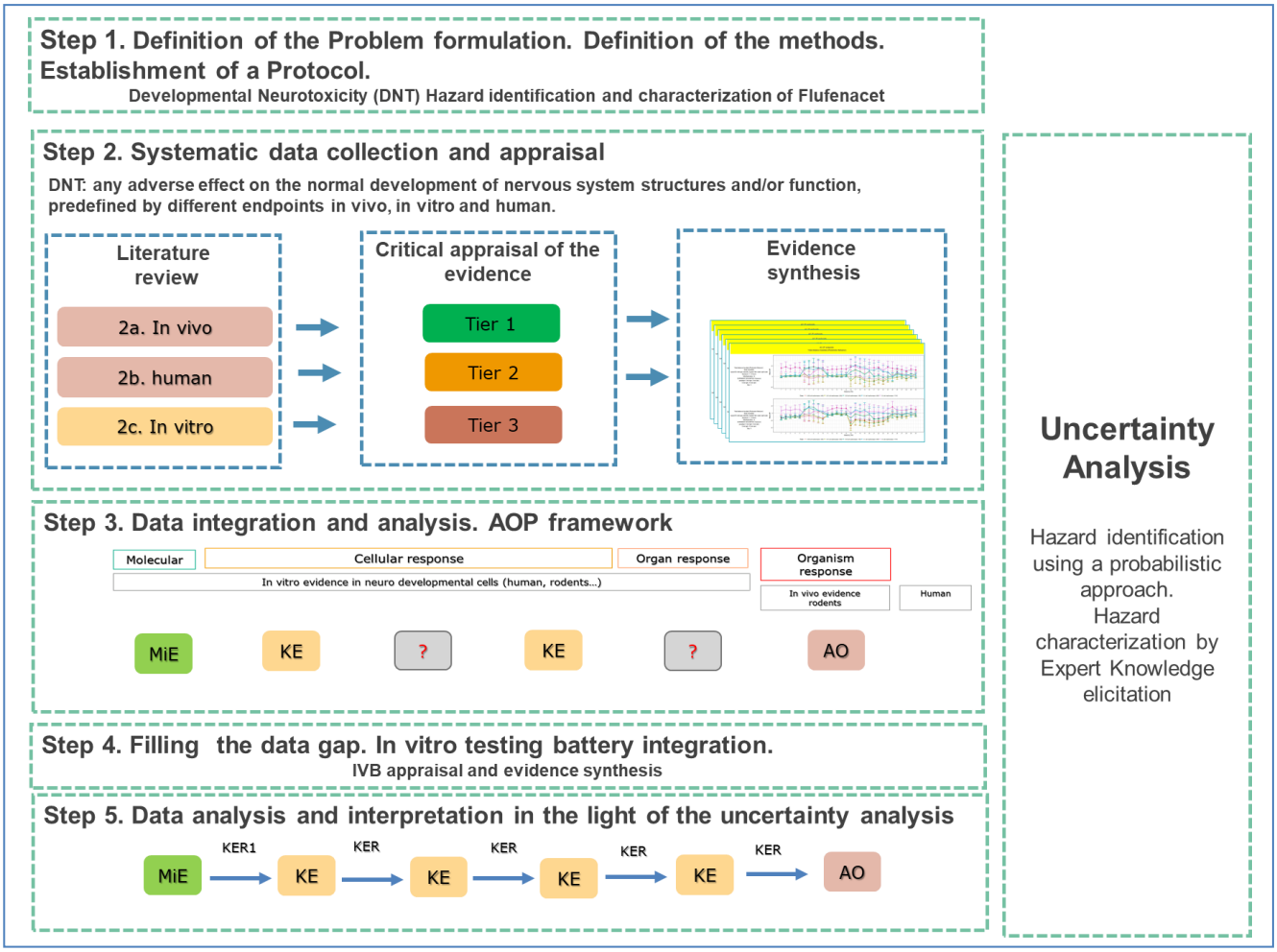
3 Rationale for performing the IATA and approach used

This IATA case study has been structured to test the hypothesis that the inclusion of in vitro data would support the DNT hazard identification and characterisation of flufenacet. The rationale applied follows the one used for developing IATA for deltamethrin and is sharing the same approach. However, based on the evidence retrieved for flufenacet, the resulting iterative approach was different.

The mechanistic understanding provided by the integration of the *in vitro* evidence in the AOP framework is expected to contribute to the WoE for DNT. The IATA workflow is presented in **Figure 2**. Therefore, this IATA is intended to address the problem formulation through an evidence-based approach and a pre-established protocol was developed covering the steps up to the evidence appraisal (Step 1 in **Figure 2**, see details in **Appendix A**). The protocol describes the strategy for the systematic literature review, including the screening for relevance, the data extraction and the critical appraisal of evidence retrieved in the scientific literature from *in vivo*, *in vitro* and human observational studies (HOS) (Step 2 in Figure 2. details of the studies retrieved, included and excluded are reported in **Appendix B1**). Only *in vivo* and *in vitro* studies from the literature ranked with low risk of bias were considered for further analysis. The AOP conceptual framework was therefore applied to integrate the evidence obtained from the systematic literature review (Step 3 in Figure 2) and identify data gaps. In the following step the results of testing in the DNT-IVB (Masjosthusmann et al., EFSA 2020) were included as KEs evidence for the postulation of the AOP in addition to the evidence retrieved from the systematic literature review (Step 4 in **Figure 2**).

An uncertainty analysis was performed for each line of evidence (*in vitro* and *in vivo*) to establish if a causal association between exposure to flufenacet and DNT can be established.

Figure 2: IATA workflow



4 Data gathering and application of IATA

4.1. Step 2 Data gathering. Literature searches and risk of bias

4.1.1. Systematic literature

The literature searches were conducted in three electronic bibliographic databases (PubMed, Web of Science, TOXNET) and three resources indexing PhD theses (DART, EBSCO and PQDR) in July 2020, and updated on 23 of November 2020 by an information specialist. Search strings are described in the protocol (**Appendix A**).

Terms for the exposure were combined with relevant terms for DNT outcomes (human and *in vivo* studies) or methods (*in vitro* studies) and a specific search string was designed to identify studies applying high-throughput methods to evaluate potential DNT without terms of exposure. The DNT outcomes were predefined by a series of toxicological *in vivo* and *in vitro* endpoints and measurements in HOS and categorised in endpoint categories translated into keywords for the searches (see **Table 1;** see also Appendix A Section 2.1.3,).

Two independent reviewers screened the literature identified through the searches; 137 unique references were identified after removing duplicates (see PRISMA Chart, **Figure 3**). The evidence was clustered as *in vivo* (containing *in vivo* experimental studies), *in vitro* (containing *in vitro* studies and behavioural studies with exposure conducted in zebrafish up to 120 hours post fertilisation) or human (containing HOS) in the title and abstract screening. The title and abstract screening left five relevant articles that underwent a full-text review, of those, three were classified as *in vitro*, one as human evidence and two as *in vivo* (of those one was also classified as *in vitro*). For the *in vivo* studies, one study was included, for *in vitro* studies a total of three studies were included (1 of them conducted in zebrafish up to 120 hours post fertilisation) and for human no relevant publications were retrieved (see full list of references included and excluded and reasons for exclusion, **Appendix B2.2**).

Figure 3: PRISMA chart systematic literature review result for flufenacet of the screening for relevance

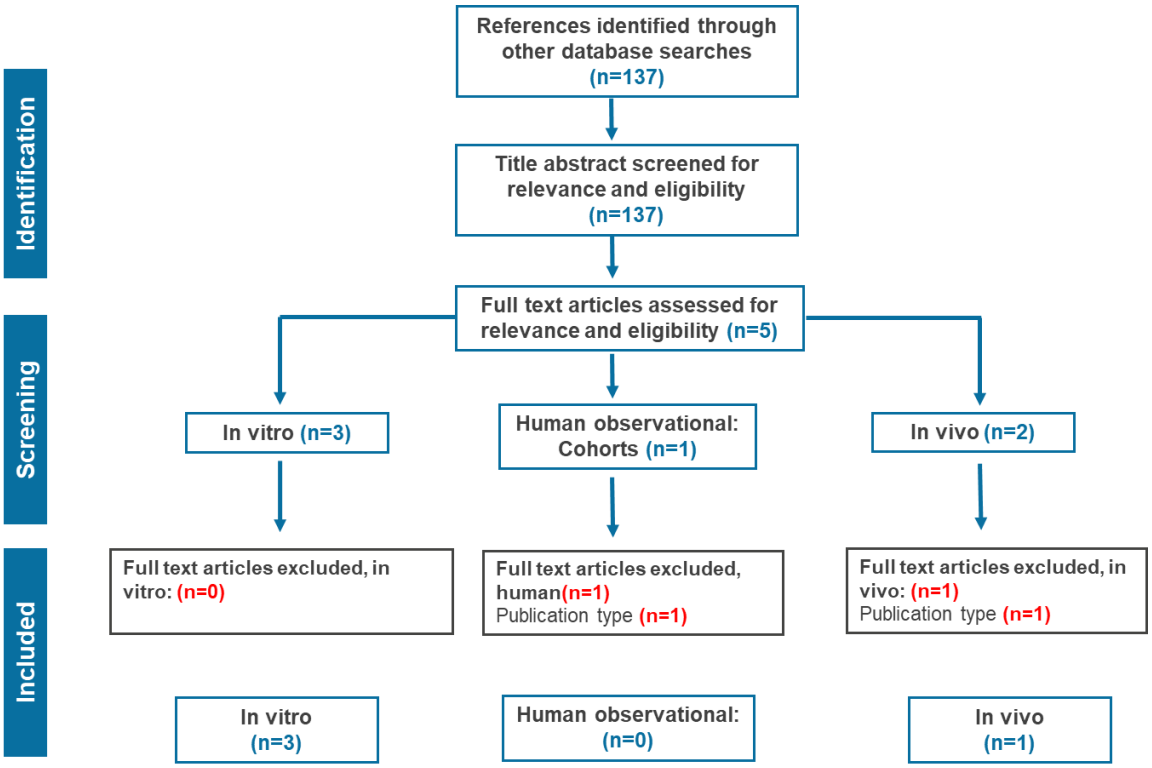


Table 1. Preliminary list of endpoints laid down during the protocol development

Developmental neurotoxicity predefined endpoints	Endpoints categories	Specific endpoints
In vivo experimental studies	Neuropathology endpoints	Brain weight Quantitative morphometric evaluation (linear measurement, areal measurements, brain morphometric landmarks, brain regions measured, stereology) Qualitative neuropathology examination (diagnostic criteria, severity score criteria, standard and special stain used, period) Neuroimaging (quantitative, e.g. MRI)
	Behavioural endpoints	Behavioural ontogeny (functional observation battery e.g. righting reflex, negative geotaxis, motor activity)…

		Motor activity Auditory startle response Learning and memory: Letter Mazes M, Y, E (Position, Discrimination: latency to escape, errors, trials to criterion). Morris Water Maze (Spatial Learning: latency to escape over trials, path length to locate hidden platform, search parameters during retention probe trials). Passive Avoidance (Associative Learning: latency, trials to criterion). Biel Water Maze (Sequential learning: latency, errors) Cincinnati Maze (Sequential/Egocentric Learning: latency, errors) Social behaviour (qualitative or quantitative, tbc test method) Elevated maze Swim test Tactile startle Conditioned freezing Amphetamine challenge MK801 challenge
	Clinical chemistry endpoints	Hormone levels Cholinesterase activity
	Endpoint category 'other'	Ophthalmological evaluation Neurotransmitter levels Long-term potentiation
In vitro experimental studies	Proliferation endpoints	
	Apoptosis endpoints	
	Differentiation	Neurogenesis Gliogenesis

		Oligodendrocyte differentiation Astrocytic differentiation
	Migration endpoints	Neuronal migration Radial migration. Glial migration
	Growth/maturation	Neurite outgrowth Neuronal morphology Synaptogenesis Neuronal cell types
	Network formation/function	Microelectrode array (MEA)
	Cytotoxicity/viability	MTT assay Lactate dehydrogenase (LDH) level Neutral red accumulation
	Channels/transporters	Sodium Calcium Chloride Potassium
	Proteins	Synaptophysin SNAP25 Synaptobrevin MAP2
	Receptors/ Neurotransmitters	GABA NMDA 5-HT
	Enzymatic activity	Calcineurin Calmodulin Dephosphorylation
	Microglia activation	TNF alpha
	Oxidative stress	ROS production Nrf2 expression/translocation Mitochondrial membrane integrity

	Cell organelles integrity	Nuclear integrity Lysosomal integrity Mitochondrial membrane integrity
	Neurophysiology/patch clamp	Membrane excitability
	Genomic	
	Behavioural endpoints (in zebrafish)	Thigmotaxis Locomotor activity Spasms Swimming activity
	Pathology (in zebrafish)	Cranio-morphological effects Curvature of the body axis Quantitative morphometric examination (body area, head area, head-body angle) Qualitative morphometric evaluation
Human observational studies	Intellectual disability	
	Communication disorders	
	Autism Spectrum Disorders (ASD)	
	Attention-deficit hyperactivity disorder (ADHD)	
	Motor disorders	
	Specific learning disorders	

4.1.2. Risk of Bias (RoB)

Risk of bias (RoB) in eligible studies on rodents (experimental toxicological studies) was appraised by endpoints using tailored versions of the OHAT-NTP RoB tool (NTP, 2015). For the *in vitro* studies, a non-validated tool developed by OHAT-NTP for a specific project (PFOA and PFOS Monograph (NTP, 2016)) was adapted. Critical appraisal tools were laid down upfront and are described in the protocol (**Appendix A**). **Table 2** shows the different questions and domains appraised for *in vivo*, *in vitro* and human lines of evidence and the three key questions selected for this specific assessment. Studies were classified as being at low (Tier 1), moderate (Tier 2) or high (Tier 3) RoB for each of the endpoints they measured. RoB tiers were derived weighing the appraisal from the individual RoB domains some of which were identified as key. The outcome of the RoB appraisal and descriptive forest plots are presented in **Appendix B.3.2** for *in vivo* and *in vitro* lines of evidence.

RoB was appraised by endpoint in the cases that it was considered that the different endpoints measured in a study used different methodology and thus had different RoB. RoB for *in vivo* studies was appraised by answering nine questions resulting in five endpoints categorised in Tier 1; RoB for *in*

vitro studies was appraised by answering 10 questions for 15 endpoints resulting in 15 endpoints in Tier 1 (summary of the studies and results of the RoB are presented in **Appendix B.3.2**).

Table 2: Critical appraisal tool questions for the RoB analysis and key questions selected for each line of evidence (see Appendix A. Protocol for further details)

Selection Bias	In vivo	In vitro
1. Was administered dose or exposure level adequately randomized?	Key Q	Yes
2. Was allocation to study groups adequately concealed?	Yes	-
3. Did selection of study participants result in appropriate comparison groups	-	Yes
Counfounding Bias		
4. Did the study design or analysis account for important confounding and modifying variables?	-	-
Performance Bias		
5. Were experimental conditions identical across study groups?	Yes	Yes
6. Were the research personnel blinded to the study group during the study?	Yes	Yes
Attrition/Exclusion Bias		
7. Were outcome data complete without attrition or exclusion from analysis?	Yes	Yes
Detection Bias		
8. Can we be confident in the exposure characterisation?	Key Q	Key Q
9. Can we be confident in the outcome assessment?	Key Q	Key Q
Selective Reporting Bias		
10. Were all measured outcomes reported?	Yes	Yes
Other Sources of Bias		
11. Were there other potential threats to internal validity?	Sistemic tox	Cytotoxic
12. Were there other potential threats to internal validity?	-	NumberRep

4.2. Step 3. Systematic review data and uncertainty analysis

The workflow shown in **Figure 2** illustrates how the *in vivo* and *in vitro* lines of evidence were extracted, appraised, weighted and integrated. This allowed for assessing whether inclusion of results generated from the IVB would strengthen the conclusion of the hazard identification and characterisation of DNT liability of flufenacet. The hazard identification uncertainty analysis was conducted by addressing the questions in **Table 3**. All the graphs containing raw data are presented in **Appendix B4.3** and all the identified sources of inconsistencies and uncertainties were listed in tables presented in **Appendix B5.2** for *in vivo* and *in vitro* lines of evidence. Then, the Working Group experts (composed by 7–12 experts) provided their final judgement on the causal association of the experimental endpoints with flufenacet in terms of bounded probability. A threshold of 66% (twice as possible as not) was used as the minimum subjective probability leading to the conclusion of a causal association. Only MIEs, KEs and adverse outcomes (AO) with an estimated probability ≥66% were considered for the next step of the AOP network postulation.

