# Modelling of quantitative Adverse Outcome Pathways

Nicoleta Spînu

A thesis submitted in partial fulfilment of the requirements of
Liverpool John Moores University for the degree of
Doctor of Philosophy.

March 2021

*To my family*

## Acknowledgements

First and foremost, I would like to thank my director of studies, Prof Mark TD Cronin, for seeing in me a potential young scientist and trusting my candidature for the position. I entered both worlds of predictive toxicology and computational modelling almost from scratch, and Mark provided valuable guidance and mentorship, and environment that enabled me to grow in these fields from the first day. Thank you in particular for allowing me to explore and overstep my boundaries, for encouraging me when I was mostly in doubt, for allowing me to go to conferences, training, summer schools, and workshops that I was drawn to, for the freedom needed to complete the research described in this thesis, and for the constant discussions and feedback received throughout my PhD.

Further thanks go to my co-supervisors, Dr Judith C Madden and Dr Steven J Enoch, for sharing their experience and expertise that helped to advance the research project as much. Thank you very much for your support, Judith and Steven.

At the Chemoinformatics Research Group, my past and current lab peers: Dr Julia Pletz, Dr Maria Sapounidou, Dr Gopal Pawar, Dr David J Ebbrell, Dr James Firman, William Masinja, Samuel J Belfield, Michelle Assante, Zuzana Hasarova, Courtney V Thompson could not have been more welcoming and friendly. I cherish the moments we shared, be it scientific and/or social events.

Dr Andrew P Worth has played an important role in my career. His support from the day we met helped to advance my PhD as well as to find my place as a pharmacist in the field of the 3Rs. Thank you very much, Andrew, for all the invaluable suggestions and advice, and for making the PhD period enjoyable.

The secondment spent at the European Union Joint Research Centre was a memorable period of my PhD. I would like specially to thank Dr Maurice Whelan for making this possible, Dr Anna Bal-Price, Dr Brigitte Landesmann, Dr Clemens Wittwehr, Dr Sandra Coecke, Dr Andrea Richarz, Dr Adelaide Dura, Dr Federica Madia and so many others for making my stay pleasant, productive and meaningful towards advancing the qAOP concept and broaden my knowledge about policy and the European Union regulatory framework of alternative methods to animal testing.

I am greatly thankful for the invitation to participate at the Lorentz workshop. The steering committee, cochairs, participants, partners, the Lorentz@Snellius Center, as well as the funding to attend the workshop, are gratefully acknowledged. The discussions with my teammates including Dr Ivana Campia, Dr David Pamies, Dr Stefan Scholz, Dr Dan Villeneuve led to the conceptual framework for quantification of an AOP network. I'm very thankful for the brainstorming sessions we had.

Dr Junpeng Lao made one of my ambitions becoming true. I have learned coding in PyMC3 while showing its relevance in toxicology, all of this being online, and it wouldn't have been possible without his technical support and patience. Thank you very much for your time, Junpeng.

My PhD was funded by the European Union Marie Skłodowska-Curie Action Innovative Training Network in3 Project under the Grant Agreement no. 721975. While I do appreciate the funding, I would like to thank the consortium and its partner network. I value each virtual and in-person shared events. Especial thanks go to Prof Paul Jennings, the Lead PI, Prof Marjana Novič that agreed to be the in3 advisor for my research, to everyone at the Instituto Mario Negri in Milan for hosting me during the secondment, and to my dearest Early Stage Researchers for the valuable friendships we established.

Lastly, and most importantly, my family and friends were my pillars of strength and motivation. Thank you for cheering me up, listening and trying to understand my research project. To my grandmother Augustina, my parents Luminița and Nicolae, and my brother Licu: Vă iubesc și vă mulțumesc pentru toată grija și nemărginitul sprijin primit de-a lungul anilor.

## Abstract

With the growth of green chemistry initiatives, there is a demand for improved regulatory assessment of human exposure to exogenous factors. Proposed a decade ago, the adverse outcome pathway (AOP) framework serves as a knowledge assembly, evaluation, interpretation, and communication tool, designed to support pathway-oriented chemical risk assessment (CRA). The increasing number of resources and advances in machine learning (ML), artificial intelligence (AI), and the quantification of AOPs (qAOPs) has allowed for the integration of a variety of data streams including new approach methodologies (NAMs). These may predict causally inferred tipping points of the relationships that characterise a disease/adverse effect across multiple levels of biological organisation. This thesis aimed to provide an in-depth analysis of the qAOP concept and reinforces the types of efforts required to achieve validation, harmonisation and regulatory acceptance of qAOP models. The first part of this thesis assesses available qAOP models against a series of predefined common features, which enabled the challenges and opportunities for improving current practices to be identified. The second part of this thesis proposes improved methodologies for qAOPs, including the derivation of a network of linear AOPs that better depicts the complexity of biological effects and quantification of a simplified mechanistic AOP network based on domain knowledge and topology analysis. The thesis ends with a case study focused on the identification of empirical quantitative data associated with a linear AOP for quantification purposes. To apply the methodologies formulated, neurotoxicity, represented by neurodegenerative diseases such as impairment of cognitive function and Parkinsonian motor deficits, was studied. Lastly, the role of causality and reasons of why pattern-recognition is not sufficient to translate qualitative/mechanistic information into predictive models are discussed. Overall, the findings contribute to the advancement of the qAOP framework by expanding the knowledge, proposing recommendations and setting future directions towards the development and regulatory and scientific consensus of causal predictive qAOP models in toxicology. Other benefits to the field of study include how to combine information from linear AOPs into a more realistic representation of biological processes for the development of predictive models and the identification of which information (from alternatives) would be required for toxicological understanding. The work underlines knowledge gaps that need to be addressed, and exemplifies how to make use of, and integrate, the variety of available evidence for more informed predictions and improved decision making.

## List of Abbreviations

| | |
|---|---|
| **3Rs** | Reduce, refine, replace |
| **ADME(T)** | Absorption, distribution, metabolism, excretion (and toxicity) |
| **AEP(s)** | Aggregate exposure pathway(s) |
| **AI** | Artificial intelligence |
| **AIC** | Akaike information criterion |
| **ALP** | Autophagy-lysosome pathway |
| **ANOVA** | One-way analysis of variance |
| **AO(s)** | Adverse outcome(s) |
| **AOP(s)** | Adverse outcome pathway(s) |
| **AOP-Wiki KB** | AOP-Wiki Knowledge Base |
| **BBB** | Blood-brain-barrier |
| **BDNF** | Human brain-derived neurotrophic factor |
| **BGM** | Binary Gibbs Metropolis |
| **BH** | Bradford Hill |
| **BISCT** | Bayesian Inference for Substance and Chemical Toxicity software |
| **BMD** | Benchmark dose |
| **CAS RN** | Chemical Abstracts Service Registry Number |
| **CI** | Credible interval |
| **CKE(s)** | Common key event(s) |
| **CNS** | Central nervous system |
| **CPTs** | Conditional probability tables |
| **CRA** | Chemical risk assessment |
| **Cyt c** | Cytochrome c |
| **DA** | Dopamine / Dopaminergic |
| **DAGs** | Directed acyclic graph(s) |
| **DNT** | Developmental neurotoxicity |
| **DTXSID** | The US EPA Comptox Chemical Dashboard substance identifier |
| **EAGMST** | Extended Advisory Group on Molecular Screening and Toxicogenomics |
| **EC50** | Half maximal effective concentration |
| **EC** | Effective concentration |
| **EC JRC** | European Commission Joint Research Centre |
| **EU** | European Union |
| **EURL ECVAM** | European Union Reference Laboratory for alternatives to animal testing |
| **FAIR** | Findability, Accessibility, Interoperability, and Reusability |
| **GIVIMP** | The OECD Guidance Document on Good *In Vitro* Method Practices |
| **GLP** | Good Laboratory Practice |
| **HDI** | Highest density interval |
| **hiPSCs** | Human induced pluripotent stem cells |
| **IATA** | Integrated Approaches to Testing and Assessment |
| **ICH** | International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use |
| **IC50** | Half maximal inhibitory concentration |
| **IRA** | Integrated risk assessment |
| **KE(s)** | Key event(s) |
| **KER(s)** | Key event relationship(s) |
| **MCMC** | Markov chain Monte Carlo |
| **MeSH** | Medical Subject Headings term |
| **MIE(s)** | Molecular initiating event(s) |
| **ML** | Machine learning |
| **MLE** | Maximum likelihood estimation |
| **MoA** | Mode of action |
| **MoE** | Margin of exposure |

| | |
|---|---|
| **MPTP** | 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine |
| **NAMs** | New approach methodologies |
| **NGRA** | Next generation risk assessment |
| **NIS** | Na+/I- symporter |
| **NMDARs** | N-methyl-D-aspartate receptors |
| **NOAELs** | No observed adverse effect levels |
| **NUTS** | No-U-Turn Sampler |
| **OECD** | Organisation for Economic Co-operation and Development |
| **OHAT** | Office of Health Assessment and Translation |
| **PBTK** | Physiologically-Based Toxicokinetic |
| **PD** | Pharmacodynamic |
| **P-gp** | P-glycoprotein |
| **PK** | Pharmacokinetic |
| **PoD** | Point of departure |
| **PPL(s)** | Probabilistic programming language(s) |
| **qAOP(s)** | quantitative AOP(s) |
| **(Q)IVIVE** | (Quantitative) *In-vitro-to-in vivo* extrapolation |
| **QSAR(s)** | Quantitative Structure-Activity Relationship(s) |
| **RA** | Risk assessment |
| **RCR** | Risk characterisation ratio |
| **REACH** | Registration, Evaluation, Authorisation and restriction of Chemicals regulation |
| **RfD** | Reference dose |
| **RIVM** | Dutch National Institute for Public Health and the Environment |
| **RLRS** | Real-life risk assessment simulation |
| **ROS** | Reactive oxygen species |
| **SciRAP** | Science in Risk Assessment and Policy framework |
| **SDGs** | Sustainable Development Goals |
| **semi-q/qWoE qAOPs** | Semi-quantitative/quantitative weight-of-evidence qAOPs |
| **SLogP** | Calculated logarithm of the octanol-water partition coefficient |
| **SMILES** | Simplified Molecular Input Line Entry System |
| **SNpc** | Substantia nigra pars compacta |
| **SR** | Systematic review |
| **TEF** | Toxicity equivalent factor |
| **TGs** | Test Guidelines |
| **TH(s)** | Thyroid hormone(s) |
| **Tox21** | Toxicology in the 21st Century |
| **UN** | United Nations |
| **UNEP** | United Nations Environment Programme |
| **UPS** | Ubiquitin-proteasome system |
| **UVB** | Ultraviolet B |
| **WoE** | Weight of evidence |

# Table of Contents

# Chapter 1. Introduction

The global production of chemicals is expected to double in the period 2015-2030 (UN 2015). Thus, we are exposed daily to chemical substances for most of which the underlying mechanisms of toxicity remain unknown and/or not sufficiently studied. As a result, a better human and/or environmental risk assessment (RA) for the continuously increasing number of chemical substances is required.

Traditionally, methods of chemical risk assessment (CRA) are time-consuming, costly and animal intensive besides the associated ethical considerations and other drawbacks that the classical toxicology testing entails (Knight et al. 2021). For instance, the relevance to humans (and most environmental species) is questionable. Hence, there is a demand for new methods and ways to implement CRA for the many existing and new chemicals.

Currently, CRA is anchored in legislation, which is clumsy and slow to change. The prescriptive and mandatory standard data requirements, stated in many regulations, are not well suited to adapt to the recent advances and rapidly evolving developments of alternative methods to animal testing such as new approach methodologies (NAMs) (Laroche et al. 2019; Punt et al. 2020). In addition, the data available for CRA are heterogeneous with an appreciation that it is challenging to analyse and interpret the diverse and complex streams of evidence (Krewski et al. 2020). Importantly, the progression of animal-free safety assessment is being hampered by the lack of emphasis on the qualitative and quantitative linkages between cellular chemical exposure and mechanistic toxicology, and a lack of integration of *in silico* and *in vitro* tools (Mahony et al. 2020). As such, the adverse effects of chemicals are, as yet, not fully understood, and initiatives aimed at strengthening knowledge of chemical hazards and chemical exposures are still urgently needed (EC 2019).

The paradigm shift *Toxicology in the 21st Century* (Tox21) and efforts in what is known as the 3Rs (reduce, refine, and replace) in research and regulation sets out a framework towards pathway-oriented mechanistic research in predictive toxicology (National Research Council 2007; Roper and Tanguay 2020). This has stimulated and advocated for the development and implementation of the Adverse Outcome Pathway (AOP) concept along with the recent advances such as AOPs networks and quantitative AOPs (qAOPs) for hypothesis-driven toxicity assessment. The value of AOPs in identifying testing and data gaps and knowledge discovery, establishing consensus on the weight of evidence (WoE) to support mechanistic pathways, and defining the quantitative relationship between key events (KEs) for the prioritisation and refinement of test methods, to name a few, are gaining momentum worldwide (Garcia-Reyero and Murphy 2018). Thus, for many reasons, there is an overwhelming demand for improved methods to exploit existing data for a better understanding by utilising AOPs, to identify and prevent adverse effects, to optimise the current process of CRA, and to inform and support the CRA and decision-making processes of exposure to environmental and industrial compounds.

## 1.1. Chemical risk assessment

CRA represents a discipline at the science-policy interface (Wittwehr et al. 2020). It is a scientific evaluation process that leads to an outcome, ideally quantitative on a risk scale, such that the risk from exposure to a chemical in a particular set of circumstances can be established. The potential risk can be communicated to policy- and safety decision-makers allowing for the formulation and implementation of an appropriate risk management strategy. The aim of the strategy is to ensure a very low probability of adverse effects following exposure to the chemical substances (Greim and Snyder 2018; van Leeuwen and Vermeire 2007). Risk is defined as a function of chemical hazard potency, which can be identified and described in toxicology studies, and of the degree of the magnitude, duration and frequency of chemical exposure. CRA involves four main steps: (1) problem formulation, (2) hazard identification and characterisation, (3) exposure assessment and (4) risk characterisation (Figure 1.1).

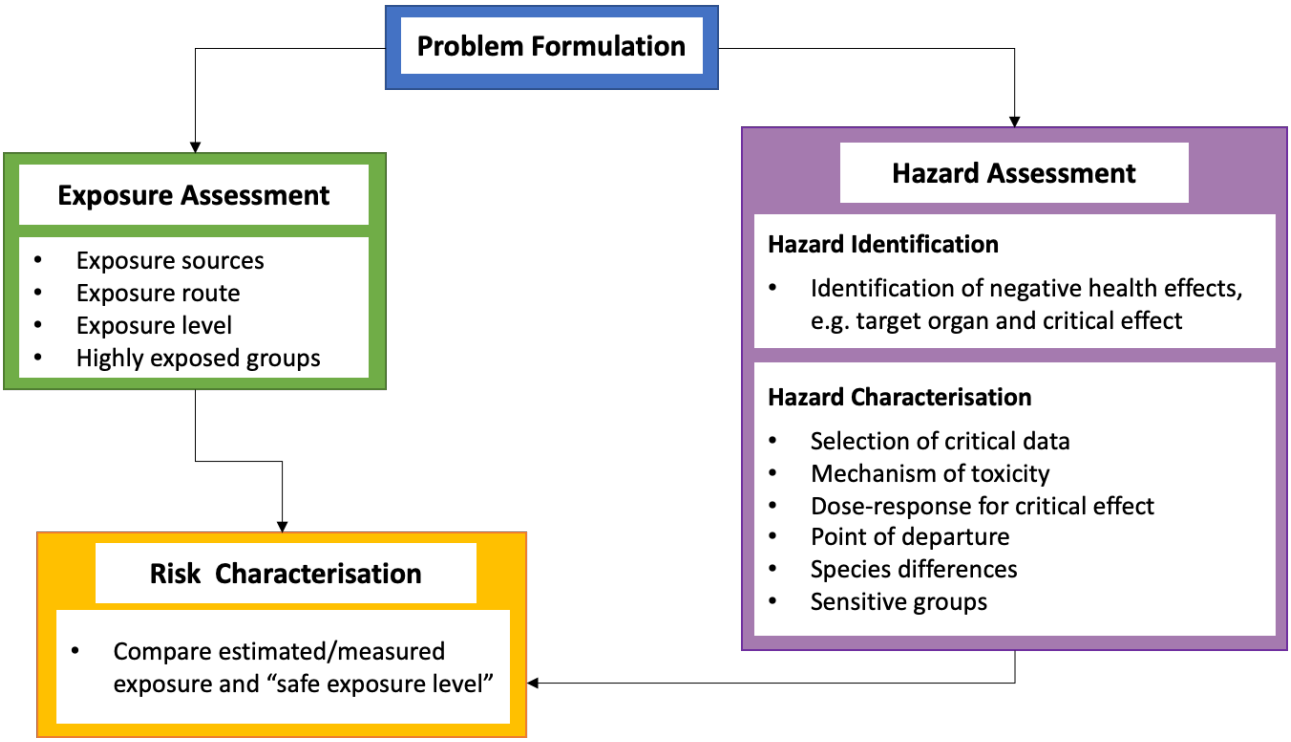


Figure 1.1. An overview of the main steps currently followed in chemical risk assessment.

Hazard assessment examines a chemical's inherent potential to cause adverse effects on human health and the environment. It uses various tools and frameworks, combining information from *in vivo*, *in vitro* and *in silico* methods and epidemiological studies. This information includes non-standard (non-guideline) sources or non-target species, to aid with screening and ranking chemicals. Exposure assessment helps to facilitate the understanding of the potential for exposure in a human population of concern. It establishes the link from (intentional or unintentional) emissions of chemicals into the environment to exposure of biological targets: ecosystems, communities, populations, whole organisms, organs, tissues, and cells. Risk characterisation encompasses both hazard and exposure assessment and serves as the intermediary between RA and risk management. It is conducted to determine the difference between anticipated human exposure and a dose/concentration that is known not to cause adverse effects. It is commonly expressed as

the risk characterisation ratio (RCR), or margin of exposure (MoE), depending on whether assessment (uncertainty) factors are included or not. As such, risk should be identified, well characterised, assessed and estimated/quantified for further risk management. Risk management should also provide an answer as to how much risk is likely and which measures are required, if any, to reduce this risk so not to represent a threat to human health or the environment (Greim and Snyder 2018; van Leeuwen and Vermeire 2007).

Today, CRA tends to be pragmatic and aims to be protective rather than objective. This is because CRA: (1) relies heavily on non-human data, the mechanistic relevance of which may not be directly attributable to humans; (2) is based on apical effects and not multifactorial disease aetiologies; (3) does not consider co-exposures (mixtures and networks of pathways); (4) typically relies on the reference dose (RfD)/point of departure (PoD) derived from no observed adverse effect levels (NOAELs) that do not enable a quantification of the associated uncertainty (more recently the benchmark dose (BMD) has been applied, but its use in CRA is still in its infancy); (5) there is a lack of data accessibility and interoperability, which makes data integration difficult; (6) there is a lack of flexibility in data requirements by regulators, which limits the uptake of new scientific developments in a timely manner (Beronius et al. 2020; Hoffmann et al. 2017; Stephens et al. 2016).

Despite, and possibly due to, all the limitations, there is an opportunity for the development of a quantitative framework appropriate to integrate the variety of streams of evidence derived from *in vitro* and *in silico* models. The aim is to go beyond the existing *in vivo* models towards a better-informed and improved assessment of human health and environment risks to chemical exposures. Thus, there is the momentum to bring CRA up to date with modern technologies and (computational) tools such as qAOPs.

## 1.2. The concept of adverse outcome pathway

The adverse outcome pathway (AOP) framework, originally described with regard to ecotoxicology (Ankley et al. 2010), has been proposed as one method to aggregate and organise relevant information. It can assist in the definition of how measurable perturbations in response to stressors, i.e., chemicals, genetic or environmental factors, lead to an adverse outcome (AO), or an event of regulatory concern. AOPs have evolved rapidly from a conceptual paradigm into qualitative and quantitative models. They describe an initial exposure, resulting in a molecular initiating event (MIE), through a series of key events (KEs) and their relationships, i.e., key event relationships (KERs) to inducing the AO observed at different levels of biological organisations, e.g., cellular, tissue, organ, organism, population levels, as outlined in Figure 1.2 (OECD 2018b).

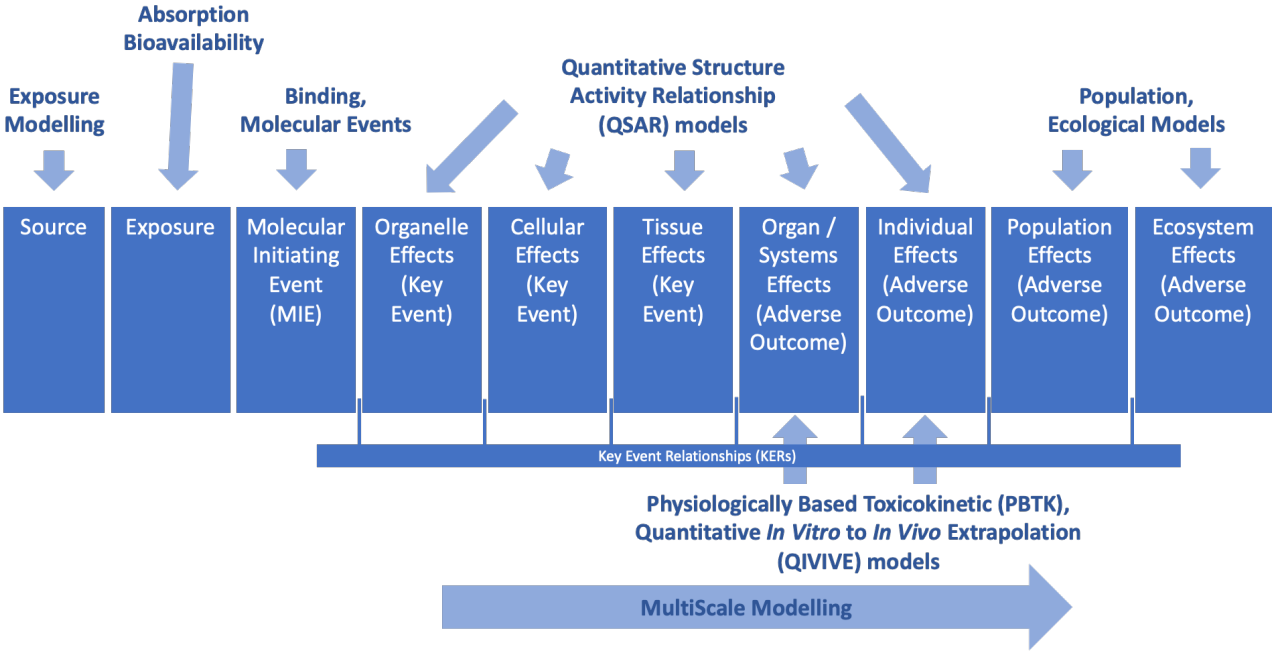


Figure 1.2. An overview of an AOP construction and the types of *in silico* models that may be associated with a generic AOP adapted from Cronin and Richarz (2017).

Since 2012, the Organisation for Economic Co-operation and Development (OECD) has been running the AOP Programme to promote the concept, to support the development of AOPs of regulatory relevance and to exploit AOP knowledge in, e.g., Integrated Approaches to Testing and Assessment (IATA) (OECD 2016). An AOP is structured and described according to a key set of principles and guidelines accepted by scientific and regulatory communities with the final goal of the endorsement of the AOP by the OECD (OECD 2016; OECD 2018b; Villeneuve et al. 2014a). Through the AOP-Wiki Knowledge Base (KB), an open-source platform, the OECD aims to encourage the transparent development and recording of AOPs to allow for global input. At the time of writing, the OECD AOP-Wiki KB comprises a list of 325 linear AOPs[1], of which 15 linear AOPs[2] are endorsed. The AOP-Wiki KB represents the most prominent and well-developed module out of several

[1] https://aopwiki.org/metrics, accessed on March 16, 2021.
[2] https://www.oecd-ilibrary.org/environment/oecd-series-on-adverse-outcome-pathways_2415170x, accessed on March 16, 2021.

ongoing projects known as AOP-Wiki KB, e.g., Effectopedia, AOP Explorer and the Intermediate Effects Data Base, all of which are being co-ordinated under the auspices of the OECD (Carusi et al. 2018).

AOPs may support various regulatory and scientific applications, including: compound prioritisation during testing and development of alternative methods to animal testing; increased collaboration across disciplines and sectors; identification of novel mechanisms and biomarkers; identification of data gaps and; increased ability to perform read-across; ultimately, contributing to the reduction of the reliance on animal models (Carusi et al. 2018; Coady et al. 2019). An example of an AOP with an application in RA is the AOP for skin sensitisation used to develop and refine chemical categories and integrated assessment and testing approaches of cosmetics ingredients (Gautier et al. 2020; Macmillan and Chilton 2019; OECD 2014a).

However, there is a need for an increased international effort for the development and implementation of AOPs as a means to increase their regulatory and other scientific applications. The recent Global Chemicals Outlook II Report of the United Nations Environment Programme (UNEP) recommended that stakeholders "*accelerate the development of the concept of Adverse Outcome Pathways (AOPs) to support hazard assessment*" (UNEP 2019). A notable effort to support this is the initiatives of the European Union's Reference Laboratory for the Validation of Alternative Methods (EURL ECVAM) that focusses attention on the validation and application of knowledge from AOPs to make better regulatory decisions. For instance, identification of common mechanisms and modes of action (MoA) of carcinogenicity, development of a battery of *in vitro* assays for developmental neurotoxicity, based on a network of AOPs, are a few of the ongoing projects (Zuang 2019).

## 1.3. Populating and assessing an AOP

AOPs are very knowledge hungry. At the same time, CRA is built on data collection, hence, there is a need for data. Systematic review (SR) methods, e.g., academic research into standardised toxicity studies or other available knowledge, represent one way of obtaining data. SR is a protocol-driven approach where data are structured, each line of evidence is rigorously documented, and the quality of data is evaluated against a pre-determined set of criteria. In addition, qualitative data analysis and quantitative meta-analysis, statistical evaluation and WoE of datasets are incorporated where possible, to answer a specific pre-defined research question (Beronius et al. 2020; Hoffmann et al. 2017). The steps in conducting SR include (1) problem formulation, (2) literature search, (3) selection of studies, (4) extraction of data, (5) quality assessment of individual studies and (6) integration of data (Hoffmann et al. 2017). Ideally, SR methods should focus on critical and controversial questions in CRA that are key to the safety decision or policy direction that needs to be set. These focused questions should, ideally, be identified as early as possible (Pease and Gentry 2016). Importantly, SR methods are acceptable for use in the European Union (EU) within the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) regulation. This is because SR methods have been leveraged successfully to increase transparency and rigour by introducing strategies that limit bias and random error while identifying the best available evidence relevant to a literature-based chemical assessment (Hoffmann et al. 2017; Whaley et al. 2016). However, the quality of a meta-analysis, i.e., synthesis of data from included studies, depends, to a large extent, on the quality of the individual studies. Even if the separate studies are of high quality, a meta-analysis may not be advisable if there is lack of compatibility among studies, e.g., differences in the study populations, doses, adverse effects, which may lead to the considerable heterogeneity in the results that challenge drawing robust conclusions (Beronius et al. 2020; Hoffmann et al. 2017). At the same time, studies published in peer-reviewed journals do not necessarily follow guidelines to ensure study quality, such as the OECD Test Guidelines (TGs), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) TGs or Good Laboratory Practice (GLP) principles. As a result, these studies need to be evaluated additionally for relevance and reliability. Other studies that follow the national and/or international guidelines can be subject to confidentiality and, therefore, have limited availability. At the same time, SR methods usually rely on *in vivo* and *in vitro* studies, taking less account of *in silico* approaches. A reason might be that there is no distinct protocol for the development and application of *in silico* models even if there have been attempts to define this (Myatt et al. 2018). This is mainly due to the diversity of models in terms of the data used, their structure, statistical analysis including sensitivity analysis, sample size etc. SR has recently evolved into a pathway-oriented framework to generate, or support, the hypothesis that is able to analyse causal relationships and assumptions that link multiple hypothesis and outcomes (Roth et al. 2020). Thus, making use of SR in an organised manner for better evidence-based and scientifically-informed decisions can help to map the available mechanistic knowledge while assessing quality and confidence of the lines of evidence with the

final goal to bring together exposure and multiscale modelling and discovery of potential hazards under a common scientific and regulatory umbrella.

The evidence derived using SR methods and mapped to an AOP, helps ground KEs and KERs. This grounding is, ideally, in a chemical agnostic manner, since the existing information in the literature is captured in an unbiased way. Additionally, the information derived from an SR is evaluated for relevance and other criteria such as using Bradford Hill (BH) considerations (Hill 1965) or by recommendations from the OECD (OECD 2018b) as presented in detail in Chapter 2. Similarly, information extracted during an SR could be used in data-driven approaches to computationally predict KEs and KERs, e.g., model parametrisation, or for designing *in vitro* experiments for quantification of the AOP as discussed in Chapter 5. Thus, the AOP approach is considered a useful knowledge management tool to direct the transition from apical endpoint driven toxicological hazard and RA towards mechanistic toxicology led by discovery and quantification of causal pathways (Bloomingdale et al. 2017; Hartung et al. 2017; Sturla et al. 2014).

## 1.4. Basis for quantitative AOPs

Classic toxicology assumes that the effect of a toxicant is a function of the dose applied – toxicology's most famous concept is the paraphrasing of Paracelsus that the "*dose makes the poison*". In reality, most toxicological testing is performed at a series of doses either to ensure safety or to determine levels at which there are no effects, the lowest dose causing an effect or a pre-defined standard effect, such as 50% of a particular biological activity. These values, as well as the nature of the dose-response correlation itself, are critical factors in guiding the RA. Considered as a novel approach in quantitative RA, qAOPs use *in silico* computational techniques that translate molecular mechanistic understanding of toxicity into safety testing strategies including estimation of potential risks, i.e., the magnitude of exposure needed to elicit an adverse effect (Schultz and Watanabe 2018). Therefore, it gives scientific confidence and real-world applications to a qualitative AOP. In the context of the present thesis, a qAOP was defined as a mathematical description of specific biological processes of an AOP from the MIE to the AO. A quantitative understanding of an AOP requires an understanding of how much change in an upstream biological response, i.e., an early KE in an AOP, is needed to cause a level of downstream biological effect, e.g. eliciting a later KE in an AOP (Schultz and Watanabe 2018). By definition, KEs are measurable experimentally - from these data dose-responses can be derived, and/or estimated. This enables KERs to be described mathematically, to define the tipping points of the transitions between KEs, ideally from the MIE up to the AO (Schultz and Watanabe 2018). This is captured as a response-response relationship, the unique feature that distinguishes a qAOP from other biologically based mathematical models to predict KEs, that are part of the qualitative AOP (Figure 1.2). However, the development of qAOPs is in its infancy and efforts to describe the state-of-the-art and progress made to develop qAOPs including methodologies, applications, and examples to address challenges, such as heterogonous and/or limited quantitative evidence, are needed.

The main advantage of qAOP models is the ability to use the pathway-derived data to extract further information and knowledge, especially when they can be formalised into computational models. When associated with chemical structure, i.e., Quantitative Structure-Activity Relationship (QSAR) models, and other approaches, e.g., Physiologically-Based Toxicokinetic (PBTK) and Quantitative *In-Vitro*-to-*In-Vivo* Extrapolation (QIVIVE) models, can provide a direct linkage between chemistry and adverse effect, leveraging the content of the AOP to support the meaning and interpretation of the model as shown in Figure 1.2. Thus, the potency of a chemical should be evaluated based on both the response-response in the earlier KEs (to anticipate the apical endpoint as informed by the qualitative description of the AOP) as well as by the QSAR and PBTK models, which account for the physico-chemical properties characteristic to the AO (hazard identification and assessment) and details about the exposure and kinetics (exposure assessment). Currently, most of the *in silico* models from AOPs are derived from the MIE (Allen et al. 2016; Cronin and Richarz 2017). Even though a qAOP model aims to be used primarily as a screening and prioritisation tool to allow for the evaluation of new compounds and understanding and/or predicting its toxicity, there is also a need to develop more qAOP models that would assess the impacts of multiple pathways and chemical mixtures.

Therefore, complex quantitative network modelling might be more adequate for RA purposes. Thus, the development of a qAOP model is a multiscale process which requires specification of chemicals while the exchange of information typically needs to occur in both directions across biological scales, e.g., feedback mechanisms and adaptation or acquired tolerance to a toxicant during long-term exposures.

A qAOP model is data-dependent. At the same time, we are gaining new data streams, e.g., NAMs, that can be directed to provide information for qAOPs – preferably on a mechanistic basis. NAMs, such as high-throughput omics technologies, *in vitro* tests and batteries of *in vitro* assays, organs-on-a-chip, deep learning (DL) and machine learning (ML) techniques, as well as protein binding 3D models have been developed, some extensively, and proposed to provide information relating to the potential of a compound to cause toxicity (Mahony et al. 2020). Hence, NAMs have the potential to populate qAOPs to derive some kind of cellular PoD in a chemical safety assessment. Collaborative projects such as the in3 project (refer to section 1.5) are ideal platforms to study and propose applications of data-derived models from NAMs. However, the main challenge remains - achieving regulatory acceptance of the novel methods and facilitating changes to the regulatory standard requirements. Ideally, RA should be based on well-understood toxicological mechanisms, and include studies carried out under controlled experimental conditions, with detailed recording of the data, model and supporting documentation. The long-term challenge is to update the regulatory requirements to accommodate the novel methods, rather than trying to match new methods with current information requirements. Thus, a qAOP model can serve as a bridge between NAMs and the current regulatory landscape that can help, eventually, to translate the new science into internationally accepted standard methods with a final utility in CRA (Figure 1.3).

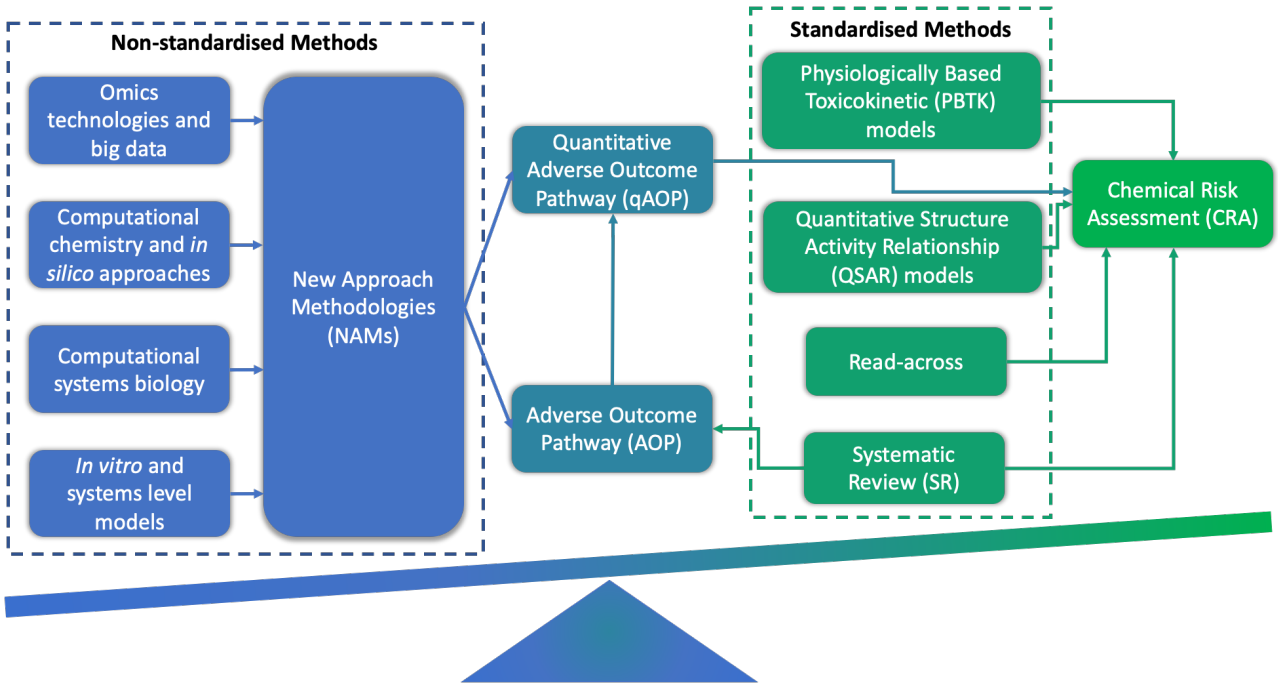


Figure 1.3. A representation of where an AOP and qAOP fit within the current landscape of novel methods to assess the safety of chemicals as well as their role in helping to reach an equilibrium between NAMs and a pathway-derived CRA much needed nowadays. (Non-)standardised methods, refers to methods accepted by a regulatory authority and used in practice for CRA that follows proposed templates and guidelines.

## 1.5. The in3 Project

Given the needs identified above, i.e., data to populate qAOPs, guidance to model and structure qAOPs, demonstration of qAOPs being fit-for-purpose etc., more scientific initiatives are required as well as greater influence from key public and private players, such as the OECD, the European Commission Joint Research Centre (EC JRC) and the United States Environmental Protection Agency (US EPA), to name a few.

The "*An Integrated Interdisciplinary Approach to Animal-Free Chemical and Nanomaterial Safety Assessment*" (in3) Project aimed to create a training network to drive the synergistic development and utilisation of *in vitro* and *in silico* tools for human chemical and nanomaterials safety assessment. The project focused on human induced pluripotent stem cell (hiPSC) derived tissues, including liver, kidney, brain, lung, and vasculature system, and utilised mechanistic toxicology, qAOPs, biokinetics, cheminformatics and modelling approaches to derive testable integrated prediction models (Figure 1.4). The project aimed to acquire a unique set of interdisciplinary knowledge and trained 15 PhD students from 11 beneficiaries, i.e., public institutions including universities and biotechnology companies while working towards the same goal, utilising the same chemicals, donor cells, assays, and software packages. The project ran between 1 January 2017 – 30 June 2021.