Tables 4 and **5** include the list of methods of the studies included in flufenacet assessment and its appraisal for *in vitro* and *in vivo* lines of evidence.

This process resulted in only one in vivo study and 3 in vitro studies.

Only MIEs, KEs and AOs with an estimated probability of at least 66% in the UA were selected for the AOP postulation based on the robustness of their empirical support (see **Appendices B5.2**). Inconsistency in the regulatory assessment of the DNT in vivo study between EFSA and US EPA was highlighted during the review. This was noted as part of the review of the OECD DNT IVB guidance were the draft version of the IATA was also included. Therefore, the authors provided an update of the UA where it was concluded that the histomorphometry changes cannot be dismissed and with the only available in vivo data, a DNT AO should be included in this IATA.

Table 3. Assessment questions for the uncertainty analysis for hazard identification (see Appendix B1, Statistical Report for further details)

Line of evidence	Question 1. Hazard identification	Expression of the uncertainty (probabilistic assessment)
In vitro experimental studies	Does exposure to flufenacet triggers the specific endpoint/KE/AO (for zebrafish) as measured in acute and developmental protocol (washout yes/no) (assuming a monotonic concentration-response relationship) and in *in vitro* studies (EFSA DNT-IVB and literature) carried out in human and/or rat and/or mouse neural cells in development?	**Bounded probability** No: Prob < 0.66 Yes: Prob ≥ 0.66
In vivo experimental studies	Does exposure to flufenacet affect this specific endpoint/ endpoint category/adverse outcome in a dose-response relationship in **experimental animal studies** exposed during pregnancy and/or post-natally until weaning [maximum up to 21 days postnatal for rats and mice]?	**Bounded probability** No: Prob < 0.66 Yes: Prob ≥ 0.66

Table 4. Summary of Material and Methods and RoB of the studies providing evidence for the Molecular Initiating Event (MIE), Key Event (KE) and adverse outcome (AO) of the postulated AOP after the uncertainty analysis (UA) for in vivo evidence

Publication	Endpoint measured (for the results, see graphs in Appendix B4.3)	Study characteristics	Exposure characteristics	RoB tier
RefID183 **A Developmental Neurotoxicity Screening Study with Technical Grade Flufenacet in Wistar Rats. OECD 426**	**Neurohistopathology** Liver, thyroid/parathyroid weight and histopathology Serum T3 and T4	Sprague-Dawley rat Males and Females Standard housing condition GLP, OECD 426 compiling Private funding source Group size: 6 rats/dose	From GD 6 to PND 11via feed to the dams at 20, 100 and 500 ppm plus control	**Tier 3**
RefID183 **A Developmental Neurotoxicity Screening Study with Technical Grade Flufenacet in Wistar Rats. OECD 426**	**Learning and memory** (Average Errors; duration; trials to criterion; average Errors) during learning and retention phases Passive Avoidance performance - learning phase during	Sprague-Dawley rat Crl:CD^{R} BR VAF7Plus Males and Females Standard housing condition GLP, OECD 426 compiling Private funding source	From GD 6 to DL 11 via feed to the dams at 20, 100 and 500 ppm plus control	**Tier 1**

	learning and retention phases **Motor activity** **Startle**	Group size: 18–19 rats/dose		

Table 5. Summary of Material and Methods and RoB of the in vitro studies

Publication	Endpoint categories, specific endpoints and measurements (for the Results, see graphs in Appendix B 4.3)	Study characteristics	Exposure characteristics	RoB tier
RefID: 3204. Screening the ToxCast phase II libraries for alterations in network function using cortical neurons grown on multiwell microelectrode array (mwMEA) plates. Strickland JD, Martin MT, Richard AM, Houck KA, Shafer TJ, 2017	Network formation (MEA) % Spikes in Bursts Burst duration Interspike Interval Intervals between bursts MBR (Mean Burst Rate) MISIB (Mean Interspike Interval In Burst) MFR (Mean Firing Rate)	Rat cortical primary cells PN 0–24 h Chemically defined medium Public funding source Number of biological replicates: 3	Duration 12 day. Concentration: 0, 0.03, 0.1, 0.3, 1, 3, 10 and 30 µM Chronic treatment	Tier 1
RefID: 3202. Evaluation of chemical effects on network formation in cortical neurons grown on microelectrode arrays. Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP	Network formation (MEA) % Spikes in Bursts Burst duration Interspike Interval Intervals between bursts MBR (Mean Burst Rate) MISIB (Mean Interspike Interval In Burst) MFR (Mean Firing Rate)	Rat cortical primary cells PN 0–24 h Chemically defined medium Public funding source Number of biological replicates: 3	Duration 12 day. Concentration: 0, 0.03, 0.1, 0.3, 1, 3, 10 and 30 µM Chronic treatment	Tier 1
RefID: 3205. Multidimensional In Vivo Hazard Assessment Using Zebrafish Truong L, Reif D, St. Mary L, Geier M, Truong H, Tanguay R, 2013	Zebrafish. Locomotor activity. Photomotor Response (EPR – Embryo photomotor response)%	Tropical 5 D wild type zebrafish cells PN 0–24 h Chemically defined medium Public funding source High-throughput, one embryo per well Number of replicates:32	Duration 120 hours. (analysis at 24 or 120 hours post fertilisation) Concentration: 0 broad concentration range spanning four orders of magnitude (6.4 nM to 64 µM)	Tier 1

4.3 Step 4. data gap. Summary of the methods and results for the DNT in vitro battery (IVB)

In vitro methods used for gathering data follow the OECD GD 211 for describing non-guideline *in vitro* test methods. However, OECD GD 211 pays little attention to the description of the test system (i.e. cell culture) and the steps leading to it being established. Therefore, OECD GD 211 was modified to adapt it to the common needs of test developers and regulators by developing an annotated toxicity test method template (ToxTemp; Krebs et al., 2019). ToxTemp for the individual test methods contributing to this IATA case study can be found in Masjosthusmann et al. (2020), Appendices C–G¡.

4.3.1. Test systems and test methods of DNT-IVB

Briefly, test methods are based on test systems geared to assess key neurodevelopmental processes that are needed for brain development and occur in a concerted, spatiotemporal fashion *in vivo.* Seven test systems, i.e. proliferating primary fetal human neural progenitor cells (hNPC; test method NPC1), differentiating hNPC (test method NPC2–5), human induced pluripotent stem cells (hiPSC)-derived neural crest cells (NCC, test method UKN2), Lund human mesencephalic (LUHMES) cells (UKN4), hiPSC-derived peripheral neurons (UKN5), primary rat cortical mixed cultures (rNNF) and the NeuCyte SynFire kit consisting of hiPSC-derived excitatory and inhibitory neurons and human primary astrocytes (hNNF) build the cellular bases of the test methods mimicking certain neurodevelopmental processes over time *in vitro*. Thus, the assays measures compounds' effects on the developmental change of a key neurodevelopmental process over time. Such processes (and corresponding test methods) are neural progenitor cell proliferation (NPC1), migration of: (a) NCC (UKN2), (b) radial glia (NPC2a), (c) neurons (NPC2b), (d) oligodendrocytes (NPC2c), neuronal differentiation (NPC3), neurite outgrowth/morphology of: (a) dopaminergic (UKN4), (b) peripheral (UKN5) and (c) young cortical (NPC4) neurons, oligodendrocyte differentiation (NPC5) as well as rat and human neuronal network formation (rNNF and hNNF, respectively). Endpoint evaluations of these DNT test methods were accompanied by viability and/or cytotoxicity analyses leading to the generation of 23 individual DNT-specific (neurodevelopmental process) and unspecific (viability/cytotoxicity) assessments. At the time, except for the hNNF assay, these test methods had high enough readiness to be used for a screening approach (Bal-Price et al., 2018).

The methodological details are fully reported in Masjosthusmann et al. (2020), in the following paragraphs, a summary of the test methods is laid out:

1) The test method NPC1 assesses NPC proliferation in a 3D neurosphere format. Therefore, individual neurospheres of defined sizes are grown in suspension culture in single wells of a 96-well plate. Spheres are treated with respective compounds for 72 hours. For the last 16 hours of proliferation, BrdU is added to the cultures and BrdU incorporation into the DNA is measured via BrdU antibody binding using luminescence. At the same time mitochondrial reductase as a possible marker of cell viability is measured via the Alamar blue assay and lactate dehydrogenase (LDH) release into the medium, an indicator for cytotoxicity, is quantified.
2) The UKN2 test method measures NCC migration into an area devoid of cells. Therefore, hiPSC-derived NCC are seeded into 96-well plates containing a stopper in the centre of each well. After stopper removal, cells are allowed for 24 hours to migrate into the cell-free area. Area covered due to cell migration is quantified via high content image analysis (HCA). In parallel, cell viability is assessed by measuring calcein-AM staining via HCA.
3) The NPC2–5 test methods run in a multiplexed fashion. Here, migrated cells are stained for Hoechst (nuclei), TUBB3 (neurons) and O4 (oligodendrocytes) simultaneously. For all test methods NPC2–5, LDH is measured after 72 and 120 hours and the Alamar blue assay is performed after 120 hours. The NPC2 migration assays consist of the three test methods, NPC2a,b and c. Each method assesses migration of an individual cell type in the neurosphere

migration area. The predominant cell type determining total migration distance of the neurospheres are radial glia cells. Therefore, NPC2a measures total migration distance via phase contrast images after 72 hours and via high content imaging HCI of Hoechst-stained nuclei after 120 hours. Simultaneously, migration distances of stained neurons (see 4., NPC3) and oligodendrocytes (see 6., NPC5) are assessed within NPC2b and c, respectively. Here, the mean distance of all neurons (NPC2b) or oligodendrocytes (NPC2c) from the edge of the sphere core to the position of each neuron is measured and plotted as a ratio of radial glia migration (NPC2a). This approach allows the evaluation of the cell type-specificity of a migratory effect.

4) Neuronal differentiation is quantified within the NPC3 test method. The number of TUBB3-positive cells is determined by immunocytochemical staining, HCA and annotation by a convolutional neuronal network (CNN) running on Keras implemented in Python 3. Neuronal differentiation is given as the percentage of TUBB3-positive cells of all Hoechst-positive nuclei in the migration area after 120 hours of differentiation.

5) Young, central neurons identified with the CNN in test method NPC3 now undergo morphological evaluation with the NPC4 test method. Therefore, neurons are skeletonised with a self-written algorithm (Schmuck et al., 2017) and neurite length (µm) and neurite area (pixel) are determined for each individual neuron.

6) Oligodendrocyte differentiation is assessed with the NPC5 assay. The number of O4-positive cells is determined by immunocytochemical staining, HCA and annotation by a CNN running on Keras implemented in Python 3. Oligodendrocyte differentiation is given as the percentage of O4-positive cells of all Hoechst-positive nuclei in the migration area after 120 hours of differentiation.

7) Neurite area of central, dopaminergic neurons is measured in LUHMES cells by the UKN4 test method 24 hours after start of compound treatment. Cells are stained with calcein-AM and H-33342 and HCA is performed. An automated algorithm is used to identify neurite area and counting double-positive cells for calcein-AM and H-33342 assesses viability.

8) Neurite area of peripheral, hiPSC-derived neurons is measured by the UKN5 test method 24 hours after start of compound treatment. Cells are stained with calcein-AM and H-33342 and HCA is performed. An automated algorithm is used to identify neurite area and counting double-positive cells for calcein-AM and H-33342 assesses viability.

9) The rNNF assay measures network formation based on primary rat cortical cells by exposing the culture to compounds 2 hours after plating on microelectrode arrays (MEAs). Cells are treated again with compounds at days 5 and 9, whereas recordings are collected on days 2, 5, 7, 9 and 12. After the last measurement at day 12, viability is assessed. Data are analysed from the area under the curve across the measurements at different points.

10) Network formation and function based on human pre-differentiated hiPSC-based excitatory and inhibitory neurons as well as primary astrocytes (SynFire Kit, NeuCyte), is assessed with the hNNF assay. The cells are plated on MEAs and exposed to compounds after 7 days of differentiation. Recordings are performed every week on day 7, 14, 21, 28 and 35. To minimise possible acute effects of compounds in the measurements, in contrast to the rNNF assay, a washout is performed one day before each recording day. Additionally, cytotoxicity is assessed after 3 days of compound exposure on days 10, 17, 24 and 35. Data are analysed from the area under the curve across the measurements at different points.

For the storage and management of all testing data a database was set up at the Ruhr University Bochum, Germany. It operates on Post-greSQL and is designed to store raw and meta data for subsequent data processing. To enable assay-specific data analysis, a data evaluation tool was developed in R. It uses the data.table package for data management, as well as the drc and bmd packages for curve fitting and plotting. Briefly, concentration-response data were normalised to the median solvent control and re-normalised to the starting point of the curve, the R software (drc package) was used to calculate the optimal curve fit for each experiment by running several non-linear models and using the Akaike´s information criteria to assess its performance, and the benchmark concentrations

(BMCs) according to individual assays' predefined benchmark responses (BMRs), as well as upper and lower confidence limits were calculated by parametric bootstrapping in combination with median resampling (200-times). BMCs were used for assay-specific decision trees for compound classification into 'DNT-specific hit', 'non-specific hit', 'borderline hit' or 'no hit'. A 'DNT-specific hit' is defined by an alteration of a DNT-specific endpoint that can be distinguished from general cytotoxicity, while an 'unspecific hit' does not allow this distinction and therefore the change in DNT endpoint is due to general cytotoxicity. A 'borderline hit' cannot clearly be classified in either a 'DNT-specific' or 'unspecific hit' and 'no hit' does not produce a BMC value (for more details on classification see Masjosthusmann et al., 2020).

4.3.2. Results summary

A summary of BMCs resulting in DNT-specific or unspecific hits for flufenacet is given in **Table 6**. All other endpoints not listed in the table did not result in BMCs and were thus not hits. Concentration-response curves underlying these BMCs can be found in Masjosthusmann et al. (2020) and Frank et al., 2017. Flufenacet did not result in any DNT-specific hits and the BMC observed for oligodendrocyte differentiation was considered as a non-specific hit.

Table 6. Summary of BMC across the positive assays of the DNT testing battery

	Flufenacet
BMC_{25} migration (UKN2)	>100*
BMC_{25} neurite area (UKN5)	>100*
BMC_{10} migration radial cell (NPC2)	>20**
BMC_{30} neurite length (NPC4)	>20**
BMC_{30} neurite area (NPC4)	>20**
BMC_{30} oligodendrocyte differentiation (NPC5)	17.8^{ns}
BMC_{50} rat neuronal network formation (rNNF)	>20
BMC_{50} human neuronal network formation (hNNF)	>20
BMC_{30} neurite maturation (rat cortical neurons 2)	–
BMC_{30} synaptogenesis (rat cortical neurons 2)	–

ns = non-specific hits (the BMC for the effect is not separated from the BMC for cytotoxicity/cell viability).
Specific hits according to respective compound are marked in bold.
Units for the numerical values given in the table correspond to µM.
*No Benchmark Response (BMR) at concentration higher than 100 µM in the absence of cytotoxicity. The BMR is a value of effect size and is defined as an effect size that is higher than the general variability of the measured endpoint and is therefore determined based on the variability of the respective endpoint. For UKN 2 and 5 the BMR is set at 20% (BMR20). If the parameter is not affected at the BMR20 the compound is classified as a 'no hit'.
**No Benchmark Response (BMR) at concentration higher than 20 µM in the absence of cytotoxicity. For NPC 2, 4 and 5 the BMR is defined as at least 1.5× the standard deviation (SD) (between experiment variation) as BMR10 for migration distance radial glia 72 h and 120 h, cytotoxicity 72 h and 120 h and BMR30 for all other endpoints.

4.4. Step 5. Data analysis and interpretation, AOP postulation

As a result of the uncertainty analysis a DNT AO was concluded in the evaluated *in vivo* study (changes in the linear measurements of the caudate putamen in females at PND72). No more in vivo studies or HOS were retrieved in the literature for flufenacet.

A larval zebrafish study was retrieved from literature measuring photomotor behavioural changes and was considered negative in the uncertainty analysis. The outcome of the systematic review on in vitro data complemented by the DNT-IVB (Masjosthusmann et al., 2020) that measured chemical-induced

changes on potential DNT KEs and MiEs indicated lack of any relevant empirical data on mechanistic information on KEs or MiEs for an DNT-AOP.

4.4.1. Adverse outcome

The AO was defined as an effect relevant for regulatory decisions. The evidence retrieval resulted in only one study being used for the description of the AO, the regulatory OECD TG 426 study presented by applicant for its registration as pesticide active substance.

The study was conducted in line with OECD TG 426. However, based on the criteria in the CAT (see Appendix A for the criteria and B3.2 for the detailed appraisal) some of the endpoints were classified in Tier 3 due to only one or two doses evaluated.

The outcome of the study was assessed differently across different regulatory bodies (EFSA conclusion and USEPA 2011 to include Canada). For the scope of this IATA a conservative approach was taken, and the adverse outcome is defined as changes in linear measurements of the caudate putamen area in female rats at PND72 from 1.7 mg/kg bw.

4.4.2. Integration of the DNT IVB results

The outcome of the DNT-IVD didn't provide any mechanistic evidence to link the AO observed in the OECD TG 426 with the key events tested in the battery and a DNT-IVB informed AOP cannot be developed. **Therefore, the iterative process was stopped and no DNT AOP was postulated.**

5 Discussion and Conclusion

The overall purpose of this IATA was to address the question: How certain are we that the active substance flufenacet is a developmental neurotoxicant in humans based on the data collected, appraised, synthesised and integrated using an operational protocol and in line with the IATA framework?

This was addressed through the identification of lines of evidence and through the the uncertainty analysis of the. The revised outcome of the analysis indicates that evidence is available for which a probability of at least 66% would support the identification of a DNT AO. This conclusion was reached after the IATA commenting period and a conservative approach was taken regarding the outcome of the OECD TG 426 study (see Appendix B5.2) This consideration is relevant, because this IATA is intended to be used to understand the contribution of the DNT-IVB to identify substances having developmental neurotoxic properties.

Regarding the analysis of the in vitro data (only data from the DNT IV-B was available), there was no evidence of any DNT KE; therefore, a DNT IV-B informed AOP, informed IATA, cannot be developed. This is based on the scientific assumption that the DNT IV-B is measuring fundamental key neurodevelopment processes that, if disrupted, would culminate in an AO. Although some limitations are recognized for the DNT IV-B (e.g. metabolic competence, hormonal based mechanism, myelination, astrocyte toxicity), the battery was considered sufficient based on current scientific knowledge and test readiness (Bal-Price et al., 2018).

Following the integration of all evidence from *in vivo* and *in vitro* studies, it is concluded that flufenacet is unlikely to be a developmental neurotoxicant in humans throughout a direct mechanism. The inclusion of the DNT-IVB data has been critical to contextualize the outcome of the in vivo study. Therefore, with the assumption that the AO observed in the in vivo regulatory study was not consequent to an AOP derived from the KEs measured by the DNT IV-B, alternative AOPs should be investigated/proposed. This is indeed the case of flufenacet, a substance able to change the circulating levels of thyroid hormones. Therefore, a causative link between the observed AO and decrease in THs cannot be excluded. As discussion on DNT effects consequent to hormonal changes is out of the scope for this IATA, no additional actions where taken to investigate this pathway. In addition, testing flufenacet metabolites in the DNT-IVB would reduce the uncertainty associated with the limited metabolic competence of the test systems.

This case study supports the applicability of the DNT IVB for data poor chemicals or to test the hypothesis that a substance is not DNT through a direct neurotoxic mechanism. The negative results in the zebrafish study were consistent with the lack of effects in the IVB. Although additional work is needed to address the ability of these new approach methodologies to capture DNT effects mediated by hormonal changes and metabolites, this case study helps in delineating the toxicological applicability domain of the DNT-IVB and supports the inclusion of the alternative model zebrafish in the DNT hazard identification and characterisation process. These considerations are relevant, and this case study should help the OECD DNT expert group in how to address them in the DNT *in vitro* guidance.

References

Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, … and Leist M, 2018. Recommendation on test readiness criteria for new approach methods (NAM) in toxicology: exemplified for developmental neurotoxicity (DNT). ALTEX – Alternatives to Animal Experimentation, 35(3), 306–352. doi: 10.14573/altex.1712081.

Frank CL, Brown JP, Wallace K, Mundy WR and Shafer TJ, 2017. From the cover: developmental neurotoxicants disrupt activity in cortical networks on microelectrode arrays: results of screening 86 compounds during neural network formation. Toxicological Sciences, 160(1), 121–135. https://doi.org/10.1093/toxsci/kfx169

Krebs, A., Waldmann, T., Wilks, M. F., van Vugt-Lussenburg, B. M., Van der Burg, B., Terron, A., ... & Leist, M. , 2019. Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. ALTEX – Alternatives to Animal Experimentation, 36, 682–699, doi: 10.14573/altex.1909271.

Makris SL, Raffaele K, Allen S, Bowers WJ, Hass U, Alleva E, Calamandrei G, Sheets L, Amcoff P, Delrue N and Crofton KM, 2009. A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. Environmental Health Perspectives, 117(1):17–25. doi: 10.1289/ehp.11447

Masjosthusmann S, Blum J, Bartmann K, Dolde X, Holzer A-K, Stürzl L-C, … and Fritsche E, 2020. Establishment of an a priori protocol for the implementation and interpretation of an in-vitro testing battery for the assessment of developmental neurotoxicity, EFSA supporting publications 2020;17(10):1938E, 152 pp. doi: 10.2903/sp.efsa.2020.EN-1938

Nguyen, L., Rigo, J. M., Rocher, V., Belachew, S., Malgrange, B., Rogister, B., ... & Moonen, G., 2001. Neurotransmitters as early signals for central nervous system development. Cell and Tissue Research, 305, 187–202. https://doi.org/10.1007/s004410000343

NTP, 2015. OHAT Risk of Bias Rating Tool for Human and Animal Studies, Office of Health Assessment and Translation, National Toxicology Program, 2015.

NTP, 2016. Monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS). National Toxicological Program, Research Triangle Park.

OECD, 2007. Test No. 426: Developmental Neurotoxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264067394-en

OECD, 2018a. Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. OECD Series on Adverse Outcome Pathways, No. 1, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jlv1m9d1g32-en

OECD, 2018b. Test No. 443: Extended One-Generation Reproductive Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264185371-en.

Paparella M, Bennekou SH and Bal-Price A, 2020. An analysis of the limitations and uncertainties of *in vivo* developmental neurotoxicity testing and assessment to identify the potential for alternative approaches. Reproductive Toxicology, 96, 327–336. https://doi.org/10.1016/j.reprotox.2020.08.002

Schmuck, M. R., Temme, T., Dach, K., de Boer, D., Barenys, M., Bendt, F., ... & Fritsche, E. (2017). Omnisphero: a high-content image analysis (HCA) approach for phenotypic developmental

neurotoxicity (DNT) screenings of organoid neurosphere cultures in vitro. Archives of toxicology, 91(4).

US EPA (US Environmental Protection Agency), 1998. Health effects guidelines OPPTS 870.6300 Developmental Neurotoxicity Study, EPA/ 712/c-98/239.Office of Prevention Pesticides and Toxic Substances.

Appendix A: Protocol of the IATA

Please refer to the separate publication for full Appendix A.

ENV/CBC/MONO(2022)24/ANN1

Appendix B1: Statistical analysis report

Please refer to the separate publication for full Appendix B1.

ENV/CBC/MONO(2022)24/ANN2

Appendix B2.2 List of included studies and excluded with reason flufenacet

DNT Flufenacet. Excluded studies

Humans

Not humans

European Food Safety Authority, 2008. Conclusion regarding the peer review of the pesticide risk assessment of the active substance. EFSA Journal, 6, 84.

In vivo

Not *In vivo*

European Food Safety Authority, 2008. Conclusion regarding the peer review of the pesticide risk assessment of the active substance. EFSA Journal, 6, 84.

In vitro

Not *In vitro*

European Food Safety Authority, 2008. Conclusion regarding the peer review of the pesticide risk assessment of the active substance. EFSA Journal, 6, 84.

DNT Flufenacet. Included studies

In vivo

[redacted] Developmental Neurotoxicity Study of Technical Grade Flufenacet Administered Orally via Diet to Crl:CD®BR VAF/Plus® Presumed Pregnant Rats

In vitro

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K and Friedman KP, 2019. Evaluation of chemical effects on network formation in cortical neurons grown on microelectrode arrays. Toxicological Sciences, 169, 436-455. doi: 10.1093/toxsci/kfz052.

Strickland JD, Martin MT, Richard AM, Houck KA and Shafer TJ, 2018. Screening the ToxCast phase II libraries for alterations in network function using cortical neurons grown on multi-well microelectrode array (mwMEA) plates. Archives of Toxicology, 92, 487–500. doi: 10.1007/s00204-017-2035-5. Epub 2017 Aug 2.

Truong L, Reif D, St Mary L, Geier M, Truong H and Tanguay R, 2013. Multidimensional *in vivo* hazard assessment using zebrafish. Toxicological Sciences, 137, 212–233.

Appendix B3.2 Outcome of the ROB flufenacet

Critical appraisal table - In vivo

Behavioural: Learning and memory

RefID:183

Study characteristics and population	Specific endpoint	Learning and memory – M-WM Learning phase (Average errors); Learning and memory – M-WM Learning phase (Duration); Learning and memory – M-WM learning phase (Trials to criterion); Learning and memory – M-WM Retention phase (Average errors); Learning and memory – M-WM Retention phase (Duration); Learning and memory – M-WM retention phase (Trials to criterion); Learning and memory – Passive Avoidance performance – learning phase (Latency); Learning and memory – Passive Avoidance performance – retention phase (Latency); Learning and memory – Passive Avoidance performance (trial to criterion – Learning phase); Learning and memory – Passive Avoidance performance (trial to criterion – Retention phase)
	Species:	Rat
	Strain:	Sprague-Dawley
	Sex:	Female; Male
	Housing condition:	Standard (GLP)
	Funding source:	Private
	Group size:	19
Exposure	Period of exposure:	from GD 6 to PDN 11
	Route of administration:	Oral (via feed)
	Vehicle:	Corn oil
	Doses:	0 ppm; 20 ppm; 100 ppm; 500 ppm
	Dose frequency:	Day
Measurement	Measurement time:	23 PDN; 23.1 PDN; 23.2 PDN; 30 PDN; 30.1 PDN; 59 PDN; 66 PDN
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Yes. 'Following at least six days of acclimation, P-generation females were weighed and those with a body weight more or less than 20% of the mean weight [18] were rejected. The remaining females were assigned to the control or an exposure group in sequence, as they were determined to be inseminated. This approach was used to ensure unbiased assignment of the animals to dose groups and that there would be approximately an equal number of litters from each dose group available for testing on a given day.'
Allocation concealment	Q2: was allocation [of animals] to study groups adequately concealed?	PLRoB	Not relevant.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes. Lots of methods descriptions illustrate this.
Blind research personnel	Q4: was the research personnel* blinded to the study group during the study?	NR	Not reported.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. All exceptions were noted and all data were presented in appendices.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Yes. Extensive analytical work was conduction on the diet and brain concentrations were measured in offspring.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	Yes. Not blinded but Well established methods, extensive validation data provided. GLP work.
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Yes. Extensive methods and results sections with appendices that contain all data.
Toxicity	Q9: were there other potential threats to internal validity?	DLRoB	Maternal body weight gain was marginally impacted, but not enough to bias results.

Behavioural: Motor activity

RefID:183

Study characteristics and population	Specific endpoint	Motor activity (Open field) – Number of movements; Motor activity (Open field) – Number of movements Habituation; Motor activity (Open field) – Time spent in movement; Motor activity (Open field) – Time spent in movement Habituation
	Species:	Rat
	Strain:	Sprague-Dawley
	Sex:	Female; Male
	Housing condition:	Standard (GLP)
	Funding source:	Private
	Group size:	
Exposure	Period of exposure:	from GD 6 to PDN 11
	Route of administration:	Oral (via feed)
	Vehicle:	Corn oil
	Doses:	0 ppm; 20 ppm; 100 ppm; 500 ppm; NA
	Dose frequency:	Day
Measurement	Measurement time:	14 PDN; 18 PDN; 22 PDN; 60 PDN
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Yes. 'Following at least six days of acclimation, P-generation females were weighed and those with a body weight more or less than 20% of the mean weight [18] were rejected. The remaining females were assigned to the control or an exposure group in sequence, as they were determined to be inseminated. This approach was used to ensure unbiased assignment of the animals to dose groups and that there would be approximately an equal number of litters from each dose group available for testing on a given day.'
Allocation concealment	Q2: was allocation [of animals] to study groups adequately concealed?	PLRoB	Not relevant
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes. Lots of methods descriptions illustrate this

Blind research personnel	Q4: was the research personnel* blinded to the study group during the study?	NR	Not reported.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. All exceptions were noted and all data were presented in appendices.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Yes. Extensive analytical work was conduction on the diet and brain concentrations were measured in offspring.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	Yes. Not blinded but Well established methods, extensive validation data provided. GLP work.
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Yes. Extensive methods and results sections with appendices that contain all data.
Toxicity	Q9: were there other potential threats to internal validity?	DLRoB	Maternal body weight gain was marginally impacted, but not enough to bias results.

Behavioural: Startle

RefID: 183

Study characteristics and population	Specific endpoint	Startle – Auditory Startle Reflex (Peak Amplitude Habituation); Startle – Auditory Startle Reflex (Peak Amplitude)
	Species:	Rat
	Strain:	Sprague-Dawley
	Sex:	Female; Male
	Housing condition:	Standard (GLP)
	Funding source:	Private
	Group size:	18
Exposure	Period of exposure:	from GD 6 to PDN 11
	Route of administration:	Oral (via feed)
	Vehicle:	Corn oil
	Doses:	0 ppm; 20 ppm; 100 ppm; 500 ppm
	Dose frequency:	Day
Measurement	Measurement time:	23 PDN; 23.1 PDN; 23.2 PDN; 23.3 PDN; 23.4 PDN; 23.5 PDN; 61 PDN; 61.1 PDN; 61.2 PDN; 61.3 PDN; 61.4 PDN; 61.5 PDN

Risk of bias appraisal	Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Yes. 'Following at least six days of acclimation, P-generation females were weighed and those with a body weight more or less than 20% of the mean weight [18] were rejected. The remaining females were assigned to the control or an exposure group in sequence, as they were determined to be inseminated. This approach was used to ensure unbiased assignment of the animals to dose groups and that there would be approximately an equal number of litters from each dose group available for testing on a given day.'
Allocation concealment	Q2: was allocation [of animals] to study groups adequately concealed?	PLRoB	Not relevant.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes. Lots of methods descriptions illustrate this.
Blind research personnel	Q4: was the research personnel* blinded to the study group during the study?	NR	Not reported.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. All exceptions were noted and all data were presented in appendices.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Yes. Extensive analytical work was conduction on the diet and brain concentrations were measured in offspring.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	Yes. Not blinded but Well established methods, extensive validation data provided. GLP work.
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Yes. Extensive methods and results sections with appendices that contain all data.
Toxicity	Q9: were there other potential threats to internal validity?	DLRoB	Maternal body weight gain was marginally impacted, but not enough to bias results.