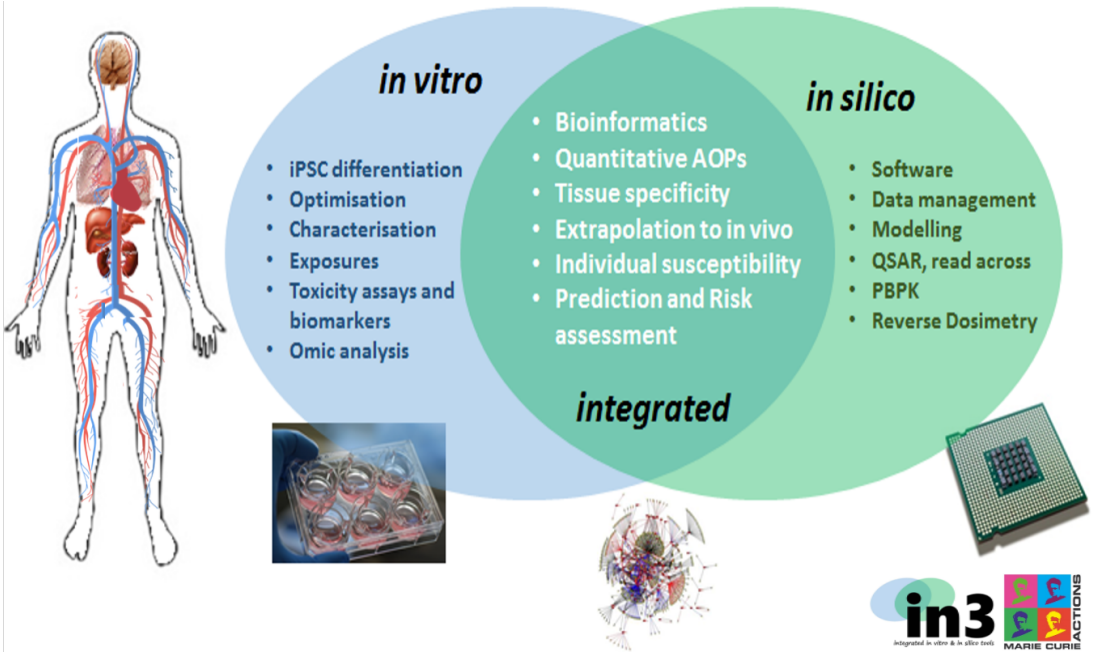


Figure 1.4. The graphical summary of the research carried by the in3 Project, adapted from the official website of the in3 Project (https://www.estiv.org/in3/about.html, accessed on March 16, 2021).

One of the core scientific activities of the in3 consortium was to develop and optimise qAOP models. The present PhD work has been undertaken as part of the in3 Project contributing with innovative approaches and solutions by the exploitation of existing *in vivo* and *in vitro* data and mechanistic knowledge, and integration into computational models. This adds value to the scientific community's efforts towards the utility of the AOP framework, enhancing the use of non-animal methods in human health RA. This additionally allows the strengthening of AOPs to achieve more efficient regulatory utility, aids the better understanding

of molecular mechanisms and provides concrete solutions and guidance on strategies and approaches for the development of qAOP models. This also underlines a new crowd-sourced way of research in predictive toxicology.

## 1.6. Research aims of this thesis

At the current time, there is considerable interest in qAOPs, but few viable qAOPs have been developed. There is also a lack of understanding of how these qAOPs could be used for regulatory purposes and in RA specifically. Therefore, the overall aim of this PhD thesis was to formulate strategies and develop qAOP models by using mechanistic knowledge, *in vitro* data and computational modelling approaches for CRA decision-making. The specific objectives to achieve this aim were:

I. To review the state-of-the-art of the qAOP concept.
    - This involved the collection of definitions proposed by the scientific community in the scientific literature as well as the assessment of available qAOP models against a series of pre-defined characteristics essential in the development and evaluation of qAOP models. The results are outlined in Chapter 2.

II. To demonstrate the potential utility of the science of networks applied to qualitative linear AOPs available in the OECD AOP-Wiki KB in order to depict better the mechanistic pathways and help to inform prioritisation strategies for testing and development of alternative methods.
    - This involved collection and curation of linear AOPs and development of AOP networks. As a case study, the neurotoxicity in humans was chosen as the endpoint to showcase the formulated methodology. The results are described in Chapter 3.

III. To design a framework for quantification of AOP networks.
    - This was applied to a simplified version of the AOP network following the results of Chapter 3 with the focus on developmental neurotoxicity. This involved investigating available empirical data and the selection of an appropriate computational approach, as described in Chapter 4.

IV. To investigate the application of information required for the review, endorsement and approval of a linear AOP by the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) available in the OECD AOP-Wiki KB for quantification purposes.
    - This involved formulation of a decision tree to select the references containing quantitative details, assessment of selected publications, data extraction, preparation and modelling. The case study was carried out for a compound known to induce Parkinsonian motor deficits. The results are outlined in Chapter 5.

The PhD thesis ends with a Chapter where future directions and implications for advancing the concept of qAOPs are discussed.

# Chapter 2. Theoretical background of quantitative Adverse Outcome Pathways

The work presented in this Chapter is based on a published review (Spinu et al. 2020) initiated during a secondment at the European Commission Joint Research Centre (EC JRC), Ispra, Italy (September 2018 – January 2019). The Chapter has been updated from the published paper with a dedicated section on semi-quantitative/quantitative weight of evidence qAOP (semi-q/qWoE qAOP) models to provide a complete overview of the state-of-the-art of the qAOP concept. In addition, two qAOP models (Rowland et al. 2019; Song et al. 2020) that were published after the initial study was conducted have been included as part of the analysis.


## Abstract

The qAOP concept is gaining interest due to its potential regulatory applications in CRA. Even though an increasing number of qAOP models are being proposed as computational predictive tools, there is no framework to guide their development and assessment. As such, the objectives of this Chapter were to: (i) analyse the definitions of qAOPs published in the scientific literature, (ii) define a set of common features of existing qAOP models derived from the published definitions, and (iii) identify and assess the existing published qAOP models and associated software tools. As a result, six probabilistic qAOPs and eleven mechanistic qAOPs were evaluated against the common features. Additionally, three semi-q/qWoE type of qAOPs were discussed. The review offers an overview of how the qAOP concept has advanced and how it can aid toxicity assessment in the future. It proposes common features that a qAOP model can be evaluated against for regulatory applicability. Further efforts are required to achieve validation, harmonisation and regulatory acceptance of qAOP models.

## 2.1. Introduction

From its establishment in 2010 (Ankley et al. 2010), the AOP framework aimed to enhance efficiency and transparency in CRA (OECD 2018b). In the context of the present thesis, the working definition of an AOP is that it represents a series of building blocks that describe biological events a chemical follows to induce an adverse event, i.e., AO at different levels of biological organisations as outlined in Chapter 1. Recent progress in the development of AOPs covers a spectrum of novel endpoints and chemicals/categories including nanoparticles and other classes of stressors, e.g., microplastics and radiation (Chauhan et al. 2019; Jeong and Choi 2019; Jeong et al. 2018). Furthermore, new ways of deriving AOPs have been proposed such as data mining, deep learning or a combination of ML techniques (Carvaillo et al. 2019; Jeong et al. 2019; Rugard et al. 2020).

In addition to the increasing numbers of linear (qualitative) AOPs, AOP networks are being extensively applied and have considerable value. An AOP network is defined as a set of linear AOPs sharing common events and, therefore, representing a better depiction of biological processes (Knapen et al. 2018; Villeneuve et al. 2018a). Examples of AOP network applications include: mapping chemicals to linear AOPs to identify common interactions (Aguayo-Orozco et al. 2019a); understanding the mechanistic pathways leading to mitochondrial dysfunction (Dreier et al. 2019); identification of common KEs for chemical screening and integrated testing strategy for developmental neurotoxicity (Li et al. 2019); chemical assessment using biologically-based testing batteries (Angrish et al. 2017); and the development of an exploratory AOP database to derive "putative" AOPs (Pittman et al. 2018). Moreover, progress has been made with regard to the use of topological features in the network, such as the degree to which the most common/highly connected paths within an AOP network can be identified (Pollesch et al. 2019). Additionally, many molecular initiating events (MIEs) have been thoroughly modelled *in silico* due to their ability to describe the interaction between the stressor and the biological receptor at the molecular level that induces adverse effects (Allen et al. 2016). *In silico* models of MIEs are represented by 2-D and 3-D structural alerts and QSARs (Allen et al. 2020; Cronin and Richarz 2017; Mellor et al. 2016) and have been incorporated in mechanistically-based PBTK models that evaluate exposure-response relationships (Gao et al. 2019; MacKay et al. 2013).

Formerly, various types of AOPs were distinguished from qualitative to semi-quantitative and quantitative AOPs (Perkins et al. 2015; Villeneuve et al. 2014a). While qualitative AOPs can be used to guide chemical decision-making during the development of novel compounds including the integration of diverse lines of evidence, prioritisation of testing strategies and screening of chemicals, design and development of fit-for-purpose assays, qAOP models are seen as tools for quantitative risk assessment for both new and existing chemicals (Carusi et al. 2018; Coady et al. 2019; Villeneuve et al. 2014b). Hence, each type of AOP has potential utility in CRA (Hecker and LaLone 2019).

The concept of a qAOP as a predictive computational model is gaining interest due to its ability to address the question of how much perturbation, at any of the upstream KEs, and under what conditions, the AO is likely to occur (Conolly et al. 2017; Patlewicz et al. 2015). CRA has long been based upon an understanding

and application of information from linear or threshold dose-response relationships. However, it is now recognised that linear and threshold relationships, i.e., NOAEL dose-response relationships are often not the norm and that non-linear dose-response modelling is more biologically plausible than low dose linear modelling. qAOP models are able to capture the dynamics underlying all kinds of a dose-response relationship, including (1) the common S-shaped (sigmoid) for toxicants having a threshold, (2) linear or non-linear curves in the low dose range for toxicants assumed to be without a threshold, e.g., genotoxic substances, (3) non-linear and non-monotonic relationships, e.g., as claimed for endocrine disruptors. Therefore, the relationship between the endpoints may be captured as a regression equation or a more complicated mathematical model or a probabilistic function. For example, the response-response relationship in a qAOP model can be conceptually represented as the magnitude of change in an upstream biological event plotted on the x-axis and the magnitude or severity of a downstream biological effect plotted on the y-axis (Conolly et al. 2017). Thus, a qAOP helps to define the biological tipping point(s) along the pathway, and the probability or magnitude with which those tipping points are exceeded (Conolly et al. 2017; LaLone et al. 2017a).

Several international workshops have identified critical aspects in developing a qAOP model including the quantification of KERs, data availability, defining the threshold for inducing an effect, incorporation of modulating factors, e.g., genetic predisposition, previous exposures, the establishment of mathematical rules for the KERs, parametrisation of non-linear models and validation and implementation, to name a few (Kleinstreuer et al. 2016; Leist et al. 2017; Wittwehr et al. 2017). However, the extent to which these challenges are addressed by available qAOP models is not covered by the scientific literature. On the other hand, whilst knowledge is being acquired and systematically captured, there is no official guidance providing a coherent and all-encompassing framework for the development and assessment of a qAOP model. The existing guidance, developed by the OECD, explains how to build evidence for an AOP and this highlights the importance of the quantitative understanding of the KER as a criterion in the assessment of the overall confidence of an AOP (OECD 2018b). In addition, the OECD guidance on the use of AOPs in the development of IATA states that a qAOP can help to target a KE and select the appropriate assays for TGs development or refinement to predict the AO (OECD 2016).

## 2.2. Aim of this chapter

This Chapter aimed to evaluate the progress made in the qAOP concept in chemical safety assessment, to identify areas of strength and weakness, as well as opportunities for developments to support regulatory acceptance of qAOPs. The specific objectives were (i) to analyse published definitions of qAOPs in the scientific literature; (ii) to summarise the current status of the semi-q/qWoE qAOPs; (iii) to formulate a set of common features of a qAOP model; (iv) to assess the types of qAOP models based on the identified features that utilise probabilistic and mechanistic approaches, as well as methods and software tools used for the modelling of qAOPs. Relevant scientific literature in the Web of Science, Pubmed and Google Scholar databases published before October 2020 was screened, as described below.

## 2.3. Computational modelling in the context of qAOPs

The OECD Guidance document on the use of AOPs in IATA (OECD 2016) defines a qAOP as "*an assembly of KEs supported by descriptions of how the KEs can be measured and the accuracy and precision with which the measurements are made along with KERs supported by a quantitative understanding of what magnitude and/or duration of change in the upstream KE is needed to evoke some magnitude of change in the downstream KE"*. Despite this clear definition, the meaning of qAOPs has often been interpreted differently, with various definitions given and, as a result, varying expectations of the scientific community. Screening the scientific literature for the Medical Subject Headings (MeSH) term "quantitative Adverse Outcome Pathways", 23 publications were found which referred to the concept of qAOP (Appendix I). The definitions identified that referred to the concept of qAOP were retrieved and analysed individually to understand and map a series of common features that the authors recognised as essential for the development of a qAOP model. Thus, a list of five common features for qAOP models was formulated encompassing the expectations of the scientific community (Table 2.1), mainly (I) problem formulation, (II) mechanistic knowledge and associated data, (III) quantitative approaches, (IV) regulatory applicability, (V) additional considerations. These features help to understand how the modelling of a qAOP has been approached as well as the opportunities for improving the modelling process.

Table 2.1. Common features of a qAOP model envisaged by the scientific community. The assessment criteria were used to characterise and evaluate available probabilistic and mechanistic qAOP models.

| Common feature | Description | Assessment criteria |
|---|---|---|
| Problem formulation | • A qAOP should answer a well-defined question relevant to the AO of interest.<br>• The purpose of the model dictates how much mechanistic understanding is required, and the way a qAOP should be developed, validated and used. | • The question addressed and/or purpose of modelling<br>• AO studied |
| Mechanistic knowledge and associated data | • The OECD AOP-Wiki KB can support the development of a qAOP model to predict an endpoint of interest. Empirical data for model parametrisation, fitting and/or testing can be obtained from the description of KERs published in the OECD AOP-Wiki KB.<br>• Whilst for quantification it is recommended to start with linear AOPs, it should not impede quantification of networks or highly connected KEs/KERs within an AOP network.<br>• A qAOP model relies heavily on data: not only bioactivity of a compound/mixtures, but also measurements of effects at relevant doses/concentrations and appropriate timescales including physico-chemical properties and molecular descriptors. Data may come from a range of *in vivo* and *in vitro* studies specifically designed to test an AOP as a hypothesis and/or retrieved from a variety of sources to assist with this process.<br>• Both adjacent and non-adjacent KEs paired as upstream-downstream in a KER should be quantified even though each of them impacts differently on the modelling process, e.g., in the context of Bayesian network modelling. Adjacency refers to whether there are other KEs positioned in between linear construction of an AOP or not.<br>• Different biological level of organisations should be quantified if this is relevant to the AO of interest and available data allowed. | • Presence of the AOP in the OECD AOP-Wiki KB<br>• Type of AOP: linear or network<br>• Single chemical(s)/mixtures<br>• Type of data: *in vivo*, *in vitro*, *in silico* and/or other variables<br>• Dose/concentration-responses (D/C-R) and time-responses (T-R)<br>• The adjacency of KERs: adjacency and non-adjacency<br>• Biological levels: cellular, tissue, organ, organism, population |
| Quantitative approaches | • The modelling approaches can vary from being probabilistic and logic-based (Boolean) to deterministic, stochastic type of modelling.<br>• The mathematical expression can take various forms including linear regressions and ordinary differential equations resulting in different graphical shapes, e.g., linear, sigmoidal, Gaussian-type plots. | • Type of quantitative approach |
| Regulatory applicability | • A qAOP model should imply various applications to regulatory decision-making and acceptance. | • Human health/ecological risk assessment |
| Additional considerations | • These considerations can influence the regulatory approval, reduce the uncertainties and extend the applicability domain of the predictions of a qAOP model.<br>• A qualitative AOP is not chemical-specific, therefore, neither should be a qAOP. However, the test data mostly come from experiments conducted with a particular interest in assessing a specific chemical or class of chemicals. Understanding this aspect allows for coupling with other *in silico* models including read-across.<br>• It is not mandatory that the test methods used (models and measured endpoints) are adopted/validated following national/international guidelines. However, they should be performed in a quality-controlled environment where relevance of the model is proved based on scientific rationale and reproducibility of data.<br>• Even though none of the definitions identified referred to uncertainty and sensitivity analysis, this aspect should be considered as well for its value in validating the predictions of a qAOP model while giving confidence in its further applications. | • Cross species extrapolation<br>• Modulating factors<br>• Positive/negative feedback loops<br>• Compensatory mechanisms<br>• Test method adopted/validated<br>• Kinetics<br>• Chemically agnostic<br>• Exposure assessment<br>• Uncertainty evaluation<br>• Sensitivity analysis<br>• Availability: open access or not |

Additionally, three conceptual classes of qAOPs have been suggested by Gust et al. (2016) and Perkins et al. (2019a):

1) Semi-q/qWoE qAOPs. These utilise quantitative weighting and numerical assessments of multiple lines of evidence to rank the confidence in KERs for further quantification (Gust et al. 2016; Perkins et al. 2019b). For example, to calculate the quantitative confidence scoring of KERs of a linear AOP, BH considerations (biological plausibility, essentiality, dose-response, temporal and incidence concordance) were proposed as a conceptual method by Becker et al. (2017), while Collier et al. (2016) additionally used metrics related to data quality for the KEs as will be described in more detail in Section 2.4.
2) Probabilistic qAOPs and qAOP networks are presented in depth in Section 2.5. These are computational models that incorporate statistical or probabilistic approaches such as Bayesian networks covering few events or an entire AOP to build predictive relationships between MIEs and/or KEs linked to apical outcomes (Gust et al. 2016; Perkins et al. 2019b).
3) Mechanistic qAOPs and qAOP networks that are analysed in Section 2.6. These are computational models defined as deterministic models where mathematical functions of the MIE, KE and KER can be used to predict the likelihood that a later event or outcome would occur based on changes in an earlier event given specified initial conditions (Gust et al. 2016; Perkins et al. 2019b).

The definitions of the qAOP concept as identified in the scientific literature support all these types of qAOP models, with only a small proportion (fewer than 10%) referring to semi-q/qWoE qAOPs, and approximately 25% to probabilistic qAOPs while all papers referred to mechanistic qAOPs. Therefore, whilst the first type of qAOP can be regarded as an extension of a qualitative AOP with empirical data, the second and third types of qAOP are mathematical models, distinguished according to the type of modelling approach. Thus, the first type of qAOP is conceptually different to the second and third, and an opportunity exists to make use of semi-q/qWoE qAOPs to develop predictive models based on probabilistic or mechanistic approaches as presented graphically in Figure 2.1.

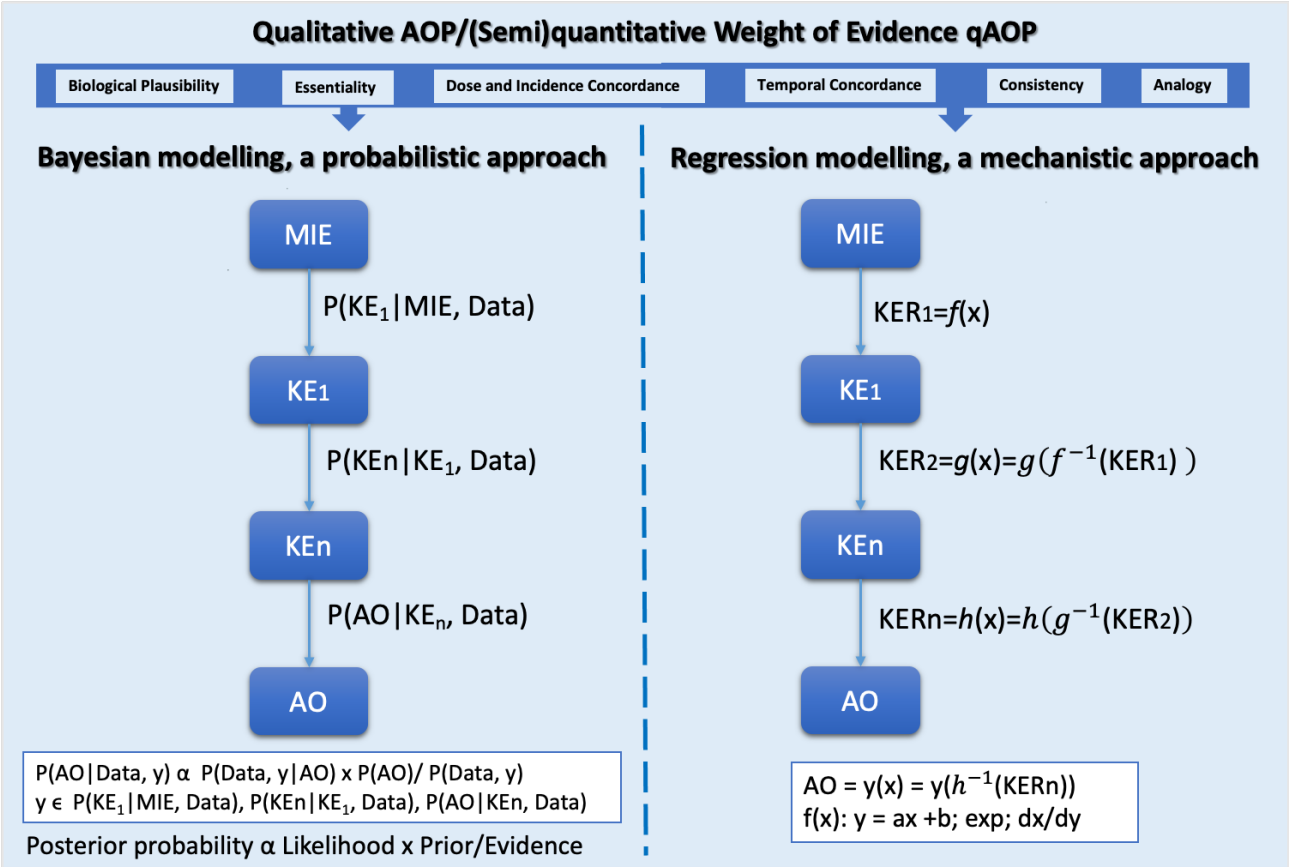


Figure 2.1. Conceptual representation of available types of qAOP models. Qualitative AOPs have an informative role for prioritisation and computational modelling of AO of interest and can additionally be quantified by a weight-of-evidence. A common approach to probabilistic modelling relies on the use of Bayes theorem as described below. Mechanistic qAOP models utilise mathematical functions, including linear regressions.

## 2.4. Overview of semi-quantitative/quantitative weight-of-evidence AOPs

Weight of evidence (WoE) analysis has been proposed as a method to assess the level of maturity and confidence in an AOP (Becker et al. 2015). It can be defined as a process in which each line of evidence, i.e., a group of evidence of similar type, is evaluated and weighted against a set of considerations to support possible answers to a research question (Hardy et al. 2017). Today, WoE evaluation is a key element that the OECD guidance on the development and assessment of AOPs requires to support the postulated biological pathway/disease processes induced by stressors (OECD 2018b).

The BH considerations were originally developed to examine the causality of associations observed in epidemiological studies (Hill 1965). Later, the evolved BH considerations were formulated to increase WoE determinations in the application of MoA/species concordance analysis (Meek et al. 2014a; Meek et al. 2014b). Only recently, the evolved modified BH considerations were adopted for the assessment of the WoE for KEs including MIEs, KERs and overall linear AOPs (OECD 2018b). Therefore, BH considerations are now widely used to assess experimental/historical/epidemiological/mechanistic types of data for the following purposes: (i) to promote consistency between different streams of evidence, (ii) to support acceptance or rejection of a hypothesised AOP, (iii) to offer recommendations for additional targeted research, (iv) to prioritise substances for further testing, (v) to guide the development of more efficient testing strategies, (vi) to identify potential gaps and critical data needs for a complete CRA. In rank order, the evolved BH considerations include biological plausibility > essentiality > empirical support (dose-response, temporality, and incidence concordance). These can each be put in the context of the various aspects of qAOPs. The biological plausibility of KERs is defined as "*an understanding of the fundamental biological processes involved and whether they are consistent with the causal relationship being proposed in an AOP*" (OECD 2018b). The essentiality of KEs refers to "*experimental data for whether or not downstream KEs or the AO are prevented or modified if an upstream event is blocked*" (OECD 2018b). The empirical support of KERs is characterised by the "*toxicological data derived by one or more reference chemicals where dose-response and temporal concordance for the KE pair can be assessed*" (OECD 2018b). Thus, the WoE analysis in support of an overall AOP can be summarised qualitatively or (semi)quantitatively.

The term "semi-quantitative/quantitative qAOPs" has been promoted to increase certainty and strength of KERs that are of particular interest for the development of probabilistic and mechanistic qAOP models (Perkins et al. 2019b). It refers to the quantification of a WoE analysis by attributing metrics of weights to the available mechanistic knowledge of an AOP based on prescriptive guidelines or expert judgement for a systematic assessment of confidence levels, which also, expresses uncertainty in a given AOP (Perkins et al. 2019b). This leads to an improved assessment of confidence in the estimation of the AO and supports decision-making transparently.

Currently, three methodologies applied to linear AOPs are described with few modifications to the OECD guidance (Table 2.2). The frameworks (Becker et al. 2015; Becker et al. 2017; Collier et al. 2016) include three BH considerations and two additional BH criteria, which were consistency, e.g., between the levels of

biological organisations, and analogy, e.g., between chemical stressors that the OECD recommends being carried through the evaluation process and not as separate entities. Becker et al. (2015) used a prototype multi-criteria decision analysis (MCDA) model for qWoE of an AOP that describes aromatase inhibition leading to reproductive dysfunction in fish. Collier et al. (2016) proposed a mathematical approach to quantify the scientific evidence of two AOPs: (1) non-competitive ionotropic GABA receptor antagonism leading to epileptic seizures, and (2) antagonist-binding and stabilisation of a co-repressor to the peroxisome proliferator-activated receptor $\alpha$ (PPAR$\alpha$) signalling complex ultimately causing starvation-like-weight loss. Becker et al. (2017) formulated a mathematical approach to compare two modes of action, a mutagenic path vs activation of PPAR$\alpha$ that were structured as KEs and KERs for clofibrate, a rodent liver carcinogen. Becker et al. (2015) conducted a sensitivity analysis, i.e., the weights associated with two of BH criteria were varied from 0% to 100%, while the weight of the other criterion was held constant to identify the most sensitive KER against weight alterations to each of the BH criteria. Collier et al. (2016) developed the qWoE based on BH criteria and in addition, evaluated data quality criteria, established by the US EPA known as General Assessment Factors (GAFs). Becker et al. (2017) advocated the need for a semi-qWoE, i.e., incorporation of narrative discussions besides the quantitative assessment. Thus, the biological plausibility criterion was not assessed numerically, and a narrative description was proposed instead. The (non)adjacency of KERs, i.e., the sequential order of the KEs, has been evaluated by Becker et al. (2015) and Becker et al. (2017). All the assessed AOPs are endorsed by the OECD. All the approaches are rank-based methods, with higher values representing greater importance of BH criteria for the specific AOP. A summary of the comparison of the technical aspects that differentiated the three methodologies, which mainly included the approach and scoring system applied and the investigated AOP, is presented in Table 2.3.

There are several opportunities to improve the currently proposed frameworks for semi-q/qWoE qAOPs that can be underlined. Software such as the Science in Risk Assessment and Policy (SciRAP)[1] tool might serve as a transparent reporting and evaluation resource. It allows for the increase/decrease of weights and for the inclusion of additional comments (narrative assessment) that leads to a personalised assessment of the selected lines of evidence and their quantification in a structured manner. However, SciRAP relies solely on two criteria: relevance, i.e., "*the contribution a piece or line of evidence would make to answer a specified question if the information comprising the evidence were fully reliable*", and reliability, i.e., "*the extent to which the information comprising of a piece or line of evidence is correct given by accuracy and precision*" (Hardy et al. 2017). The inclusion into a semi-q/q WoE qAOP tool of BH considerations and data quality criteria following the characteristic building blocks of an AOP can align the requirements of all stakeholders towards a better-informed testing strategy. Also, an ideal qWoE tool should allow for the evaluation if not only *in vivo* and *in vitro* studies, but also *in silico* and epidemiological studies in addition to its being applicable to any type of stressors, e.g., chemicals, nanoparticles, genetic and environmental factors, in a structured

[1] www.scirap.org, accessed on March 16, 2021.

manner informed by an AOP. Furthermore, limitations and uncertainties should be quantified, for which probabilistic approaches seem most appropriate. For example, Jannicke Moe et al. (2020a) developed a Bayesian network to predict the acute toxicity of chemicals to juvenile fish based on a combination of information, mainly data from fish embryo toxicity, physico-chemical properties, chemical category and toxicity data for other species, with the goal of supporting the WoE to replace the OECD TG 203 on acute fish toxicity assay with non-animal methods. Another approach to assessing qAOPs was provided by Kleinstreuer et al. (2016) who proposed a decision formula to quantify the significance of a biological event by taking into account its position within an AOP. Expanding on the proposals of Kleinstreuer et al. (2016), Chapter 3 illustrates how the position of a KE/KER in an AOP network can be quantified by applying a topology analysis and how the results can have an essential role in decision-making processes. An example of the application of these findings is provided in Chapter 4 through the quantification of a simplified AOP network following the available mechanistic knowledge.

Table 2.2. An overview of the defining questions used in the assessment of an AOP in comparison to the OECD Guidance that follow the Bradford Hill modified considerations and consistency and analogy criteria by three published methodologies.

| Bradford Hill Criteria | OECD Guidance (OECD 2018b) | Becker et al. (2015) | Collier et al. (2016)[1] | Becker et al. (2017) |
|---|---|---|---|---|
| **Biological concordance / plausibility of KERs** | *"Is there a mechanistic, i.e. structural or functional relationship, between $KE_{up}$ and $KE_{down}$ consistent with established biological knowledge?"* | *"Is there a mechanistic, i.e., structural or functional relationship between Key $Event_{upstream}$ and Key $Event_{downstream}$ consistent with established biological knowledge?"* | *"Does the hypothesised KER conflict with broader biological knowledge? How well established is the KER in the context of the AOP?"* | *"Does the hypothesised MOA (AOP) conflict with broader biological knowledge? How well established is the MOA (AOP)?"* |
| **Essentiality of KEs** | *"What is the impact on downstream KEs and/or the AO if an upstream KE is modified or prevented?"* | *"Are downstream Key Events and/or the Adverse Outcome prevented if an upstream Key Event is blocked?"* | *"Is the sequence of events reversible if dosing is stopped or a key event is prevented?"* | *"Is the sequence of events reversible if dosing is stopped or a KE prevented?"* |
| **Empirical observations / evidence of KERs** | *"Are there inconsistencies in empirical support across taxa, species and stressors that do not align with expected pattern for hypothesised AOP?"*<br>*"Dose and temporal concordance: Does $KE_{up}$ occur at lower doses and earlier time points than $KE_{down}$?"*<br>*"Incidence concordance: At the same dose of stressor, is the incidence of $KE_{up}$ greater than that for $KE_{down}$?"* | *"Does the empirical evidence support that a change in Key $Event_{upstream}$ leads to an appropriate change in Key $Event_{downstream}$?"*<br>*"Dose and temporal concordance: Does Key $Event_{upstream}$ occur at lower doses and earlier time points than Key $Event_{downstream}$?"*<br>*"Incidence concordance: Is the incidence of Key $Event_{upstream}$ greater than that for the Key $Event_{downstream}$?"* | *"Dose concordance: Is the upstream key event observed at doses below or similar to those associated with the downstream key event?"*<br>*"Incidence concordance: Is the occurrence of the downstream key event effect less than that for the preceding key events?"*<br>*"Temporal concordance: Are the key events observed in hypothesised order?"* | *"Dose concordance: Are the KEs observed at doses below or similar to those associated with the adverse effects?"*<br>*"Incidence concordance: Is the occurrence of the adverse effect less than that for the preceding KEs?"*<br>*"Temporal concordance: Are the KEs observed in hypothesised order?"* |
| **Additional considerations** | | | | |
| **Consistency (among different biological contexts)** | NA | *"Are there inconsistencies and uncertainties in empirical support across taxa, species and stressors that do not align with an expected pattern for the hypothesised AOP?"* | *"Is the pattern of observations across species/strains/organs/test systems what would be expected based on the hypothesised AOP?"* | *"What is the pattern of observations across species/strains/organs/test systems? What would be expected based on the hypothesised MOA (AOP)?"* |
| **Analogy (consistency across chemicals)** | NA | *"Is the MIE affected by a variety of chemicals or by a single/very limited number of chemicals?"* | *"Would the AOP be anticipated based on broader chemical specific knowledge, e.g., the chemical is a member of a category for which related chemicals have known or strongly suspected AOP?"* | *"Would the MOA (AOP) be anticipated based on broader chemical-specific knowledge, e.g., the chemical is a member of a category for which related chemicals have known or strongly suspected MOA (AOP)?"* |

[1]It used the United States Environmental Protection Agency (US EPA) General Assessment Factors for WoE evaluation of the KEs in addition to the BH criteria.

Table 2.3. A comparison of three published methodologies for modelling semi-q/qWoE qAOPs.

| Methodology | Becker et al. (2015) | Collier et al. (2016) | Becker et al. (2017) |
|---|---|---|---|
| **Adverse Outcome Pathway(s)** | Aromatase inhibition leading to reproductive dysfunction in fish (AOP ID 25[2]) | Two AOPs:<br>• AOP for ionotropic g-aminobutyric acid (GABA) receptor antagonism induced by non-competitive channel blockers leading to epileptic seizures (AOP ID 10[3])<br>• Antagonist-binding causing stabilisation of co-repressor (SMRT or N-CoR) to PPAR$_\alpha$ ligand binding domain causing downstream starvation-like body-weight loss (AOP ID 6[4]) | Rodent hepatocarcinomas induced by clofibrate through two different MoAs, i.e., activation of PPAR$_\alpha$ (AOP ID 37[5]) vs mutagenic path. |
| **Software** | DECERNS software[6] | Not mentioned | Not mentioned |
| **Approach** | Multi Criteria Decision Analysis (MCDA) | Assessment of data quality and causality | Comparative analysis of two MoAs for a chemical |
| **Steps** | 1. Define KERs as alternatives for which despite unknown relative confidence, need to be prioritised;<br>2. Map out the criteria and metrics based on BH considerations and KERs as a value tree;<br>3. Assign weights to indicate the importance of each of the considerations and metrics associated with each KER;<br>4. Score each KER based on each metric;<br>5. Integrate scores and weights for each KER to assess overall confidence level;<br>6. Conduct sensitivity analysis. | 1. Prepare the AOP (assembling evidence);<br>2. Prepare criteria weighting and scoring (weighting evidence);<br>3. Aggregate lines of evidence (weighing the body of evidence). | 1. Identify postulated MOA(s) (AOPs);<br>2. Qualitatively evaluate the evidence in support/inconsistent with the KE/KER;<br>3. Quantitatively rate each KE/KER using the BH considerations;<br>4. Assign weights to each of the BH considerations;<br>5. Derive the composite score for each KE and KER;<br>6. Integrate the evidence of causality for the MOA (AOP);<br>7. Compare the quantitative scores for the hypothesised MOA (AOP). |
| **Scoring system (assigned weight)** | Low/Weak (1)<br>Moderate (2)<br>High/Strong (3) | The criterion does not apply (0)<br>The criterion applies very weakly (1)<br>The criterion applies weakly (2)<br>The criterion applies moderately (3)<br>The criterion applies strongly (4)<br>The criterion applies very strongly (5) | Strong Evidence (3)<br>Moderate Evidence (2)<br>Weak Evidence (1)<br>No evidence (0)<br>Weak Counter Evidence (-1)<br>Moderate Counter Evidence (-2)<br>Strong Counter Evidence (-3) |

[2] https://aopwiki.org/aops/25, accessed on March 16, 2021.
[3] https://aopwiki.org/aops/6, accessed on March 16, 2021.
[4] https://aopwiki.org/aops/10, accessed on March 16, 2021.
[5] https://aopwiki.org/aops/37, accessed on March 16, 2021.
[6] http://www.decerns.com/, accessed on March 16, 2021.

## 2.5. Overview of probabilistic quantitative linear AOPs and AOP networks

Another type of qAOP model, to maximise existing resources and facilitate toxicity assessment, is represented by the probabilistic approaches. Bayesian networks use a directed acyclic graph (DAG) to form conditional probability relationships. Each node in the network corresponds to a KE or additional variable, e.g., physico-chemical properties, while edges show the conditional dependencies between two KEs that form a KER. In other words, the Bayesian network uses conditional probability tables (CPTs) for each KE (node) to determine the probability of activity for parent and child nodes, i.e., an upstream KE leading to a downstream KE based on the Bayes' rule, which is the unique mathematical equation for this type of modelling. Whilst the choice of KEs in the DAG is informed by the structure of the AOP, a Bayesian network can be entirely data-driven and may, or may not, be consistent with the topology of the AOP. Therefore, the Bayesian network approach has other applications in predictive toxicology in addition to qAOP development. These include: identification of the best biomarkers to characterise chemical exposure using the dose-response analysis to determine the PoDs (Hack et al. 2010); development of an efficient testing strategy (Jaworska et al. 2015); classification of chemicals based on a MoA (Carriger et al. 2016); classification of the cellular effects of nanoparticles (Furxhi et al. 2019); and prediction of the severity level of drug-induced liver injury (Williams et al. 2020). Currently, six qAOP models have been identified that follow the Bayesian approach and were assessed for the common features: (I) problem formulation, (II) mechanistic knowledge and associated data, (III) quantitative approaches, (IV) regulatory applicability, (V) additional considerations. They are described in turn below and listed in Tables 2.4 and 2.6. A further qAOP model represents a combination of both probabilistic and mechanistic approaches (Zgheib et al. 2019), and it is discussed in Section 2.6.

### 2.5.1. Problem formulation

A variety of purposes can be recognised across the available probabilistic qAOPs models. The AOs covered by these models include organ failure or ecotoxicological population level endpoints.

### 2.5.2. Mechanistic knowledge and associated data

Seven probabilistic qAOPs are available in the AOP-Wiki KB (AOPs IDs 7, 23, 25, 30, 207, 245, 284). Two probabilistic qAOPs utilised AOP networks (Burgoon et al. 2020; Chu 2018). One probabilistic qAOP represented a curated version of a series of linear AOPs (AOPs IDs 7, 23, 25, 30) (Rowland et al. 2019). The qAOP of Jannicke Moe et al. (2020b) included a linear AOP with KEs represented by multiple measurements, e.g., oxidative phosphorylation and formation of reactive oxygen species (ROS) to describe the first KE. All probabilistic qAOP models incorporated various types of data, including experimentally and/or judgement based derived results. Jannicke Moe et al. (2020b), Jeong et al. (2018) and Rowland et al. (2019) quantified AOPs of interest using experimental data, while Chu (2018) conducted specific experiments and Perkins et al. (2019a) used a combination of *in vitro* data and expert judgment. Importantly, probabilistic approaches are flexible and can estimate predictions for both single chemicals and mixtures more easily than mechanistic

approaches, e.g., binary assumption of a state of a KER, while taking into account the randomness given by the chosen type of distribution. As a result, Perkins et al. (2019a) quantified liver steatosis caused by both individual, and a mixture of, chemicals. In addition, a qAOP model can predict many other endpoints for which data may be available, while these endpoints may not necessarily be identified as KEs. For example, Burgoon et al. (2020) quantified genes potentially considered as biomarkers for liver steatosis. Likewise, Chu (2018) analysed the exposure to single organophosphate pesticides and binary and tertiary mixtures (a synergistic effect). However, not all of the probabilistic qAOPs assessed this aspect, i.e., mixture vs individual chemicals. For example, Jannicke Moe et al. (2020b) quantified the linkage between exposure to 3,5-dichlorophenol to a reduced number of fronds in the aquatic plant *Lemna minor*. The data included in the model of Rowland et al. (2019) were represented by a series of single chemicals tested at various concentrations, collected from the literature known to have an impact on the reproduction function in fathead minnow. Interestingly, nanoparticles were assessed in addition to single (small) organic compounds. As such, Jeong et al. (2018) quantified the reproductive toxicity of silver nanoparticles induced via oxidative stress in the nematode *Caenorhabditis elegans*. All probabilistic qAOPs made an attempt to link molecular/cellular effects to organ effects through adjacent KERs, and thus, to model biological processes at a multiscale level. However, not all probabilistic qAOP models accounted for dose and time responses. Whilst all included dose-responses, the models developed by Chu (2018), Jeong et al. (2018), Perkins et al. (2019a), Zgheib et al. (2019) and Rowland et al. (2019) made predictions related to time.

### 2.5.3. Quantitative approaches

Jannicke Moe et al. (2020b) formulated CPTs based on the count of observations and statistical analysis. Comparing these two CPTs, those based on the count of observations gave more accurate predictions at high and low stressor concentrations, while CPTs based on statistical models gave better predictions at intermediate stressor concentrations. When no information is available, the probability of activation can be set at 50%, for example, the qAOP model developed by Perkins et al. (2019a). Another important aspect is the type of variables used to define the nodes, in discrete or continuous forms. Most qAOP models defined the nodes as discrete states: intervals (Jannicke Moe et al. 2020b), yes/no and decrease/stable/increase (Jeong et al. 2018), active/inactive (Perkins et al. 2019a) and categories/groups of intervals or periods of time (Chu 2018). Depending on its scope, the Bayesian network can have different outputs: the probability of a compound being active at a given concentration (Perkins et al. 2019a); the prediction of responses of each KE at different concentrations (Jannicke Moe et al. 2020b); the calculation of relative risk (Chu 2018); or the analysis of causal relationships between KEs (Jeong et al. 2018). Importantly, Rowland et al. (2019) formulated a data-driven methodology in order to capture the variability of the experimental data utilised assuming a linear relationship between the experimental input, e.g., tested doses, fold-changes and the measurable output, e.g., responses. The parameters values were estimated from a Gaussian distribution given by the mean and standard deviation from the confidence intervals. A sigmoidal function was used to fit the dose-response and bootstrapping was applied to capture the noise, e.g., given by measurement errors.

Thus, the algorithm developed by Rowland et al. (2019) suits sparse datasets perfectly allowing for the calculation of thresholds of the activity comparing the response of each sample population to the response of each activity at the lowest and highest levels.

### 2.5.4. Regulatory applicability

The regulatory applicability was defined as any relevant potential use in CRA that the authors of the qAOP models have proposed. Thus, it does not necessarily follow a specific regulation. Two of the qAOP models (Burgoon et al. 2020; Zgheib et al. 2019) are applicable in human health risk assessment, three qAOP models (Chu 2018; Jannicke Moe et al. 2020b; Rowland et al. 2019) in ecological risk assessment and a single qAOP model (Jeong et al. 2018) in nanoparticle risk assessment. The qAOP by Rowland et al. (2019) can be used for hazard assessment and prioritisation.

### 2.5.5. Additional considerations

None of the qAOP models included kinetic considerations, non-adjacent KERs, details about compensatory mechanisms or feedback loops. However, the qAOP model developed by Chu (2018) considered modulating factors such as environmental stressors. Furthermore, this qAOP integrated probability, risk, and exposure responses to assess the population size of Chinook salmon. In addition, for experimentally derived data, none of the tests or assays are formally validated or nationally/internationally adopted. However, Jannicke Moe et al. (2020b) performed tests using the aquatic plant *Lemna minor*, which is widely accepted in guidance for toxicity testing (OECD 2006). Nevertheless, as the authors pointed out, *Lemna minor* is used for the analysis of an endpoint, which is an AO in an AOP rather than an entire AOP. Sources of uncertainty were listed by Chu (2018), Jannicke Moe et al. (2020b), Zgheib et al. (2019) and Rowland et al. (2019), while sensitivity analysis was conducted for all the qAOPs. These types of qAOPs have been modelled using existing software and/or coded in programming languages, i.e., R[1].

[1] https://www.R-project.org/, accessed on March 16, 2021.

Table 2.4. Characterisation of six probabilistic models that use the Bayesian network approach and an AOP construct based on directed acyclic graphs.

| Model purpose | Adverse outcome | Mechanistic knowledge and associated data | | | | | | | | Quantitative approach | Regulatory applicability | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OECD AOP-Wiki KB[1] | Type of AOP[2] | Single chemical(s)/ Mixtures | Data type | Adjacent KERs | Biological level(s) | D/C-R[3] | T-R[4] | | | |
| The risk posed by pesticides and environmental stressors to population size of Chinook salmon | Alteration of population dynamics | No[5] | AOPN | Mixtures | *In vitro* experimental data, literature data, AOP construction, environmental factors, population characteristics | √ | Molecular, cellular, organ, organism, population | √ | √ | Bayesian Network-Relative Risk type of model | Ecological risk assessment | Chu (2018) |
| Effects on reproduction of *Lemna minor* (duckweed) | Reduced number of fronds | AOP ID 245 | LAOP | Single chemical | *In vitro* experimental data, AOP construction | √ | Molecular, cellular, organism | √ | - | Bayesian network type of model (discrete states as three intervals) | Ecological risk assessment | Jannicke Moe et al. (2020b) |
| Toxicity of silver nanoparticles, linking MIE to the AO | Reproduction failure | AOP ID 207 | LAOP | Nanoparticles | *In vitro* experimental data, literature data, AOP construction | √ | Molecular, cellular, organ, organism | √ | √ | Bayesian network type of model (discrete states as yes/no, and decrease/stable/ increase), Bootstrapping | Ecological risk assessment | Jeong et al. (2018) |

*To be continued*

[1]Model follows an AOP structure, the MIE (ID 12) can be found in the AOP-Wiki KB, however, the AOP itself is not yet published.
[2]Linear AOP (LAOP), AOP Network (AOPN).
[3]Dose/Concentration-Responses (D/C-R).
[4]Time-Responses (T-R).
[5]Numbers represent the indices (XXX) of the AOP in the AOP-Wiki KB available at https://aopwiki.org/events/XXX.

Table 2.4. Continued

| Model purpose | Adverse outcome | Mechanistic knowledge and associated data | | | | | | | | Quantitative approach | Regulatory applicability | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OECD AOP-Wiki KB[6] | Type of AOP[7] | Single chemical(s)/ Mixtures | Data type | Adjacent KERs | Biological level(s) | D/C-R[8] | T-R[9] | | | |
| Occurrence of steatosis under different chemical exposures | Hepatic steatosis | No[10] | AOPN | Mixtures | Expert judgement, literature data, AOP construction | √ | Molecular, cellular, tissue, organ | √ | - | Bayesian network type of model (discrete states as active or inactive) | Human health risk assessment | Burgoon et al. (2020); Perkins et al. (2019a) |
| Comparison between probabilistic and mechanistic approaches | Nephron attrition leading to chronic kidney disease | AOP ID 284 | LAOP | Single chemical | *In vitro* experimental data on human RPTEC/TERT1 cells, AOP construction | √ | Molecular, cellular, tissue, organ | √ | √ | Dynamic Bayesian network model | Human health risk assessment | Zgheib et al. (2019)[11] |
| Data-driven approach to predict the level of reproduction of a network of AOPs | Decrease in population trajectory of fathead minnow | AOP IDs 7, 23, 25, 30 | AOPN | Single chemicals | Empirical data | √ | Molecular, cellular, organism, population | √ | √ | Stochastic modelling, Bootstrapping | Hazard assessment and prioritisation | Rowland et al. (2019) |

[6]Model follows an AOP structure, the MIE (ID 12) can be found in the AOP-Wiki KB, however, the AOP itself is not yet published.
[7]Linear AOP (LAOP), AOP Network (AOPN).
[8]Dose/Concentration-Responses (D/C-R).
[9]Time-Responses (T-R).
[10]Model is included in the AOPXplorer tool (http://apps.cytoscape.org/apps/aopxplorer, accessed on March 16, 2021) as it follows the structure of an AOP network.
[11]Model represents a combination of both probabilistic and mechanistic approaches.

## 2.6. Overview of mechanistic quantitative linear AOPs and AOP networks

A mechanistic qAOP model is driven by hypothesis testing and utilises a series of deterministic and stochastic techniques that are discussed briefly below. Eleven qAOP models were identified that follow a mechanistic approach, which were assessed for the common features: (I) problem formulation, (II) mechanistic knowledge and associated data, (III) quantitative approaches, (IV) regulatory applicability, (V) additional considerations. They are described below and summarised in Tables 2.5 and 2.6.

### 2.6.1. Problem formulation

The focus of this type of qAOP relies mainly on understanding the mechanism of toxicity and associated relevant taxonomic domain. The AOs are represented by effects at the ecotoxicological population level and organ toxicity, e.g., chronic kidney disease, neurodegenerative diseases.

### 2.6.2. Mechanistic knowledge and associated data

Eleven mechanistic AOPs currently available in the AOP-Wiki KB were quantified, five being endorsed (AOPs IDs 25, 42, 48, 150, 284). These models were developed using a variety of types of data, including experimental/empirical dose- and time-responses. qAOP models derived empirically utilise expert judgement for their development. For instance, Foran et al. (2019) proposed a modular approach for qAOPs with limited mechanistic data and extensive time required for modelling. The approach focused on making use of the existing data while informing where further tests are needed to provide information for the quantification of all KERs. Conversely, qAOP models derived purely experimentally have the advantage of verifying the feasibility of the development of AOP-based assays, e.g., a battery of *in vitro* tests. For example, to quantify the AOP for developmental neurotoxicity following the inhibition of acetylcholinesterase, Yozzo et al. (2013) studied different levels of biological organisation during zebrafish embryogenesis. Importantly, if data are not available or inappropriate to compute the KER that links the final KE to an AO at an organ level, but a series of KERs can be quantified at cellular and tissue levels, the endpoint can become a KE that occurs earlier than an AO. For example, *in vitro* data were employed by the computational model of Zgheib et al. (2019) that quantified chronic kidney injury in a dose- and time-response manner with the endpoint represented by oxidative stress, a cellular effect. qAOP models derived from a combination of both empirical and experimental data will often predict the outcome better and increase the overall confidence in the applicability of the qAOP model. For instance, Muller et al. (2015) described the impact of engineered nanoparticles on hatching of zebrafish eggs using high-throughput data at different time points. Model performance showing the experimental differences between the data sources has also been evaluated, e.g., Margiotta-Casaluci et al. (2016) investigated *in vivo* fish egg production following exposure to a chemical class of interest at various concentrations. The final model included data from other studies, and the results were compared with human data. At the same time, empirical data are suitable for the optimisation and validation of the predicted response-response relationships as illustrated by Hassan et al. (2017) who optimised the quantification of a classic thyroid hormone (TH) synthesis inhibitor in developmental

neurotoxicity in a rodent model using data from the literature. Likewise, Doering et al. (2018) investigated the activation of the aryl hydrocarbon receptor leading to early life stage mortality and validated the resultant qAOP model with empirical evidence. Integration of *in silico*, *in vitro* and *in vivo* data was employed to model the teratogenicity of single and mixture azole fungicides by Battistoni et al. (2019). Integration of data mining and quantitative assessment was used by Song et al. (2020) to investigate an AOP network that describes the impact of Ultraviolet B (UVB) radiation on crustaceans. At the same time, not all quantified AOPs accounted for both dose- and time-scales. Foran et al. (2019) and Doering et al. (2018) focused primarily on predictions based on the tested concentrations. Importantly, most of the published qAOP models utilised linear AOPs, with the exception of Margiotta-Casaluci et al. (2016) who described chronic exposure to synthetic glucocorticoids leading to perturbation in egg production linking three AOPs in a network: disruption of glucose homeostasis, effects on the immune system and androgenic effects. This integration of evidence shows the complexity of different pathways and their different sensitivities to chemicals.

### 2.6.3. Quantitative approaches

Several quantitative approaches were applied for the development of the existing qAOP models. The qAOPs of Muller et al. (2015), Hassan et al. (2017), and Foran et al. (2019) were quantified using purely mathematical equations. Battistoni et al. (2019) developed a multistage dose-response model applying Bayesian statistical analysis. Besides empirical dose-response, systems biology models were used as a quantitative approach by Battistoni et al. (2019) and Zgheib et al. (2019). Importantly, not all quantified AOPs follow every level of biological organisation. For example, the qAOP formulated by Zgheib et al. (2019) focused on the cellular level due to limited data for the other potential downstream KEs. However, full quantification was undertaken by Muller et al. (2015), Margiotta-Casaluci et al. (2016), Doering et al. (2018), Hassan et al. (2017), Battistoni et al. (2019) who conducted experiments to fill the gaps beyond the available empirical evidence. The qAOP model developed by Conolly et al. (2017) linked multiple models to create a mechanistic qAOP model for aromatase inhibition leading to reproductive dysfunction: a mechanistic hypothalamus-pituitary-gonad model, a vitellogenin liver compartment model, a statistical model relating vitellogenin levels to fecundity and a density-dependent population matrix model. It was later extended from fathead minnow (*Pimephales promelas*) to two other species (female zebrafish (*Danio rerio*) and female Japanese medaka (*Oryzias latipes*)) to broaden the taxonomic domain of applicability and therefore, its potential regulatory applications (Doering et al. 2019). Thus, AOP ID 25 has three associated qAOP models (Conolly et al. 2017; Doering et al. 2019; Foran et al. 2019).

Regarding the mathematical expressions, linear regression was used by Doering et al. (2018) and Foran et al. (2019), while exponential equations were used by Foran et al. (2019) and by Hassan et al. (2017) for the computational prediction of TH disruption on the developing brain in rats as described. Elsewhere, Battistoni et al. (2019) used kinetic equations adapted from a published systems mathematical biology model to simulate the kinetics of single chemicals and mixtures to simulate the perturbation which may lead the co-exposure of chemicals. A systems biology model was also employed by Zgheib et al. (2019) that used over 50

differential equations and, as a result, showed the need for extensive parametrisation (335 parameters). A combination of linear models, kinetic equations and statistical analysis was considered by Muller et al. (2015) in a study of copper nanoparticles. The qAOP models of Margiotta-Casaluci et al. (2016) and Yozzo et al. (2013) applied statistical analysis, i.e., one-way analysis of variance (ANOVA) to the experiments conducted to evaluate the pathway of interest quantitatively. Quantitative KERs of the model of Song et al. (2020) were developed based on the experimental data using the maximum likelihood estimation (MLE) functions in the Benchmark Dose Analysis (BMDS) Software of the US EPA. The experimental data were fitted to five types of frequentist inference models, Exponential, Hill, Linear, Polynomial and Power, selected based on a combination of visual inspection of model fit, goodness-of-fit and Akaike information criterion (AIC).

### 2.6.4. Regulatory applicability

The regulatory applicability in regard to a potential CRA has been analysed following the suggestions of the authors of the qAOP models. All qAOPs have applications in ecological risk assessment, while the qAOP model developed by Foran et al. (2019) is intended for screening and/or prioritisation purposes and that developed by Zgheib et al. (2019) is proposed for human health risk assessment. The qAOP of Conolly et al. (2017) showed additional potential applications: comparing the qAOP simulations to empirical data, how a response-response function can be derived and how to estimate the BMD for an untested chemical using toxicity equivalent factor (TEF).

### 2.6.5. Additional considerations

The adjacency and non-adjacency of KERs were considered by Hassan et al. (2017), Doering et al. (2018) and Foran et al. (2019). Hassan et al. (2017) developed the non-adjacent KER using literature data to model the gaps. Doering et al. (2018) used non-adjacent KERs to check and verify the linkage between KEs and the AO. Foran et al. (2019) proposed a modular approach as a feasible solution to the AOPs lacking empirical dose- and time-response data. Zgheib et al. (2019) used a mathematical inversion by reversing the exponential equation based on the initially given function for a pair of KEs (an upstream KE leading to a downstream KE) to describe dose-time-response relationships ensuring the qAOP model was not chemical-specific. Four qAOPs also incorporated kinetics: Battistoni et al. (2019); Hassan et al. (2017); Margiotta-Casaluci et al. (2016); Muller et al. (2015). Furthermore, Battistoni et al. (2019) included a modulating factor, i.e., identifying that ethanol can also inhibit retinoic acid synthesis, and a negative feedback loop, i.e., regulation of retinoic acid resulting from increased synthesis of CYP26A1. Doering et al. (2019); Doering et al. (2018) developed a qAOP that is applicable across species. Uncertainty of the model was considered by Hassan et al. (2017), Doering et al. (2018), Battistoni et al. (2019) and Foran et al. (2019). Sensitivity analysis was performed by Margiotta-Casaluci et al. (2016) and Zgheib et al. (2019). The mathematical equations and/or the code of the qAOP models of Hassan et al. (2017), Doering et al. (2018), Zgheib et al. (2019), Muller et al. (2015) are available in the supplementary information of the publications.

Table 2.5. Characterisation of the existing eleven mechanistic qAOPs.

| Model purpose | Adverse outcome | Mechanistic knowledge and associated data | | | | | | | | Quantitative approach | Regulatory applicability | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OECD AOP-Wiki KB[1] | Type of AOP[2] | Single chemical(s)/ mixtures | Data type | Adjacent KERs | Biological level(s) | D/C-R[3] | T-R[4] | | | |
| Association of MIE to AO at higher level of biological organisations | Increased frequency of spontaneous tail contractions | No | LAOP | Single chemical | *In vivo* experimental data | √ | Molecular, tissue, organ | √ | √ | Statistical analysis | Ecological risk assessment | Yozzo et al. (2013) |
| Mechanism of CuO engineered nanoparticles toxicity | Mortality | No | LAOP | Nanoparticles | *In vitro* experimental data | √ | Molecular, cellular, organ, organism | √ | √ | Linear regression, one-compartment toxicokinetic model | Ecological risk assessment | Muller et al. (2015) |
| Development of a qAOP network | Egg production | No | AOPN | Single chemical | *In vitro* and *in vivo* experimental data | √ | Molecular, cellular, tissue, organ, individual | √ | √ | Statistical analysis | Ecological risk assessment | Margiotta-Casaluci et al. (2016) |
| Development of a qAOP and potential applications | Population declining trajectory (reproductive dysfunction) | AOP ID 25 | LAOP | Single chemical | Empirical data | √ | Molecular, cellular, tissue, organ, individual, population | √ | √ | A mechanistic model, a compartment model, a statistical model, a density-dependent population matrix model | Ecological risk assessment | Conolly et al. (2017) |
| Development of a qAOP on developmental neurotoxicity | Brain malformation | AOP ID 42 | LAOP | Single chemical | *In vivo* experimental data | √* | Molecular, cellular, tissue, organ | √ | √ | Mathematical equations (exponential regression) | Human risk assessment | Hassan et al. (2017) |

*To be continued*

Table 2.5. Continued

| Model purpose | Adverse outcome | Mechanistic knowledge and associated data | | | | | | | | Quantitative approach | Regulatory applicability | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OECD AOP-Wiki KB[1] | Type of AOP[2] | Single chemical(s)/ mixtures | Data type | Adjacent KERs | Biological level(s) | D/C-R[3] | T-R[4] | | | |
| Development of a cross-species qAOP | Mortality increase, population declining trajectory | AOP ID 150 | LAOP | Mixtures | *In vitro* experimental data on COS-7 cells | √* | Molecular, organism, population | √ | - | Linear regression, statistical analysis | Ecological risk assessment | Doering et al. (2018) |
| Simulation of the mechanism of toxicity | Abnormalities at facial primordia branchial arches | No | LAOP | Single chemicals | *In vitro* experimental data, *in vivo* and *in silico* data | √ | Molecular, cellular, tissue, organ | √ | √ | Multistage dose-response model, Bayesian analysis | Ecological risk assessment | Battistoni et al. (2019) |
| Define the taxonomic domain of applicability of an existing qAOP | Decreased fecundity | AOP ID 25 | LAOP | Single chemical | *In vivo* experimental data | √ | Cellular, tissue, organ, individual | √ | √ | Regression, statistical analysis | Ecological risk assessment | Doering et al. (2019) |
| Quantification of KERs with available data in a modular manner | Decrease in population; Impairment of memory and learning | AOPs IDs 25 and 48 | LAOP | Single chemicals | Empirical data | √* | | √ | - | Linear regression (response-response function) | Screening or prioritisation | Foran et al. (2019) |
| Toxicity pathway assembly using data mining and quantification of KERs | Increased mortality | AOPs IDs 327-330 | AOPN | Ultraviolet B (UVB) radiation | *In vitro* experimental data on *Daphnia magna* | √ | Cellular, tissue, organ, individual | √ | √ | Maximum Likelihood Estimation (MLE) | Ecological risk assessment | Song et al. (2020) |

*To be continued*

Table 2.5. Continued

| Model purpose | Adverse outcome | Mechanistic knowledge and associated data | | | | | | | | Quantitative approach | Regulatory applicability | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OECD AOP-Wiki KB[1] | Type of AOP[2] | Single chemical(s)/ mixtures | Data type | Adjacent KERs | Biological level(s) | D/C-R[3] | T-R[4] | | | |
| Comparison between probabilistic and mechanistic approaches | Nephron attrition leading to chronic kidney disease | AOP ID 284 | LAOP | Single chemicals | *In vitro* experimental data on human RPTEC/TERT1 cells, AOP construction | √ | Molecular, cellular, tissue, organ | √ | √ | Empirical dose-response model, systems biology model | Human health risk assessment | Zgheib et al. (2019)[5] |

[1]Numbers represent the indices (XXX) of the AOP in the AOP-Wiki KB available at https://aopwiki.org/events/XXX.
[2]Linear AOP (LAOP), AOP Network (AOPN).
[3]Dose/Concentration-Responses (D/C-R).
[4]Time-Responses (T-R).
[5]Model represents a combination of both probabilistic and mechanistic approaches.
*Non-adjacent KERs were modelled as well.

Table 2.6. Characterisation of the available qAOP models based on the additional considerations listed in Table 2.1.

| Reference | Cross species extrapolation | Modulating factors | Feedback loops | Compensatory mechanisms | Test method adopted/ validated | Kinetics | Chemical agnostic | Exposure assessment | Uncertainty evaluated | Sensitivity analysis | Publicly available |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Probabilistic qAOP models** | | | | | | | | | | | |
| **Chu (2018)** | - | √ | - | - | - | - | - | √ | √ | √ | √ |
| **Jannicke Moe et al. (2020b)** | - | - | - | - | √[1] | - | - | - | √ | √ | √ |
| **Jeong et al. (2018)** | - | - | - | - | - | √ | - | - | - | √ | √ |
| **Burgoon et al. (2020); Perkins et al. (2019a)** | - | - | - | - | - | - | - | - | - | √ | √ |
| **Rowland et al. (2019)** | - | - | - | - | - | - | - | - | √ | √ | - |
| **Zgheib et al. (2019)** | - | - | - | - | - | - | √ | - | √ | √ | √ |
| **Mechanistic qAOP models** | | | | | | | | | | | |
| **Yozzo et al. (2013)** | - | - | - | - | - | - | - | - | - | - | - |
| **Muller et al. (2015)** | - | - | - | - | - | √ | - | - | - | - | √ |
| **Margiotta-Casaluci et al. (2016)** | - | - | √ | - | - | √ | - | - | - | √ | - |
| **Conolly et al. (2017)** | - | - | √ | √ | - | √ | √ | - | √ | - | - |
| **Hassan et al. (2017)** | - | - | √ | √ | - | √ | - | √ | √ | - | √ |
| **Doering et al. (2018)** | √ | - | - | - | - | - | - | - | √ | - | √ |
| **Battistoni et al. (2019)** | - | √ | √ | - | - | - | - | √ | √ | - | - |
| **Doering et al. (2019)** | √ | - | - | - | - | √ | √ | - | √ | √ | - |
| **Foran et al. (2019)** | - | - | - | - | - | - | - | - | - | - | - |
| **Song et al. (2020)** | √ | - | - | - | - | - | NA | √ | √ | - | √[2] |
| **Zgheib et al. (2019)** | - | - | - | - | - | - | √ | - | √ | √ | √ |

[1]The *in vitro* measurements were conducted on a plant recognised by the OECD TGs for toxicity testing of the endpoint.
[2]https://www.niva.no/en/projectweb/radb, accessed on March 16, 2021.

## 2.7. Software tools

A variety of software tools used for the development of the qAOPs analysed were identified in this study (Table 2.7). In total, 21 tools were distinguished, with twelve of them being publicly available. The range of software tools can be classified into tools used for (i) data analysis, (ii) modelling, simulation and calibration and (iii) model storage. The most common tools used were Microsoft Excel, the *drc* R package for writing the mathematical functions of dose-responses, MC Sim for statistical analysis and BayesiaLab for probabilistic modelling. A unique tool is the Bayesian Inference for Substance and Chemical Toxicity (BISCT) software developed specifically to help to predict quantitative estimates based on the toxicological evidence. Another important tool used is Effectopedia, an open platform that allows qAOP models to be stored in a central location. This compilation of software shows the vast potential in the development of appropriate tools to help advance and apply the qAOP concept.

Table 2.7. List of software used for modelling probabilistic and/or mechanistic qAOPs.

| Tool Name | Functionality | URL[1] | Use Rights | Example of qAOPs |
|---|---|---|---|---|
| **BayesiaLab** | Model generation, analysis, simulation, and optimisation | https://www.bayesialab.com/ | Restricted by licence | Carriger et al. (2016); Jaworska et al. (2015) |
| **BISCT** | Prediction of an adverse event likely to occur given the evidence | https://github.com/DataSciBurgoon/bisct | Open access | Perkins et al. (2019a) |
| ***bootstrap* R package** | Bootstrap, cross-validation, jackknife | https://cran.r-project.org/web/packages/bootstrap/ | Open access | Jeong et al. (2018) |
| ***brms R package*** | Bayesian generalised nonlinear multivariate multilevel models | https://cran.r-project.org/web/packages/brms/ | Open access | Jannicke Moe et al. (2020b) |
| ***drc* R package** | Analysis of dose-response data | https://cran.r-project.org/web/packages/drc/index.html | Open access | Chu (2018); Jannicke Moe et al. (2020b) |
| **Effectopedia[2]** | Storage of a qAOP model | https://sourceforge.net/p/effectopedia/wiki/Home/ | Open access | Zgheib et al. (2019) |
| **GraphPad Prism** | Analysis and plotting the data | https://www.graphpad.com/scientific-software/prism/ | Restricted by licence | Doering et al. (2018) |
| **Lmtest** | A collection of tests, data sets, and examples for diagnostic checking in linear regression models | https://cran.r-project.org/web/packages/lmtest/index.html | Open access | Chu (2018) |
| **Matlab** | Analysis and design processes | https://www.mathworks.com/products/matlab.html | Restricted by licence | Hassan et al. (2017) |
| **MC Sim** | Bayesian statistical inference | https://www.gnu.org/software/mcsim/ | Open access | Battistoni et al. (2019); Hack et al. (2010); Zgheib et al. (2019) |
| **MC Stan** | Model generation, simulation, calibration | https://mc-stan.org/ | Open access | Zgheib et al. (2019) |

*To be continued*

[1]Last accessed on March 16, 2021.
[2]Planned for release at: https://effectopedia.org/.

Table 2.7. Continued

| Tool Name | Functionality | URL[3] | Use Rights | Example of qAOPs |
|---|---|---|---|---|
| **Microsoft Excel** | Statistical analysis | https://products.office.com/en-us/excel | Restricted by licence | Foran et al. (2019); Hack et al. (2010); Hassan et al. (2017) |
| **Netica** | Model construction, optimisation | https://www.norsys.com/download.html | Restricted by licence | Chu (2018); Jannicke Moe et al. (2020b) |
| ***PerformanceAnalytics R package*** | Correlation analysis | https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html | Open access | Jeong et al. (2018) |
| **Samiam** | BN modelling and reasoning | http://reasoning.cs.ucla.edu/samiam/ | Open access | Jeong et al. (2018) |
| **SAS software** | Statistical analyses | https://www.sas.com/en_us/software/sas9.html | Restricted by licence | Hassan et al. (2017) |
| ***SigmaPlot*** | Graphs plotting | http://sigmaplot.co.uk/products/sigmaplot/ | Restricted by licence | Jeong et al. (2018) |
| **SigmaStat software** | Statistical analyses | https://systatsoftware.com/products/sigmastat/ | Restricted by licence | Margiotta-Casaluci et al. (2016) |
| **SPSS** | Statistical analysis | https://www.ibm.com/uk-en/products/spss-statistics | Restricted by licence | Jeong et al. (2018); Yozzo et al. (2013) |
| **WEKA** | Collection of machine learning algorithms for data mining tasks | https://www.cs.waikato.ac.nz/~ml/weka/index.html | Open access | Furxhi et al. (2019) |
| **US EPA Benchmark Dose (BMDS) Software** | Analysis of dichotomous (quantal) data, continuous data, nested developmental toxicology data, and multiple tumour analysis | https://www.epa.gov/bmds/benchmark-dose-software-bmds-version-311-download | Open access | Furxhi et al. (2019); Song et al. (2020) |

[3]Last accessed on March 16, 2021.

## 2.8. Conclusions

This Chapter has summarised the recent progress made in the development of qAOP models. Additionally, a list of common features essential for developing qAOP models has been outlined, i.e., problem formulation, mechanistic knowledge and associated data, quantitative approaches, regulatory applicability and additional considerations derived from published definitions in the scientific literature. These features still need to be harmonised and gain general acceptance. Hence, following the conceptual classes of qAOP models as proposed by Gust et al. (2016) and Perkins et al. (2019b), existing probabilistic and mechanistic qAOPs were identified and characterised according to the predefined common features. Additionally, the semi-q/qWoE AOPs were reviewed for their applicability to increase confidence in qualitative AOPs besides the assessment of AOPs required by the OECD. The qAOPs discussed herein illustrate a range of computational techniques and software tools applicable to such modelling. Importantly, these examples highlight the powerful capability of a qAOP model to integrate diverse types of data (physico-chemical, *in silico*, *in vitro*, *in vivo*) for different stressors including chemicals, nanoparticles and radiations.

As mentioned above, there is currently no regulatory guidance on how to develop and evaluate a qAOP model. As more examples of qAOPs become available, there will be an increasing need to provide a coherent framework to support the evaluation and purpose-specific application of qAOPs in a regulatory context. While it is beyond the scope of this Chapter to outline such a framework, a number of elements (principles) can be identified, some of which may be essential, and others desirable, depending on the application.

On the basis of what has been determined in this Chapter, an ideal qAOP should:

- Predict a defined AO (defined endpoint).
- Address a specified regulatory question and context of use (problem formulation).
- Be consistent with the qualitative description of the AOP of interest.
- Have a clear domain of applicability (including species, taxa, modulating factors).
- Be characterised in terms of its predictive performance and robustness (uncertainty and sensitivity analysis).
- Be transparent and traceable, to allow independent evaluation and verification of the qAOP model (including input data, simulated outputs, and correct implementation of the mathematical equations).
- Be understandable and user-friendly, to ease its interpretability and application.
- Be flexible, to allow the analysis of both existing and new molecules.
- Be updateable, to readily incorporate new data from diverse sources once it becomes available.
- Be reproducible, to enhance the confidence in the consistency and accuracy of the qAOP model output.
- Be portable, so that the qAOP model can be integrated with other mathematical models, such as kinetic models.

- Be publicly available, either in the form of a working platform, or availability of code.

Although current efforts in qAOP modelling are limited, the field is gaining momentum. This Chapter can therefore serve as a starting point to formulate formal guidance on the development, assessment and application of probabilistic and mechanistic qAOPs following, or not, the assessment of qualitative AOPs informed by semi-q/qWoE qAOPs together with the OECD assessment in CRA. In the context of the present thesis, this points the way for future research to investigate networks of linear AOPs and the usefulness of probabilistic modelling to evaluate pathway-driven and chemically-induced toxicological effects.

# Chapter 3. Development and analysis of an Adverse Outcome Pathway network for human neurotoxicity

As Chapter 2 showed, a network of linear AOPs can depict the complexity of biological effects better than single linear AOPs. Notably, a network can serve as the grounds for quantitative modelling of chemically-induced adverse events. The work presented in this Chapter is based on a methodology published as part of the doctoral work (Spinu et al. 2019). The Chapter was enriched with details regarding the topology parameters and a dedicated description on the potential use of the ToxCast™ data for modelling quantitative KERs of the AOs of the AOP network for neurotoxicity.


## Abstract

An AOP network is an attempt to represent the complexity of systems toxicology. This Chapter illustrates how an AOP network can be derived and analysed in terms of its topological features to guide research and support regulatory assessment. A four-step workflow describing general and applied design principles was established and implemented. An AOP network linking nine linear AOPs describing neurotoxicity was mapped and made available in AOPXplorer. The resultant AOP network was modelled and analysed in terms of its topological features, including level of degree, eccentricity and betweenness centrality. Several well-connected KEs were identified, and cell injury/death was established as the most hyperlinked KE across the network. The derived network expands the utility of linear AOPs to better understand signalling pathways involved in developmental and adult/ageing neurotoxicity. The results provide a solid basis to guide the development of *in vitro* test method batteries, as well as further quantitative modelling of KEs and KERs in the AOP network, with an eventual aim to support hazard characterisation and CRA.

## 3.1. Introduction

The science of networks is defined as the collection, management, analysis, interpretation and presentation of relational data (Brandes et al. 2013). The investigation of networks is spread widely throughout all branches of biology and chemistry, from neurobiology (Bassett and Sporns 2017) to genomics (Li et al. 2017). For example, in biology, the application of networks has made advances towards uncovering the organising principles of various complex systems, e.g., protein-protein interactions, metabolomics, signalling and transcription-regulatory networks (Barabasi and Oltvai 2004). On the other hand, systems toxicology, considered as an application of systems biology, aims to describe the perturbation by toxicants and the resilience of the essential defence and adaptive mechanisms across multiple levels of biological organisations (Hartung et al. 2017; Sturla et al. 2014). In other words, systems toxicology helps to identify meaningful disease-specific biomarkers as opposed to systems biology, where the purpose is to discover the underlying molecular and cellular mechanisms (Aguayo-Orozco et al. 2019b). Systems biology captures interactions between biological entities, while systems toxicology focuses on the temporal/spatial relationships between processes/events, triggered by exposure to a stressor(s), particularly chemicals.

Although AOPs are linear constructs and thus a simplification of complex physiological and toxicological processes (Vinken et al. 2017), it is well appreciated that AOPs are interconnected and potentially share the same processes or KEs (Knapen et al. 2018). As such, network science provides an appealing framework to better represent the complexity of biological processes by studying relationships among interconnected linear AOPs. The term "AOP network" can be defined as a set of individual AOPs sharing at least one common element represented by a KE, including an MIE and an AO (Villeneuve et al. 2014a). Different AOPs diverging from a single MIE, or converging to a single AO, also form AOP networks, even if they do not have any other KE in common (Knapen et al. 2018). An individual AOP can be considered as a pragmatic unit of development and evaluation, while an AOP network can be seen as the functional unit of prediction (Villeneuve et al. 2014a; Villeneuve et al. 2014b). Hence, an individual AOP should be treated as a building block within a larger AOP network that more comprehensively describes the biological processes involved in real-world scenarios. This does, however, imply that it will become increasingly important to move away from viewing single linear AOPs in isolation and to consider instead non-linear and branched AOPs within the broader context of AOP networks, as acknowledged in recent guidance (OECD 2017). The challenge is to integrate individual AOPs into a network for a predefined application and to characterise the network in quantifiable terms.

The OECD AOP project, especially the AOP-Wiki KB module[1], brings together the scientific community to develop, share and discuss AOP-related knowledge while accelerating and facilitating AOP development in a central location, allowing the connectivity of AOPs to be explored (Villeneuve et al. 2014a). By developing an AOP network, all the possible AOPs in the AOP-Wiki KB that are relevant to the specific question may be examined. As AOPs are living documents (Villeneuve et al. 2014a), capable of accommodating updates to the

[1] https://aopwiki.org/, accessed on March 16, 2021.

description of KERs and the addition of new KEs, AOP networks should also be regarded as living documents. The AOP-Wiki KB module is designed to automatically generate AOP networks through the identification of common KEs involved in multiple AOPs. This allows information that has been curated for one AOP to be reused in another, avoiding duplication of effort. Recently, a "global" AOP network was developed in order to evaluate the overall connectivity and structural features of existing linear AOPs in the AOP-Wiki KB module (Pollesch et al. 2019). This illustrated the possibility of deriving AOP networks for toxicological applications. However, the development and use of AOP networks are still in its infancy, and further proof-of-concept examples are needed, including approaches for characterising the underlying uncertainties and limitations (Edwards et al. 2016).

To describe and analyse an AOP network, a range of network analytics can be used to identify and investigate specific network properties, such as topological features or interactions between linear AOPs (Knapen et al. 2018). Although the visual examination of the AOP network graph is compelling, the use of techniques from graph theory facilitates the interpretation of a network in terms of its quantitative topological characteristics. To analyse the topology of an AOP network, many metrics can be calculated to describe the overall shape and structure of the network. Several parameters were identified and described by Villeneuve et al. (2018a), such as level of degree (also known as valency, here the number of KERs linked to a KE), betweenness centrality, path occurrence, eccentricity, topological sorting, connectivity, contraction and matching index (defined in section 3.3.4). Using these kinds of metrics helps to identify the upstream or downstream KEs, points of convergence and divergence, positive and negative feedback loops, etc. The topological parameters allow not only the characterisation of AOP networks but also the identification of the most common/highly connected KEs. In the absence of empirical information on the toxicological relevance of individual KEs, it seems reasonable to prioritise the most highly connected KEs for testing and quantification. These topological parameters can be grouped further depending on the question of interest and network size so that not all of them need to be measured at once. For example, an AOP network can be tested for the degree, path occurrence, betweenness centrality and eccentricity in order to identify points of convergence and/or divergence. Thus, a key challenge is how to establish which parameters are most relevant for the specific question and context of use.

Developmental and adult/ageing neurotoxicity are important endpoints in CRA and are emerging fields for method development and use in regulatory decision-making (Bal-Price et al. 2018a; Fritsche et al. 2018). Early life exposures to certain chemicals, such as pesticides, may have long-term adverse health consequences for the developing brain. In addition, adult/ageing neurotoxicity, e.g., Alzheimer's and Parkinson's diseases, pose significant challenges for societies with rapidly ageing populations. Various test systems are used to evaluate the neurotoxicity of a chemical, including cell lines, primary rodent cells, hiPSC-derived mixed neuronal/glial cultures in 2-D and 3-D cultures etc. (Bal-Price et al. 2015; Schmidt et al. 2017). While none of these are currently validated for regulatory use (e.g., as OECD TGs), they provide relevant information for AOP development (Bal-Price et al. 2015). Furthermore, evaluating and mapping available linear AOPs for

neurotoxicity into a network helps to understand the causative linkages between KEs in terms of mechanistic knowledge supported by empirical evidence while identifying knowledge gaps, limitations and opportunities related to pathophysiological pathways involved.

## 3.2. Aim of this chapter

The main objective of this Chapter was to develop an AOP network for human neurotoxicity and characterise the network by using the analytics proposed by Knapen et al. (2018) and Villeneuve et al. (2018a). A workflow to guide scientists interested in the development of AOP networks was formulated and utilised. In addition, a further aim was to analyse the neurotoxicity AOP network to identify the most common/highly connected KEs and KERs as the basis for quantitative modelling.