Neuropathology: Brain weight

RefID: 183

Study characteristics and population	Specific endpoint	Brain weight – Absolute; Brain weight – Relative
	Species:	Rat
	Strain:	Sprague-Dawley
	Sex:	Female; Male
	Housing condition:	Standard (GLP)
	Funding source:	Private
	Group size:	6
Exposure	Period of exposure:	From GD 6 to PDN 11
	Route of administration:	Oral (via feed)
	Vehicle:	Corn oil
	Doses:	0 ppm; 20 ppm; 100 ppm; 500 ppm
	Dose frequency:	Day
Measurement	Measurement time:	74 PDN
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Yes. 'Following at least six days of acclimation, P-generation females were weighed and those with a body weight more or less than 20% of the mean weight [18] were rejected. The remaining females were assigned to the control or an exposure group in sequence, as they were determined to be inseminated. This approach was used to ensure unbiased assignment of the animals to dose groups and that there would be approximately an equal number of litters from each dose group available for testing on a given day.'
Allocation concealment	Q2: was allocation [of animals] to study groups adequately concealed?	PLRoB	Not relevant.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes. Lots of methods descriptions illustrate this

Blind research personnel	Q4: was the research personnel* blinded to the study group during the study?	NR	Not reported.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. All exceptions were noted and all data were presented in appendices.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Yes. Extensive analytical work was conduction on the diet and brain concentrations were measured in offspring.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	Yes. Not blinded but automated assessments.
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Yes. Extensive methods and results sections with appendices that contain all data.
Toxicity	Q9: were there other potential threats to internal validity?	DLRoB	Maternal body weight gain was marginally impacted, but not enough to bias results.

Neuropathology: Quantitative morphometric evaluation

RefID: 183

Study characteristics and population	Specific endpoint	Quantitative morphometric evaluation – Cerebrum length (gross measurement); Quantitative morphometric evaluation – Caudate Putamen; Quantitative morphometric evaluation – Caudate Putamen (Diagonal); Quantitative morphometric evaluation – Caudate Putamen (Transverse); Quantitative morphometric evaluation – Cerebellum; Quantitative morphometric evaluation – Cerebellum (gross measurements); Quantitative morphometric evaluation – Corpus Callosum; Quantitative morphometric evaluation – Dentate Gyrus; Quantitative morphometric evaluation – Frontal Cortex; Quantitative morphometric evaluation – Parietal Cortex
	Species:	Rat
	Strain:	Sprague-Dawley
	Sex:	Female; Male
	Housing condition:	Standard (GLP)
	Funding source:	Private
	Group size:	6
Exposure	Period of exposure:	From GD 6 to PDN 11

	Route of administration:	Oral (via feed)
	Vehicle:	Corn oil
	Doses:	0 ppm; 100 ppm; 500 ppm
	Dose frequency:	Day
Measurement	Measurement time:	83 PDN
Risk of bias appraisal		Tier: 3

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Yes. 'Following at least six days of acclimation, P-generation females were weighed and those with a body weight more or less than 20% of the mean weight [18] were rejected. The remaining females were assigned to the control or an exposure group in sequence, as they were determined to be inseminated. This approach was used to ensure unbiased assignment of the animals to dose groups and that there would be approximately an equal number of litters from each dose group available for testing on a given day.'
Allocation concealment	Q2: was allocation [of animals] to study groups adequately concealed?	PLRoB	Not relevant.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes. Lots of methods descriptions illustrate this.
Blind research personnel	Q4: was the research personnel* blinded to the study group during the study?	NR	Not reported.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. All exceptions were noted and all data were presented in appendices.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PH RoB	Yes. Extensive analytical work was conduction on the diet and brain concentrations were measured in offspring.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PHRoB	Only 1 or 2 doses depending on the measurement. This is line with the guidance when no effect was observed at the top dose it is expected that measurement of the endpoint would be only triggered if effect is observed at the top dose.
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Yes. Extensive methods and results sections with appendices that contain all data.
Toxicity	Q9: were there other potential threats to internal validity?	DLRoB	Maternal body weight gain was marginally impacted, but not enough to bias results.

Critical Appraisal table-in vitro

Network formation/function: #ABE. Number of electrodes with burst rates of above 0.5 bursts per minute

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	#ABE. Number of electrodes with burst rates of above 0.5 bursts per minute
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	cell system and method uniform and fully automated, with negative and positive control added to each plate

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in ToxCast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: % Spikes in Bursts

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	% Spikes in Bursts

characteristics and test system	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high-throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: BR-Burst rate

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	BR-Burst rate
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All th condition well described in the paper and characteristic of the high throughput methodology
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: Burst duration

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	Burst duration

characteristics and test system	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: correlation

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Correlation
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the condition are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: IBI-Mean time interval between bursts (sec.)

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	IBI-Mean time interval between bursts (sec.)

characteristics and test system	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high-throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: Interspike Interval

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Interspike Interval
	Test system:	Primary cells
	Species:	rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: Mean Interspike Interval in Burst (MISIB)

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	Mean Interspike Interval in Burst (MISIB)

characteristics and test system	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	number of replicates is described and definition of biological replicate from which the assessment is performed is reported

Network formation/function: MFR

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	MFR
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: Mutual information

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	Mutual information

characteristics and test system	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: network spike duration

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	network spike duration
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: network spike duration standard deviation

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study	Specific endpoint	network spike duration standard deviation

characteristics and test system	Test system:	Primary cells
	Species:	rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: network spike peak

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Network spike peak
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: NS-ISI. Mean interspike interval for spikes in network spikes

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	NS-ISI. Mean interspike interval for spikes in network spikes
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.

Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: number of active electrodes

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Number of active electrodes
	Test system:	Primary cells
	Species:	rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement

Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: Number of spikes in bursts

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Number of spikes in bursts
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day
	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.

Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Network formation/function: number spikes in network spikes

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, Friedman KP, 2019 – RefID: 3202

Study characteristics and test system	Specific endpoint	Number spikes in network spikes
	Test system:	Primary cells
	Species:	Rat
	Origin of the test system:	Cortical
	Stage of development of the primary cells:	Post-natal
	Medium:	Chemically defined medium (CDM)
	Funding source:	Public
	Number of biological replicate:	3
Exposure	Exposure duration:	12 Day

	Treatment frequency:	Chronic
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Cell system and method uniform and fully automated, with negative and positive control added to each plate.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	DLRoB	Random answer
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	All the conditions are well described in the paper and are characteristic of the high-throughput methodology.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	The method is fully automated and, although not declared in the paper, it is very likely that the compound is blinded up to the elaboration of the data as described in toxcast testing.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Criteria are applied to each step of the protocol across all the 17 assays and for all chemicals in the battery. The high throughput approach guarantee that all outcome are treated equally and the same criteria for the assessment are applied.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Commercial source declared and concentrations are prepared and administered automatically.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	The prediction model is well described and in line with experiments having a similar approach.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Raw data are available as supplement and summary table are presented in the paper.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Two cytotoxicity methods are applied, one metabolic and one directly measuring biomarker of cell death.

Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	Number of replicates is described and definition of biological replicate from which the assessment is performed is reported.

Critical Appraisal table-zebrafish

Photomotor Response (EPR – Embryo photomotor response)

Truong L, Reif D, St Mary L, Geier M, Truong H, Tanguay R, 2013 – RefID: 3205

Study characteristics and population	Specific endpoint	Photomotor Response (EPR – Embryo photomotor response); Photomotor Response (LPR – Larval photomotor response)
	Experiment description:	Static condition exposure
	Strain:	Tropical 5 D wild-type zebrafish
	EmbryWell:	1
	Water:	Standard
	Funding source:	Public
	Number of biological replicate:	32
Exposure	Exposure duration:	6 – 120 Hours
	Concentration:	0 μM; 0.0064 μM; 0.064 μM; 0.64 μM; 6.4 μM; 64 μM
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Zebrafish embryos were loaded into a 96 well plate by an automated embryo placement system. Tropical 5 D wild-type adult zebrafish
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Direct evidence in the paper, same plate.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	Automatic method.

Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Automatic analysis. 120 hpf, larval behaviour is the in vitro endpoint.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Purity
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	Automatic
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Data available in the supplementary material.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	No cytotoxic concentrations
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 replicates

Photomotor Response (LPR – Larval photomotor response)

Truong L, Reif D, St Mary L, Geier M, Truong H, Tanguay R, 2013 – RefID: 3205

Study characteristics and population	Specific endpoint	Photomotor Response (EPR – Embryo photomotor response); Photomotor Response (LPR – Larval photomotor response)
	Experiment description:	Static condition exposure
	Strain:	Tropical 5D wild-type zebrafish
	EmbryWell:	1
	Water:	Standard
	Funding source:	Public
	Number of biological replicate:	32
Exposure	Exposure duration:	6–120 hours
	Concentration:	0 μM; 0.0064 μM; 0.064 μM; 0.64 μM; 6.4 μM; 64 μM
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Zebrafish embryos were loaded into a 96 well plate by an automated embryo placement system. Tropical 5D wild-type adult zebrafish.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	To be deleted.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Direct evidence in the paper, same plate.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	DLRoB	Automatic method.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Automatic analysis. 120 hpf, larval behaviour is the in vitro endpoint.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	DLRoB	Purity
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	Automatic
Selective outcome report	Q8: were all measured outcomes reported?	DLRoB	Data available in the supplementary material.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	No cytotoxic concentrations.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 replicates

Appendix B3.3 Outcome of the ROB IVB (In vitro battery)

DNT Appraisal In vitro battery table

Proliferation in primary human NPC

Area/BrdU incorporation–proliferation

Study characteristics and test system	Specific endpoint	Area/BrdU incorporation–proliferation
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	Lonza Verviers SPRL (Verviers, Belgium).
	Stage of development of the primary cells:	Embryonic
	Medium:	Proliferation medium (B27), consisting of DMEM (Life Technologies, Darmstadt, Germany) and Hams F12 (Life Technologies) (3:1) supplemented with 2% B27 (Life Technologies), 1% penicillin and streptomycin (Pan Biotech, Aidenbach, Germany), 20 ng/ml epidermal growth factor (EGF, Life Technologies) and 20 ng/ml recombinant human fibroblast growth factor (FGF, R&D systems, Wiesbaden, Germany).
	Funding source:	Public
	Number of biological replicate:	5 replicate wells per condition
Exposure	Exposure duration:	72 hours
	Concentration:	7 concentrations 1/3 dilution up to 20 μM
	Treatment frequency:	Single
Measurement	Measurement time:	7 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automated.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	Using a luminescence-based BrdU Assay (from Roche) and a luminometer (Baumann et al., 2014, 2015; Nimtz et al., 2019). Data are analysed through a classification model reported in Section 8.3.9 of the IVB report.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity and cell viability are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 biological replicates per condition

Migration assay (NPC2 test)

NPC2a migration distance radial glia (72 h)/NPC2a migration distance radial glia (120 h)/NPC2b migration distance neurons (120 h)/NPC2c migration distance oligodendrocytes (120 h)

Study characteristics and test system	Specific endpoint	NPC2a migration distance radial glia (72 h)/NPC2a migration distance radial glia (120 h)/NPC2b migration distance neurons (120 h)/NPC2c migration distance oligodendrocytes (120 h)
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	Lonza Verviers SPRL (Verviers, Belgium).
	Stage of development of the primary cells:	Embryonic
	Medium:	The differentiation medium consists of DMEM (Life Technologies) and Hams F12 (Life Technologies) at a ratio of 3 to 1 supplemented with 1% N2 (Life Technologies) and 1% penicillin and streptomycin (Pan Biotech). NPCs migrate for 5 days and so differentiate into radial glia cells, neurons, oligodendrocytes and astrocytes (Baumann et al., 2015; Nimtz et al., 2019; Bal-Price et al., 2018). For letting spheres migrate and differentiate, they are incubated at 37°C and 5% CO2 at a pH between 7.2 and 7.6. Cells are fed on day 3 by replacing half the medium with fresh medium.
	Funding source:	Public
	Number of biological replicate:	5 replicate wells per condition
Exposure	Exposure duration:	72 hours/120 hours
	Concentration:	7 concentrations 1/3 dilution up to 20 μM
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	72 hours/120 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.

Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.

Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	PLRoB	NPC2a manually measuring the radial migration from the sphere core on brightfield images (72 h). NPC2a assessed by automatically identifying the migration area of each sphere of Hoechst stained nuclei on fluorescence images (120 h). Migration distance neurons (120 h) (NPC2b) is the mean distance of all neurons from the edge of the sphere core to the position of each neuron (see 'neuronal differentiation') and is given as ratio of the 'migration distance radial glia (120 h)'. Migration distance oligodendrocytes (120 h) (NPC2c) is the mean distance of all oligodendrocytes from the edge of the sphere core to the position of each oligodendrocyte (see 'oligodendrocyte differentiation') and is given as ratio of the 'migration distance radial glia (120 h)'. Cell number (120 h) is measured as the number of Hoechst-positive objects in the migration area detected on the fluorescence images from each sphere after 120 h. Nuclei detection is carried out using the HCS Studio: Cellomics Scan software Data are analysed through a classification model reported in Section 8.3.9 of the IVB report.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity and cell viability are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 biological replicates per condition

Neuronal differentiation (NPC3 Test)

NPC3 neuronal differentiation

Study characteristics and test system	Specific endpoint	NPC3 neuronal differentiation
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	Lonza Verviers SPRL (Verviers, Belgium).
	Stage of development of the primary cells:	Embryonic
	Medium:	The differentiation medium consists of DMEM (Life Technologies) and Hams F12 (Life Technologies) at a ratio of 3 to 1 supplemented with 1% N2 (Life Technologies) and 1% penicillin and streptomycin (Pan Biotech). NPCs migrate for 5 days and so differentiate into radial glia cells, neurons, oligodendrocytes and astrocytes (Baumann et al., 2015; Nimtz et al., 2019; Bal-Price et al., 2018). For letting spheres migrate and differentiate, they are incubated at 37°C and 5% CO2 at a pH between 7.2 and 7.6. Cells are fed on day 3 by replacing half the medium with fresh medium.
	Funding source:	Public
	Number of biological replicate:	5 replicate wells per condition
Exposure	Exposure duration:	120 hours
	Concentration:	7 concentrations 1/3 dilution up to 20 μM
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	120 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.

Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	Number of all TUBB3-positive cells in per cent of the amount of Hoechst-positive nuclei in the migration area after 120 h of differentiation. Automatised. The annotation of neurons is carried out by a CNN running on Keras implemented in Python 3.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity and cell viability are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 biological replicates per condition

Neurite length and neurite area (NPC4 Test)

Neurite length (120 h) (NPC4) and neurite area (120 h) (NPC4)

Study characteristics and test system	Specific endpoint	Neurite length (120 h) (NPC4)/Neurite area (120 h) (NPC4)
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	Lonza Verviers SPRL (Verviers, Belgium).
	Stage of development of the primary cells:	Embryonic
	Medium:	The differentiation medium consists of DMEM (Life Technologies) and Hams F12 (Life Technologies) at a ratio of 3 to 1 supplemented with 1% N2 (Life Technologies) and 1% penicillin and streptomycin (Pan Biotech). NPCs migrate for 5 days and so differentiate into radial glia cells, neurons, oligodendrocytes and astrocytes (Baumann et al., 2015; Nimtz et al., 2019; Bal-Price et al., 2018). For letting spheres migrate and differentiate, they are incubated at 37°C and 5% CO2 at a pH between 7.2 and 7.6. Cells are fed on day 3 by replacing half the medium with fresh medium.
	Funding source:	Public
	Number of biological replicate:	5 replicate wells per condition
Exposure	Exposure duration:	120 hours
	Concentration:	7 concentrations 1/3 dilution up to 20 μM
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	120 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.

Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	N Length, is assessed as neurite length in µm based on the skeletonisation of each annotated neuron (see 'neuronal differentiation') that reaches a predefined intensity threshold. N Area is assessed as the area in amount of pixel for each skeletonised neuron (see neurite length)
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity and cell viability are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 biological replicates per condition

Oligodendrocyte differentiation (120 h) (NPC5)

Study characteristics and test system	Specific endpoint	Oligodendrocyte differentiation (120 h) (NPC5)
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	Lonza Verviers SPRL (Verviers, Belgium).
	Stage of development of the primary cells:	Embryonic
	Medium:	The differentiation medium consists of DMEM (Life Technologies) and Hams F12 (Life Technologies) at a ratio of 3 to 1 supplemented with 1% N2 (Life Technologies) and 1% penicillin and streptomycin (Pan Biotech). NPCs migrate for 5 days and so differentiate into radial glia cells, neurons, oligodendrocytes and astrocytes (Baumann et al., 2015; Nimtz et al., 2019; Bal-Price et al., 2018). For letting spheres migrate and differentiate, they are incubated at 37°C and 5% CO2 at a pH between 7.2 and 7.6. Cells are fed on day 3 by replacing half the medium with fresh medium.
	Funding source:	Public
	Number of biological replicate:	5 replicate wells per condition
Exposure	Exposure duration:	120 hours
	Concentration:	7 concentrations 1/3 dilution up to 20 μM
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	120 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.

Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.
Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	It is determined as the number of all O4-positive cells in per cent of the amount of all Hoechst-positive nuclei in the migration area after 120 h of differentiation. The annotation of oligodendrocytes is carried out by a CNN running on Keras implemented in Python 3.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity and cell viability are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	5 biological replicates per condition

The cMINC neural crest cell migration assay (UKN2 Test)

Study characteristics and test system	Specific endpoint	Human induced pluripotent stem cell (hiPSC)
	Test system:	Pluripotential stem cells
	Species:	Human
	Origin of the test system:	Line IMR90_clone #4 (WiCell, Wisconsin)
	Stage of development of the primary cells:	Stem cells
	Medium:	The medium is fully described in Section 8.26 in the report.
	Funding source:	Public
	Number of biological replicate:	At least 4 technical replicates and 3 biological replicates.
Exposure	Exposure duration:	24 hours
	Concentration:	At least 5 serial dilutions ½ or 1/3
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	24 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.

Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	For image acquisition, an automated microplate reading microscope (Array-Scan VTI, Cellomics, Thermo Fisher Scientific) equipped with a Hamamatsu ORCA-ER camera (resolution 1024 × 1024) is used. Images were recorded in 2 channels at excitation/emission wavelengths of 365 ± 50/535 ± 45 to detect H-33342 in channel 1 and 474 ± 40/535 ± 45 to detect calcein in channel 2. In both channels, a fixed exposure time and an intensity histogram-derived threshold is used for object identification (Stiegler et al., 2011).
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	3 biological replicates per condition

NeuriTox neurite outgrowth of CNS neurons test (UKN4 test)

Study characteristics and test system	Specific endpoint	Immortalised primary (LUHMES) cells derived from an 8-week-old mesencephalon were used.
	Test system:	Primary cells
	Species:	Human
	Origin of the test system:	LUHMES
	Stage of development of the primary cells:	Fetal
	Medium:	The medium is fully described in Section 8.2.7. Lund human mesencephalic (LUHMES) cells are handled in culture as described before (Lotharius et al., 2005; Scholz et al., 2011).
	Funding source:	Public
	Number of biological replicate:	At least 3 technical replicates and 3 biological replicates.
Exposure	Exposure duration:	24 hours
	Concentration:	At least 5 serial dilutions ½ or 1/3
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	24 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.

Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	Cells are stained with H-33342 and calcein-AM and imaged via a high-content imaging microscope (Cellomics ArrayScanVTI; Thermo Fisher Scientific) to assess viability and neurite area (NA) i.e. neurite outgrowth of the cells as described before in detail (Krug et al., 2013; Stiegler et al., 2011). An imaging algorithm is used to distinguish between live and dead cells. All stained nuclei (H-33342 positive) are used to calculate the number of cells in total in channel 1. All H-33342 and calcein double-positive cells are then defined as viable objects in channel 2. Viability in each well was estimated by dividing the amount of viable cells to total cell number. To assess the NA, the algorithm calculates a likely area of the cell soma and subtracts this area from all calcein-positive pixels imaged, resulting in the NA (Figure 17C). As the seeded precursor cells show no such area, this calculated NA represents the neurite outgrowth of the cells within 24 h. In a final step for data display, the single well data were always normalised to the mean of untreated controls.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	3 biological replicates per condition

PeriTox neurite outgrowth of neural crest cell test (UKN5 test)

Study characteristics and test system	Specific endpoint	hiPSC line SBAD2. SBAD2 cells were derived and characterised at the University of Newcastle from Lonza fibroblasts CC-2511,
	Test system:	hiPSC
	Species:	Human
	Origin of the test system:	Lot 293971 with the tissue acquisition number 24245 (Baud et al., 2017).
	Stage of development of the primary cells:	Stem cells
	Medium:	The medium is fully described in Section 8.2.8
	Funding source:	Public
	Number of biological replicate:	At least 3 technical replicates and 3 biological replicates.
Exposure	Exposure duration:	24 hours
	Concentration:	At least 5 serial dilutions ½ or 1/3
	Treatment frequency:	Single (replacement of half of the medium is carried out on day 3)
Measurement	Measurement time:	24 hours
Risk of bias appraisal		Tier: 1

Bias domain	Question	Score	Judgement
Randomisation	Q1: was administered dose or exposure level adequately randomised?	DLRoB	Exposure is conducted automatised.
Allocation concealment	Q2: was allocation [of e.g. cells] to study groups adequately concealed?	PLRoB	Lack of adequate blinding during the study would not appreciably bias results for cell maintenance and cell dosing personnel.
Exposure condition across groups	Q3: were experimental conditions identical across study groups?	DLRoB	Yes, conducted at the same time with the same conditions.
Blind research personnel	Q4: was the research personnel blinded to the study group during the study?	PLRoB	Not reported, but unlikely to impact the automated procedures or the growth of cultures.

Attrition or exclusion	Q5: were outcome data complete without attrition or exclusion from analysis?	DLRoB	Yes. Acceptability parameters were predefined and criteria for exclusion of single wells, concentration, conditions and endpoints were pre-established and reported in the acceptability criteria of the report.
Exposure characterisation	Q6: can we be confident in the EXPOSURE characterisation?	PLRoB	No analytical work was carried out on source, stock or medium. Source was ToxCast analytical grade. Exposure, culture, times etc. are all considered adequate.
Outcome assessment	Q7: can we be confident in the OUTCOME assessment?	DLRoB	Cells are stained with H-33342 and calcein-AM and imaged via a high-content imaging microscope (Cellomics ArrayScanVTI; Thermo Fisher Scientific) to assess viability and neurite area (NA) i.e. neurite outgrowth of the cells as described before in detail (Krug et al., 2013; Stiegler et al., 2011). An imaging algorithm is used to distinguish between live and dead cells. All stained nuclei (H-33342 positive) are used to calculate the number of cells in total in channel 1. All H-33342 and calcein double-positive cells are then defined as viable objects in channel 2. Viability in each well was estimated by dividing the amount of viable cells to total cell number. To assess the NA, the algorithm calculates a likely area of the cell soma and subtracts this area from all calcein-positive pixels imaged, resulting in the NA (Figure 17C). As the seeded precursor cells show no such area, this calculated NA represents the neurite outgrowth of the cells within 24 h. In a final step for data display, the single well data were always normalised to the mean of untreated controls.
Selective outcome report	Q8: were all measured outcomes reported?	PLRoB	Yes.
Cytotoxicity	Q9a: were there other potential threats to internal validity? CYTOTOXICITY	DLRoB	Cytotoxicity are tested at the same time and reported for each condition of the experiment.
Replicates and Repetitions	W9b: were there other potential threats to internal validity? REPETITIONS/REPLICATES	DLRoB	3 biological replicates per condition

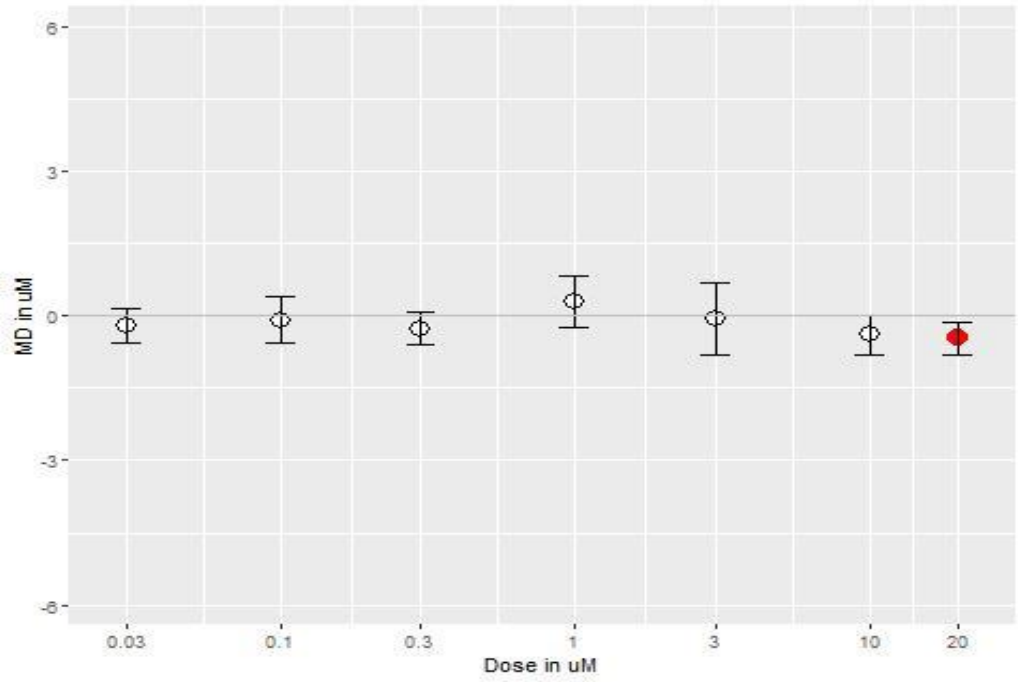
Appendix B4.3 Graph report *In vivo* and *in vitro* flufenacet