## 3.3. Methodology

### 3.3.1. Data Set

Linear AOPs from the OECD AOP-Wiki 2.0 were investigated manually to develop the derived AOP network following the criteria described in Section 3.3.3. The following information about the status of individual AOPs was extracted and collected in an Excel spreadsheet available as supplementary information in the GitHub repository: progress through the OECD review and endorsement processes (e.g., under development, endorsement by the Working Party on Hazard Assessment (WPHA)/Working Group of the National Coordinators of the Test Guidelines Programme (WNT), approval of the EAGMST), KE title, KE type (i.e., MIE, KE, AO), KER (i.e., linkage between upstream and downstream KEs), adjacency of the relationship between a pair of KEs, and qualitative WoE. The linear AOPs were collected in December 2018.

### 3.3.2. Stressors

The OECD AOP-Wiki 2.0 was also used to extract the stressors (chemical initiators and/or non-chemical stressors) triggering KEs, including MIEs and AOs of the collected linear AOPs, together with the available unique identifier number used in PubMed (PMIDs) listed in the stressor's description page. The data for the stressors (in this case, all chemicals) were compiled in an Excel spreadsheet included in the GitHub repository. In addition, the Chemical Abstracts Service Registry Number (CAS RN), Simplified Molecular Input Line Entry System (SMILES) strings, and details on the industrial and therapeutic uses were retrieved from the PubChem database[2] to understand the nature of the chemical stressor responsible for the initiation of the linear AOPs. At the same time, the ToxCast™ Dashboard[3] was investigated for relevant data relating to assays that may be associated with the KEs, including MIEs and AOs of the collected linear AOPs.

### 3.3.3. Network construction

The process of developing a so-called "derived" AOP network (i.e., derived from existing AOPs) followed the four steps that are illustrated in Figure 3.1. Initially, the "General Design Principles" were formulated. These principles are intended to be generic in nature and can be applied to any other question of interest. The "Applied Design Principles" are an illustrated version of the General Design Principles followed for the development of an organ-specific AOP network – in Figure 3.1, this is shown for neurotoxicity. The methodology of each part of the four-step process is described below.

[2] https://pubchem.ncbi.nlm.nih.gov/, accessed on March 16, 2021.
[3] https://comptox.epa.gov/dashboard/chemical_lists/TOXCAST, accessed on March 16, 2021.

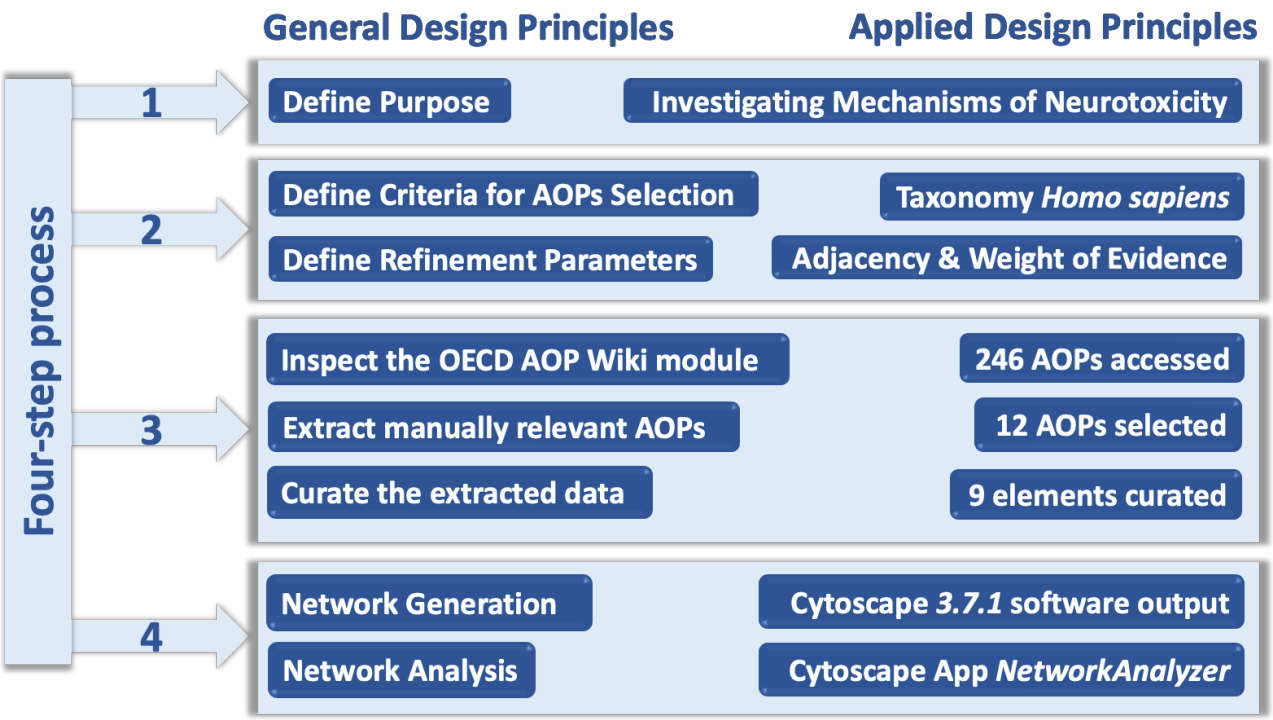


Figure 3.1. The General Design Principles of the four-step workflow for developing a derived AOP network illustrated by Applied Design Principles with regard to human neurotoxicity.

#### *3.3.3.1. Step 1*

Step 1 of the workflow is the definition of the purpose of an AOP network to be modelled. In this investigation, the purpose was to identify the most common/highly connected KEs and KERs in a neurotoxicity AOP network as the basis for quantitative modelling. Accordingly, the scope of the exercise included the linear AOPs known for human neurotoxicity formulated and published in the AOP-Wiki KB module.

#### *3.3.3.2. Step 2*

Step 2 of the workflow is the definition of the criteria for the selection of the AOPs for the development of the network. In this investigation, the criteria included:

- The AOP development stage in terms of the progress of the AOP through the OECD review and endorsement processes,
- The life stage applicability, and
- The taxonomic applicability.

Of these, taxonomy was chosen to be the main criterion for the collection of individual AOPs, i.e., those for human toxicology. The AOP development stage was investigated to evaluate the level of maturity of the AOPs used to derive the AOP network. It gives an indication of uncertainties in the AOP network and shows where further efforts are needed to elucidate the underlying mechanisms.

As one of the uses of an AOP network is quantitative modelling, developing a network with a high level of qualitative and quantitative evidence will give confidence to the model applicability. As such, refinement of the parameters for the initially collected linear AOPs following Step 1 was considered: adjacency and non-

adjacency of relationships between a pair of KEs, i.e., if a KER requires intermediate KEs or not, and the WoE, specifically the qualitative level of understanding for the relationships between a pair of KEs: high, medium or low. Considerations of the BH criteria for WoE assessment were out of the scope of this modelling exercise. This study relied on the assessments performed by the authors of the AOPs, which are summarised in dedicated tables in the AOP-Wiki KB module.

#### *3.3.3.3. Step 3*

Step 3 of the workflow is the identification of appropriate AOPs from the AOP-Wiki KB module. The AOPs identified, according to the criteria in Step 2, were inspected and collected manually in an Excel spreadsheet. The information contained in the AOPs was subsequently curated using the ontology annotations of KEs titles as presented in the supplementary material in the GitHub repository.

#### *3.3.3.4. Step 4*

Step 4 of the workflow is the generation and analysis of the network. *Cytoscape 3.7.1*[4], an open-source software platform, was used to model the AOP network, and *NetworkAnalyzer 3.3.2* App[5] (Assenov et al. 2008), a pre-installed application of the Cytoscape software, was used to analyse the resulting AOP network. The nodes were manually positioned as needed to conserve space and maximise readability. Additional annotation information (e.g., WoE, adjacency and type of KE) was used to further define the visual attributes of the AOP network. The KEs shared by more than one AOP are shown graphically as non-repetitive (i.e., represented by a single arrow), while the duplication of a relationship between a pair of KEs was taken into account when calculating the network analytics.

### 3.3.4. Network analysis

The level of degree, betweenness centrality and eccentricity were chosen to characterise the derived AOP network analytically due to their ability to quantify the position of a KE in relation to its neighbour KEs in the network using Cytoscape *NetworkAnalyzer 3.3.2* App[5]. The level of degree refers to the number of edges linked to nodes of interest (Barabasi and Oltvai 2004). Directed AOP networks can distinguish two types of degrees: an incoming degree ($degree_{in}$) representing the number of links that point to upstream KEs, and an outgoing degree ($degree_{out}$) that denotes the number of links that start from downstream KEs (Villeneuve et al. 2018a). Thus, the level of degree allowed for the identification of points of convergence and divergence and to analyse the overall connectivity of the KEs across the AOP network. The betweenness centrality and eccentricity helped to assess the furthermost upstream and downstream KEs across the AOP network. Mathematically, betweenness centrality is defined as the sum over all pairs of nodes $n$, i.e., KEs, of the network $N$:

$$BC(n) = \sum_{\substack{s \neq n \neq t \\ s,t \in N}} \frac{\sigma_{st}(n)}{\sigma_{st}}$$

[4] https://cytoscape.org/, accessed on March 16, 2021.
[5] http://apps.cytoscape.org/apps/networkanalyzer, accessed on March 16, 2021.

where $\sigma_{st}(n)$ denotes the total number of shortest paths from $s$ to $t$ that pass through the node $n$, i.e., KE, and $\sigma_{st}$ denotes the total number of shortest paths from $s$ to $t$ (Cytoscape 2018).

The eccentricity is a centrality measurement given by the maximum non-infinite length of the shortest path between the nodes $n$, i.e., KEs, that takes the inverse of the maximum of the distances between the nodes $n$:

$$Ecc = \frac{1}{dist_{max}(n)}$$

A node is more central if the maximum of the distances is smaller. If the node is isolated, the value of this parameter becomes zero (Netzwerkerin). KEs with a higher score for the degree and betweenness centrality and a lower score for the eccentricity were considered the most common/highly connected KEs. Comparative analysis of multiple parameters provided the centrality score more efficiently and, therefore, less uncertainty in defining the most common/highly connected KEs.

## 3.4. Results

### 3.4.1. Development of the AOP network for neurotoxicity

Published AOPs in the AOP-Wiki KB were used to develop a derived AOP network for human neurotoxicity. Initially, twelve linear AOPs relevant to human neurotoxicity were identified in accordance with the methodology outlined in Section 3.3. Table 3.1 provides details of the AOPs selected, including their stage of development.

Table 3.1. Twelve linear AOPs for neurotoxicity available in the OECD AOP-Wiki KB module used for modelling the AOP network for human neurotoxicity.

| AOP Wiki ID | AOP Title | Additional Taxonomy | Qualitative Weight of Evidence | OECD Status (Last checked November 2020) |
|---|---|---|---|---|
| 3 | Inhibition of the mitochondrial complex I of nigrostriatal neurons leads to Parkinsonian motor deficits | *Homo sapiens, Rattus norvegicus* | High | **WPHA/WNT Endorsed |
| 10 | Binding to the picrotoxin site of ionotropic GABA receptors leading to epileptic seizures in adult brain | *Homo sapiens, Mus musculus, Rattus norvegicus, Colinus virginianus, Danio rerio* | High | **WPHA/WNT Endorsed |
| 12 | Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development leads to neurodegeneration with impairment in learning and memory in aging | *Homo sapiens,* Monkey, *Rattus norvegicus, Mus musculus, Danio rerio* | Low | **WPHA/WNT Endorsed |
| 13 | Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities | *Homo sapiens, Mus musculus,* Monkey, *Rattus norvegicus* | High | **WPHA/WNT Endorsed |
| 17 | Binding of electrophilic chemicals to SH (thiol)-group of proteins and /or to seleno-proteins during brain development leads to impairment of learning and memory | *Homo sapiens, Rattus norvegicus, Mus musculus* | Moderate | EAGMST Under Review |
| 26 | Calcium-mediated neuronal ROS production and energy imbalance | - | - | Under Development |
| 42 | Inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes in mammals | *Homo sapiens, Rattus norvegicus, Mus musculus* | High | **WPHA/WNT Endorsed |
| 48 | Binding of agonists to ionotropic glutamate receptors in the adult brain causes excitotoxicity that mediates neuronal cell death, contributing to learning and memory impairment. | *Homo sapiens, Rattus norvegicus, Mus musculus* | High | **WPHA/WNT Endorsed |
| 54 | Inhibition of Na+/I- symporter (NIS) leads to learning and memory impairment | *Homo sapiens, Rattus norvegicus* | High | **WPHA/WNT Endorsed |
| 134 | Sodium Iodide Symporter (NIS) Inhibition and subsequent adverse neurodevelopmental outcomes in mammals | *Homo sapiens, Rattus norvegicus* | High | Under Development |
| 152 | Interference with thyroid serum binding protein transthyretin and subsequent adverse human neurodevelopmental toxicity | - | - | *Under Development |
| 260 | CYP2E1 activation and formation of protein adducts leading to neurodegeneration | *Homo sapiens* | - | Under development |

*AOP included in the OECD work plan

**The OECD changed the name from TFHA (Task Force on Hazard Assessment) to WPHA (Working Party on Hazard Assessment

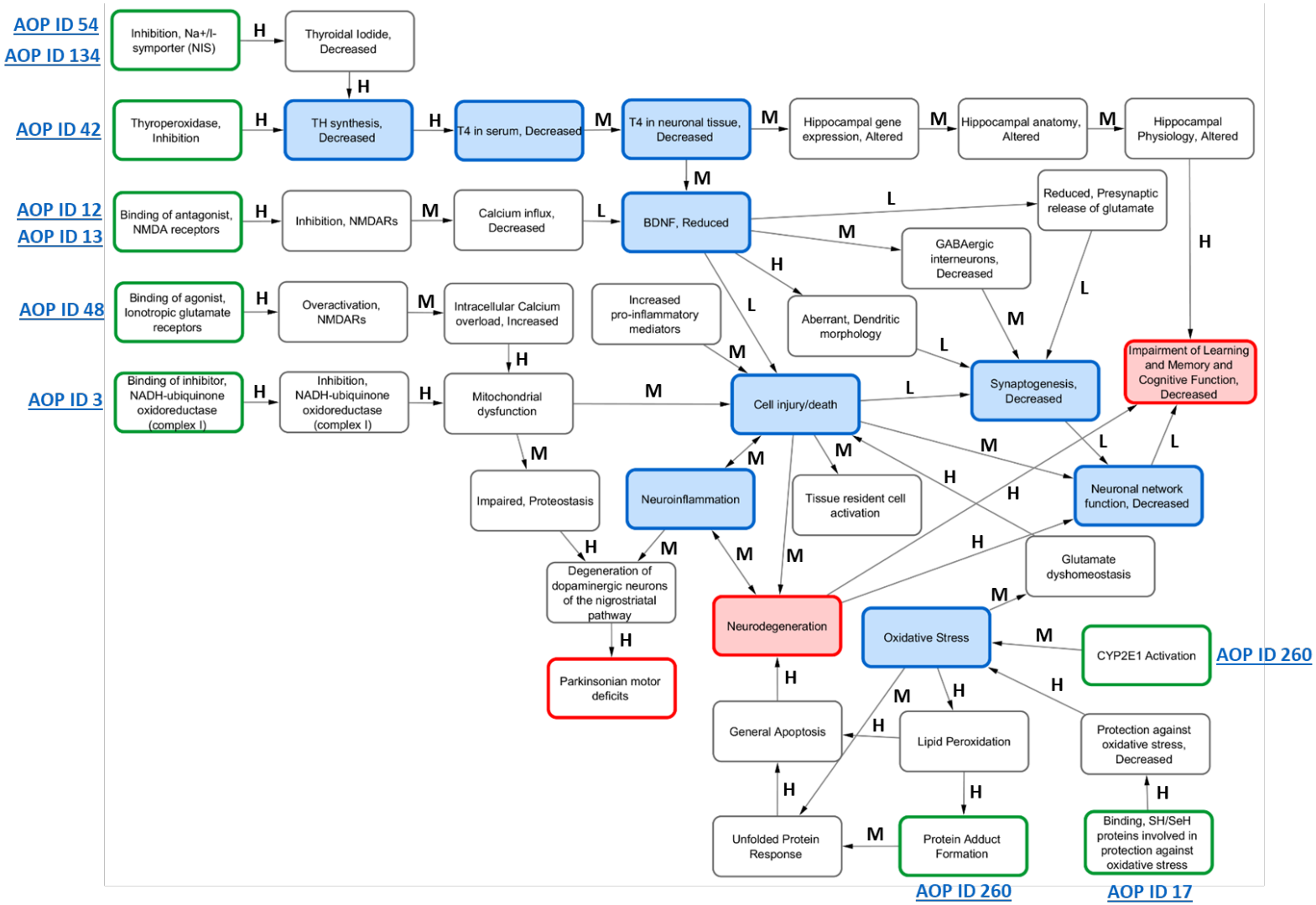


Figure 3.2. Derived network of nine AOPs for neurotoxicity containing adjacent KERs. Green squares indicate an MIE, blue squares indicate the most common/highly connected KEs, red squares indicate an AO, and red squares filled with red colour indicate the most common/highly connected AOs. Solid arrows indicate relationships between KEs that are adjacent. The KEs shared by more than one AOP are shown as non-redundant, i.e., represented by a single arrow. The qualitative WoE between two KEs is annotated as H for high, M for medium and L for the low level of evidence. The KE *cell injury/death* is a common KE across the AOP network, being the most centrally located and most highly connected.

After curation, with the exception of AOPs IDs 10, 26, and 152, all the nine other AOPs were found to share common KEs and were mapped in a network. The developed AOP network is shown graphically in Figure 3.2. The MIE defined as *binding of the antagonist, N-methyl-D-aspartate receptors (NMDARs)* is common to two AOPs (AOP ID 12 and AOP ID 13), and similarly, the MIE *inhibition of Na+/I- symporter (NIS)* linked two other AOPs (AOP ID 54 and AOP ID 134). Three AOs defined as *neurodegeneration*, *Parkinsonian motor deficits* and *impairment of learning and memory/decrease of cognitive function* connected all nine AOPs.

The common KEs across the network are represented by the *reduction of the human brain-derived neurotrophic factor (BDNF)*, *mitochondrial dysfunction*, *oxidative stress*, *neuroinflammation*, *cellular injury/death*, *degeneration of dopaminergic neurons of the nigrostriatal pathway*, *decrease of neuronal network function*, *decrease in the synthesis of the thyroid hormones (TH)*, *decrease in thyroxine (T4) in serum and neuronal tissue*. Interestingly, different upstream KEs contribute to the same common KEs triggering different downstream KEs. For example, *oxidative stress* is initiated by two MIEs: *activation of CYP2E1* and *binding to SH/SeH proteins* of two different AOPs. At the same time, once triggered, oxidative stress leads to several other downstream KEs, such as *dyshomeostasis of glutamate*, *lipid peroxidation* and *unfolded protein response*.

The most centrally located KE across the network is *cell injury/death* triggered by several mechanisms. For instance, *reduced levels of BDNF*, which is widely expressed in the developing and mature central nervous system (CNS), cause aberrations in neuronal morphology and function, including neuronal cell death. Since it is a neurotrophic factor, it plays an essential role in neuronal survival, proliferation, differentiation (synaptogenesis) and maturation (AOP ID 13). Another crucial KE is *neuroinflammation* which triggers *cell injury/death* and *neurodegeneration* through the increased release of different pro-inflammatory mediators from activated microglia and astrocytes exacerbating neurodegeneration which potentiates neuroinflammation (AOP ID 17). Therefore, *cell injury/death* is involved in a feedback loop mechanism of cellular injury/death-neuroinflammation-neurodegeneration. *Impaired proteostasis* through the dysregulation of the ubiquitin-proteasome system (UPS) and the autophagy-lysosome pathway (ALP) increases the accumulation of certain proteins (e.g., α-synuclein), contributing to the degeneration of dopaminergic (DA) neurons of the nigrostriatal pathway that further leads to the motor deficits observed in Parkinson's disease (AOP ID 3). *Cell injury/death* also leads directly to the *decrease of the neuronal network function* implied in the *impairment of learning and memory/decrease of cognitive function* (AOPs IDs 13, 17, 48, 54). *Dyshomeostatis of glutamate* and *mitochondrial dysfunction* are other mechanisms associated with neuronal cell injury/death (see references for each AOP in AOP-Wiki KB).

The AOP network for neurotoxicity relies solely on KERs established between adjacent KEs. The use of the adjacent relationships between KEs shows the biological plausibility of triggering *neurodegeneration* as one of the most common/converging KEs (AOP ID 48) and AOs (AOP ID 12 and AOP ID 260) through different signalling pathways. Depending on brain structure and the sub-type of neurons undergoing neurodegeneration, different AOs can be triggered. Indeed, as illustrated through this network, degeneration

of DA neurons in substantia nigra pars compacta (SNpc) leads to the motor deficit, the AO in Parkinson's disease (AOP ID 3). However, *neurodegeneration* in the hippocampus or cortex leads mainly to *impairment of learning and memory/decrease of cognitive function* (AO of AOPs IDs 12, 13, 17, 42, 48, 54, 134).

The empirical evidence supporting KERs in this AOP network is mainly described in a qualitative or semi-quantitative manner. The WoE supporting KERs varies from low to high; with low WoE possibly based on poor/insufficient empirical data or contradictory information. To increase the empirical evidence supporting the KER, more experiments designed for such a purpose may be required. The development of such networks of biological paths can help to identify and develop appropriate non-animal alternatives for the safety testing of chemicals. This is thoroughly discussed within the thesis. In addition, to improve confidence in an AOP network, for instance, to support regulatory use, e.g., CRA, a better quantitative definition of the thresholds to trigger respective KEs within each KER is needed. In other words, the availability of quantitative KERs should enable an assessment of the likelihood, and under what conditions of chemical concentration and exposure duration, a cascade of KEs triggered by an MIE will lead to an AO.

### 3.4.2. Analytical characterisation of the AOP network for neurotoxicity

The analysis was performed on the derived AOP network for neurotoxicity that contains KEs and adjacent relationships between KEs of nine linear AOPs. As a result, network analytics confirmed that the most hyperlinked KE across the network is *cell injury/death*, followed by *neuroinflammation*, *reduction of BDNF*, *neurodegeneration*, and *decrease in neuronal network function*, with a level of degree of 13, 10, 8, 8 and 8, respectively. The least connected KEs were the MIEs with a level of degree 1, such as *binding of agonist to the ionotropic glutamate receptors*. The AO of *Parkinsonian motor deficit* also has a level of degree of 1, as one linear AOP is currently developed for this AO. The overall connectivity of the KEs is shown in Figure 3.3. However, it is important to outline that the results represent the scenario at the time the linear AOPs were investigated. As more linear AOPs are added, the extent of connection to a particular KE may change.

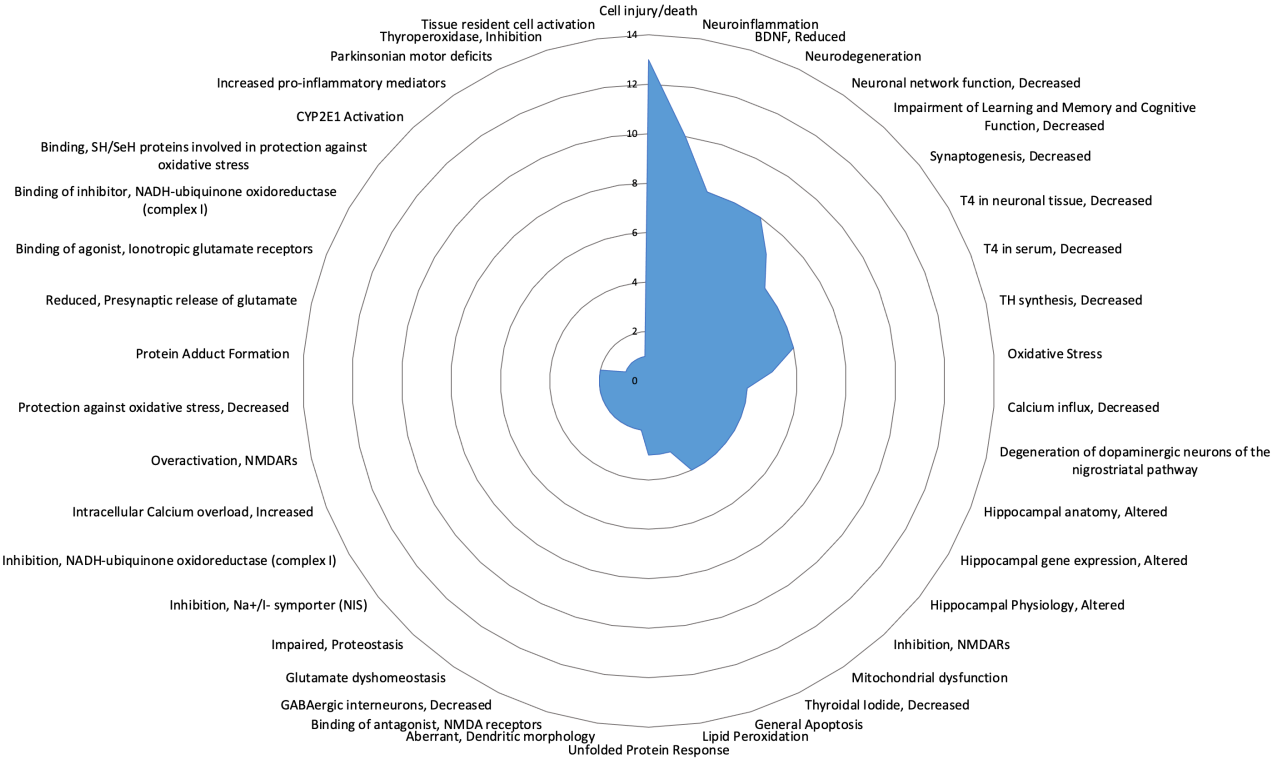


Figure 3.3. The overall connectivity of KEs used to develop the derived AOP network for neurotoxicity. The score indicates the number of the KERs associated with a KE. *Cell injury/death* has the highest score, which means that is the most interconnected KE across the network. The least connected KE was the AO of *Parkinson's motor deficits* due to the fact that only one AOP is currently available for this outcome.

The level of degree of KEs helped to identify points of convergence (common KEs) and divergence across the network, as listed in Table 3.2. Seven convergent KEs and twelve divergent KEs were identified following the score of the $degree_{in}$ and $degree_{out}$. Points of convergence are defined as KEs linked to more upstream than downstream KEs (Villeneuve et al. 2018a), while points of divergence are defined as KEs linked to more downstream than upstream KEs (Villeneuve et al. 2018a). For example, *oxidative stress* is linked to three downstream KEs (*glutamate dyshomeostasis*, *unfolded protein response*, *lipid peroxidation*) and two upstream KEs defined as MIEs (*CYP2E1 activation*, *binding to SH/SeH proteins*). The AO the *impairment of learning and memory/decrease of cognitive function* has the highest number of incoming KERs ($degree_{in}$)

with a score of seven, while the highest number of outgoing KERs ($degree_{out}$) is *cell injury/death* with a score of seven. This indicates that *the impairment of learning and memory/decrease of cognitive function* is a point of high convergence and *cell injury/death* is a point of high divergence. Furthermore, all the MIEs of the network were identified as points of divergence, except the MIE *protein adduct formation* described in the AOP ID 260, which was linked to one upstream and one downstream KEs. *Protein adduct formation* leads to the accumulation of unfolded proteins in the endoplasmic reticulum (downstream KE), but also lipid peroxidation (upstream KE) contributes to the formation of protein adducts through one of its main products 4-hydroxynonenal[1].

Table 3.2. The list of identified seven convergent and 12 divergent KEs for the AOP network for neurotoxicity.

| **Convergent Key Events** | |
|---|---|
| **KE Type** | **KE Title** |
| KE | General apoptosis |
| AO | Impairment of learning and memory/Cognitive function, Decreased |
| KE | Neuroinflammation |
| AO | Parkinsonian motor deficits |
| KE | Synaptogenesis, Decreased |
| KE | Tissue resident cell activation |
| KE | Unfolded protein response |
| **Divergent Key Events** | |
| **KE Type** | **KE Title** |
| KE | BDNF, Reduced |
| MIE | Binding of agonist, Ionotropic glutamate receptors |
| MIE | Binding of antagonist, NMDA receptors |
| MIE | Binding of inhibitor, NADH-ubiquinone oxidoreductase (complex I) |
| MIE | Binding, SH/SeH proteins involved in protection against oxidative stress |
| KE | Cell injury/death |
| MIE | CYP2E1 activation |
| KE | Increased pro-inflammatory mediators |
| MIE | Inhibition, Na+/I-symporter (NIS) |
| KE | Lipid peroxidation |
| KE | Oxidative stress |
| MIE | Thyroperoxidase, Inhibition |

[1] https://aopwiki.org/aops/260, accessed on March 16, 2021.

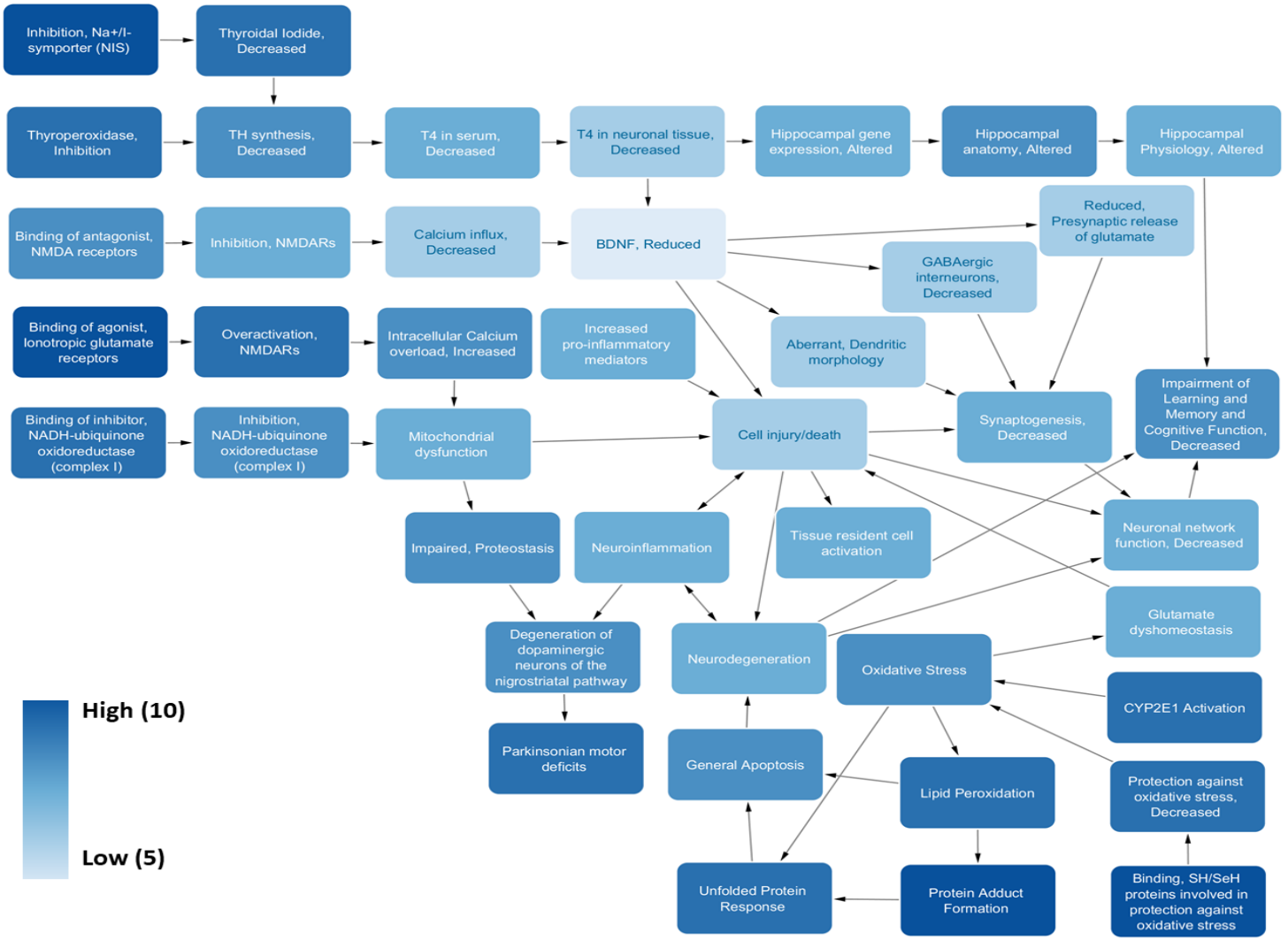


Figure 3.4. The eccentricity of KEs across the AOP neurotoxicity network. The darker the colour, the more upstream is the KE positioned across the network, being less influenced by other KEs and less involved in the network. *Reduction of BNDF* is the lightest in colour, which means that it is the most connected KE. This parameter also confirms that the *reduction of BDNF* can be listed as one of the most common/highly connected KEs. The visual representation of the eccentricity was set from the score of 5 because this was the lowest score calculated by the NetworkAnalyzer App of Cytoscape software and therefore, served as the comparator for differences in scoring metrics for the individual KEs.

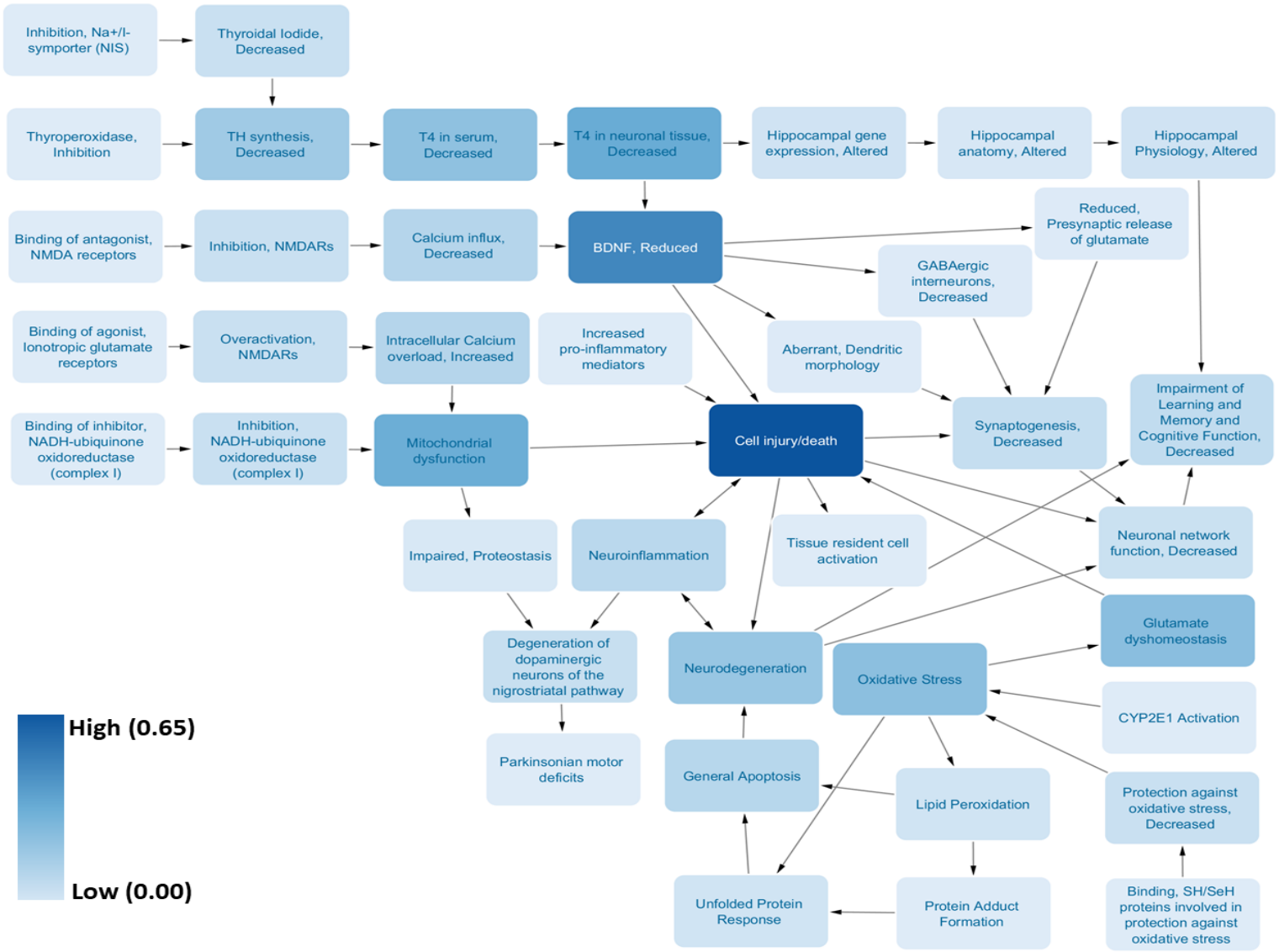


Figure 3.5. The betweenness centrality of the KEs across the AOP neurotoxicity network. The darker the colour, the more centrally located is the KE in comparison with the other KEs. This parameter confirms that *cell injury/death* can be considered as the most common/highly connected KE, having the highest score of 0.65.

Another critical analytical measure is the eccentricity, which is a node centrality index that helps to sort the KEs into upstream and downstream KEs. A low score of eccentricity shows that the KE is more centrally located within the network and can be easily influenced by other KEs with which it is interconnected. The most centrally located KE, according to the eccentricity, is the *reduction of BDNF* with a score of 5. The most upstream KEs, according to their eccentricity, are associated with the two MIEs: *inhibition, Na+/I- symporter (NIS)* and *binding of agonist, Ionotropic glutamate receptors*, with a score of 10 which indicates the maximum distance to the other KEs. These results are represented in Figure 3.4.

Betweenness centrality measures the number of shortest paths between any two KEs in the AOP network that passes through the KE of interest (Villeneuve et al. 2018a). The KE with the highest betweenness centrality score was *cell injury/death,* which means that it is located most centrally within the network and confirms the assumptions made based on the graphical representation. This information complements the results given by the level of degree. These results are represented in Figure 3.5.

The statistical distribution of the number of KEs, in relation to the level of $degree_{in}$ and $degree_{out}$, shows that the majority of the KEs are associated with at least other two KEs, with almost 50% of KEs for both the $degree_{in}$ and the $degree_{out}$ without considering the presence of redundant, i.e., repetitive KERs (Figure 3.6. A and B). The number of shared AOPs by a KE varies between one and seven AOPs (Figure 3.6.C). This has a tremendous impact on the development and analysis of an AOP network, as a network can be modelled once a KE shares at least two linear AOPs. The eccentricity parameter reveals that almost 67% of the KEs are so interconnected that they cannot be categorised as upstream or downstream KEs (Figure 3.6.D).