In vitro

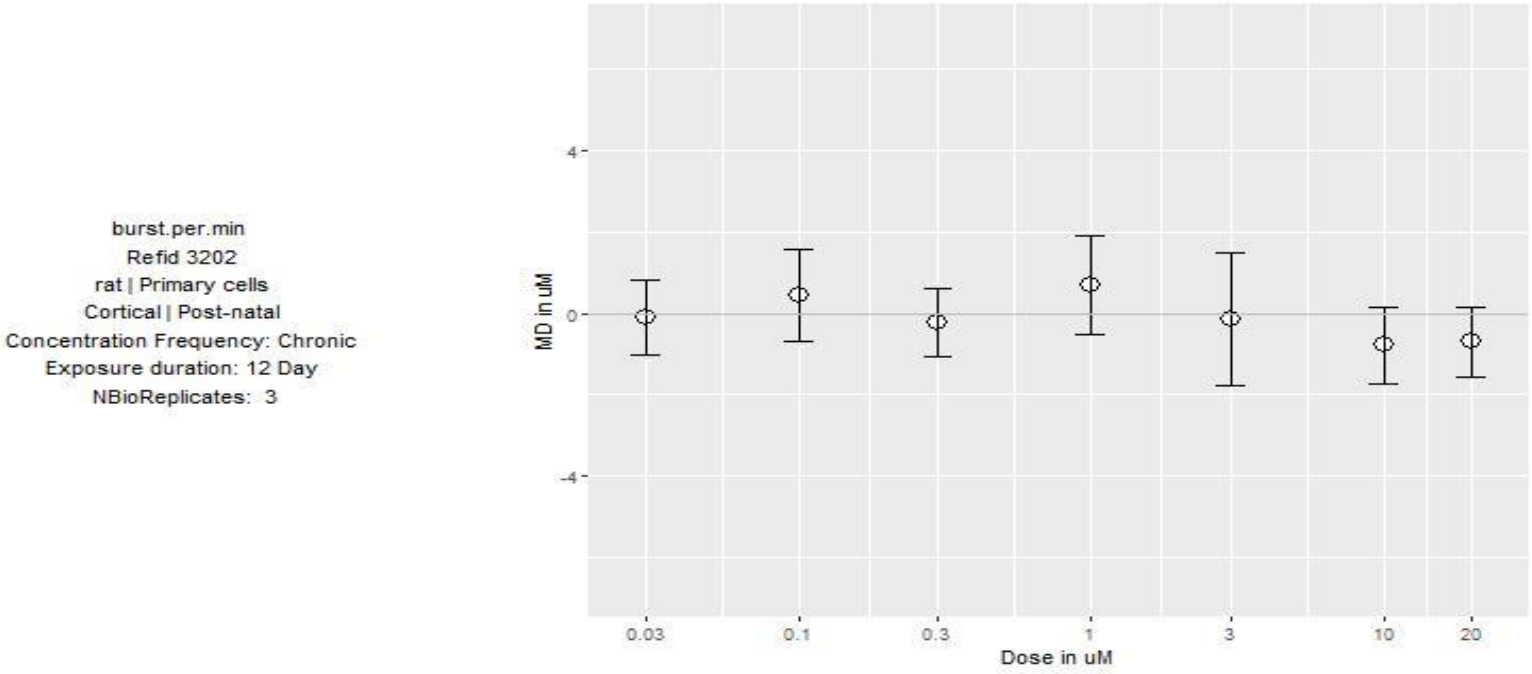
MEA

1. KE *in vitro*. Altered Neuronal Network Function. MEA – General activity. MFR (Mean Firing Rate)

meanfiringrate
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3

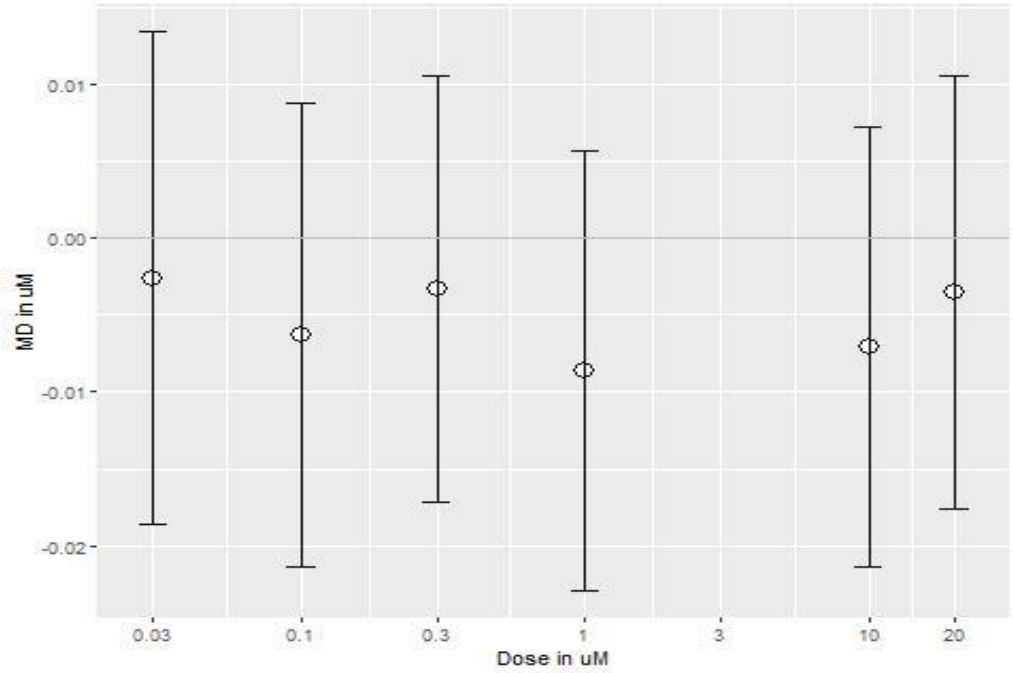


2. KE *in vitro*. Altered Neuronal Network Function. MEA – General activity. MBR (Mean Burst Rate)



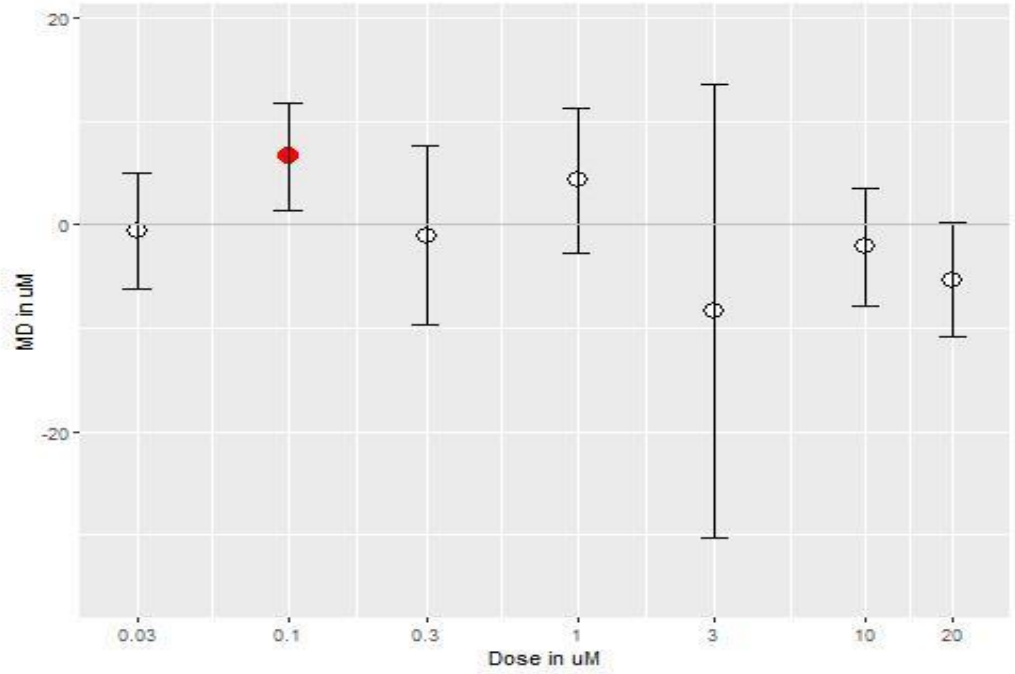
3. KE *in vitro*. Altered Neuronal Network Function. MEA – Bursting activity. Mean Interspike Interval (MISB)

mean.isis
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



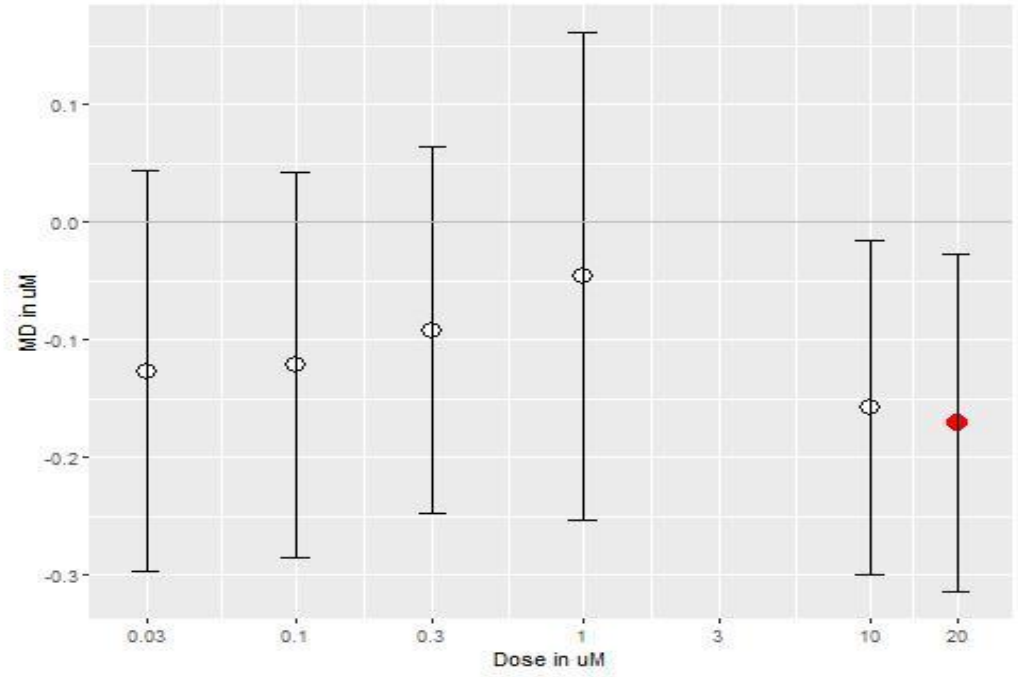
4. KE *in vitro*. Altered Neuronal Network Function. MEA – Bursting activity. % spikes in burst

per.spikes.in.burst
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



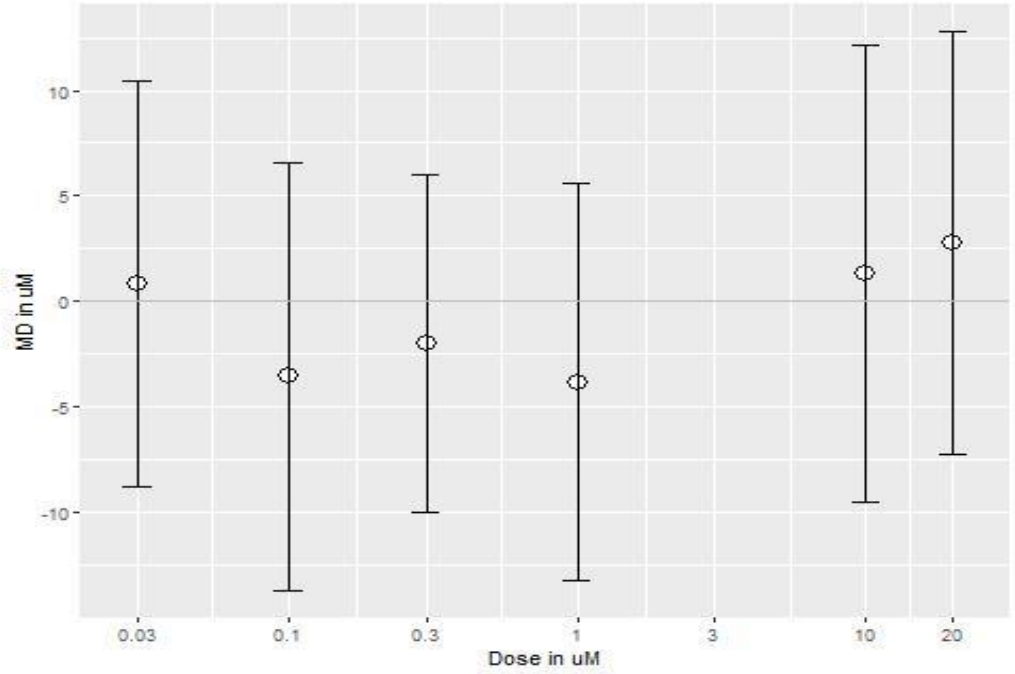
5. KE *in vitro*. Altered Neuronal Network Function. MEA – Bursting activity. Burst duration BD

mean.dur
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



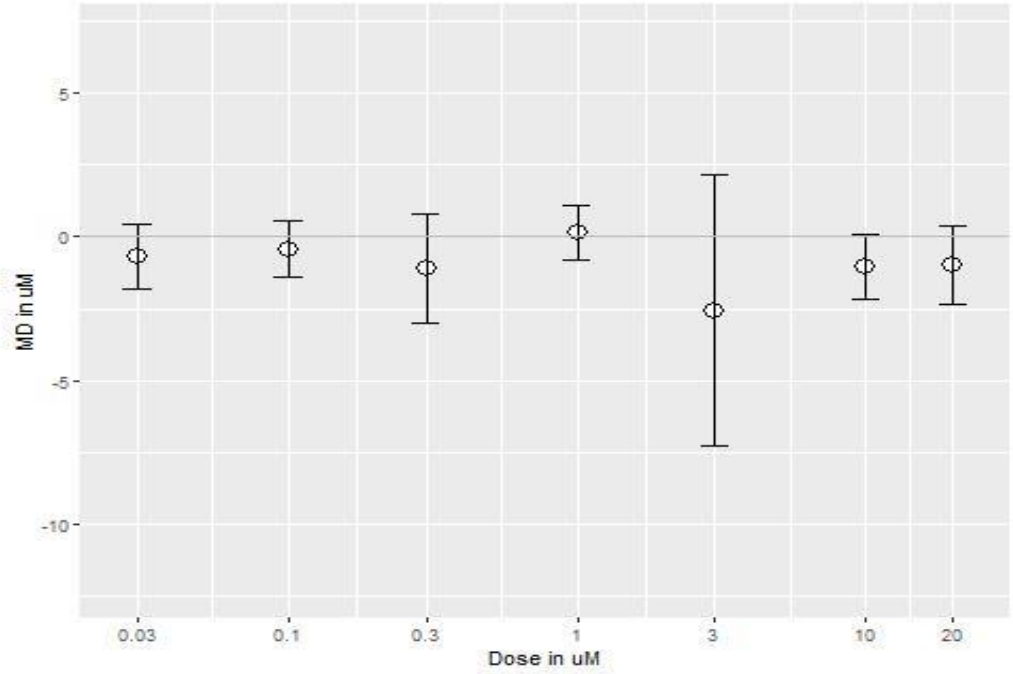
6. KE *in vitro*. Altered Neuronal Network Function. MEA – Bursting activity. Intervals between Burst

mean.IBIs
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



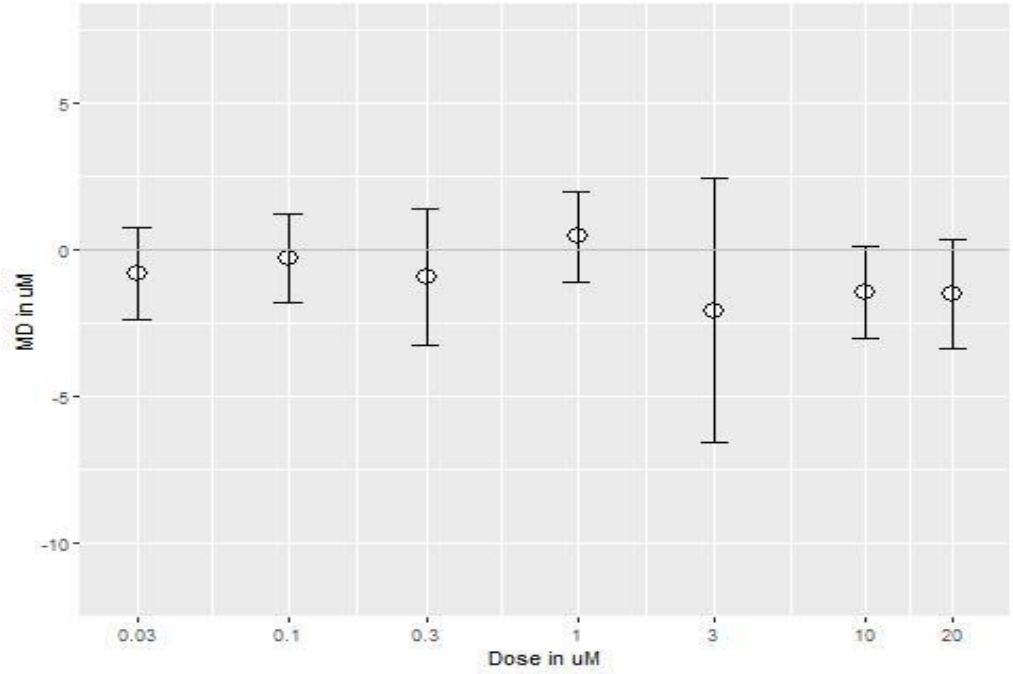
7. KE *in vitro*. Altered Neuronal Network Function. MEA – Bursting activity. Number of active electrodes

nAE
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



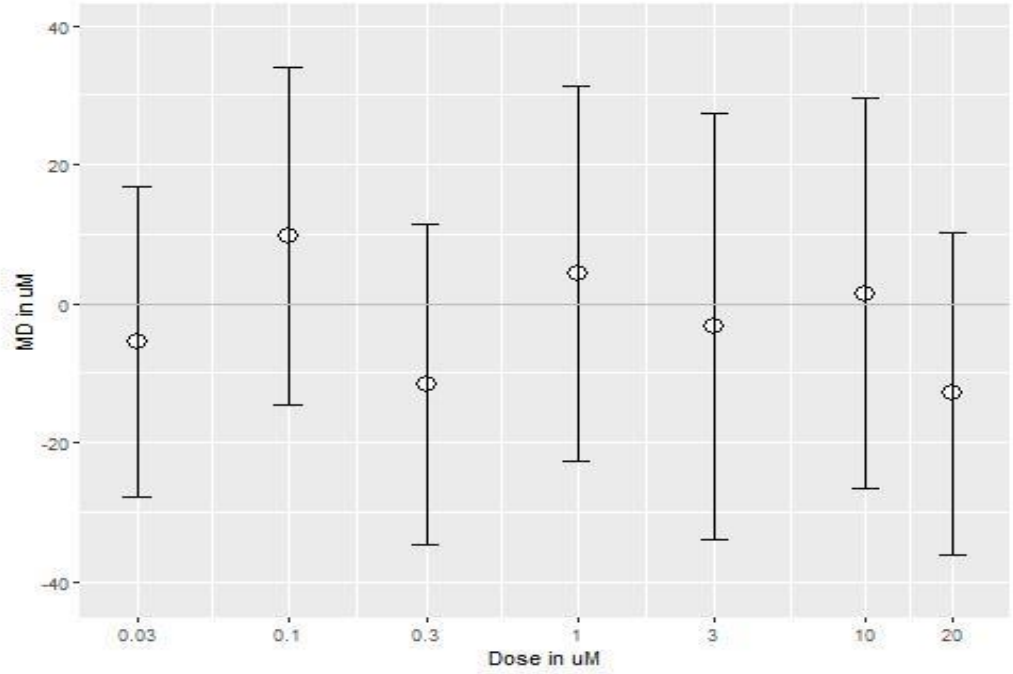
8. KE *in vitro*. Altered Neuronal Network Function. MEA – General activity. Number of actively Bursting electrodes (#ABE)