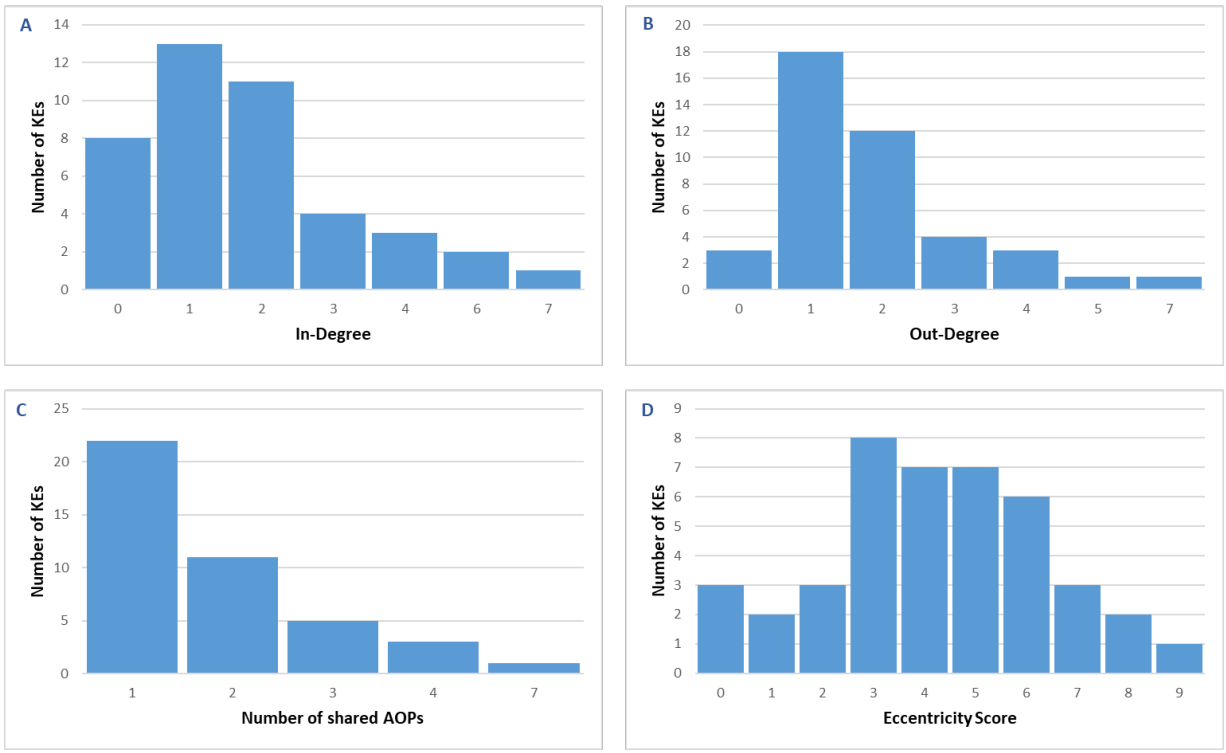


Figure 3.6. The statistical distribution of analytical parameters in relation to the number of KEs in the AOP neurotoxicity network. (A) represents the number of KEs and the associated number of incoming KEs. (B) shows the number of KEs and associated number of outgoing KEs. (C) indicates the distribution of KEs in shared AOPs. (D) shows the distribution of KEs according to the eccentricity score (a measure of how upstream or downstream a KE is).

The common KEs identified based on both the graphical representation and analytics of the AOP network for neurotoxicity could serve as a basis for developing/selecting *in vitro* assays. These *in vitro* test methods could be included in an IATA for evaluating neurotoxicity induced by individual chemicals and mixtures (Bal-Price and Meek 2017). For example, Li et al. (2019) propose an IATA for the assessment of developmental neurotoxicity by selecting a set of assays that can be used to assess common KEs. The work here also supports the common KEs identified by Li et al. (2019) as testing endpoints. Such IATA offers the possibility of addressing different regulatory needs including screening and prioritisation, hazard identification/characterisation or even risk assessment if combined with exposure and absorption, distribution, metabolism, excretion, and toxicity (ADMET) data (Aschner et al. 2017; Bal-Price et al. 2018a; Bal-Price et al. 2018c).

### 3.4.3. Adjacency and non-adjacency in the context of the AOP network for neurotoxicity

One aspect considered while developing the AOP network was the inclusion of solely adjacent relationships between KEs, i.e., pair of KEs connected via a KER and not through intermediate KEs. Relying only on directly connected KEs facilitates the quantitative simulation of the AOP network. The topological difference between AOPs networks consisting of adjacent and both adjacent and non-adjacent relationships were evaluated by comparing the analytical parameters of the two AOP networks. The AOP networks containing both types of interactions are represented graphically in Figure 3.7.

Four out of nine AOPs included in the AOP network contain solely adjacent relationships (AOPs IDs 12, 13, 48, and 260) and five of nine AOPs contain both types of relationships (AOPs IDs 3,17,42, 54, and 134). Several KEs were involved in non-adjacent relationships, including *reduction of BDNF*, *decrease in TH synthesis*, a *decrease of T4 in serum*. *Cell injury/death* and *reduction of BDNF* remain the most connected/common KEs across the network. The AO defined as the *impairment of learning and memory/decrease of cognitive function* is involved in five non-adjacent relationships, besides the other seven adjacent relationships, and therefore, becoming the most connected KE across the network.

Since a non-adjacent relationship is likely to be associated with more biological processes, an AOP network containing both types of relationships implies more connections, representing a higher level of biological complexity. Network analytics show differences in terms of distance and path length. At the same time, a given stressor might trigger all kinds of relationships, and the AOP network serves as a suitable platform for such evaluations. For the initial development of a qAOP it is easier to model adjacent KEs only, rather than include both types. However, this also depends on data availability and the scope of quantitative modelling.

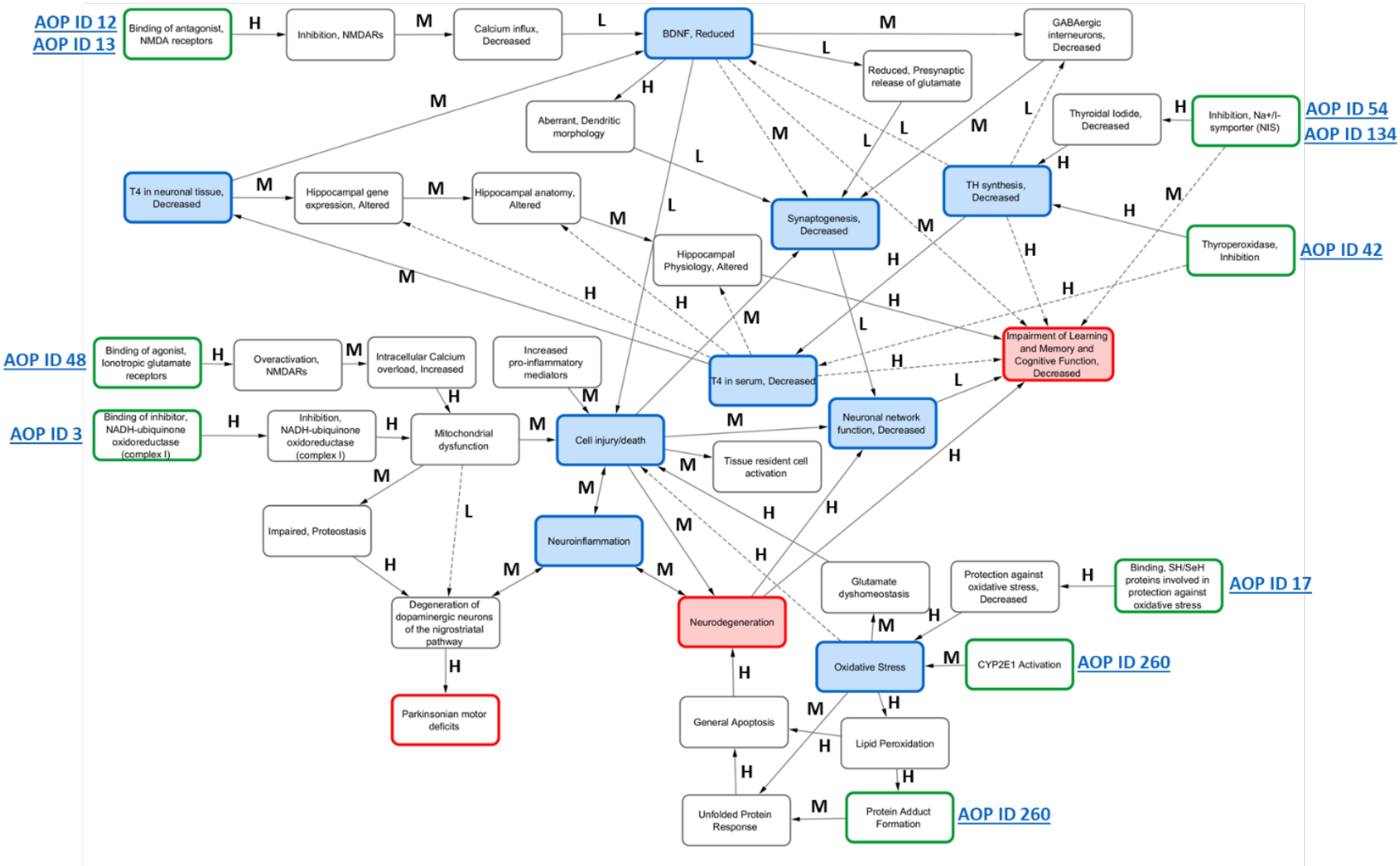


Figure 3.7. Derived AOP network for neurotoxicity containing adjacent and non-adjacent relationships. Green squares indicate an MIE, blue squares indicate the most common/highly connected KEs, red squares indicate an AO, and red squares filled with red colour indicate the most common/highly connected AOs. Solid arrows indicate adjacent relationships between KEs. Dashed arrows indicate non-adjacent relationships between KEs. The qualitative WoE between two KEs is annotated as H for high, M for medium and L for the low level of evidence.

### 3.4.4. Mapping stressors to the AOP network

Chemicals either individually or in combinations (mixtures), as well as other factors such as particles and infectious agents, represent stressors responsible for the initiation of an AOP (OECD 2017). Mapping stressors to linear AOPs in an AOP network allows for the evaluation of interactions between co-occurring stressors. Furthermore, for the purpose of an IATA, it is essential to derive a clear relationship between MIEs and AOs and whether there are interactions between AOPs (OECD 2016). Different (types of) stressors may interact at the MIE or downstream KEs common to multiple AOPs.

The AOP network for neurotoxicity is represented by chemicals as stressors, with no additional types being mentioned in the AOP-Wiki KB module at the time of retrieval. Based on an understanding of the nature of MIEs, *in silico* models can be derived and as a result, inform IATA and read-across. For example, several types of MIEs with associated AOPs have been distinguished and described by Cronin and Richarz (2017), including covalent reactivity, changes in receptor or enzyme activity. The different types of MIEs are identified in the AOP network for neurotoxicity including chronic receptor inhibition (binding of antagonist to NMDA receptors) and activation (binding of agonist to ionotropic glutamate receptors, binding of inhibitor to NADH-ubiquinone oxidoreductase (complex I), binding to SH/SeH proteins involved in protection against oxidative stress, inhibition of thyroperoxidase, inhibition of Na+/I- symporter (NIS)), covalent reactivity (protein adduct formation) and enzyme activation (CYP2E1 activation). These examples could serve as starting points in the development of *in silico* models for neurotoxicity, i.e., (Q)SAR models.

AOP networks are critical for addressing exposures to multiple stressors that lead to the same AO or to individual stressors that perturb multiple MIEs (Knapen et al. 2015; Villeneuve et al. 2018a). For example, the *inhibition of thyroperoxidase* (MIE ID 279) is induced by chemicals with industrial and therapeutic uses, such as antifungal agents (e.g., 2(3H)-benzothiazolethione, mercaptobenzothiazole), antithyroid agents (e.g., thiouracil, propylthiouracil, methimazole), pesticides (e.g., ethylenethiourea), industrial agents (e.g., 4-nonylphenol) and cosmetic ingredients (e.g., resorcinol). On the other hand, acrylamide, with multiple chemical and industrial applications and also being a widely occurring food contaminant from cooking, binds to SH/Seleno proteins, an MIE in the AOP ID 17 that leads to the impairment in learning and memory through neuronal degeneration. Acrylamide also induces protein adduct formation, an MIE in AOP ID 260, that leads to neurodegeneration, an AO in AOPs IDs 12 and 260 and a KE in AOP ID 48. This is because of the electrophilic nature of acrylamide it reacts covalently with nucleophilic sulfhydryl groups on certain proteins that are critically involved in membrane fusion of the nerve terminals (Lopachin 2004; Lopachin and Decaprio 2005).

For qAOP modelling purposes, in its initial phase of development, evidence that a chemical can induce an entire AOP is of great help. However, the AOPs evaluated herein lacked stressors known to be active across all the biological levels of the AOP. This might be due to the fact that no compounds were yet tested for those AOPs or were not tested at high enough concentrations. The only AOP that has chemicals associated with all KEs is AOP ID 42 “Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals”. Therefore, a qAOP could be derived and was modelled by Hassan et al. (2017) for 6-propyl-2-

thiouracil, which is an enzyme inhibitor that is known to trigger AOP ID 42. Information on all chemicals collected for the AOPs used for modelling the AOP network is provided in the supplementary material in the GitHub repository.

The current description of stressors in the AOP-Wiki KB is lacking in detail, so it would be very valuable to include more information such as mechanistic knowledge related to the kinetics, existing QSAR models and read-across predictions, as well as other data sources. This would make the AOP-Wiki KB module not only a repository but also a resource for modelling qAOPs.

One of the critical requirements for *in silico* modelling is the availability of reliable data. A data repository that could be used for *in silico* modelling is the ToxCast™ dashboard[1]. It contains 68 high-throughput screening (HTS) assays conducted on the brain tissue for several endpoints, including oxidative stress, binding to DA or GABAergic neurons, that characterise KEs of a linear AOP. Following the modelled AOP network for neurotoxicity, the list of assays was investigated for its potential utility in developing quantitative KERs (qKERs) that induce any of the AOs. To achieve this objective, the list was narrowed down to assays conducted on rat for the same time point of one hour (Table 3.3).

[1] https://comptox.epa.gov/dashboard/chemical_lists/TOXCAST, accessed on March 16, 2021.

Table 3.3. Eight assays, conducted on brain tissue in rats for 1 hour identified in the ToxCast™ dashboard, shown to be anchored to a key event of the AOP network for neurotoxicity for the potential utility for modelling quantitative KERs of qAOP models. The targets were receptors measured in tissue-based cell free assays.

| AOP(s) ID(s) | Event Type & ID | Event Name | Gene Name | Gene Symbol | ToxCast™ Assay Title | Assay Function | Signal direction | Component Source |
|---|---|---|---|---|---|---|---|---|
| AOP 13, 12 | KE 52 | Decreased, Calcium influx | Calcium channel, voltage-dependent, N type, alpha 1B subunit | Cacna1b | NVS_IC_rCaChN | Binding | Loss | Rat cortical membranes |
| AOP 48 | KE 177 | Oxidative stress | Nitric oxide synthase 1, neuronal | Nos1 | NVS_ENZ_rCNOS_Activator | Enzymatic activity | Gain | Rat brain membranes |
| AOP 13 | MIE 201 | Activation of NMDA | Glutamate receptor, ionotropic, N-methyl D-aspartate 1 | Grin1 | NVS_LGIC_rGluNMDA_Agonist | Binding | Loss | Rat forebrain membranes |
| AOP 54 | KE 851 | Decrease of GABAergic interneurons | Gamma-aminobutyric acid (GABA) A receptor, alpha 1 | Gabra1 | NVS_LGIC_rGABAR_NonSelective | Binding | Loss | Rat whole brain |
| AOP 54 | KE 851 | Decrease of GABAergic interneurons | Gamma-aminobutyric acid (GABA) A receptor, alpha 6 | Gabra6 | NVS_LGIC_rGABARa6 | Binding | Loss | Rat cerebellar membranes |
| AOP 48 | MIE 875 | Binding of agonist, Ionotropic glutamate receptors | Glutamate receptor, ionotropic, AMPA 1 | Gria1 | NVS_LGIC_rAMPA | Binding | Loss | Rat forebrain membranes |
| AOP 48 | MIE 875 | Binding of agonist, Ionotropic glutamate receptors | Glutamate receptor, ionotropic, kainate 1 | Grik1 | NVS_IC_rKAR | Binding | Loss | Rat forebrain membranes |
| AOP 13 | KE 383 | Reduced, Presynaptic release of glutamate | Glutamate receptor, metabotropic 1 | Grm1 | NVS_GPCR_rmMGluR1 | Binding | Loss | Rat cerebellum |

A qAOP model, as a screening and predictive model, should ideally allow evaluating both active (positive) and inactive (negative) compounds. The available data contained unbalanced categories of these two types of compounds, as shown in Figure 3.8. Thus, methodologies that can deal with such a challenge are needed to be developed to improve the predictive power of a qAOP model, as discussed in the subsequent Chapter 4.

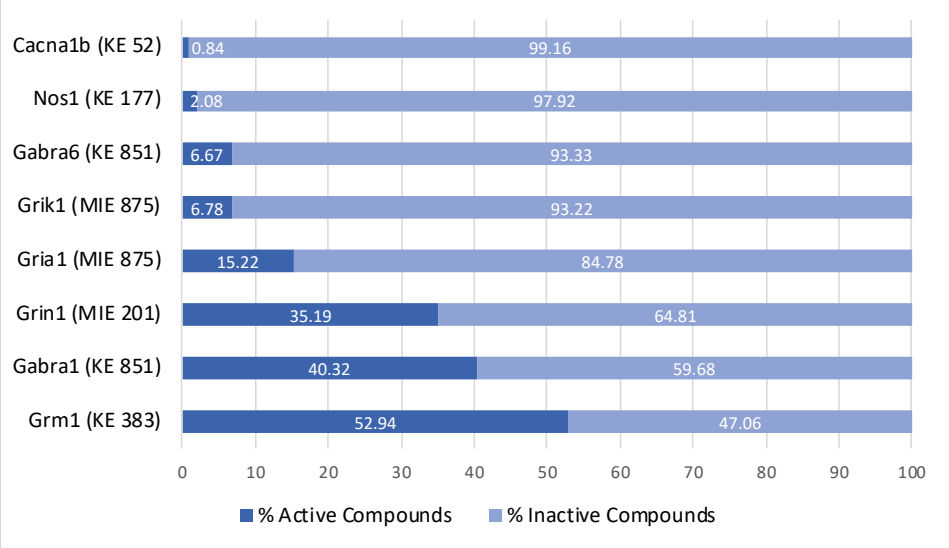


Figure 3.8. The percentage of active vs inactive compounds tested for multiple concentrations for the identified assays in the ToxCast™ dashboard linked to KEs of the AOP network for neurotoxicity.

To model qKERs, multiple concentrations are necessary in order to compute dose-responses as previously described. Most of the compounds were tested at a single concentration (Figure 3.9). Additionally, the identified assays assess binding to receptors and enzymatic activity of compounds. This allows mapping of molecular and cellular effects for neurotoxicity as shown by a recent study (Chappell et al. 2020) that investigated synthetic food colours to alter signalling pathways related to neurodevelopmental processes. Thus, such data are suitable for the discovery of mechanistic pathways at molecular and cellular levels, and efforts are needed for appropriate data generation pipelines for modelling qAOPs.

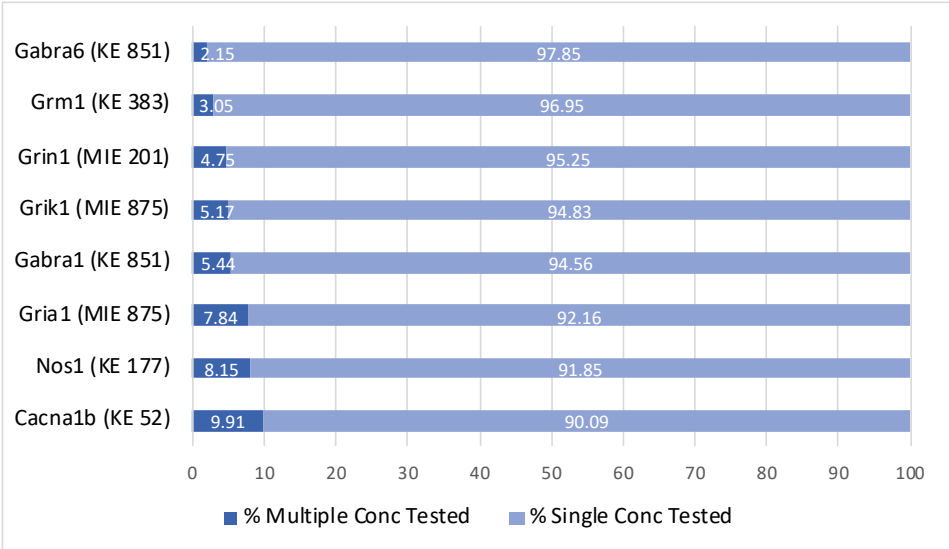


Figure 3.9. The percentage of compounds tested for multiple concentrations vs single concentration for the identified assays in the ToxCast™ dashboard linked to KEs of the AOP network for neurotoxicity.

An overview of the compounds that showed activity for multiple concentrations tested in one of the assays is presented in Figure 3.10. Nine compounds showed activity across at least two assays that can potentially allow development of different kinds of response-response relationships:

- CP-457920 – Two genes of the same KE.
- Didecyldimethylammonium chloride - Two KEs of the same linear AOP.
- Emamectin benzoate – MIE to KE of two different linear AOPs.
- Mancozeb – Two different MIEs of two different linear AOPs.
- Phenylmercuric acetate – MIE to KE of two different linear AOPs.
- Sodium dodecylbenzenesulfonate – MIE to KE of the same linear AOP.
- SSR 241586 HCl – Multiple KERs:
    - MIE to KE of the same linear AOP.
    - MIE to KE of two different linear AOPs.
    - Two KEs of two different linear AOPs.
    - Two MIEs of two different linear AOPs.
- Tributyltin chloride – MIE to KE of two different linear AOPs.
- Tributyltin methacrylate – MIE to KE of two different linear AOPs.

The lack of response in the other assays might suggest that the effects may be non-specific responses. Also, none of KERs can allow the modelling of the relationship of the KEs previous to the AOs of the AOP network for neurotoxicity, instead they model randomly connected KEs. A qKER should be seen as the core element of a qAOP model to predict an AO/apical endpoint of research and regulatory interest while making the qAOP a multiscale model to assess causally linked changes at different biological levels and account for time differences.

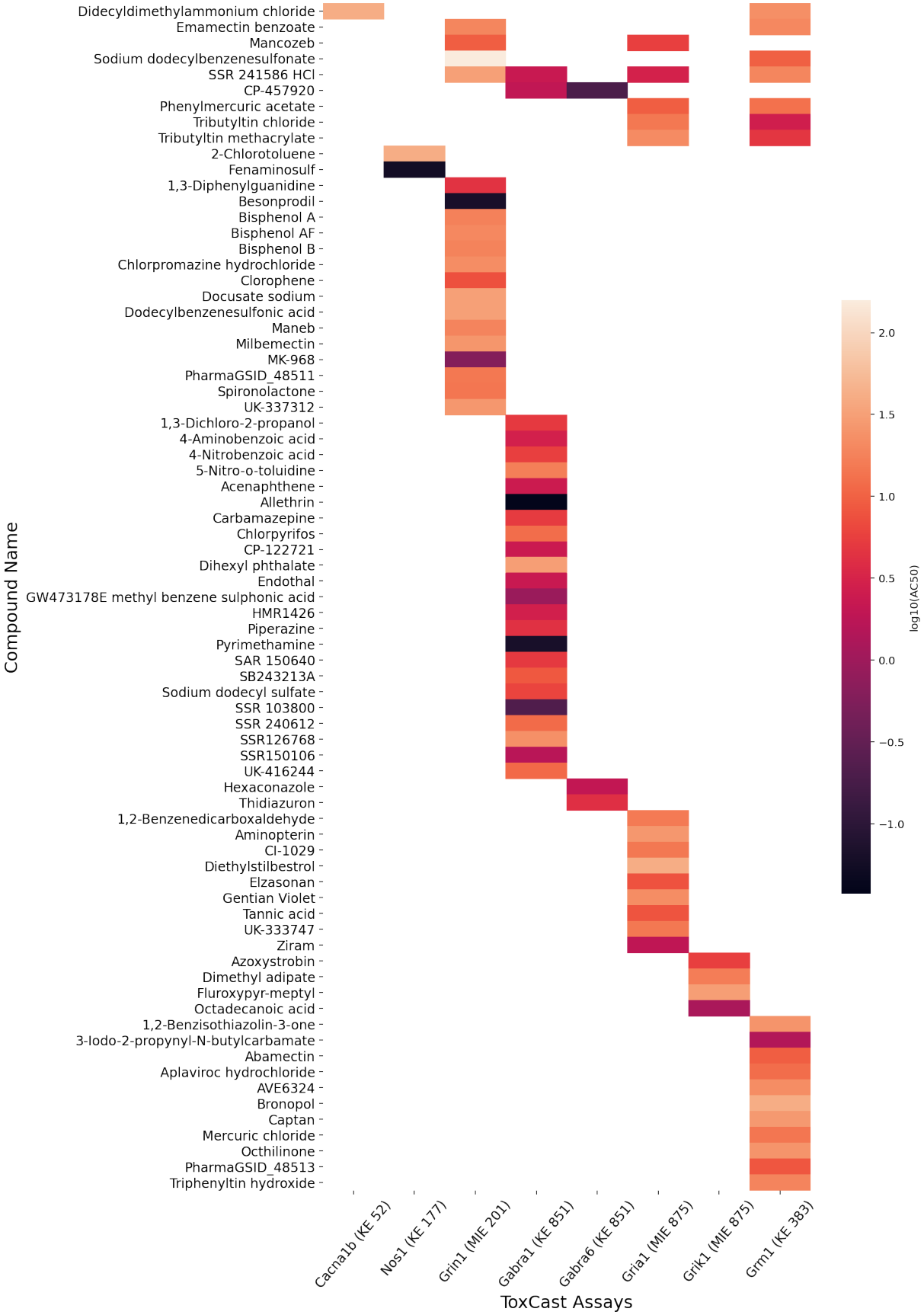


Figure 3.10. An overview of the active compounds ($log_{10}$AC50) across the identified assays in the ToxCast™ dashboard associated with the KEs of the AOP network for neurotoxicity ordered alphabetically. The first nine compounds showed activity across multiple assays.

### 3.4.5. The use of AOPXplorer

The AOPXplorer module[1] of the AOP-Wiki KB was designed to visualise and explore AOP networks for a given AO. It also allows the uploading of additional data, e.g., high-throughput screening, omics and dose-response data that can be used to predict the AO. The AOPXplorer was developed as an App of the Cytoscape software that can be easily downloaded and installed. This facilitates the development of AOP networks while making them living documents. Currently, the AOPXplorer repository contains 18 AOP networks developed for different endpoints, including coagulopathy, ulcer gastric, steatosis, lung fibrosis, skin sensitisation, epilepsy, to name a few. The AOP network modelled herein was also included in the AOPXplorer repository, as shown in Figure 3.11 (Burgoon 2019). This allows further improvement and refinement of the AOP network by the scientific community that can view and further enrich the AOP network, thereby contributing to a better assessment of neurotoxicity.

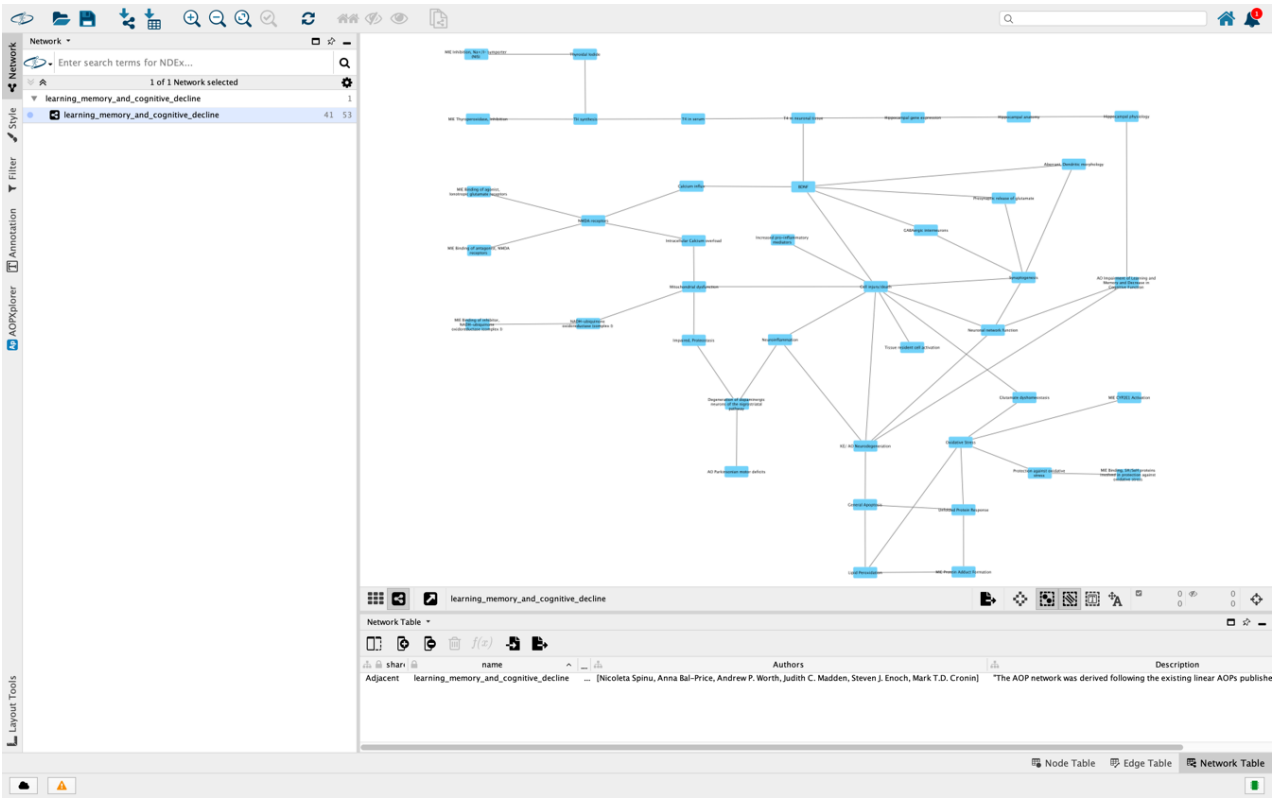


Figure 3.11. The interface of the AOPXplorer tool where the AOP network for neurotoxicity has been stored, which is presented graphically in the right-hand side of the window.

[1] http://apps.cytoscape.org/apps/aopxplorer, accessed on March 16, 2021.

### 3.4.6. Applications of the derived-AOP network

Even though the concept of the AOP network is still in its infancy, there are examples of AOP networks applied to both human and other species toxicological endpoints (Table 3.4). Because of its advantages, such as the use of analytics to characterise the position of a KE (including MIE and AO) within a network, the concept of a derived-AOP network has a plethora of applications in predictive toxicology, which are exemplified below.

Table 3.4. Summary of eight derived-AOP networks currently developed and publicly available*.

| AOP Network Title | Taxonomy | Aim | Reference |
|---|---|---|---|
| **AOP network related to reproductive and developmental toxicity in fish** | Fish | Toxicity assay development | Knapen et al. (2015) |
| **AOP network for disrupted androgen-and insulin-like hormone 3 (INSL-3)-dependent in male rats** | Rodents | Evaluation of chemical mixtures | Howdeshell et al. (2017) |
| **AOP network linking activation of the nicotinic acetylcholine receptor in honeybees to colony death** | Bees | Evaluation of biological plausibility and empirical support to identify knowledge gaps | LaLone et al. (2017b) |
| **AOP network for metabolic disorders mediated by hepatic steatosis** | *Homo sapiens* | Identification of critical paths | Knapen et al. (2018) |
| **Decreased serum thyroid hormone AOP network** | Rodents<br>Amphibians<br>Fish | The use of layers and the identification of points of convergence and/or divergence | Knapen et al. (2018) |
| **Hub KEs for inflammation-related AOP network** | Not specified | Connection of AOPs that previously had no shared KEs | Villeneuve et al. (2018b) |
| **Cytochrome P45019 [CYP19]-AOP network** | Not specified | Linking all possible AOPs to an AOP | Villeneuve et al. (2018a) |
| **Thyroxine [T4]-AOP network** | Not specified | Linking all possible AOPs to a biological process | Villeneuve et al. (2018a) |

*The review was conducted before July 2019.

## 3.5. Discussion

More generally, AOP networks strengthen the utility of the OECD AOP-Wiki KB by increasing scientific confidence in the application of individual AOPs, facilitating a better understanding of their roles as individual blocks in the network of complex biological interactions. Network analytics can be utilised to analyse multiple perturbations and complex interactions across the biological and time scales of interconnected AOPs. As additional mechanistic details enrich the existing AOPs, it is envisaged that AOP networks will become more complete and more informative for predictive toxicology and regulatory decision making.

There are several challenges in the development of an AOP network:

i. The ontology annotations influence the construction of an AOP network. There are still KEs titled differently while having the same meaning and/or referring to the same process. For example, the AOP network on neurotoxicity contains KEs related to the mitochondrial dysfunction: KE ID 177, KE ID 1185, KE ID 1186. All of these KEs can be grouped or renamed under a common KE umbrella. This would also facilitate further quantification in terms of response-response relationships useful for systems toxicology. Following an expert review, such annotations can be easily amended. Slenter et al. (2018) evaluated the WikiPathways database[2] in terms of the content and curation of metabolic pathways and showed the benefits of harmonising the annotation of metabolism and metabolic pathways.

ii. As the information on the compensatory mechanisms for neurotoxicity is missing and not included in the OECD AOP-Wiki KB pages, such networks do not represent the entire complexity of biological processes (feedback and feedforward loops, etc.) and research in this sense is urgently needed. Understanding possible compensatory mechanisms and adaptive changes, which take place before the first KE is triggered, may moderate or contribute to the observed AO. If compensatory and adaptive mechanisms are effective, the cell is coping, and toxicity is not taking place. At present, this kind of information is captured in the KER descriptions, but perhaps the AOP template should also be modified permitting visualisation and description of these processes.

iii. The uncertainty of the network model partly arises from the stage of development of an AOP. An AOP network is built on individual AOPs that ideally follow the OECD requirements, hence giving confidence to the use of an AOP network for predicting toxicity and assessing chemical safety. Even though seven out of twelve individual AOPs published in the AOP-Wiki KB module are endorsed at the time of preparation of this Chapter, this should not limit the use of all selected relevant linear AOPs as the basis of an AOP network. For example, four out of the initial 12 AOPs collected for the development of the AOP network for neurotoxicity were under development, from which one AOP was included in the OECD work plan. Therefore, it will be increasingly possible to develop AOP

[2] https://www.wikipathways.org/, accessed on March 16, 2021.

networks once the OECD AOP-KB becomes more populated with linear AOPs and associated mechanistic information.

iv. When a linear AOP is updated, the AOP network should also include the changes. The AOPXplorer serves as a tool that can encompass such changes.

v. There are "orphan" AOPs that were not linked to the AOP network for neurotoxicity (AOPs IDs 10, 26, 152) and were therefore excluded in the current work. However, these may eventually provide additional KEs and KERs for expanding the network.

vi. AOP networks can be derived for different applications, including the development of toxicity assays (Knapen et al. 2015), evaluation of chemicals mixture (Howdeshell et al. 2017), evaluation of biological plausibility and empirical support to identify knowledge gaps (LaLone et al. 2017b) etc. Herein, linear AOPs were investigated for the AO of developmental and adult/ageing neurotoxicity induced in *Homo sapiens*.

vii. To maximise the application of AOP networks, the use of network analytics provides an essential instrument for characterising the network and identifying common KEs and KERs. Several parameters defined by Villeneuve et al. (2018a), including the level of degree, betweenness centrality and eccentricity, were applied in this work. Information regarding the centrality and connectivity of a KE, the most upstream or downstream KEs across the network, is of great value for finding gaps in knowledge and unforeseen paths. The choice of metric(s) depends on the intended purpose, e.g., development of a battery of *in vitro* tests.

viii. The concept of AOP network should not be confused with the concept of a hub. A hub consists of several KEs closely linked and involved in the same biological process. Therefore, a hub can be part of an AOP network. For example, one of the common KEs identified was neuroinflammation. However, the increase of pro-inflammatory mediators is also known to contribute to the cell apoptosis and necrosis, as AOP ID 17 illustrates. Since inflammation is a complex process, a KE describing it might not be identified as a connected node across the network as is the case of the increased pro-inflammatory mediators. Villeneuve et al. (2018b) developed a hub, which links different MIEs to distinct inflammation-mediated AOs or to AOPs where inflammation is an essential exacerbating element. Such a hub allows interconnectivity with other AOPs that were previously disconnected, independent of the tissue.

## 3.6. Conclusions

An AOP network for human neurotoxicity was developed in Chapter 3 using the principles of the derived AOP network. A workflow was formulated which was adapted for the purpose of the present investigation. Even though the developed AOP network is simplistic and would require updating as more information becomes available, the results provide a solid basis for prioritising the testing of KEs, for quantifying KEs and KERs, and for quantitative modelling of the AOP network. In addition, the work could be useful for identifying biomarkers of toxicity such as BDNF at different biological levels and for further developing *in silico* and *in vitro* test methods, thereby contributing to the assessment of neurotoxicity without animal testing. Future prospects include investigation of when the addition of more linear AOPs to an AOP network (once novel AOPs are discovered and formulated) no longer affects the outcome significantly - in terms of identifying the most or least connected KEs, i.e., at which point do the KEs considered as most/least connected become stable? This would allow for understanding if the variability measured by topology parameters decreases and stabilises the network construction by considering more structured knowledge.

# Chapter 4. Probabilistic modelling of a simplified Adverse Outcome Pathway network for developmental neurotoxicity

The conceptual framework presented herein reflects the discussions of an expert panel that addressed the challenge of quantifying a simplified AOP network for neurotoxicity at an international workshop "e-Resources to Revolutionise Toxicology: linking Data to Decisions", that took place between 7-11 October 2019 at the Lorentz Centre in Leiden, the Netherlands. The manuscript based on this Chapter is under preparation for publication. The *in silico* predictions for P-glycoprotein were provided by the in3 early stage researcher Liadys Mora Lagares, National Institute of Chemistry, Ljubljana, Slovenia.


## Abstract

In a century where toxicology and chemical risk assessment has been devoted to the development of alternative methods to animal testing in predictive toxicology, there is an opportunity to understand the causal factors of neurodevelopmental disorders such as learning and memory disabilities in children. New testing paradigms, along with the advances in probabilistic programming, can help with the assessment of mechanistically-driven hypotheses on the exposure to environmental chemicals that could potentially lead to developmental neurotoxicity (DNT). This investigation aimed to develop a Bayesian hierarchical model of a simplified AOP network for DNT. The model predicted the probability that a compound induces each of three selected common key events of the AOP and the AO of DNT, taking into account correlations and causal relations informed by the key event relationships. A dataset for 97 compounds representing pharmaceuticals, industrial chemicals and pesticides was compiled including physico-chemical properties as well as *in silico* and *in vitro* information. The Bayesian model was able to predict DNT with good accuracy (73%), and helped classify the compounds into low, medium and high probability classes for the potency of DNT. The conceptual framework achieved three further goals: it dealt with missing values; accommodated unbalanced and correlated data; and learned the structure of a directed acyclic graph to simulate a simplified version of an AOP network. Overall, the model demonstrated the utility of Bayesian hierarchical modelling for the development of quantitative AOP (qAOP) models and will contribute to the adoption of new approach methodologies in chemical risk assessment.

## 4.1. Introduction

Neurodevelopmental disorders such as impairment of learning, and memory and cognitive dysfunction, are of serious concern due to the health risks and consequences on the developing brain resulting from the exposure of the young to exogenous chemicals (Bennett et al. 2016; Crofton et al. 2012; Grandjean and Landrigan 2006). Furthermore, it is recognised that there is a lack of adequate information and insufficient toxicity data for most compounds, e.g., limited epidemiological evidence relating to the possible effects of developmental neurotoxicity (DNT), such that DNT is considered to be a "silent pandemic" (Grandjean and Landrigan 2006). When testing is possible, the available DNT testing guidelines for *in vivo* methods (OECD 2007; OECD 2018a; USEPA 1998) are time- and resource-consuming, besides having other drawbacks, such as being recognised as being of little predictive value for the protection of human health and the environment (Paparella et al. 2020). Importantly, the assessment of DNT is not a mandatory requirement in the European Union and the United States of America, and it is usually undertaken when data from developmental and/or reproductive toxicity studies on adult animals indicate a possible concern for neurotoxicity (Bal-Price et al. 2018c). With the limitations of *in vivo* testing for DNT, there is an opportunity to consider NAMs, such as a battery of *in vitro* test methods, omics technologies and *in silico* models as a viable alternative. Whilst NAMs for DNT are not yet standardised or approved by regulatory authorities, they can provide valuable mechanistic insights regarding potential developmental neurotoxicants (Bal-Price et al. 2015; Fritsche et al. 2018; Sachana et al. 2019).

Frameworks are required to organise information from NAMs to predict complex toxicological endpoints, including neurotoxicity and implicitly DNT, as explained throughout the thesis. Recent progress in the development of qualitative and quantitative AOPs underlines their utility to design appropriate experiments and computational simulations following the organisation of evidence into a series of building blocks represented by the MIEs, KEs and AOs at corresponding molecular, cellular, tissue, organ, organism and population levels (Mahony et al. 2020). However, as DNT is a complex process with multiple molecular and cellular paths, no single AOP will be able to explain it in isolation. Thus, a network of linear AOPs allows for the better depiction of the overall mechanistic understanding of DNT than a single linear AOP, as shown in Chapter 3. To illustrate this point, given the limited available resources for quantifying biological paths of DNT, identification of common key events (CKEs) that intersect the individual paths can assist in the prioritisation of testing strategies, i.e., assays based on CKEs. A CKE is considered a key event with high connectivity and is located centrally within the network of AOPs, and that it is essential to link multiple linear AOPs, i.e., MIE(s) to the AO(s) (Knapen et al. 2018; Villeneuve et al. 2018a). However, one of the limitations in the application of both linear and network of AOPs is quantification, specifically to determine the tipping points/thresholds that stressors, e.g., compounds, genetic and environmental factors, require to elicit an adverse response.

Linear and network qAOPs are increasingly being modelled using probabilistic methods. These have become a core approach in computational modelling due to the ease of computing predictions and associated uncertainties relating to variables of interest (Perkins et al. 2019a; Perkins et al. 2019b). Key amongst the probabilistic approaches are Bayesian models that fit, and make inferences from, data using Bayes' theorem for variables, about which probability distributions on the parameters of the model are derived (Gelman et al. 2013). Bayes' theorem is a mathematical formula that transforms the prior distributions into posterior distributions based on the evidence provided (Bishop 2006). The biggest advantage of this approach lies in the reproducibility of the predictions; once the prior/beliefs (i.e., our knowledge about the data before seeing the data) are defined, similar values will be obtained each time the model is run, unless new evidence is added (Gelman et al. 2013; McElreath 2016). Also, probability theory can be used to represent the uncertainty, e.g., errors arising from measurements and the finite size of the datasets etc. (Bishop 2006). An AOP network can also be regarded as a DAG i.e., a graph with no cyclic paths (the loops are removed). Hence, Bayesian networks provide an appropriate framework to compute the joint probability distributions and most importantly, causal inferences (Needham et al. 2007). There are three ways known to assign probabilities in a Bayesian model depending on the research question and available evidence: making subjective assessments, using empirical probabilities based on observed data, and constructing a parametric probability model (Gelman et al. 2013). In addition, a Bayesian model allows for data analysis to be performed using a variety of mathematical functions, linear and nonlinear regression and/or multilevel regression (Gelman et al. 2013). Hierarchical models, also known as multilevel models, random-effects, mixed-effects or varying-effects models are a generalisation of regression defined by the parameters of a model. Such hierarchical models are structured into exchangeable levels/groups, e.g., categories of chemicals, taking into account the independencies and interactions between those groups leading to improved inferences (Gelman 2006; Gelman et al. 2013). These models are ideal for the complex structure of data involving hierarchical organisation similar to what the simplified AOP network for DNT represents.

## 4.2. Aim of this chapter

Given the availability of heterogeneous and limited information about DNT, i.e., *in vivo* and *in vitro* data, and the potential for Bayesian machine learning to investigate chemical-induced DNT, this Chapter aimed to quantify a simplified, reduced version of the AOP network for neurotoxicity developed in Chapter 3 (and published by Spinu et al. (2019)). The main objective was to predict the probability that a compound induces each CKE independently and the AO, taking into account potential correlations and causal relations given by the KERs and additional details such as physico-chemical properties, *in silico* and *in vitro* information. This investigation aimed also to demonstrate whether Bayesian hierarchical modelling is fit-for-purpose in CRA informed by qAOP models.

## 4.3. Methodology

### 4.3.1. Conceptual framework

A conceptual framework was formulated to quantify a simplified version of the AOP network for neurotoxicity with the focus on the DNT that was previously established (in Chapter 3). Based on the resultant AOP network, the topology analysis combined with the domain knowledge allowed for the selection of the most promising CKEs for quantification, i.e., highly connected KEs across the AOP network that describe the AO of DNT. Thus, the reduction of brain-derived neurotrophic factor (BDNF) that leads to a decrease of synaptogenesis and a decrease of neural network formation involved in the impairment of cognitive function represented the selected biological path, as summarised in Figure 4.1. Following Bayes' theorem, the likelihood was informed by the mechanistic knowledge of the simplified AOP network. Thus, the framework helped to define the regression equations encoding the knowledge about the variables of relevance, i.e., CKEs and KERs, as well as the collection of appropriate data to fit the model from available and trustworthy resources, including peer-reviewed publications. The prior was set herein to be noninformative, as explained below. The Bayesian analysis was conducted applying the Markov chain Monte Carlo (MCMC) algorithm to perform the inferences to the unknown parameters and missing data to obtain posterior distributions given by the likelihood multiplied by the prior distributions. The posterior distributions for each chemical and each CKE, including the AO, were obtained and summarised accordingly taking into account the correlations and causal interdependencies between variables included in the model.

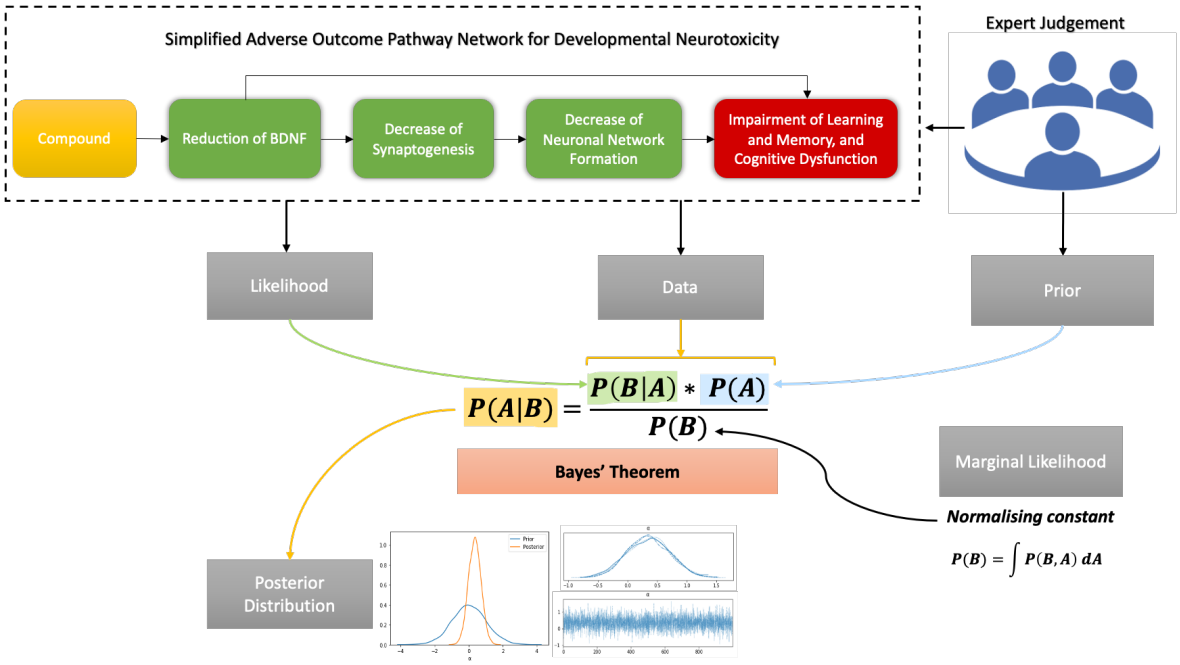


Figure 4.1. The conceptual framework used to develop the Bayesian hierarchical model. The model included three CKEs of the AOP network for DNT shown in green that served as a basis in defining the likelihood functions and collecting appropriate data to fit the model.

### 4.3.2. Data description

Two *in vitro* studies (Frank et al. 2017; Harrill et al. 2018) that tested compounds for their potential DNT adverse effect were chosen as a primary data source. The compounds were merged into a single list and aligned with information referring to compound name, CAS RN, SMILES and the US EPA Comptox Chemical Dashboard substance identifier (DTXSID). In total, 97 compounds served as a starting point to collect additional and informative details to improve the modelling. The list contained different types of compounds: pharmaceuticals, pesticides and industrial chemicals (Supplementary material available on GitHub repository). Fifty-six compounds were tested in both *in vitro* studies, while 11 compounds were tested only for synaptogenesis and 30 compounds were tested only for neural network formation. Importantly, ten compounds were tested as different salts and hence, these compounds were kept and treated individually. These two *in vitro* studies differ in the concentration ranges used, type of the viability assay applied, calculation of the effective concentration (EC) and the temporal exposure (five days vs over 12 days). Also, valproate was tested in two ranges of concentration in both studies: lower and higher. The range of higher concentrations showed response/activity in both studies and, thus, the ECx values of the range of higher concentrations tested, and associated responses, for valproate were kept as part of the data curation process. The difference in the range of concentration and time of exposure shows that synaptogenesis occurs at lower levels than neural network formation and aligns perfectly with the principles that a qAOP model is expected to demonstrate, i.e., quantitative predictions underlying the transition from one KE to the next KE as described in Chapter 2.

Other data collected included the calculated logarithm of the octanol-water partition coefficient (SLogP), a measure of lipophilicity; permeability to blood-brain-barrier (BBB); and the capability to bind to the P-glycoprotein (P-gp) transporter, i.e., if the compound acts as P-gp inhibitor, substrate and is (non)activ against P-gp. Predictions for BBB and P-gp were calculated using *in silico* models and available tools based on curated SMILES as listed in Table 4.1. BBB permeability is essential in understanding if a compound crosses into, and has the possibility to act on the CNS (Kaplan et al. 2020). P-gp is a transmembrane protein belonging to the ATP-binding cassette family of transporters (ABC-transporters), highly associated with the ADMET properties of compounds. It may contribute to a decrease in toxicity by eliminating the compound from cells and preventing its intracellular accumulation (Mora Lagares et al. 2019). Lastly, the compounds were grouped into positive/negative for DNT induction based on *in vivo* studies as summarised in Mundy et al. (2015), and positive/negative for the reduction of BDNF, informed by the literature review performed to evaluate the peer-reviewed publications that studied the impact of compounds for the reduction of BNDF (Supplementary material available on GitHub repository). The compounds were classified as active/inactive for the CKEs of decrease of synaptogenesis and neural network formation based on the selectivity identified by the corresponding *in vitro* studies, taking into account viability and the results of toxicity assays. Compounds that showed no activity were assigned zero as a numerical value, whilst any compounds not tested in one of the

*in vitro* assays were treated as missing variables. An overview of the type of data utilised for modelling DNT is presented in Figure 4.2.

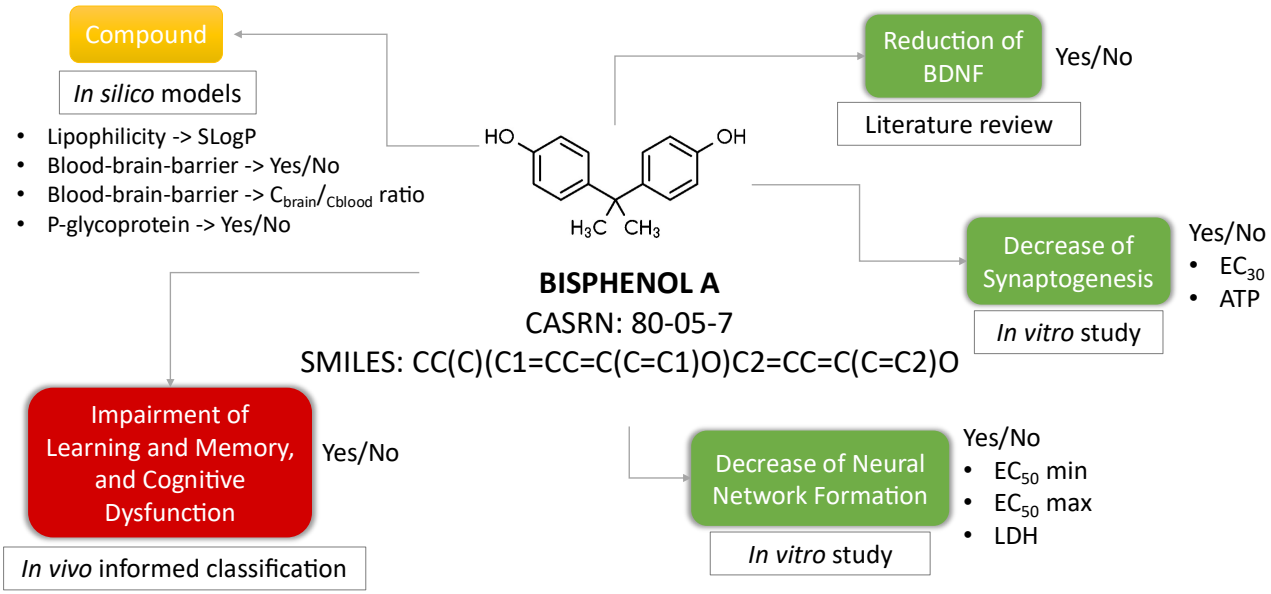


Figure 4.2. Types of information collected and curated for model development exemplified for bisphenol A. It underlines how different streams of data can be integrated for improved causal predictions.

### 4.3.3. Exploratory data analysis

Exploratory data analysis (EDA) was applied to analyse the data collected and to summarise their main characteristics including the types of variables (e.g., continuous or discrete); the shape of the distributions of individual variables; correlations between the variables using the Pearson correlation coefficient; missing values; chemical characteristics described by physico-chemical properties associated with a category of one of the CKEs and DNT; and the presence of (un)balanced categories in the dataset. EDA was conducted to help choose the appropriate priors and regressions and define the overall strategy for computational modelling.

### 4.3.4. Bayesian hierarchical modelling

A single nested partial pooled Bayesian hierarchical model was formulated taking into account data issues such as: correlations and interactions between variables; missingness at both levels' predictors and outcomes and learning from the DAG structure of the simplified AOP network for DNT. This helped capture causal relationships and additional dependencies given by the CKEs and KERs as outlined in Figure 4.3. The continuous variables were standardised prior to fitting the model. The causal graph is described as a set of linear regressions with the latent variable, e.g., Y of CKE 1 reduction of BDNF, as a predictor that goes into the next linear regression, e.g., Y of CKE 2 decrease of synaptogenesis.

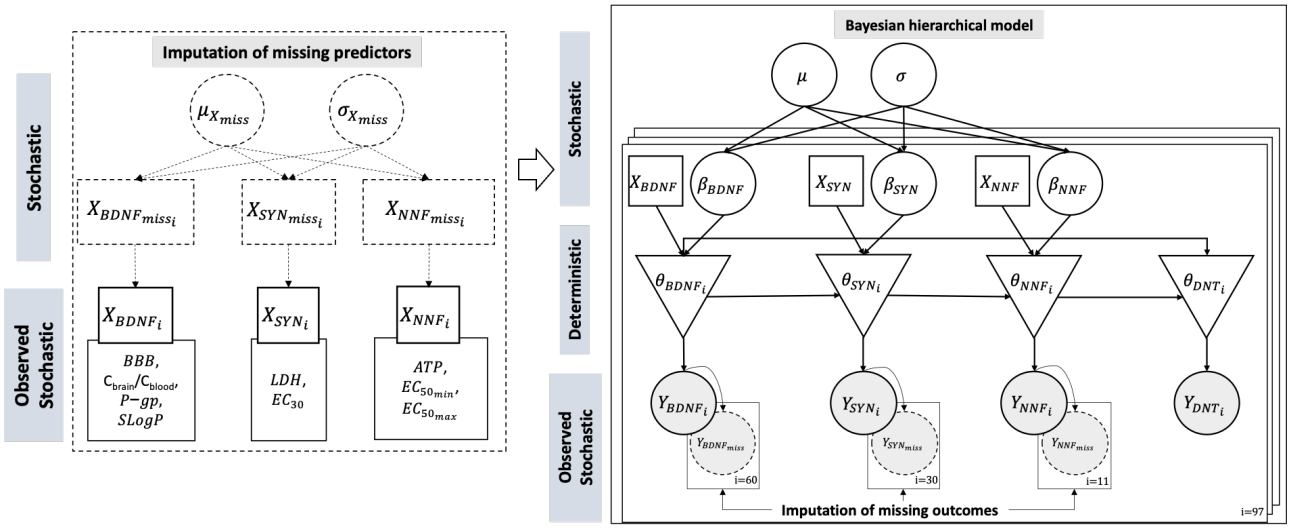


Figure 4.3. A simplified graphical representation of the proposed Bayesian hierarchical model utilised to assess individual compounds for their potency in inducing DNT. The model follows a specific biological path of the AOP network for DNT. BDNF: reduction of brain-derived neurotrophic factor; SYN: decrease of synaptogenesis; NNF: decrease of neural network formation; DNT: developmental neurotoxicity; miss: missing values; i: number of compounds; X: predictors, independent variables; Y: outcomes, dependent variables; $\beta, \mu, \sigma$: parameters of the model.

The Bayesian hierarchical model consisted of nine unknown parameters: hyperpriors $\mu$ and $\sigma$, priors $\beta$s, and the likelihood functions given by the $\theta$s. In a Bayesian framework, all the unknown parameters must have predefined distributions, which represent our belief before evaluating the data, and that will be estimated from the data. Because of the hierarchical type of modelling, the parameters were sampled and partial-pooled from a common global distribution given by the hyperpriors that were defined as noninformative. The sign tilde "∼" indicates the type of distribution the parameter was generated from. Herein, a common mean $\mu$ and standard deviation $\sigma$ were defined to describe the global distribution of the entire dataset and were generated from normal and half-normal distributions.

$$\mu \sim Normal(0,\ 0.01)$$

$$\sigma \sim HalfNormal(5)$$

Each $\beta$ varied per CKE and, thus, each group-level CKE was generated from a normal distribution shrank towards the hyperpriors specified above.

$$\beta_{BDNF}, \beta_{SYN}, \beta_{NNF} \sim \text{Normal}(\mu, \sigma)$$

The independent variables, which consisted of both discrete and continuous types of variables and their potential interactions, were grouped per CKE to model the binary predictions, as shown below.

$$X_{BDNF}: SLogP,\ C_{brain}/C_{blood},\ BBB,\ Pgp_{inhibition},\ Pgp_{substrate},\ Pgp_{active}$$

$$X_{SYN}: EC_{30},\ ATP$$

$$X_{NNF}: EC_{50_{min}}, EC_{50_{max}},\ LDH$$

The missingness has been treated by Bayesian imputation where each missing data point was masked in advance. If X predictors contained missing data, the imputation was sampled from the prior distribution as shown below, while the missing Y outcomes have been imputed from the posterior predictive distributions. Noninformative priors were defined to sample the missingness for all CKEs.

$$\mu_{Xmiss} \sim Normal(0,\ 1)$$

$$\sigma_{Xmiss} \sim HalfNormal(2)$$

$$X_{BDNF} \sim Normal(\mu_{Xmiss}, \sigma_{Xmiss}, X_{miss})$$

$$X_{SYN} \sim Normal(\mu_{Xmiss}, \sigma_{Xmiss}, X_{miss})$$

$$X_{NNF} \sim Normal(\mu_{Xmiss}, \sigma_{Xmiss}, X_{miss})$$

The likelihood function, herein defined as $\theta$s represented the deterministic relationship between the matrix of X predictors and the parameters $\beta s$ describing each CKE and were estimated from data. Besides the multiplication, the causal relationship was considered by adding the previous $\theta$ to the subsequent $\theta$ to follow the DAG structure of the AOP network. Importantly, the likelihood function for the AO of DNT was informed solely by the sum of likelihood functions of the CKEs. The subscript i index indicates that each compound had its probability estimate.

$$\theta_{BDNF_i} = \beta_{BDNF} * X_{BDNF}$$

$$\theta_{SYN_i} = \beta_{SYN} * X_{SYN} + \theta_{BDNF_i}$$

$$\theta_{NNF_i} = \beta_{NNF} * X_{NNF} + \theta_{SYN_i}$$

$$\theta_{DNT_i} = \theta_{BDNF_i} + \theta_{SYN_i} + \theta_{NNF_i}$$

The three CKEs were generated from a Bernoulli distribution of the deterministic relationships estimated from the data. The AO of DNT was generated as well from a Bernoulli distribution of the deterministic relationship that summed up the logistic regression of all $\theta$s that described independently the CKEs.

$$Y_{BDNF_i} \sim Bernoulli(\theta_{BDNF_i})$$

$$Y_{SYN_i} \sim Bernoulli(\theta_{SYN_i})$$

$$Y_{NNF_i} \sim Bernoulli(\theta_{NNF_i})$$

$$Y_{DNT_i} \sim Bernoulli(\theta_{DNT_i})$$

The model has been developed in PyMC3 version 3.9.3. Two MCMC methods were used to obtain the estimates: the No-U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo for continuous responses, and the Binary Gibbs Metropolis (BGM), a special case of the Metropolis-Hastings algorithm for binary responses, with four chains and 4000 samples for tuning step and 40000 draws in total that runs in 5 min and

23 sec. Notably, the model fitting had two objectives: (1) to make inferences about the relationships between variables, and (2) to make predictions based on the model estimates using collected and curated data.

Model fitting was evaluated for its convergence. The Gelman-Rubin diagnostics, also known as the R-hat statistic, measures how similar different chains are, i.e., within and between chains, and so, if the chains converged to the same distribution (Gelman and Rubin 1992). The Monte Carlo standard error is another measure of the accuracy of the chains, given by the posterior standard deviation divided by the square root of the number of the effective samples (Gelman et al. 2013). The smaller it is, the closer the posterior mean is expected to be to the actual value. The effective sample size estimates the independent draws, because samples will typically be autocorrelated within a chain and can increase the uncertainty in estimates (Gelman et al. 2013). It should be at least the same as the actual number of the samples. The Bayesian credible interval (CI) of 95%, also known as the highest density interval (HDI), an interval within which an unobserved parameter value falls with a particular probability, was applied. A 95% CI has the upper and lower 2.5% percentiles of the posterior distribution as its bounds.

The Bayesian hierarchical model had 657 parameters in total for a dataset of 97 compounds. The number of parameters represented a risk of overfitting; this means that a model learns too much from the sample (Gelman et al. 2013; McElreath 2016). In a Bayesian framework, a significant source of overfitting is given by choice of priors that have a vital role in normalising the likelihood functions and thus, being carried along with the model (Gelman et al. 2013; McElreath 2016). Hence, the overfitting risk depends on both structural details of the model and the composition of the sample. To prevent the risk of overfitting, we opted to shrink the $\beta$s parameters towards a common hyperprior that controlled the distribution of them and, hence, opted for a partial pooling of the level of details to produce estimates for each of CKEs and the AO, and chose "sceptical" hyperpriors and priors to regularise the inferences.

Table 4.1. List of sources and associated data collected for the development of the Bayesian hierarchical model. The table summarises all variables, i.e., predictors and outcomes defined as features included in the model for the type of data and performance where applicable.

| Feature | Description/Relevance | Data Type | Performance | Source |
|---|---|---|---|---|
| Chemical Name | The names used to define the compounds tested in both *in vitro* studies. | Not Applicable | Not Applicable | Frank et al. (2017); Harrill et al. (2018) |
| CASRN | Chemical Abstracts Service Registry Number associated with the tested compounds used to identify and track them during the modelling. | Not Applicable | Not Applicable | Frank et al. (2017); Harrill et al. (2018) |
| DNT Classification | Each compound was classified as either positive, known or potential inducing DNT/negative, safe or without evidence for inducing DNT based on *in vivo* studies. | Binary, i.e., positive (i.e., associated with DNT) or negative | Not Applicable | Frank et al. (2017); Harrill et al. (2018); Mundy et al. (2015) |
| SLogP | The logarithm of the octanol/water partition coefficient calculated based on the SMILES of the compounds. | Continuous, i.e., unitless values | Not Applicable | KNIME RDKit v.3 |
| BBB | Each compound was classified for its capability to permeate the blood-brain-barrier (BBB) based on curated SMILES. Predicting BBB permeability means indicating whether compounds pass across the BBB. The CNS-active compounds must pass across it, and CNS-inactive compounds must not pass across it to avoid CNS adverse effects. | Binary, i.e., positive (BBB permeable) and negative | The *in silico* model available in admetSAR v2.0 had the area under the curve (AUC) with a range from 0.625 to 0.99. | Literature review,<br>Online BBB Predictor v.0.9,<br>admetSAR v.2.0,<br>Liu et al. (2014); Yang et al. (2019) |
| Cbrain/Cblood | *In vivo* blood-brain-barrier penetration represented as BB = [Brain]/[Blood], where [Brain] and [Blood] are the steady-state concentration of radiolabelled compounds in the brain and peripheral blood. High absorption to CNS had a value of more than 2.0, medium absorption: 2.0 - 0.1, and low absorption: less than 0.1. The predictions are based on *in vivo* data on rats. | Continuous, i.e., unitless values | The QSAR model of Ma et al. (2005) had R=0.955 with s=0.232, used by PreADMET v.2.0 to model the predictions. | Ma et al. (2005)<br>PreADMET v.2.0 |
| P-glycoprotein Status | Each compound was classified based on curated SMILES as a substrate or not, inhibitor or not, active or inactive for P-glycoprotein (P-gp) transporter using an *in silico* model. | Binary, i.e., yes or no for a compound to act as a substrate, an inhibitor or showing or not activity against P-gp | The non-error rate and the average precision was 0.70 for the external validation set. | Mora Lagares et al. (2019) |

*To be continued*

Table 4.1. Continued.

| Feature | Description/Relevance | Data Type | Performance | Source |
|---|---|---|---|---|
| BDNF, Reduction | Each compound was classified as either positive, inducing the reduction of BDNF levels, or negative based on a literature search for *in vivo/in vitro* studies. | Binary, i.e., positive (evidence showing alterations of BDNF), or negative | Not Applicable | Literature review of historical and peer-reviewed studies that evaluated compounds for their potency of inducing reduction of BDNF |
| Activity Synaptogenesis | Selectivity and potency of a chemical were kept the way the reference classified a compound based on results for viability and effective concentrations. | Categorical, i.e., inducing or not alterations of synaptogenesis | The battery assay had a sensitivity of 87% and a specificity of 71%. | Harrill et al. (2018) |
| Viability Synaptogenesis | The amount of ATP present in each well was calculated to assess compounds for their viability. | Continuous | Not Applicable | Harrill et al. (2018) |
| Synaptogenesis, EC30 (µM) | 30% change compared to control expressed as an effective concentration EC30 (µM) for puncta per total dendrite length (the most sensitive endpoint) measured in rat primary cortical cells for five days using an imaging assay. | Continuous | Not Applicable | Harrill et al. (2018) |
| Neural Network Formation Activity | Selectivity and potency of a chemical were kept the way the reference classified a compound based on results for viability and effective concentrations. | Categorical, i.e., inducing or not alterations of neural network formation | The model had a mean accuracy of 80.2%. | Frank et al. (2017) |
| Neural Network Formation Viability | Total lactate dehydrogenase (LDH) release upon cell lysis was calculated to assess compounds for their viability. | Continuous, i.e., unitless values | Not Applicable | Frank et al. (2017) |
| Neural Network Formation, EC50min and EC50max (µM) | 50% change compared to control expressed as an effective concentration EC50 (µM) with a minimum and maximum values of all 17 parameters measured in rat primary cortical cells over 12 days using microelectrode array (MEA) recordings. | Continuous | Not Applicable | Frank et al. (2017) |

## 4.4. Results

The EDA allowed for a better understanding of the data collected and utilised for the Bayesian analysis. As a result, the final dataset was composed of unbalanced discrete variables describing P-gp inhibition, substrates and activity against P-gp, permeability to BBB along with the classifications for CKEs and AO (Appendix II, Figure S1). The distribution of continuous variables showed the presence of leverage points i.e., outliers, for SLogP, and the presence of missing values for other properties given by the long tails (Appendix II, Figure S2). Although the EDA confirms that the compiled data set is not ideally suited for modelling purposes, Bayesian modelling can adapt the different streams of data. The process of using the data in Bayesian modelling requires the data to be defined appropriately with regard to mathematic transformations.

The correlation matrix of the entire dataset allowed for the evaluation of the correlation coefficients between the variables included in the model, i.e., predictors and outcomes (Appendix II, Figure S3). It gave an indication of how to better define the regression functions and make them more informative to model the AOP network. In contrast, the correlation matrix of the missing values indicates how strongly the presence or absence of a variable affects the presence or absence of another variable (Appendix II, Figure S4). It highlighted the relevance of not discarding the missing values as the "missingness" was informative. Notably, the missingness was present at all levels: *in silico* predictions, since the dataset contained organic and inorganic compounds, and *in vitro* predictions, as not all compounds were in both *in vitro* systems. Thus, probability distributions are the key ingredients of Bayesian modelling and, hence, there is a need to assess the dataset utilised through EDA to define and fit the model.

The statistical parameters of the model did not show the presence of any divergences (Appendix II, Figures S5 and S7). The Gelman-Rubin diagnostics of the model had a value below 1.005 for all the parameters of the model. The Monte Carlo standard error of the inferences did not exceed 10% of the posterior standard deviation. The effective sample size was greater than 10% of the number of iterations. In addition, the mean value of all the inferences of four chains obtained from the data for each hyperprior and prior to eleven X predictors are presented in Figure S6 (Appendix II).

The posterior distributions and posterior predictive distributions were obtained from the Bayesian analysis. The posterior distributions represent the evidence provided by the data combined with the prior that incorporates our knowledge before analysing the data. Thus, the conclusions were derived from the likelihood functions $\theta$s yielded to predictions of the observed data, i.e., binary classifications of each CKE including the AO, defined as posterior predictive distributions. The distribution of the posterior probability of the three categories of compounds for each CKE including the AO, i.e., inducing the CKE (yes), not causing the CKE (no) and compounds with a missing level of details (missing), is shown in Figure 4.4. The results for the distribution of the posterior predicted probability of the three categories of compounds for each CKE including the AO, i.e., inducing the CKE (yes), not causing the CKE (no) and compounds with a missing level of details (missing), are represented in Figure S8 (Appendix II).

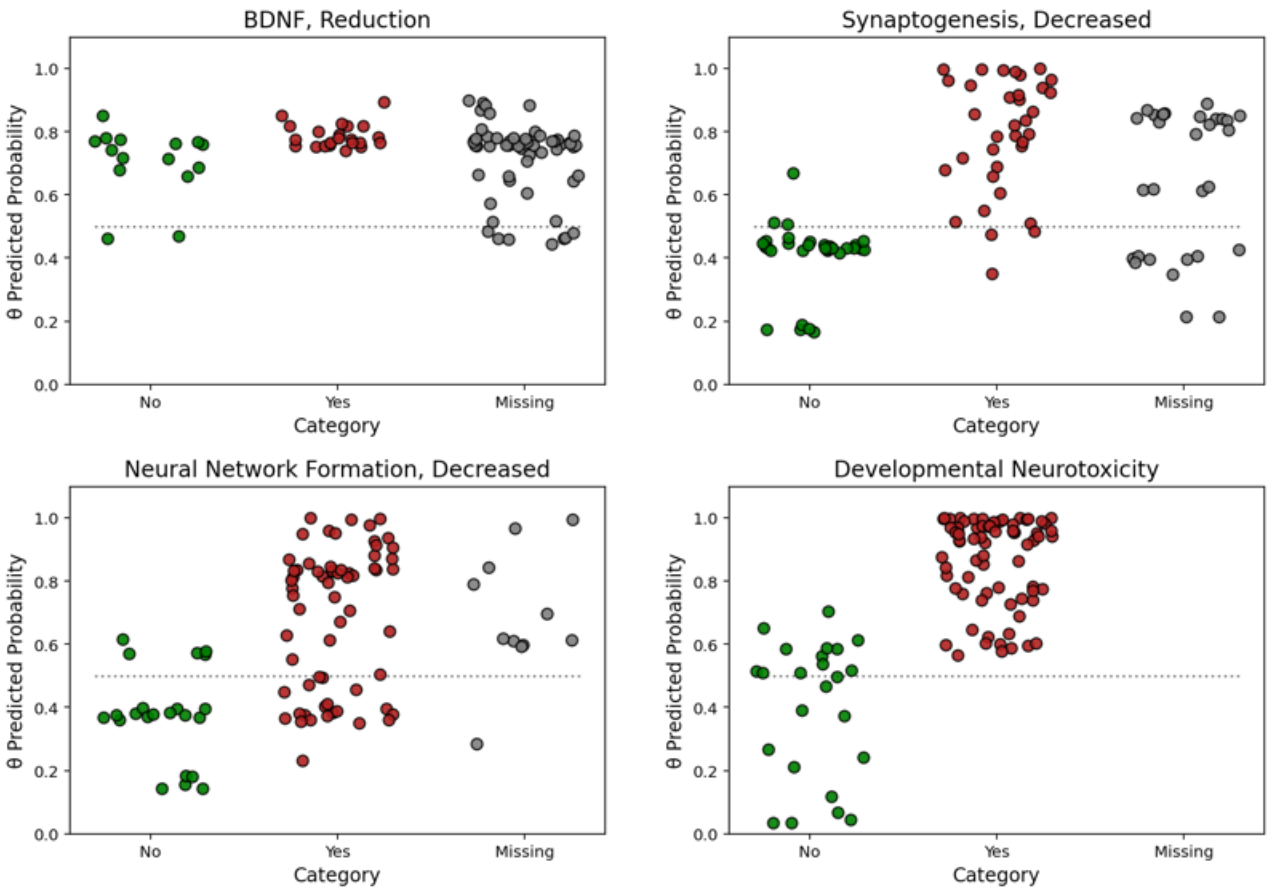


Figure 4.4. The distribution of the posterior probability of the $\theta$ likelihood functions of the three CKEs and the AO for positive and negative compounds, and compounds with a level of missingness. The compounds analysed for the AO of DNT did not contain any missing data; the classification is made based on *in vivo* studies.

An overview of the posterior probability of each compound is presented in Figure 4.5. It shows that the Bayesian credible interval is broader for compounds with missing information. Thus, the HDI captures the uncertainty given by the sources of variability and missingness in a quantified manner. For instance, colchicine, an alkaloid analogue used as an anti-inflammatory medicine, that had data across all levels, had a 95% HDI of 0.76 to 1.0 with a mean probability of 0.94 for the induction of DNT. On the contrary, lactofen, a nitrophenyl ether selective herbicide, that had missing data for the first two CKEs had a 95% HDI of 0 to 0.95 with a mean probability of 0.46 for induction of DNT.

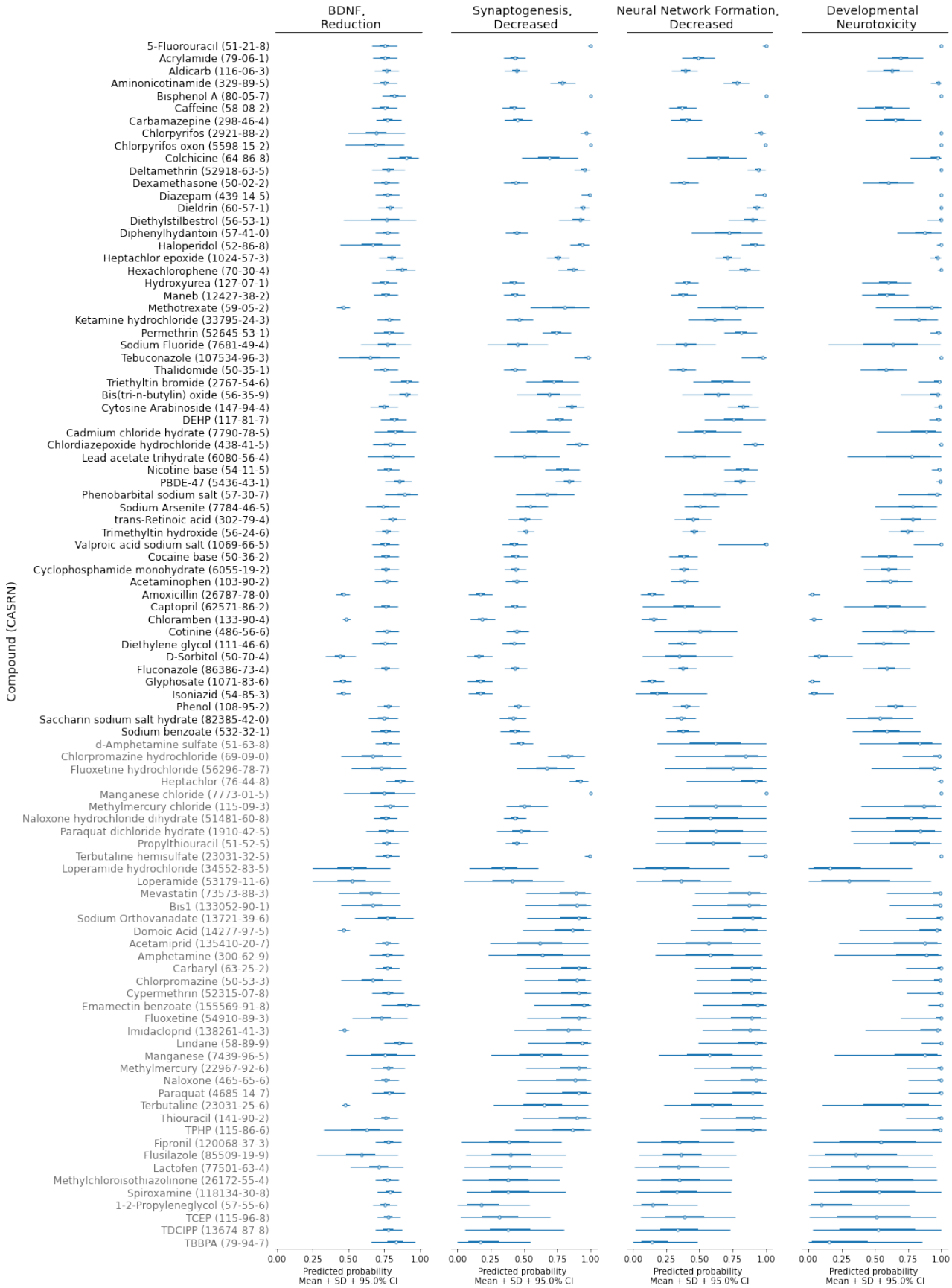


Figure 4.5. An overview of the predicted posterior probability of the θ of each independent key event that takes into account the causal relationship. Compounds in black are compounds tested in both *in vitro* studies, while the compounds in grey are compounds tested in one of the *in vitro* studies, and thus, with a higher level of missingness overall. The dot represents the mean, the thicker line is the standard deviation, and the thin line represents the 95% of the Bayesian credible interval (CI), also known as the highest density interval (HDI).