nABE
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



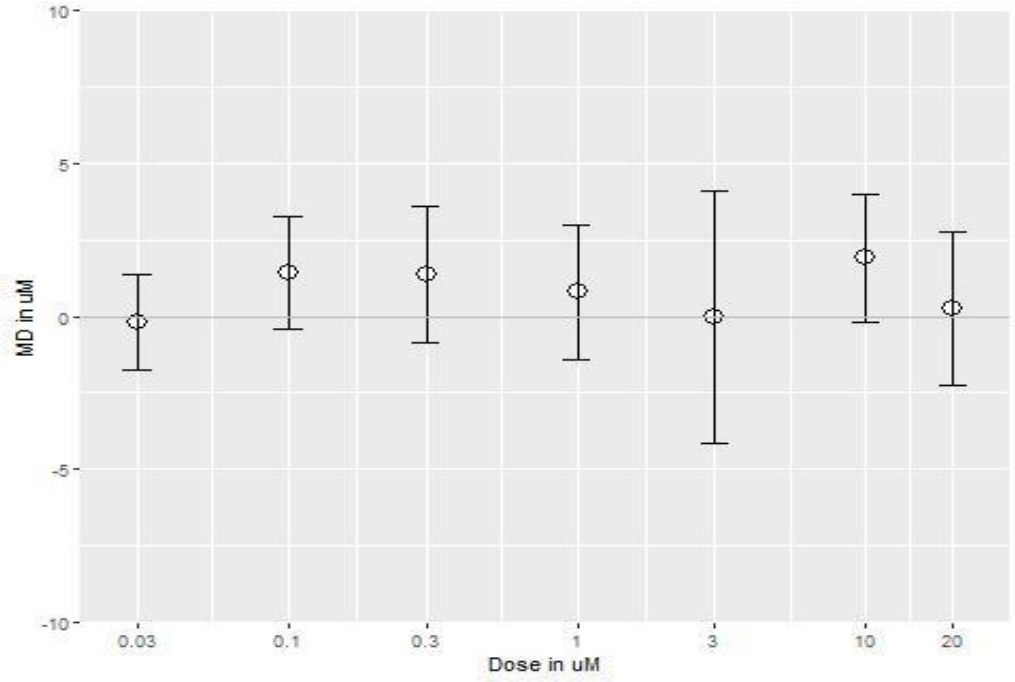
9. KE *in vitro*. Altered Neuronal Network Function. MEA – Network Connectivity. Number of Network Spikes

ns.n
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



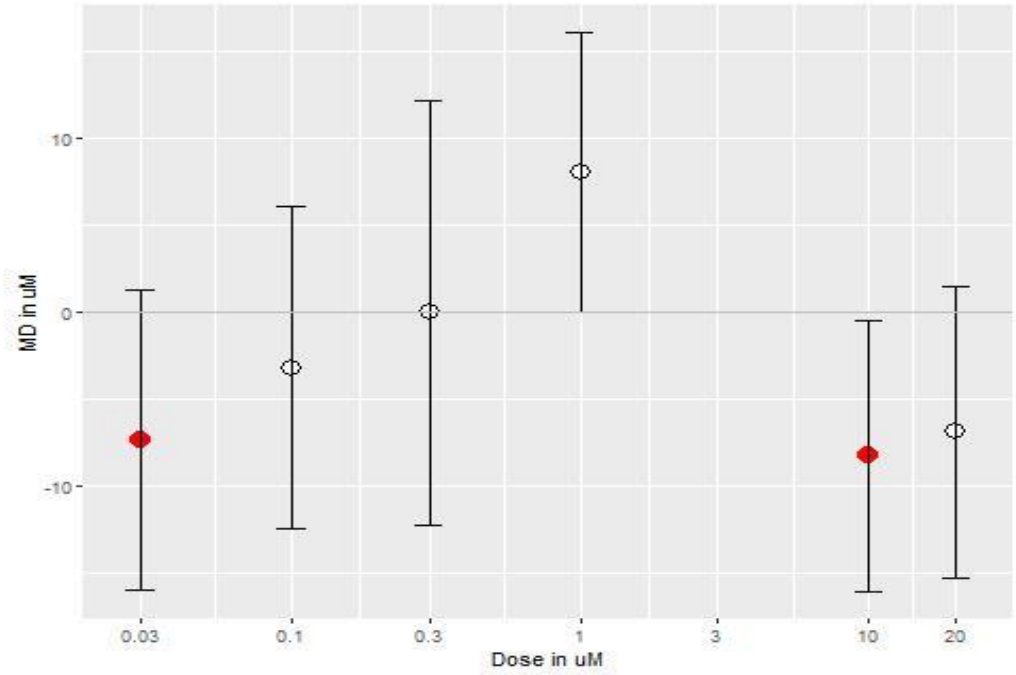
11. KE *in vitro*. Altered Neuronal Network Function. MEA – Network Connectivity. Number of Network Spikes

ns.percent.of.spikes.in.ns
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



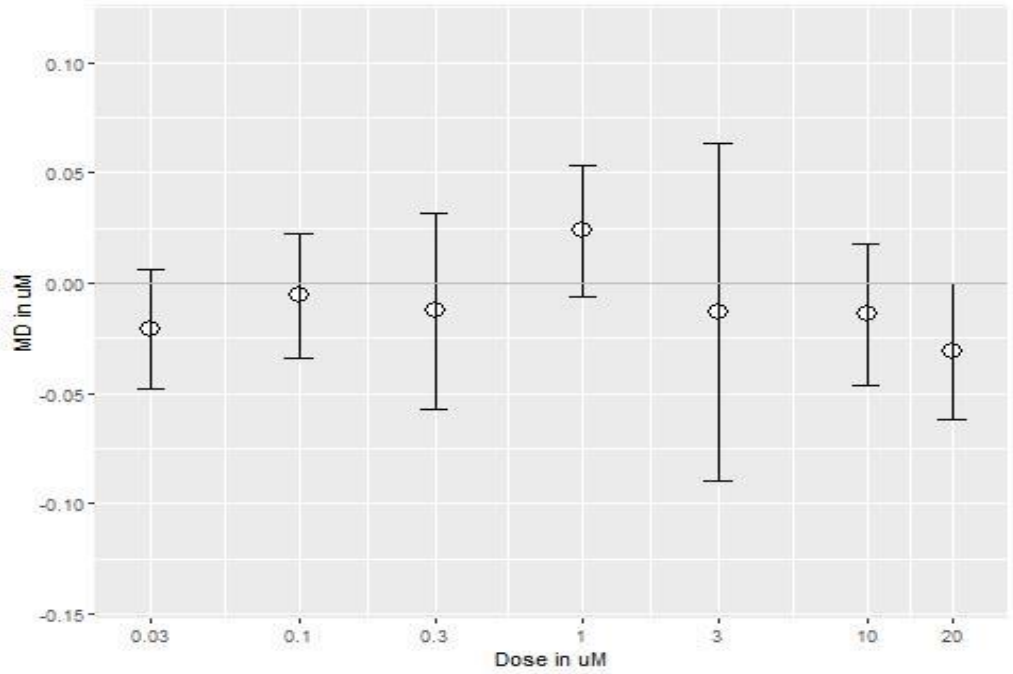
13. KE *in vitro*. Altered Neuronal Network Function. MEA – Network Connectivity. Mean number of spikes per network spikes

ns.mean.spikes.in.ns
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



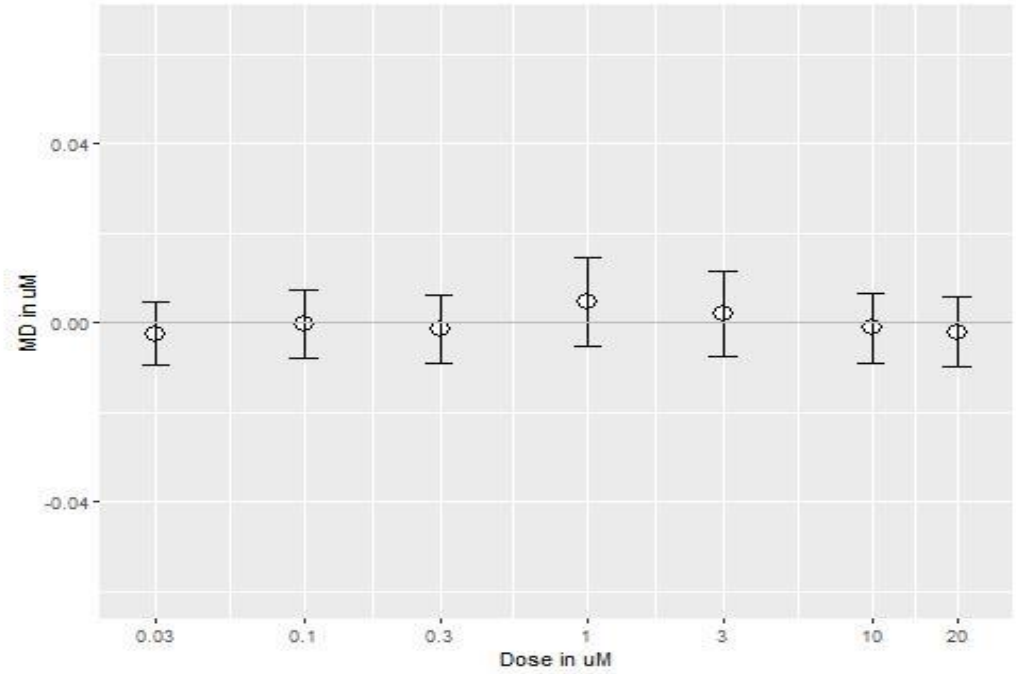
15. KE *in vitro*. Altered Neuronal Network Function. MEA – Network Connectivity. Correlation Coefficient

r
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3



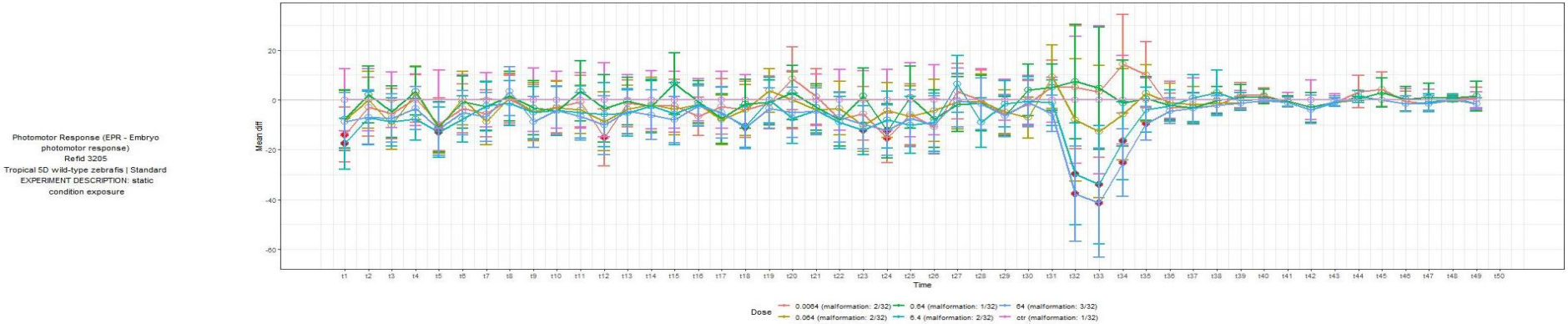
17. KE *in vitro*. Altered Neuronal Network Function. MEA – Network Connectivity. Normalised Mutual Information

Mutual.Information
Refid 3202
rat | Primary cells
Cortical | Post-natal
Concentration Frequency: Chronic
Exposure duration: 12 Day
NBioReplicates: 3

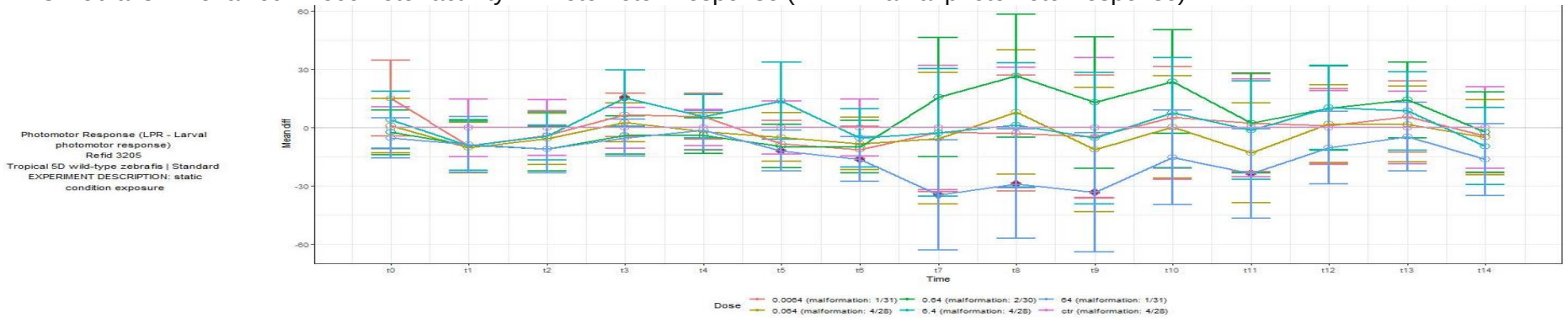


AO Zebrafish

1. AO Zebrafish. Behaviour. Locomotor activity – Photomotor Response (EPR – Embryo photomotor response)



2. AO Zebrafish. Behaviour. Locomotor activity – Photomotor Response (LPR – Larval photomotor response)

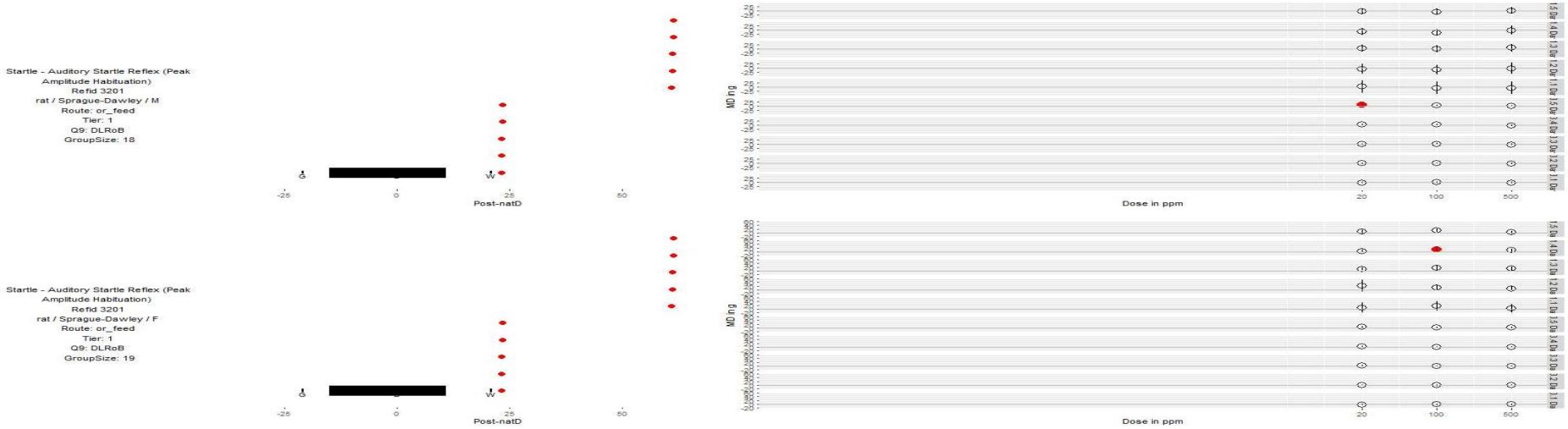


In vivo

In vivo – AO Behaviour

1. AO *In vivo*. Impairment of behavioural function. Startle acoustic – Auditory Startle Reflex (peak amplitude)

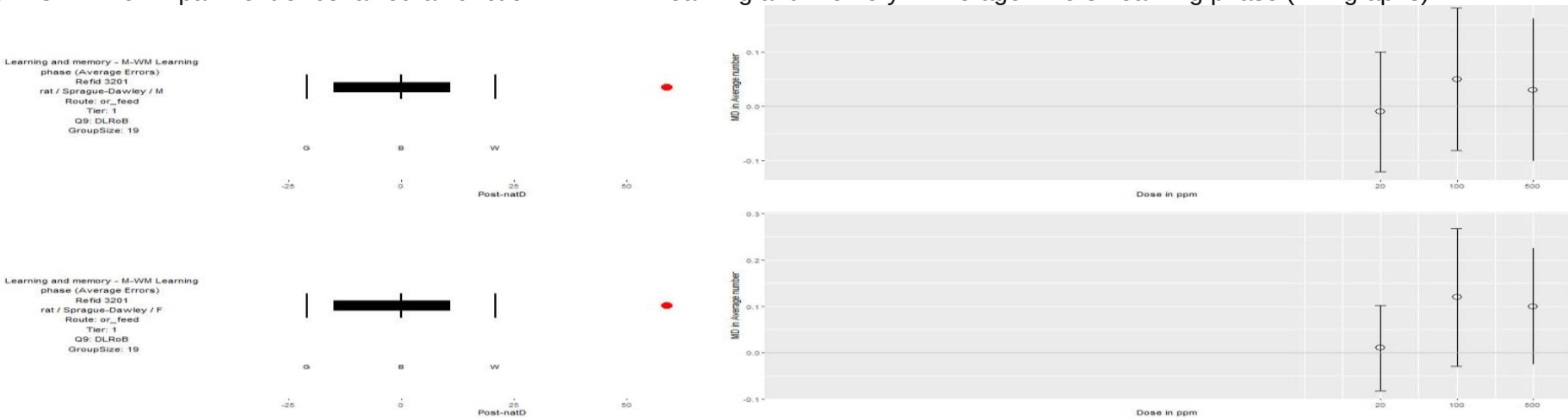
2. AO *In vivo*. Impairment of behavioural function. Startle acoustic – Auditory Startle Reflex (peak amplitude habituation)



3. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Trial to criterion , learning phase (1 – 2 graphs)

4. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Trial to criterion , retention phase (1–2 graphs)

5. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Average Errors Learning phase (1–2 graphs)



6. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Average Errors Retention phase (1–2 graphs)

7. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Average Duration learning phase (1–2 graphs)

8. AO *In vivo*. Impairment of behavioural function. M-WM – Learning and memory – Average Duration retention phase (1–2 graphs)

9. AO *In vivo*. Impairment of behavioural function. Learning and memory – Passive avoidance performance (Trial to criterion) learning phase

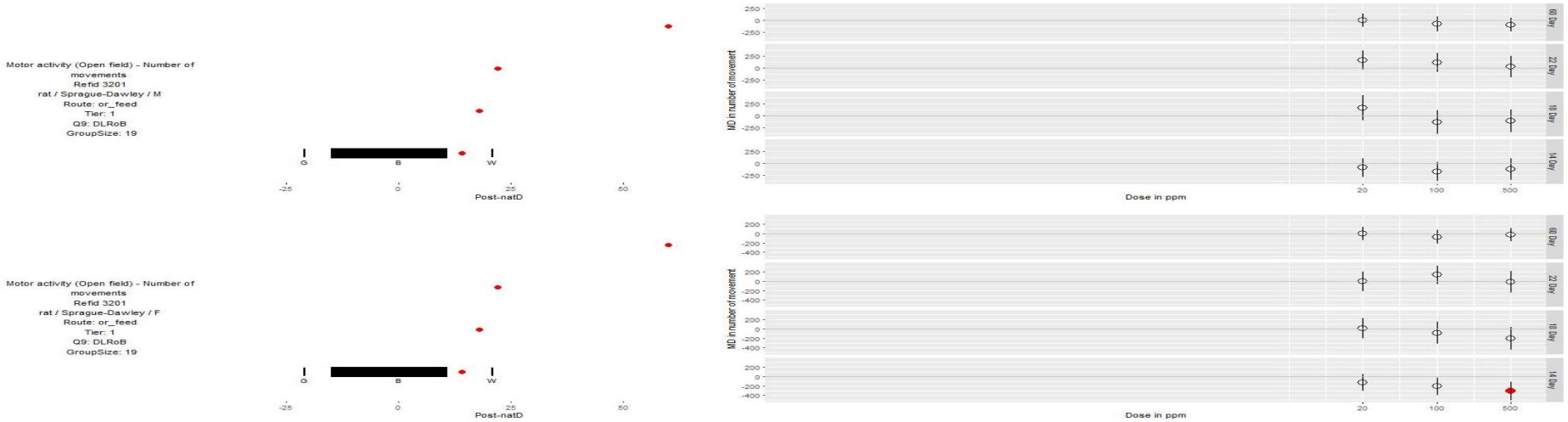
10. AO *In vivo*. Impairment of behavioural function. Learning and memory – Passive avoidance performance (Trial to criterion) retention phase

11. AO *In vivo*. Impairment of behavioural function. Learning and memory – Passive avoidance (Latency) learning phase

12. AO *In vivo*. Impairment of behavioural function. Learning and memory – Passive avoidance (Latency) short-term retention

13. AO *In vivo*. Impairment of behavioural function. Learning and memory – Passive avoidance (Latency) long-term retention

14. AO *In vivo*. Impairment of behavioural function. Motor Activity – (Open Field) – number of movements

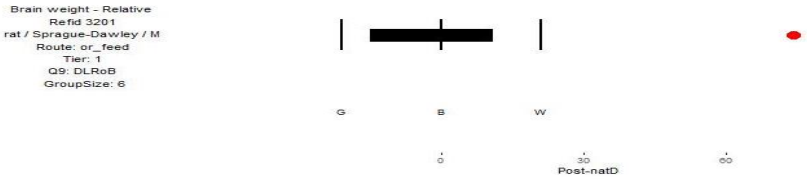


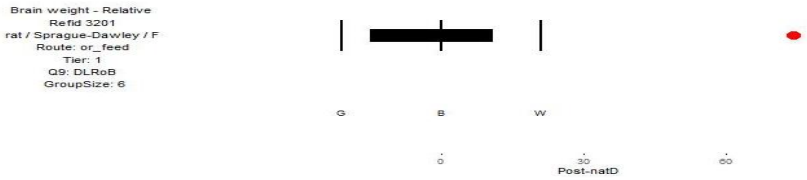
15. AO *In vivo*. Impairment of behavioural function. Motor Activity – (Open Field) – time spent in movement

In vivo – AO Neuropathology

1. AO *In vivo*. Neuropathology. Brain weight absolute

2. AO *In vivo*. Neuropathology. Brain weight relative



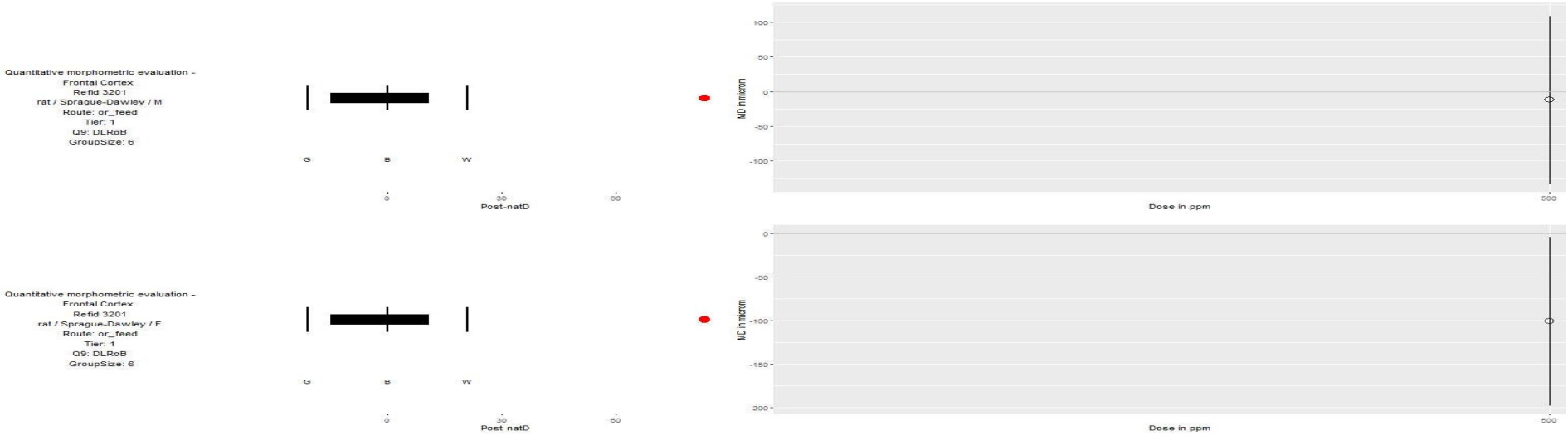


3. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – cerebellum (gross measurement)

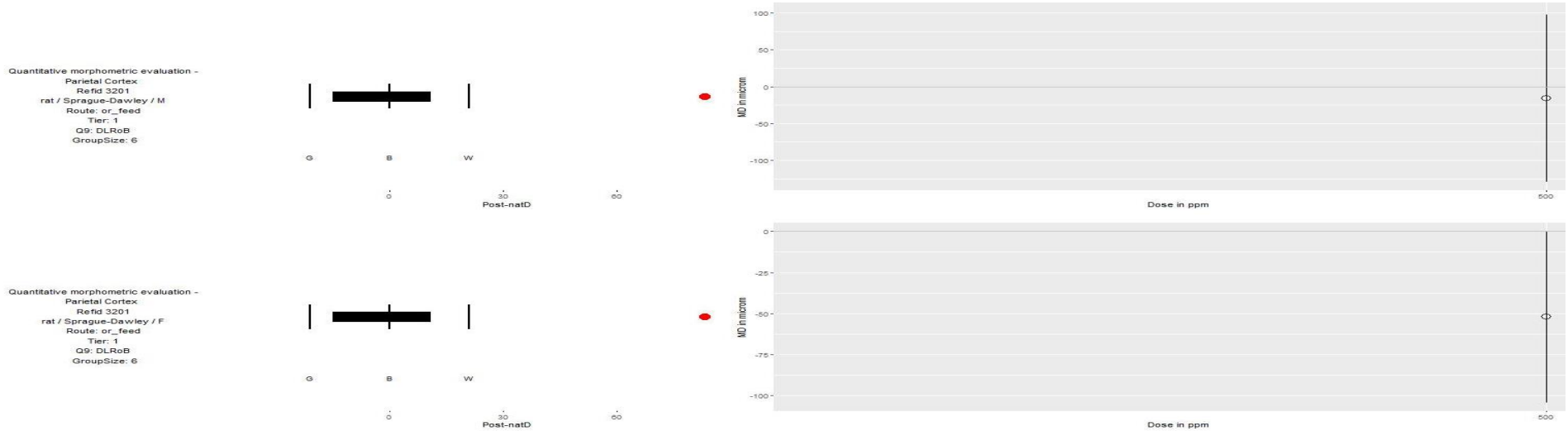
4. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – cerebellum

5. AO *In vivo*. Neuropathology. Quantitative morphometric cerebrum length (gross measurement)

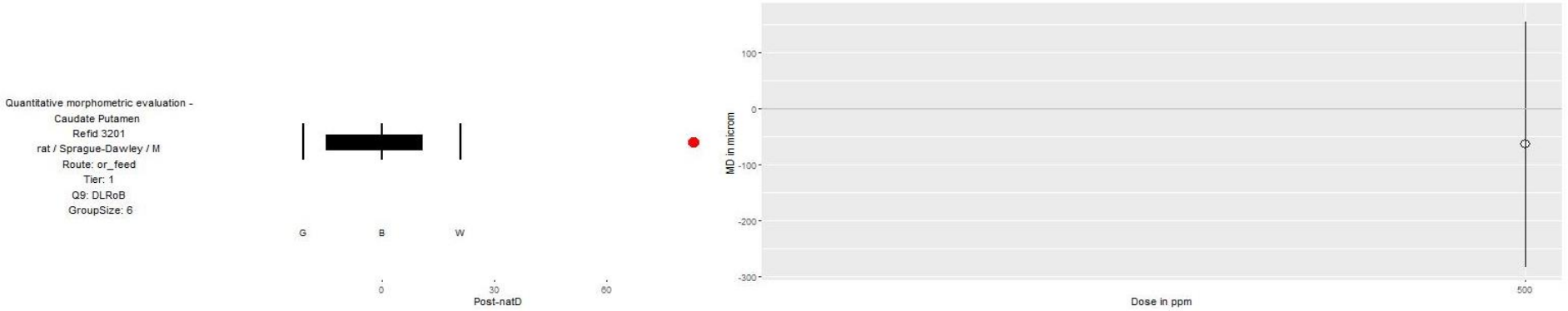
6. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Frontal cortex



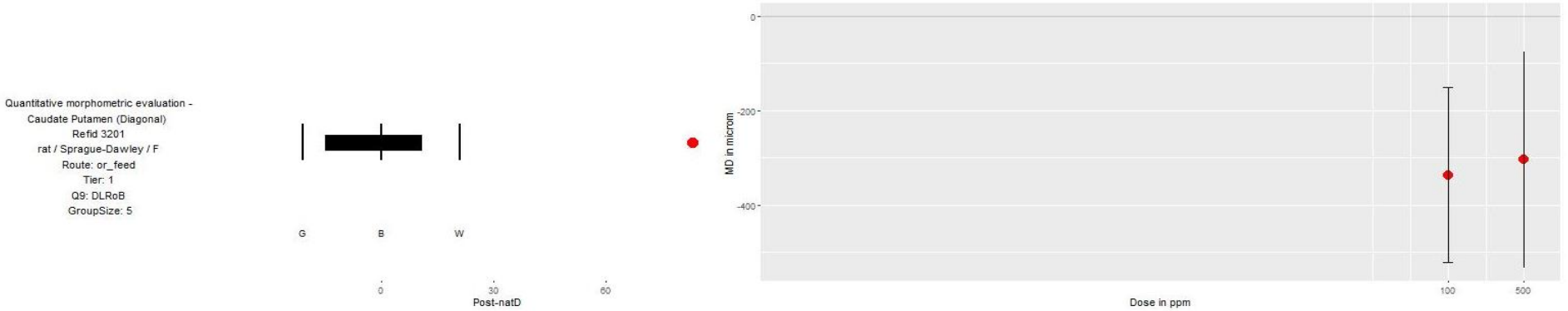
7. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Parietal Cortex



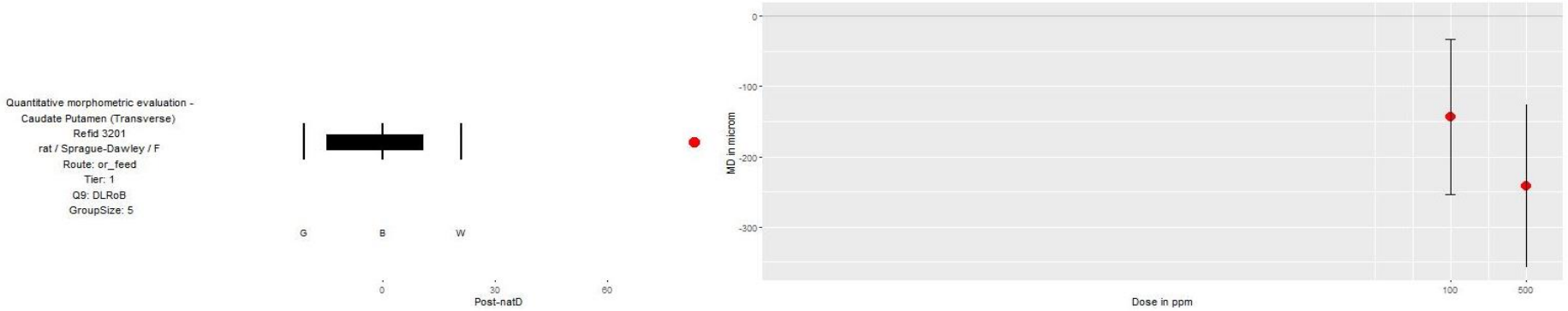
8. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Caudate Putamen



9. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Caudate Putamen (diagonal)



10. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Caudate Putamen (transverse)



11. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Dentate Gyrus

12. AO *In vivo*. Neuropathology. Quantitative morphometric evaluation – Corpus Callosum

Appendix B5.2. Uncertainty analysis tables for flufenacet

Please refer to the separate publication for full Appendix B5.2.

ENV/CBC/MONO(2022)25/ANN3