Two thresholds derived from the results of the predicted posterior distributions themselves were used to classify the compounds for high, medium and low probability of inducing a CKE and the AO. Such classification might be used for prioritisation and additional screening purposes. Therefore, all compounds had a probability higher than 40% for posterior distributions and a probability of 60-65% for posterior predicted distributions, being classified overall as having a high probability to induce a reduction of BDNF (Figures S9-S11). The results for the decrease of synaptogenesis varied for posterior distributions, while the posterior predictive distributions showed a probability of between 60-65% (Figures S12-S14). Except for d-amphetamine sulfate, d-sorbitol, bisphenol A and tetrabromobisphenol A, that were classified as having a high probability overall, the remaining compounds had a medium level of probability to decrease synaptogenesis. For the decrease of neural network formation, the posterior probability varied across the compounds, while the posterior predicted probability was between 60-65% for all compounds that were classified as having a medium level of probability overall to induce this CKE (Figures S15-S17). Finally, the predictions for the AO, which is of most interest for research and regulatory purposes, helped to classify the compounds into the three levels, with fewer in the low level of probability and most of the compounds with a high level of probability to induce DNT (Figure 4.6). The results for DNT for posterior probability and posterior predicted probability is shown in Figures S18-S19. This unbalanced classification is explained by the initial list of compounds that had 72 out of 97 compounds with known positive DNT.

The overall accuracy of the model is 86%, and the overall balanced accuracy of the model that takes into account the unbalanced categories is 73%, with a sensitivity and a specificity of 46% and 100%, respectively for the prediction of the DNT category. Two compounds have been misclassified with a high level of probability for inducing DNT: cotinine and phenol. These were identified as inactive in both of the *in vitro* studies and without *in vivo* evidence for causing DNT. A probability of 70% for posterior distribution and a probability of 71% for posterior predicted distribution for cotinine, and a probability of 65% for posterior distribution and a probability of 66% for posterior predicted distribution for phenol have been computed. The potential reasons for this misclassification can be in part because of the *in silico* model used for the prediction of BBB, which classified the compounds as positive. However, the large credible interval associated to the predicted value underlines the precautious of making decisions for these compounds. Also, the ten compounds tested under different CAS RN (salts vs base form of a chemical) showed variability in the predicted probability for DNT (Figure S20). All the pairs of compounds were classified in the same category except that a different classification has been associated to loperamide (low) and loperamide hydrochloride (medium), and terbutaline (medium) and terbutaline hemisulfate (high).

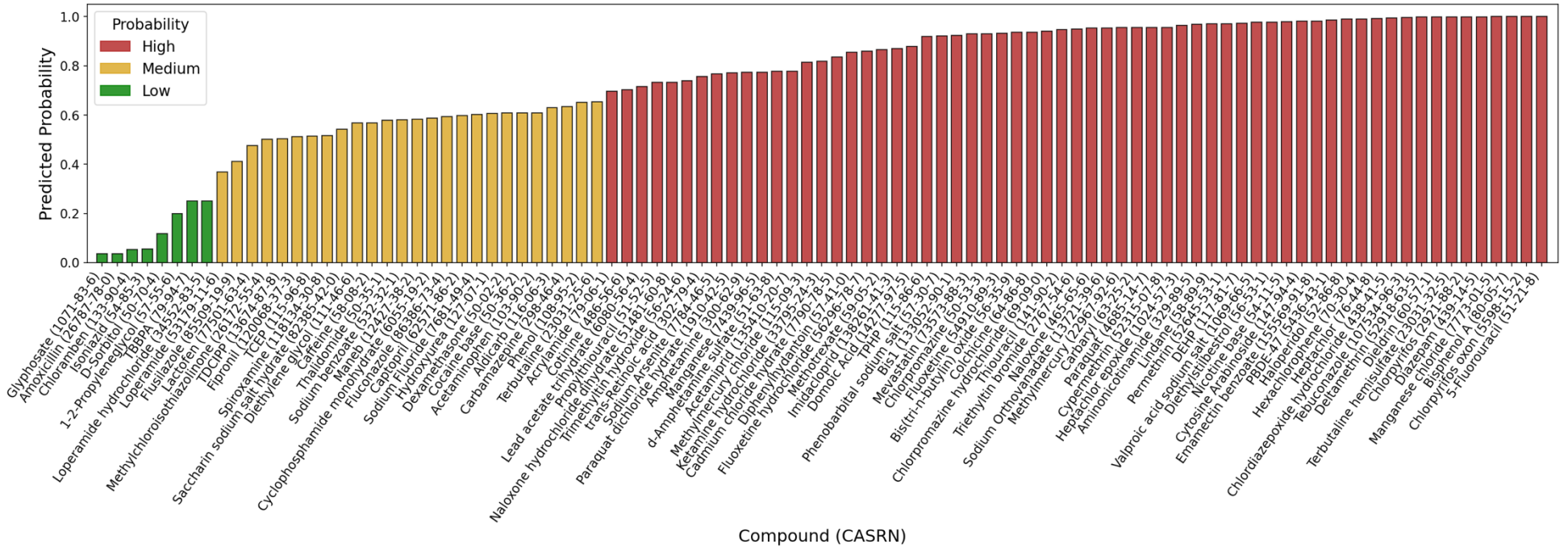


Figure 4.6. The predicted probabilities (high, medium and low) of inducing DNT colour-coded based on two thresholds estimated from the results.

### 4.4.1. Case study

A possible means of analysing and interpreting the results for screening and prioritisation as well as potential regulatory purposes, was investigated. For this, five compounds were chosen to cover different probabilities of predictions, these were: bisphenol A (BPA), glyphosate, dexamethasone, fipronil and lead acetate trihydrate (lead). All compounds are known to promote DNT, possibly through varying mechanisms. BPA has been suspected as an endocrine disruptor that can potentially lead to altered reproduction and neurodevelopmental toxicity (Mustieles et al. 2020). Glyphosate had no neurodevelopmental effects in several *in vitro* studies without effects on neurodevelopmental processes (Frank et al. 2017). Dexamethasone showed *in vitro* implications of neurobehavioral deficits (Jameson et al. 2006). Fipronil is known to disrupt the CNS by blocking GABA-gated chloride channels and glutamate-gated chloride channels (Lassiter et al. 2009). Lead has been thoroughly studied for its implications in inducing DNT and its activity has been summarised into a linear AOP (OECD AOP-Wiki KB ID 12) by Tschudi-Monnet and FitzGerald (2018). In the context of the present Bayesian model: (i) BPA contained complete details for the variables of interest and it was classified as inducing all CKEs, as well as the AO; (ii) glyphosate had missing information regarding the potency of reducing BDNF and it has been identified as negative for the other CKEs and the AO; (iii) dexamethasone showed no activity for the CKEs; however, it was identified as causing DNT based on the *in vivo* studies; (iv) fipronil had missing details regarding the CKE of the decrease of synaptogenesis, and it was identified as negative for the other CKEs and the AO; (v) lead did not show activity for the CKE relating to the decrease of synaptogenesis and it contained missing information derived from *in silico* models.

The proposed framework is comprised of three phases, as shown in Figure 4.7:

- Phase I: checking the posterior predictive distributions for the AO of DNT and the associated predicted level of probability: low, medium or high.
- Phase II: checking the compounds for their posterior distributions for the CKEs and the AO of DNT for a better understanding of classifications including the uncertainty given by the HDI.
- Phase III: translating the observations into a conclusion for decision-making, e.g., the requirement of additional *in vitro* screening.

A concluding remark is that: (i) BPA was predicted to have a high probability of inducing DNT and can be considered as positive control compounds in any future wet and/or *in silico* experiments, (ii) glyphosate showed a low probability for inducing DNT and, thus, it can be classified as a negative control compound for the same purposes mentioned above, (iii) dexamethasone, with a medium level of probability might need additional data or investigation, e.g., testing other biological paths involved in inducing DNT, (iv) fipronil, also with a medium level of probability, has to be evaluated for the CKE relating to the decrease of synaptogenesis as well as for potential other CKEs of the AOP network, and (v) lead presented a high probability of inducing DNT. Thus, the Bayesian hierarchical model formulated for a simplified AOP network was able to prioritise compounds for testing based on a variety of data, including missing data. The Bayesian hierarchical model

will help direct what further information, i.e., *in vitro*, *in silico* data will be required to strengthen an opinion or a decision about potential DNT toxicants.

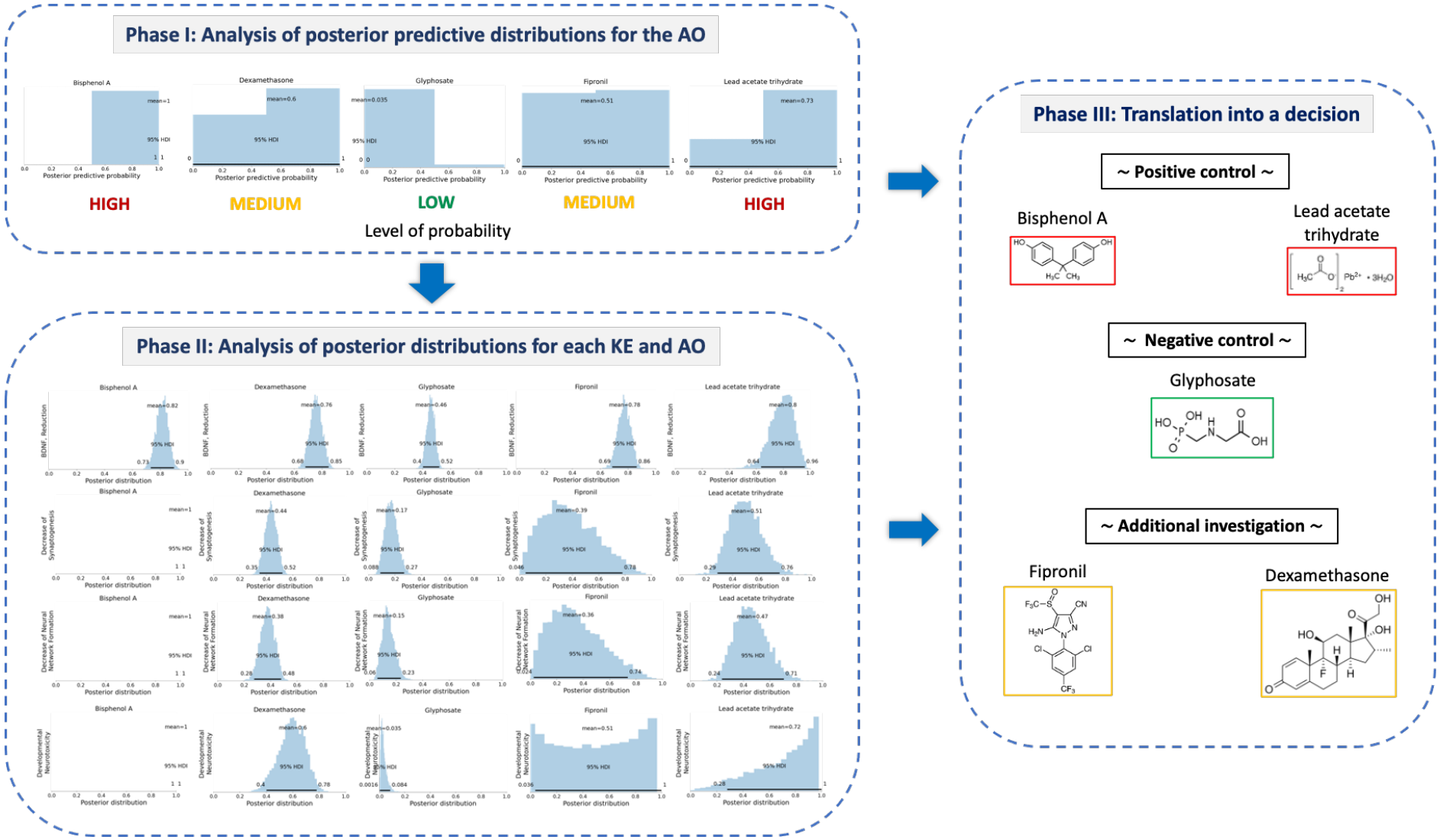


Figure 4.7. A potential utility of the results of the proposed Bayesian hierarchical model for the simplified version of the AOP network for DNT. A final decision can be informed by grouping the compounds into three classes for the level of predictions given by the probabilities of posterior distributions on both observed and unobserved variables.

## 4.5. Discussion

Bayesian machine learning has been increasingly applied in toxicology, from improving the modelling of physiologically-based pharmacokinetic (PBPK) models of inter- or intra-individual variability across a population (Bois et al. 1996; Bois et al. 2010), dealing with pseudo-replications (Lazic et al. 2020) or a combination of Bayesian statistics and deep learning to the investigation of hepatotoxicity (Semenova et al. 2020). However, only recently, has it been applied to the concept of modelling of an AOP, as reviewed in Chapter 2. To our knowledge, the Bayesian hierarchical model proposed in this Chapter is the first to model a linear path within an AOP network derived from topology analysis, combined with the expert judgement, while being able to deal with missing data. Additionally, it is a parametric model of continuous distributions as opposed to discrete Bayes nets, also known as Bayesian networks or belief networks, that imply conditional probability tables (CPTs) and that have been proposed as an option for the probabilistic quantification of AOPs (Perkins et al. 2019a).

A qAOP model is data-driven, and thus, initiatives that focus on data generation and collection are crucial for their development. The requirement for reliable and meaningful data is especially relevant with regard to the mechanistic understanding of the endpoints related to DNT. The Programme on the development of AOPs co-ordinated by the OECD has been instrumental in creating a knowledge management database of AOPs. The database plays a pivotal role in supporting research projects and improving the predictions of toxic responses for humans (OECD 2017; OECD 2018b). However, in practice, few AOPs exist compared to the abundant and increased mechanistic toxicological knowledge and those that do exist are often not sufficiently quantified for practical applications. In order to address this, the collaborative project Developmental Neurotoxicity Data Integration and Visualisation Enabling Resource (DNT-DIVER) of the National Toxicology Program (NTP) by the National Institute of Environmental Health Sciences (NIEHS) aims to jointly generate and collect data for the assessment of DNT toxicants (Behl et al. 2019). As a result, screening libraries consisting of known and suspected developmental neurotoxicants, and negative controls classified based on *in vitro* studies conducted on cell culture systems, zebrafish and planarian models are becoming available. These data could be utilised for prioritisation and additional investigation and, potentially, for the future development of qAOP models incorporating further other levels of details than those included herein. Importantly, as DNT effects occur primarily in the offspring, following exposure to chemicals, it is essential to study metabolism and other kinetic and dynamic behaviours that a compound may exhibit in order to determine the *in vivo* exposure and possible accumulation during years of exposure e.g., incorporating PBTK models adapted for pregnancy. Thus, once other important kinetic and dynamic types of information, as well as tools to facilitate this understanding, becomes available, the Bayesian model can be updated accordingly.

In the context of Bayesian modelling applied to toxicology, an *in silico* model should answer to causal rather than associative questions that cannot be computed from data alone. Causal inference can be defined as a

way of predicting what would happen, or what might have occurred, to an outcome y given a set of predictors X as a result of a treatment, intervention or exposure z (Gelman et al. 2020). Causal effects can take both linear and nonlinear functions. Herein, the causal effects given by the KERs were treated as linear models for simplicity. It was also interesting to capture the variations between the effects, and the interactions between the variables, to predict the effects of individual compounds. The causal regression model followed a hierarchical structure to cover the above aspects. Thus, hierarchical (multilevel) models are extensions of regression in which data are structured in groups and coefficients can vary by group (Gelman 2006). Consequently, hierarchical models can be used for a variety of inferential objectives, including causal inference, prediction and descriptive modelling. For example, the causal inference was simulated using linear regressions and the prediction aim was achieved by the two-class logistic regression of the dependent variables implemented. The descriptive understanding helped to underline the main features of the data and where additional efforts are needed.

Probabilistic programming languages (PPLs) have only recently been applied in ML and artificial intelligence (AI) following the development of several approximate integration algorithms, e.g., MCMC methods (Ghahramani 2015). PPLs are domain-specific languages utilised to formalise a Bayesian model, they help to automate the process of inferring unobserved variables in the model, i.e., outcomes of interest conditioned on the observed data. Thus, PPLs rely on combining the inference capabilities of probabilistic methods with the representational power of programming languages (Ghahramani 2015). The ecosystem of PPLs has become more and more available through various examples, e.g., BUGS, Edward, Infer.NET, Nimble, Pyro, Tensorflow, Turing, Stan, to name a few. In this Chapter, the Bayesian model was coded in PyMC3 (Salvatier et al. 2016), a Python package. It is acknowledged that the choice of PPL is rather subjective in that it often depends on the user's experience as to how the model can be coded.

There are several advantages in modelling a qAOP in a Bayesian (hierarchical) manner, which include:

1. A qAOP model must allow for an objective scientific judgement of potential toxicity of chemicals (Spinu et al. 2020). A Bayesian model aims to determine the posterior distribution for the model parameters allowing for the quantification of a response, the magnitude of response or the effect of the size of single KE or a pair of KEs in a probabilistic manner. Notably, the associated credible interval incorporates the uncertainty of both dependent and independent variables and the different sources of variability as is the case for DNT. Thus, a Bayesian model becomes more informative than a single best estimate that a frequentist model provides, i.e., "statistically significant" and "non-significant" and, thus, contributes to the paradigm shift in statistical thinking and decision-making as argued by Amrhein et al. (2019) while leading to a transparent, traceable, reproducible and reliable assessment. For example, Hothorn and Ralph (2020) advocated the use of compatibility intervals as an alternative to formal statistical testing in toxicology, i.e., confidence intervals, which describe the compatibility of the data with the hypothesis and the model to evaluate endpoints of regulatory interest. Thus, providing confidence in an assessment of potential DNT toxicants, by accounting for uncertainty in a Bayesian manner, can better inform

decision making while giving an understanding of the importance of the resultant predictions, especially in data-sparse situations such as DNT.

2. There is a demand for strategies that would better integrate the variety of information available, including NAMs or combinations of non-animal methods known as defined approaches, for an informed risk assessment and a structured decision-making process (Avila et al. 2020), e.g., IATA for DNT screening and prioritisation purposes (Bal-Price et al. 2018a; Paparella et al. 2020; Worth and Patlewicz 2016). Additionally, a battery of DNT *in vitro* test methods is advocated to generate valuable mechanistic data. Data from such *in vitro* test methods can be analysed to identify reliable indicators of DNT, leading to an analysis of greater complexity, but increased relevance (Paparella et al. 2020). A Bayesian model can accommodate any type of data essential for the assessment of a chemical, including missing information, be they organic or inorganic; it can also cope with the complexity of the mechanistic knowledge such as involved in DNT endpoint as well as the combination of multiple sources of information. Thus, such modelling represents a means to determine if the available information is sufficient to address a question and what kind of additional efforts and where such efforts are needed. It can also help to screen a large number of compounds and identify tailored toxicological paths for an individual compound or a chemical class, including mixtures, while offering an understanding of the likelihood of effects and the level of perturbations at different biological levels.
3. A linear AOP is considered to be a unit of development and is evaluated in a structured and transparent way, based on the available mechanistic knowledge. At the same time, networks of linear AOPs should serve as the basis for the development of predictive models (Villeneuve et al. 2014a). One reason for using linear AOPs in this manner is that they can interact between themselves, e.g., single chemicals can activate multiple paths leading to an AO and, thus, a reliable qAOP model must meet such challenges (Schultz and Watanabe 2018). A Bayesian hierarchical model can adapt to the complex mechanistic structure, informed by the AOPs/AOPs network, taking into account the interactions, correlations, and causal relations. In addition, it can simultaneously handle predictors on multiple levels i.e., KEs, individual chemicals/class of chemicals, as they are informed by the global distribution of the entire dataset. This is especially useful when the datasets are too small to be analysed separately, as is the case with DNT. The hierarchical approach is better than treating each KE independently since the data from different KEs inform one another meaningfully and learn from each other, even without adding the causal regressions. For example, underrepresented categories of chemicals will borrow strength from well-represented chemicals and, thus, the hierarchical approach can deal with the unbalanced classifications under a unified statistical framework. Therefore, a Bayesian hierarchical model can contribute to the paradigm shift towards a mechanistically-driven assessment in modern toxicology and translate a qualitative AOP into a quantitative computational predictive model for potential use in CRA. Best practices of data analysis under the umbrella of the qAOP concept might need to consider the usefulness of Bayesian models.

## 4.6. Conclusions

In conclusion, a Bayesian hierarchical multiparameter model was developed for a simplified AOP network derived based on a topology analysis, combined with expert judgement. The conceptual framework achieved three goals: it dealt with missing values; accommodated unbalanced and correlated data; and learned the structure of a DAG to simulate a simplified version of an AOP network. The model can be used to predict the potency of a compound that follows a specific biological path of inducing DNT with the associated uncertainty given by different sources of data variability. Also, the model can guide the data generation processes to better understand DNT mechanistically and support decision-making in CRA. The methodology can be applied to other endpoints of interest and can accommodate new evidence and other types of data. Future directions include the addition of other biological paths and details about kinetics and dynamics to extend the applicability domain and usage as a qAOP model.

# Chapter 5. Development of a quantitative linear Adverse Outcome Pathway for Parkinsonian motor deficits

As Chapter 2 summarised, in addition to a network, a qAOP model can also be derived from a linear qualitative AOP. However, such a quantitative approach is context- and data-dependent. The results presented herein reflect a series of discussions with Prof Mark Cronin, Dr Andrew Worth and Dr Anna Bal-Price on the use and applicability of the OECD AOP-Wiki Knowledge Base for the development of qAOPs during the secondment at the EC JRC, Ispra, Italy (September 2018 – January 2019) with the scope to evaluate the potential empirically-derived quantification of a linear AOP for Parkinsonian motor deficits.


## Abstract

The pathogenesis of Parkinson's disease (PD) is multi-factorial and industrial chemicals are associated with the risk to induce its symptoms. The AOP concept represents an excellent tool to summarise and inform quantitative modelling of causal relationships of an underlying mechanistic path for PD induced by stressors. The aim of this Chapter was to assess the OECD publication of the well-developed and endorsed qualitative AOP for Parkinsonian motor deficits in order to determine if the references cited could provide a basis for quantification. Additionally, a case study for rotenone was undertaken to extract data and model qKERs for the PD qAOP. As a result, 13 peer-reviewed publications representing *ex vivo*, *in vivo*, and *in vitro* studies, were identified as being potentially useful for quantification. In addition, three KEs at the cellular level for three time points, i.e., four hours, 24h, 48h, and 11 KERs were quantified. The response-response functions were fitted using the *drc* R package. Additionally, the concentration-responses and the effect size of KERs were modelled probabilistically in PyMC3. The results provide a series of thoughts, recommendations and opened questions on how to address the existing issues to facilitate and increase the development of qAOP models.

## 5.1. Introduction

As outlined throughout the thesis, the AOP framework has emerged as an essential instrument to synthesise the available scientific evidence regarding chemical toxicity, at different levels of biological organisation, in a transparent, structured and comprehensive manner. The development of an AOP relies on the SR methodology, which is a protocol-driven approach and utilises available data derived mainly from the scientific literature. Each line of evidence is rigorously documented and efficiently assembled following the five fundamental principles that guide the AOP development (Villeneuve et al. 2014a; Villeneuve et al. 2014b). The quality of data can be evaluated against BH considerations allowing assessment of the level of maturity in the overall AOP (Becker et al. 2015; Meek et al. 2014b), with the ultimate goal of the endorsement of the AOP by the OECD. The OECD TGs help to generate relevant and reliable information. Despite this, not all KEs of an AOP or endpoints have an approved standard method and accessible data for chemical safety assessment. Hence, academic research represents an important source of data, even though it might be insufficiently reported and not fully detailed.

An AOP encompasses both descriptive and empirical studies. However, a qualitative AOP should be quantified to make the mechanistic knowledge applicable for RA, as outlined throughout the thesis. A qAOP aims to simulate and predict the threshold, or magnitude, of chemical-induced toxicity in a KE. This quantification is based on the response of the prior KE with the focus to develop qKERs as stated in Chapter 2. The quantification of an AOP can take many forms from simple linear regression to sophisticated biologically-based computational models, depending on the level and nature of the empirical data (Perkins et al. 2019a). At the same time, a qAOP model is usually developed for a chemical, referred to as a reference compound, even though an AOP must be chemically agnostic (Conolly et al. 2017; Zgheib et al. 2019). The empirical data are usually generated independently for an individual KE and not necessarily integrated into an AOP framework. Therefore, the descriptions of an AOP, which are based on peer-reviewed publications and compiled as part of the OECD AOP Development Programme, can be explored to inform the qAOP model, this has several advantages, including:

- Original articles are cited. The reported findings are of high quality following a rigorous and transparent evaluation.
- The references are relevant to human or ecotoxicology, and associated with an AOP, allowing for the understanding of all factors involved in predicting the AO, and where additional efforts are required.
- The references are assessed for the WoE analysis, and thus, support the applications of AOPs.

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease (Dexter and Jenner 2013; Esteves et al. 2011; Glass et al. 2010; Poewe et al. 2017). Historically, it was first described by James Parkinson in his work "An Essay on The Shaking Palsy" on observations of six individuals with "paralysis agitans" in 1817 (Betarbet et al. 2005; Bove and Perier 2012; Segura-Aguilar and Kostrzewa 2015). Although PD is an age-related disorder with prevalence in industrialised countries affecting 1% of

people over 60 years and 5% of people over 80 years, 10% of cases are classified as young-onset, occurring between 20 and 50 years of age (Betarbet et al. 2005; Dexter and Jenner 2013; Esteves et al. 2011; Whitton 2007). The pathogenesis of PD is multi-factorial implicating various genetic as well as environmental factors (Dexter and Jenner 2013; Fujita et al. 2014). The hypothesis that exposure to environmental toxicants may increase the risk of developing PD has subsequently attracted interest. For example, epidemiological studies showed the association between the exposure to pesticides such as organochlorine insecticides used in agriculture with the increased risk of developing PD (Dias et al. 2013). In addition, a meta-analysis of 19 studies evaluating the potential impact of pesticide exposure found an estimated doubling of disease risk (Priyadarshi et al. 2000). As a result, several *ex vivo*, *in vivo* and *in vitro* models have been developed to produce and evaluate histological, molecular, and behavioural mechanisms induced by neurotoxicants such as rotenone, 6-hydroxydopamine, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), paraquat and lipopolysaccharides (Bove and Perier 2012; Segura-Aguilar and Kostrzewa 2015).

The pathogenesis of PD is characterised clinically by progressive impaired motor function (Dexter and Jenner 2013; Lin and Beal 2006). The main features include tremor, rigidity, bradykinesia and postural instability (Bove and Perier 2012; Dexter and Jenner 2013; Fujita et al. 2014; Zaltieri et al. 2015). Nonmotor-related symptoms are olfactory deficits, autonomic dysfunction, depression, cognitive deficits, and sleep disorders (Glass et al. 2010). The pathological hallmark of PD is represented by the loss of DA neurons in SNpc and the presence of intracellular inclusions containing aggregates of the alpha-synuclein protein, called Lewy bodies (Betarbet et al. 2005; Glass et al. 2010; Lin and Beal 2006). The symptoms of PD are only apparent when the loss of at least 50% of the DA neurons in the SNpc occurs, leading to over 80% reduction in dopamine levels in the striatum (Whitton 2007). The disease map developed by Fujita et al. (2014) offers a comprehensive overview of the molecular interactions and pathways involved in the pathogenesis of PD structured into synaptic and mitochondrial dysfunction, impaired protein degradation, alpha-synuclein pathobiology and neuroinflammation. However, the AOP construction offers a clear-cut mechanistic representation and summary of the toxicity induced, chemically, that leads to an AO. It can be used as an informative tool for computational modelling. The AOP developed by Terron et al. (2018) explains the causatively linked cellular KEs between the inhibition of mitochondrial complex I and the manifestation of Parkinsonian motor deficits for rotenone, which was investigated herein.

## 5.2. Aim of this chapter

The aim of Chapter 5 was to investigate the possibilities of developing a qAOP for PD on the basis of existing knowledge and data. Specifically, the appropriateness of the references cited in the OECD peer-reviewed publication (Bal-Price et al. 2018b) relating to adjacent KERs were evaluated against a predefined list of assessment criteria for modelling quantitatively AOP ID 3 for Parkinsonian motor deficits[1] induced by rotenone. Secondly, quantitative data, i.e., dynamic and kinetic types of information, from two *in vitro* human-based models, were extracted and mapped to model qKERs. Additionally, a comparison of frequentist and Bayesian approaches was conducted. This activity was performed as a case study, contributing overall to the advancement of the qAOP framework.

## 5.3. Methodology

### 5.3.1. Data collection and evaluation

The references cited in the OECD peer-reviewed publication for the linear AOP on Parkinsonian motor deficits (Bal-Price et al. 2018b) were collected in an Excel file and investigated for: (i) the presence of experimental studies, (ii) studies conducted for rotenone, (iii) studies conducted in human and rodent type of *in vivo*/*ex vivo*/*in vitro* models, (iv) appropriate extractable quantitative data, and (v) toxicity assessment performed by functional assays. This preliminary selection of scientific publications was conducted solely for adjacent KERs to ease the decision for data extraction for modelling qKERs. In addition, it allowed for the understanding of the type of scientific evidence required to build a qualitative AOP, the variety of stressors the AOP covers and where additional efforts are needed. The final studies selected were evaluated against a series of predefined criteria for a complete overview of their potential application for qAOP modelling purposes listed in Table 5.1. These criteria were formulated together with two experts with experience in *in vitro* and *in silico* toxicology and policy-decision making. Additionally, these criteria can guide the efforts of the scientific community towards designing testing strategies to comply with AOP construction, as the references were not primarily intended to assess an AOP. As such, they can provide suggestions on how to better improve the OECD AOP-Wiki KB in order to make it a useful and informative data repository for computational modelling.

[1] https://aopwiki.org/aops/3, accessed on March 16, 2021.

Table 5.1. A list of 16 criteria proposed to be used for the assessment of peer-reviewed publications for qAOP modelling purposes.

| No | Evaluation criteria | Justification |
|---|---|---|
| **1.** | The study is relevant to toxicology or human disease/symptom of a disease defined as an AO. | This provides information on the AO or underlying mechanism of toxic action. The same disease, e.g., PD, can have a diverse degree of different symptoms in different patients. AOP ID 3 is developed for one of the PD's symptoms: motor deficit. |
| **2.** | The study is associated with a well-developed and mature AOP. | A sufficiently mature AOP will allow understanding of all factors involved in predicting the AO. |
| **3.** | The study should provide measurements of a relevant endpoint e.g., changes of tyrosine hydroxylase culture should be evaluated in DA neurons. | The biological plausibility should follow the AO or endpoint of research interest. |
| **4.** | The study includes data for KEup and KEdown measured in the same experimental design. | This will help building quantitative KERs and decrease the uncertainty of the model. |
| **5.** | The study provides information on the directness or indirectness of the KER. Primarily studies performed to support direct (adjacent) KERs should be selected. | Allows for modelling of the quantitative relationships of adjacent KERs in order to facilitate development of the final qAOP model. Indirect (non-adjacent) relationships, feedback loops, compensatory mechanisms can be also checked for existing publications at a later stage. |
| **6.** | *In vitro* study concentrations of a chemical inducing the KER should be relevant to human exposure. | *In vitro* concentrations should be compared with human plasma level or if available tissue levels. This information will be useful later in judging the likelihood of an AO, for a given exposure scenario. |
| **7.** | The quantitative measurements are functional assay-based. If gaps are identified, omics types of data can be accepted if they are informative for dose-response modelling. | Assay-based quantitative data type should include experimental data relevant to KEup and KEdown of the same KER e.g., ATP production, necrosis/apoptosis, changes in protein levels (quantitative immunocytochemistry), uptake, activation/inhibition of receptor/ion channel function. |
| **8.** | The study includes a range of concentrations and various times of exposure (dose and time responses) to define the thresholds of KEs upstream which will trigger relevant KEs downstream in each KER. | This will help defining the exposure required in risk assessment under which AO might be triggered. |
| **9.** | The study mentions the nominal and free chemical concentration, especially in the context of *in vitro* experiments to facilitate data interpretation. | This information can be coupled later with PBPK/(Q)IVIVE models. |
| **10.** | The study provides data for cell/tissue viability in addition to the cell specific toxicity response (KE) to ensure that cell type specific endpoints are not induced by cytotoxic effects of a chemical, unless the KE of interest is cytotoxicity. | These data will indicate that the toxic response is mediated by specific mechanisms, i.e., other than non-specific cytotoxic mechanisms. |
| **11.** | The study includes data on kinetics relevant to the MIE/KEs, especially in *in vitro* experiments. | This is vital for a robust qAOP model. |
| **12.** | The test methods used (models and measured endpoints) have been adopted as national or international guidelines, e.g., OECD, ICH. | Reliability of the test methods used is vital to evaluate the uncertainty of the model. |
| **13.** | The test methods used (models and measured endpoints) have been validated in a formal validation study, or in accordance with accepted validation principles. | Reliability of the test methods used is vital to evaluate the uncertainty of the model. |
| **14.** | The study has been performed in a quality control environment (i.e., GLP). | This gives confidence in the reproducibility of data and serves as a scientific rationale for the qAOP model. |
| **15.** | Data can be retrieved from graphs/plots. | Quantitative data are needed, unless other types of data are suitable. Pictures/images (i.e., activity detection, optical density etc.) without quantitative evaluation are not accepted, unless transformed in a scoring or binary system e.g., based on an expert judgement. |
| **16.** | The study design is based on at least three measurements for the same endpoint, in three independent experiments. | Adds confidence in the reproducibility of the data. Statistical significance diminishes the uncertainty of the model. |

### 5.3.2. Data extraction and analysis

Two *in vitro* studies (Barrientos and Moraes 1999; Chou et al. 2010) were selected. They provided data for the modelling of causal linkages between three KEs at the cellular biological level, specifically the inhibition of complex I, mitochondrial dysfunction and impairment of proteostasis. Data, i.e., concentration and time responses (% of control) were extracted from figures using GetData Graph Digitizer v.2.26 software[2] and are available on the GitHub repository. The extracted data represented the mean of the *n* replicates. Data were arranged per time point and KEs since a KE could have been tested for multiple biological effects, as shown in Figure 5.1. There are missing values for the data due to the different ranges in concentration tested for the same time point. Therefore, a harmonisation was necessary to fit the responses between the effects. After the organisation of the extracted data, the estimation of missing values for the time points of four and 48 h was performed from a normal distribution using the *mice* R package v. 3.12.0[3]. The modelling of concentration-response and response-response functions was undertaken using the *drc* R package v. 3.0-1[4]. The choice of the fitted line was made based on the Akaike information criterion (AIC) of the evaluated logistic regressions. The best-fitted curves were selected to describe each causal relationship for rotenone mathematically.

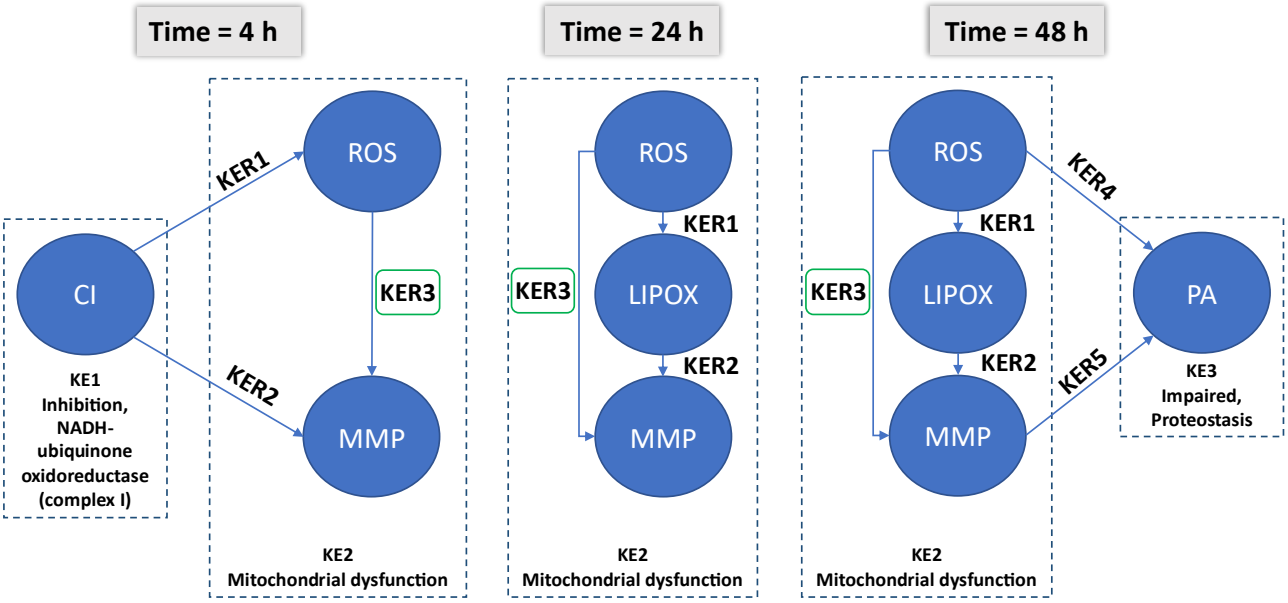


Figure 5.1. A graphical representation of the quantified causal relationships for PD informed by the KEs and specific measurements that characterise the KEs based on the available extracted data of human-based *in vitro* studies. KER3 represents the common KER across the three time points tested. CI, Complex I, Inhibition; ROS, Production of reactive oxygen species, Increase; LIPOX, Lipid peroxidation, Increase; MMP, Mitochondrial Membrane Permeability, Decrease; PA, 20S Proteasome Activity, Decrease.

[2] http://www.getdata-graph-digitizer.com/index.php, accessed on March 16, 2021.
[3] https://CRAN.R-project.org/package=mice, accessed on March 16, 2021.
[4] https://CRAN.R-project.org/package=drc, accessed on March 16, 2021.

### 5.3.3. Bayesian modelling

Extracted data were also used for the probabilistic modelling of concentration-responses and response-responses. As argued elsewhere in this thesis, a response-response relationship is causal by nature and does not imply associations. It should evoke the magnitude (threshold) of/between two key events (upstream to downstream effects). To follow the causal principles of modelling KERs, appropriate experiments are needed to be conducted. However, this is not the case for this example where data were not produced with the intention of being used for such an application. However, these data were used for demonstration purposes to showcase the importance of Bayesian modelling of both concentration-responses and response-responses for toxicity assessment.

The workflow of the Bayesian model formulation was described in depth in Chapter 4. It consists of specifying the prior distributions and likelihoods that reflect the assumptions about the data and how the model should be structured. It is followed by fitting the observed data to the model to estimate the unknown parameters of the model sampled from the posterior distributions. These samples can be used to quantify the uncertainty of the model parameters in the form of distributions, which underlines the uniqueness of Bayesian approaches to provide probability statements rather than points of estimates, taking into account sources of variances propagated along with the model.

The likelihood in the Bayesian model is defined as a log-logistic function, which is by far the most used regression to analyse a concentration-response curve (Labelle et al. 2019; Shao and Shapiro 2018). Depending on the type of effect (inhibition vs activation), it takes one of the following mathematical forms.

For inhibition type of effects:

$$f(X) = Y_{min} + \frac{Y_{max} - Y_{min}}{1 + 10^{S*(log_{10}X - log_{10}IC_{50})}}$$

For activation type of effects:

$$f(X) = Y_{min} + \frac{Y_{max} - Y_{min}}{1 + 10^{S*(log_{10}EC_{50} - log_{10}X)}}$$

From the likelihood, five parameters are derived:

$Y_{min}, Y_{\max}$ : Minimum (low) or maximum (high) concentration-response.
$IC_{50}, EC_{50}$: Half maximal inhibitory (effective) concentrations.
$S$: Hill slope or steepness of the transition of the response between two levels.

The priors of the parameters were chosen as non-informative and were described as normal distributions:

$$Y_{min} \sim Normal\ (0, 10)$$

$$Y_{\max} \sim Normal\ (100, 10)$$

$$IC_{50}, EC_{50} \sim Normal\ (1, 5)$$

$$S \sim Normal\ (0.5, 10)$$

As summarised in Chapter 2, a response-response relationship can be modelled in a variety of different ways. Herein, the effect size is proposed for the examination of the magnitude of the treatment effect between two events, i.e., a KER. One way to measure the effect size is by Cohen's index, which represents the difference between two means of responses divided by a standard deviation $s$ (pooled standard deviation) of the responses.

$$Effect\ size = \frac{M_{KE_{up}-} M_{KE_{down}}}{s}$$

$$s = \sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}$$

The effect size is widely used in meta-analysis studies and Cohen's index is the most reported statistic used to evaluate and summarise the findings (Fritz et al. 2012). The effect size was previously modelled in a Bayesian framework (Kruschke 2013). However, it was not applied to evaluate a chain of events such as of an AOP in toxicology, but rather to evaluate the difference between two groups of treatments. An interesting question to be explored is that if the effect size can be used as a means to understand how the key events differ between each other, which can implicitly describe the response-response relationships of the chain of events in a concentration and time dependent manner.

The likelihood of the effect size between two KEs was informed by Cohen's index, and the priors for the parameters of the model were defined as normal distributions for mean μ and half normal distributions for standard deviation σ.

$$\mu \sim Normal\ (0, 100)$$
$$\sigma \sim HalfNormal\ (100)$$
$$KE_{up/down} \sim Normal\ (\mu, \sigma)$$

The Bayesian model describing concentration-responses followed the DAG structure as presented in Figure 5.1. The Bayesian model of the effect size described solely the KER3 that is common for the three time points for which the same ranges of concentration were tested.

## 5.4. Results

### 5.4.1. Data assembly for qKERs

In total, 430 peer-reviewed publications were analysed for their relevance for computational modelling of qAOPs. Following the initial screening process, 56% of references were experimental studies that tested neurotoxicants other than rotenone, e.g., MPTP and its metabolite MPP+, paraquat etc., and 26% of references were descriptive studies represented by systematic review and qualitative summaries associated to the mechanistic path. 15% of references tested rotenone on species other than humans and rodents and included phenotypic and imaging analysis of any endpoints of KEs. Only 3% out of the total number of studies were conducted for the measurement of effects induced by rotenone to characterise adjacent KERs of the AOP in a quantitative manner (Figure 5.2).

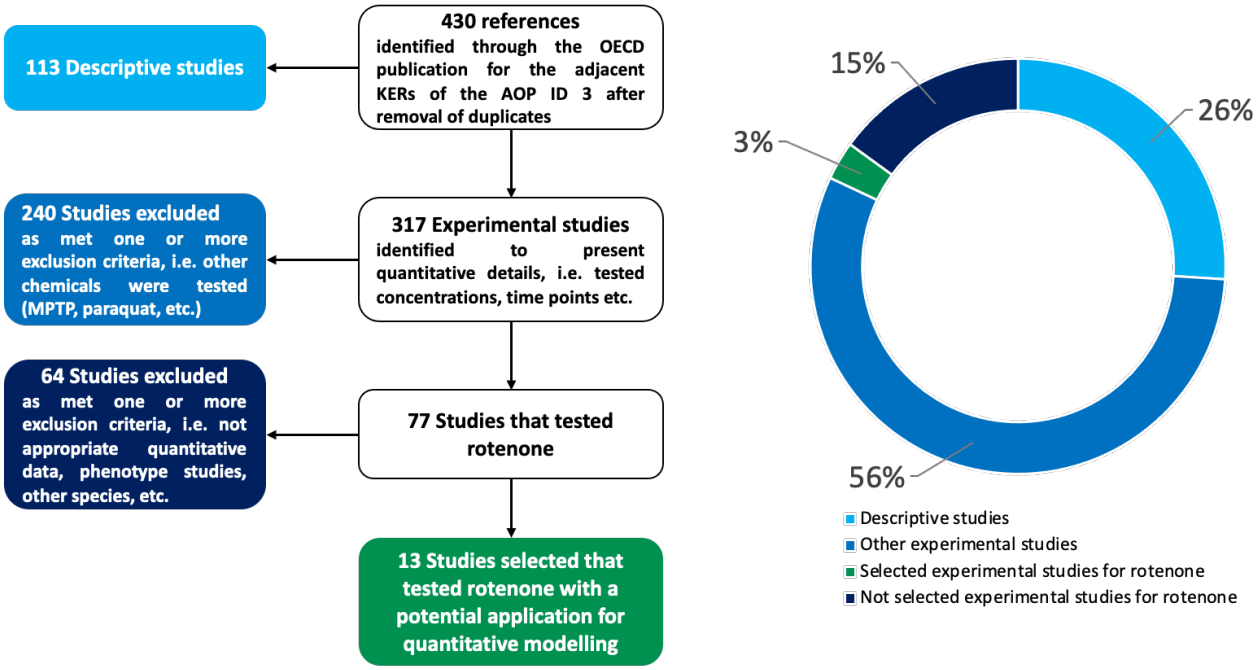


Figure 5.2. The decision tree to select peer-reviewed publications for rotenone for quantification purposes as well as the percentage of evaluated references.

The selected studies were further grouped per KEs and type of model utilised (Figure 5.3). The MIE was tested in an *in vitro* bovine heart-derived model (Grivennikova et al. 1997; Ramsay and Singer 1992), while the other KEs and AO were measured in human- and rat-based *in vivo*/*ex vivo*/*in vitro* systems. The inclusion of these two studies was because there were no other published methods that measured the ability of rotenone to bind to complex I at the time of analysis. Importantly, five studies investigated multiple KEs. For example, Saravanan et al. (2005) measured the inhibition of complex I and mitochondrial dysfunction. Such studies represent the ideal type of quantitative data that helps to reduce the potential variabilities imposed by experiments.

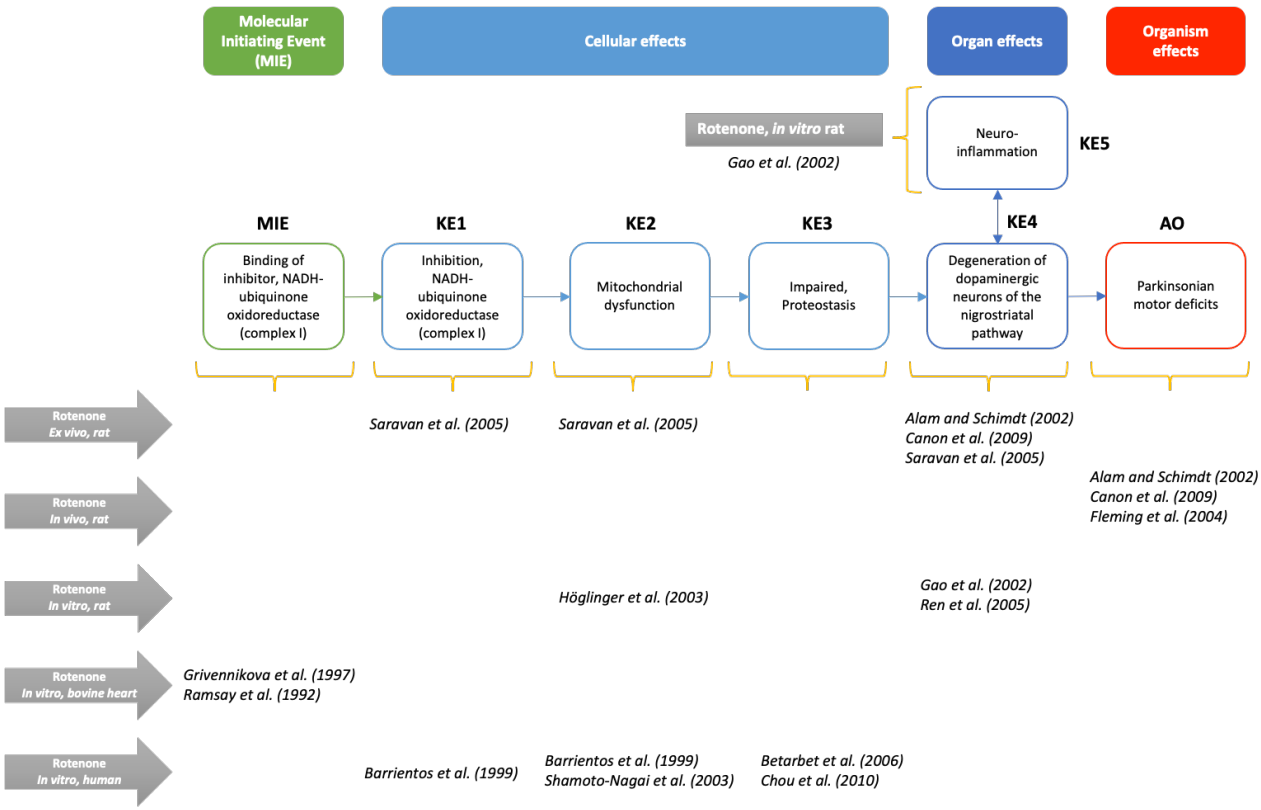


Figure 5.3. The final 13 selected references grouped per experimental type with the potential utility for their quantification purpose.

The studies selected were assessed against the criteria proposed in the Methodology section. The quality assessment yielded: (i) acute (minutes, hours) and chronic (weeks) types of exposure, (ii) that not all studies were tested in a minimum of three concentrations/doses and at a minimum of two timepoints for the development of corresponding curves and investigation of dynamics, (iii), none of the studies followed standardised test guidelines, and few were conducted in accordance with GLP conditions, (iv) data were stored mostly in figures, and none of the studies provided the raw results, and (v) cell viability/cytotoxicity and kinetics were not examined by all studies. A complete overview of the results of the evaluation can be found in the GitHub repository.

### 5.4.2. Frequentists-derived results

The best fitted mathematical equations for KERs included: (i) a three-parameter log-logistic regression calculated for the KER increase of ROS production leading to an increase in lipid peroxidation at 24h and 48h, and (ii) a four-parameter log-logistic regression computed for the remaining KERs similarly to the regression used for Bayesian modelling. The equations and associated parameters for KERs are summarised in Appendix III Table S1. The equations and associated parameters for KEs can be found in the GitHub repository.

The fitted curves for KEs and KERs presented sigmoid and exponential trends for each timepoint and for the concentration-responses. Most of the response-response functions showed that a reduction in a prior KE led to a non-linear reduction in the subsequent KE (Appendix III Figures S1-S3). Thus, the qKERs established presented continuous responses allowing for the estimation of the level of the response for each timepoint.

### 5.4.3. Bayesian-derived results

The Bayesian model allowed for the computation of the concentration-responses as a log-logistic function (Appendix III Figures S4-S8). The parameters of the concentration-response curve were defined by the mean and associated credible interval that described the uncertainty, as exemplified in Figure 5.4. Such an approach becomes more informative for decision-making comparing to single point estimates that might lead to incorrect conclusions.

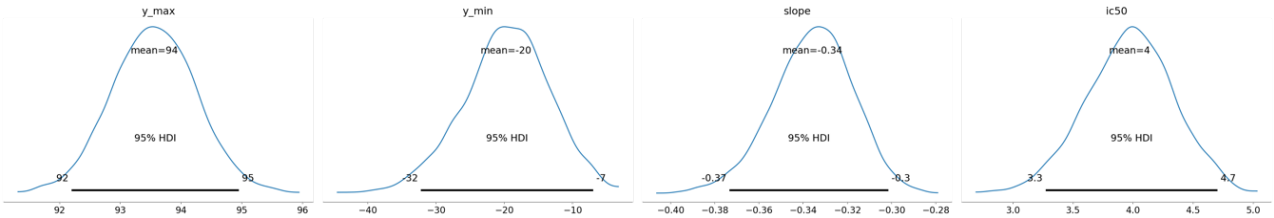


Figure 5.4. The mean and associated 95% credible interval (HDI) of each of the parameters that describe the log-logistic regression of a concentration-response. Herein, the example of the KE inhibition of complex I induced by rotenone at four hours was taken. The four individual distributions represent the uncertainty associated to the predicted parameter.

The effect size of the KER for the tested concentrations modelled probabilistically was small, almost equal to zero. It showed a slight decrease/remained unaffected over time (Figure 5.5). This underlines the importance of considering the effect size and associated differences in the mean between the KEs to identify potential tipping points.

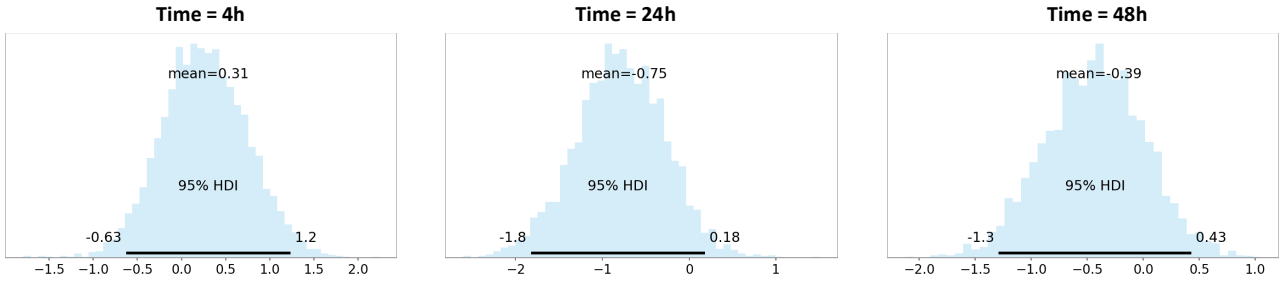


Figure 5.5. The mean and associated 95% credible interval (HDI) of the posterior distributions of the effect size of the KER3 that describes the increase in ROS production leading to decrease in mitochondrial membrane permeability at three different time scales given the available data.

### 5.4.4. Comparison between Bayesian vs frequentist modelling

This case study allowed for the comparison of Bayesian and frequentist modelling approaches used to fit concentration-response and response-response relationships. Additionally, a comparison between the concentration-response for imputed vs non-imputed missing values modelled probabilistically vs modelled in a frequentist way was conducted, as exemplified for a KE in Figure 5.6. The imputed data may lead to conclusions not supported by evidence that require justification. The frequentist-derived curve does not account for the uncertainty. This demonstrates the importance of designing appropriate experiments to generate relevant data to investigate a chemically-induced causal effect that rather be modelled probabilistically.

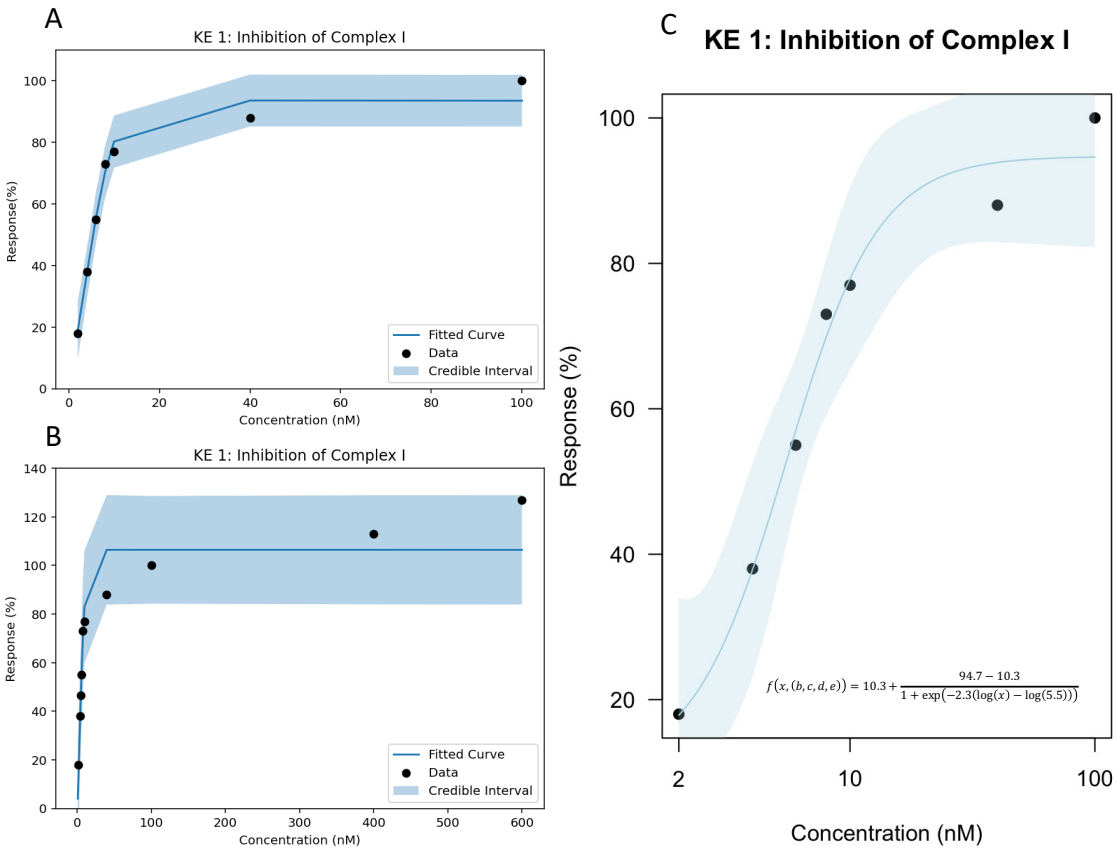


Figure 5.6. A comparison of concentration-responses fitted probabilistically in PyMC3 for the non-imputed data (A), fitted probabilistically in PyMC3 for imputed data to harmonise the ranges of concentrations with the other KEs (B), and fitted deterministically in the *drc* R package (C). The plots represent the KE 1 of inhibition of complex I by rotenone.

A summary of the main characteristics that describe the pros and cons of choosing one or other of the modelling approaches is presented in Table 5.2.

Table 5.2. A list of characteristics specific to Bayesian vs frequentist modelling.

| Bayesian modelling | Frequentist modelling |
|---|---|
| Uses probabilities for both hypotheses and data. | Never uses or gives a probability of a hypothesis. |
| Depends on the prior and likelihood of observed data. | Depends on the likelihood for both observed and unobserved data. |
| Requires one to know or construct a 'subjective prior', e.g., how to reflect the data structure in the model. | Does not require a prior, thus, it is objective. |
| There is no single method for choosing a prior. However, by trying different priors, the sensitivity of the results can be evaluated against the choice of the prior. | The p-value depends on the exact experimental setup, which needs to be fully specified ahead of analysis. |
| 95% credible interval: there is a 95% probability that the true (unknown) estimate would lie within the interval, given the evidence provided by the observed data. | 95% confidence interval: we can be 95% confident that the true (unknown) estimate would lie within the lower and upper limits of the interval, based on hypothesised repeats of the experiment. |
| Tends to be computationally intensive with the increased number of parameters. | Tends to be less computationally intensive. |

## 5.5. Discussion

The OECD publications of the endorsed AOPs, as part of the OECD AOP Development Programme, can serve as a starting point in understanding if a preliminary quantification of AOPs is possible and/or what additional efforts are required to accomplish such an objective. As shown in Chapter 2, most of the currently available probabilistic (four out of six) and mechanistic (nine out of eleven) qAOPs were developed based on experimental data. At the same time, available extensive databases, such as ToxCast$^{TM}$, are not suitable to provide data to model response-response functions of causal relationships which are informed by AOPs for neurotoxicity, as outlined in Chapter 3. There is thus an opportunity to investigate potential utility of the AOP-Wiki KB to explore quantitative causal linkages. However, as the SR methodology underlines, the selection of appropriate studies requires a transparent and reproducible workflow in decision-making. In this investigation, 16 criteria were formulated and proposed to guide the choice of peer-reviewed articles for data extraction and quantification purposes on the basis of their having appropriate quality (Table 5.1).

The AOP network presented in Chapter 3 described two adverse outcomes related to neurodegeneration: impairment of learning and memory, and cognitive dysfunction quantified as developmental neurotoxicity in Chapter 4, and Parkinsonian motor deficits. It is recommended to start the quantification of an AOP with a linear construction and subsequently introduce complexity step by step. This is because: (i) a simplified linear construction is more easily understood by stakeholders, including regulators, as opposed to systems biology type of computational models that require extensive expertise, (ii) complexity in a qAOP needs to be justified for its inclusion, e.g., if making a qAOP is more/less informative in assessing a chemically induced adverse effect, (iii) simple constructs are less time consuming, as well as being (iv) easier to design with regard to testing strategies for data generation purposes and (v) less data hungry than systems biology tools.

As a result of the work undertaken in the current investigation, several challenges were identified in the quantification of the linear AOP for Parkinsonian motor deficits for rotenone. One of these challenges included the way the data were presented. Many data, i.e., concentration- and time-responses, were displayed in graphs, and fewer in tables. This makes the retrieval process complex and potentially inaccurate. In addition, the raw results were not available for the scrutiny. For the full reuse of data, as well as ensuring they are reproducible, there should be a full access to the data. The availability of specific protocols, the open science movement and better ability to disseminate the results will assist in this process of modelling qAOPs. For instance, the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles are becoming widely accepted (Wilkinson et al. 2016). The scientific evidence, e.g., peer-reviewed publications, as well as the statistical analysis, e.g., the statistical significance informed by the chosen p-value, or the mean value calculated for three independent experiments, are sparse. Their relevance relies in understanding the testing hypothesis to evaluate the chemical toxicity. The lack of appropriate data is dragging down the development of qAOP models while available resources that are mostly inadequate have a consequence in data interpretation and for the utilisation of the published results for modelling qAOPs. Even for rotenone, a

commonly used chemical for the development and validation of *in vivo*/*in vitro* models, it was not possible to build an ideal data set. A potential open question remains if we should count on these types of data collected and mapped from available sources or focus the efforts towards producing appropriate data from specifically designed experiments and conducted with AOP development in mind, or a combination of both.

A KE can be measured by many assays. This is partially because a KE is named rather generically, which allows the integration and development of networks of KEs. This is also because a KE implies several biomarkers, which can be assessed by different assays. Thus, it becomes disputable which of the measured effects to rely on (to select and/or prioritise) as being sufficiently sensitive to capture the tipping points, or should results of all assays be considered. Herein, three measurements that were available for the KE mitochondrial dysfunction, e.g., mitochondrial membrane permeability, ROS production, and lipid peroxidation, were included and organised according to the qualitative description of the AOP. Choosing or selecting measurements representative of a single KE should be established based on a domain knowledge combined with an *in vitro* model and ideally recognised/validated assays to ensure that the qAOP model addresses the research question of interest in regard to the chemical potency of inducing the toxic effect.

If relevant, a qAOP model should reflect aspects of disease as it progresses over time. For instance, the AOP quantified herein should ideally reflect the progressive nature of cell loss in Parkinson's disease. Thus, kinetics and dynamics of chemical behaviour have to be considered for the predictions of causal response-response tipping points. A study conducted in a bovine heart-derived model (Grivennikova et al. 1997) and a study conducted in a human cell-derived *in vitro* model (Barrientos and Moraes 1999) analysed chemical kinetics that concluded a concentration below 5nM of rotenone led to a 35-40% inhibition of complex I after four hours of cell treatment. However, the inclusion of kinetics for rotenone makes the AOP chemical-specific and thus, does not follow the proposed best practices (Perkins et al. 2019a; Villeneuve et al. 2014b). A solution proposed to tackle this challenge is kinetic to remain a constant parameter in a Bayesian model comparing to the other variables (Bois 2013). Calculation of toxicity equivalent factor (TEF) to derive an equivalent concentration for another chemical, and mathematical inversion are other promising solutions in broadening the chemical domain of a qAOP model (Conolly et al. 2017; Zgheib et al. 2019).

Without consistent and appropriate quantitative data, an endpoint of interest cannot be modelled in the context of the qAOP framework. Thus, several opportunities to conduct additional research in this area can be proposed. For instance, although dose/concentration-response curves are regularly and routinely employed, they are rarely analysed further than the determination of a PoD or the concentration required to induce a 50% response, i.e., IC50, EC50 in the assay of choice. Several software tools that are available, e.g., the US EPA Benchmark Dose Software (BDMS)[5], the Dutch National Institute for Public Health and the Environment (RIVM) PROAST[6], Bayesian inference for Dose-Response Analysis (BiDRA) (Labelle et al. 2019),

[5] https://www.epa.gov/bmds, accessed on March 16, 2021.
[6] https://www.rivm.nl/en/proast, accessed on March 16, 2021.

allow for incorporation of the uncertainty in the assessment of dose/concentration-responses captured by the Bayesian approach. However, mathematically encoded non-linear response-response relationships represent the core components of a qAOP model, and efforts in modelling and software development are imperative to achieve the acceptance, use and advancement of the qAOP models in CRA. An AOP can assist the CRA process by setting the directions/guidance to organise the available mechanistic knowledge and determine what kind of new data to generate to fill the gaps for causally-informed decisions. A qAOP model quantifies the causal path/network of paths taking into account uncertainty and additional decisive factors to achieve the end goal, such as of human safety assessment.

The AO, i.e., motor deficits induced by chemical exposure, which is of regulatory interest, was measured solely *in vivo* in rats using non-validated/non-standard behavioural tests. This underlines the demand for the development of alternative methods as traditional *in vivo* testing may not identify potential neurotoxicants. This need, as well as the integration of multiple types of evidence and modelling techniques to fill the data gap, were exemplified in the recent OECD Testing and Assessment publication on the application of an AOP-based testing strategy in a read-across safety assessment of a complex I inhibitor (OECD 2020).

The integration of mechanistic pathways (AOPs) with exposure pathways, known as aggregate exposure pathways (AEPs) that allows for the incorporation of kinetics details can contribute to a complete and conclusive RA and improve confidence in the decision-making. A qAOP model uses quantitative data to predict the risk of an AO under specific exposure conditions. The use of biomonitoring and epidemiological types of evidence could help to decide on the particular exposure scenarios, which to be examined by a qAOP model.

The AOP studied herein provides compelling evidence for rotenone seconded by MPTP. Other chemicals for which the AOP for Parkinsonian motor deficits was formulated showed activity for specific KEs, but not for the entire linear construction. For example, paraquat showed neuroinflammation and dopaminergic neuron loss (Sandstrom von Tobel et al. 2014). Additional investigation is needed for the evaluation of the effects it has on mitochondria and motor deficits, or other paths. Thus, it becomes essential to evaluate both single, and a mixture of, potential inhibitors of complex I and their individual and/or combined mechanistic behaviour. The real-life exposures humans experience is to mixtures rather than single chemicals. Available mechanistic data are generated for one compound at a time. Hence, quantification of common/similar and specific biological paths to multiple chemicals ensures better depiction of aggregated coexposure/sequential exposure in humans.

As data are not available in a raw format, tools for data extraction from graphical plots represent an opportunity to be developed, validated and recognised by the scientific community. Herein, free-available software was used (GetData Graph Digitizer v.2.26 software). However, it operates solely on Windows, it is not widely accepted, and it was not designed to inform RA. An open question remains if there is a need for standardised methods to extract data for modelling of dose-responses given the importance of standardised methods to report the data utilised to generate predictions. As more standardised tools become available,

this leads to greater uptake by the industry and acceptance from regulatory agencies. It could help experimental scientists measure the correct data and report them in easily (re)usable formats.

This Chapter allowed for the development of recommendations for the generation of an ideal dataset for qAOP modelling to be proposed. However, the recommendations need to remain flexible and adaptable depending on the AOP to be quantified and available resources. The following recommendations were identified as part of this study:

- Ideally, chemicals should be assessed against a battery of *in vitro* tests and/or human iPSC-derived models to capture variability relating to the concentration tested and time-points selected.
- The *in vitro* system(s) utilised should be characterised according to the OECD Guidance Document on Good *In Vitro* Method Practices (GIVIMP) principles.
- *In vitro* biokinetics should be evaluated for experimental reliability, e.g., chemical solubility and stability, cross-contamination among wells, adsorption to plastic etc.
- The range of concentrations for testing should be based on viability and cytotoxicity assay parameters. In addition, the concentrations should be representative to human exposure.
- The time points to be tested should be based on the available mechanistic knowledge and adapted, if appropriate, to the *in vitro* model.
- A full list of the tested compounds should be reported along with relevant identifiers e.g., for model compounds, negative compounds, assay-specific positive compounds, studied compounds.
- The metabolomic consequences of compounds' exposure should be assessed, if possible.
- A preliminary set of experiments should be performed to act as a range finder and fine-tune the experimental conditions.
- The AO of the AOP should be measured. However, this remains an open question that depends on whether there are available assays to quantify the AO of the qualitative description of the AOP besides the regulatory applicability of the AO. For example, the PD AOP case that refers to motor deficits does not have a (validated) *in vitro* model to allow for the evaluation of such a symptom. Coupling with epidemiological and biomonitoring human data can become informative to evaluate and extrapolate predictions of the qAOP.
- Measure all KEs proposed by the qualitative description of the AOP. A qualitative AOP includes KEs at different biological levels, i.e., cellular, tissue, organ, organism levels. If it is not possible to measure KEs at all biological levels, KEs at different biological levels should ideally be considered. This will fulfil one of the expectations of the utility of qAOP to indicate the biological dimensions of the studied stressor.
- Evaluate a KE for essential effects, i.e., the activity of biomarkers, and justify the choice. For example, mitochondrial dysfunction can be measured for several biomarkers, e.g., ATP production, mROS levels, MPT levels, as well as nuclear mitochondrial genes. The choice of biomarkers that characterise

a disease/adverse outcome is crucial and requires knowledge and prioritisation schemes to establish testing strategies for minimising the costs, time and additional efforts.

- Map the assays and measured effects to the building blocks, i.e., KEs of the AOP.
- Provide raw results for all tests including for the replicates.

## 5.6. Conclusions

Chapter 5 investigated the scientific literature cited in the OECD publication of the AOP for Parkinsonian motor deficits in order to determine their potential for quantification. A number of references were identified as being suitable. These were assessed further against predefined criteria to establish the appropriateness of the available data. A case study to demonstrate the possibility of quantification was applied to rotenone, used as a prototypic mitochondrial toxin to evaluate PD. Data from two *in vitro* studies allowed for the mapping of the quantitative results in a concentration and time-dependent manner. The qKERs described the biological behaviour of rotenone at the cellular level resulting in the induction of a decrease of proteasome activity involved in the degeneration of DA neurons and the accumulation of alpha-synuclein protein responsible for Parkinson motor deficits. Two approaches, frequentist and Bayesian, were used to model KEs and KERs quantitatively. Best-fitted curves were obtained to describe the KEs and response-response relationships between the KEs. The effect size of KERs was calculated probabilistically. In total, three KEs at the cellular level and 11 KERs for AOP for PD were modelled. To advance the development of qAOP models, a multidisciplinary team effort is required in addition to having appropriate data and a well-proposed scenario. An advantage of the AOP framework is that it forces the developer to think causally about the biological effects following exposure to a chemical that results in the induction of an adverse effect. Despite the difficulties and complexities, quantification of the causally-chained relationships can strengthen the applications of qAOP models in CRA.

# Chapter 6. Discussion

The first section of the final Chapter of this thesis summarises and discusses the main conclusions of the research presented within Chapters 2 to 5, including the main takeaway messages (what has been learned through this research). A full discussion of the results can be found within each of the respective chapters. The second section focuses on the future work and remaining open problems (what is still to be learned, food for thought) that require solving in order to progress the development and applications of qAOP models. This Chapter is completed with a commentary on the implications of causality for the development of qAOP relevant computational models in predictive toxicology.

## 6.1. Summary of the research findings

In the present dissertation, the concept of qAOPs has been thoroughly studied following the objectives defined in Chapter 1. The main results were:

I. **The state-of-the-art of the qAOP concept was reviewed.** In total, six probabilistic qAOPs and eleven mechanistic qAOPs were identified at the time of the investigation. The qAOPs were evaluated against five common features (problem formulation, mechanistic knowledge and associated data, quantitative approaches, regulatory applicability, additional considerations) informed by the collected definitions of the qAOP formulated by the scientific community. The evaluation revealed the diverse types of qAOP models that utilised various techniques and tools.

II. **An AOP network for neurotoxicity in humans was developed and analysed.** The development of the AOP network followed the principles of the science of networks while making use of the mechanistic knowledge available in the OECD AOP-Wiki KB. It connected nine linear AOPs based on a methodology formulated as part of this research. The proposed workflow can be used for the investigation of other endpoints and research purposes. The results allowed for the identification of the most common/highly connected KEs, to be further prioritised for quantification.

III. **A conceptual framework for the quantification of a simplified AOP network for developmental neurotoxicity (DNT) was proposed.** The quantification used empirical data and Bayesian machine learning that accounted for correlations, causal relationships and missing information. The model was able to capture three KEs (reduction of BDNF, decrease of synaptogenesis, decrease of neural network formation) to predict DNT with good accuracy (73%). It also helped to classify the compounds into low, medium and high probability classes for their potency of inducing DNT.

IV. **A linear AOP for Parkinsonian motor deficits induced by rotenone was quantified for its cellular effects.** The AOP utilised empirical data extracted from scientific publications. 16 criteria to evaluate experimental studies were formulated in order to guide the *in vitro* testing strategies for quantification purposes. Three time points (four hours, 24h, 48h) and 11 KERs were modelled mathematically to fit the best curve to describe the magnitude of the transition from a downstream to an upper KE. The KEs were quantified using frequentist and Bayesian approaches.

## 6.2. Main takeaway messages of the thesis

The most significant findings of the present thesis included:

- An appreciation that existing qAOP models comprise a variety of structures and modelling techniques, as summarised in Chapter 2. Both linear and networks of AOP(s) may be quantified by applying probabilistic (stochastic) and/or deterministic *in silico* methods. Thus, there is no "one-size-fits-all" approach for qAOP development. In addition, it was determined that probabilistic approaches are not as data-dependent as deterministic approaches.
- The available biological mechanistic information can be better captured by an AOP network for further quantification and additional investigation purposes, than by a linear AOP, as outlined in Chapter 3. KEs and KERs are shared by multiple individual AOPs. Identification of points of convergence and divergence leads to a representation of real-world exposure scenarios to single chemicals and/or mixtures. A mechanistic qAOP model is pathway-driven and, hence, it should simulate changes/transitions that propagate through a network of biological paths.
- The heterogeneity and inconsistency of data should be captured by a qAOP model. Bayesian modelling was shown to be able to cope with this challenge, as demonstrated in Chapter 4, this being in addition to the other advantages that it presents. Shifting CRA towards probabilistic thinking allows details regarding the level of uncertainty and confidence in the decision-making and hypothesis testing to be provided. A qAOP model can also become an informative tool to communicate the risk arising from chemical hazards to the public.
- There is a need for appropriate data (dose/concentration- and time-responses) for computation of qKERs as exemplified in Chapter 5. A battery of *in vitro* tests, rather than *in vitro* tests conducted in isolation, and appropriate experimental design strategies, are essential to reduce data variability and facilitate the development of qAOP models. Thus, the ability to measure at a high degree of precision is essential to advance the qAOP framework.

## 6.3. Prospects for future work

Several challenges and open problems with regard to the development, validation and implementation of qAOPs for full acceptance and use have been identified throughout this thesis that can direct future research. In the pages that follow, I discuss how to embrace the identified weaknesses and obstacles as opportunities towards the mission of improved decision making in CRA.

### 6.3.1. Development of qAOPs

Computational models, i.e., ML/AI technologies, for early safety evaluation, have been increasingly employed to predict adverse effects to safeguard human health and the environment. However, the current (novel) methodologies fail to be fully utilised for CRA, e.g., as alternatives for evaluating untested compounds. This is because it is not sufficient to measure and analyse data but rather to have meaningful quantitative information that provides a mechanistic understanding of how a chemical causes an adverse effect, also acknowledged as a "signal in the data" by Bender and Cortes-Ciriano (2021).

A qAOP model can be considered a knowledge assembly process that involves (I) descriptive/causal knowledge, (II) pattern recognition knowledge, and (III) predictive knowledge. The descriptive/causal knowledge is given by qualitative AOPs that set the research direction, e.g., testing strategies and design. It should be recognised that technology and modelling are not driving the innovation, but rather the understanding of cause and effects. This is emphasised in depth in Section 6.5. Pattern recognition knowledge is represented by bioactivity type of data combined with molecular descriptors. These data matrices represent the most used data for modelling at present. Such data can allow for the extraction of features to learn more about the associations between quantitative variables. This leads to improved prediction accuracy, greater chemical space coverage, identification of better-informed target entities and interpretability of quantitative features. However, it is the predictive knowledge that makes a computational model applicable to solve real-world situations that is required by regulatory agencies. Despite all of the progress, the question remains open about how to reach the required level i.e., when is the knowledge considered sufficient to evaluate complex toxicity mechanisms of potential toxicants?

A qAOP model is both data- and mechanism-driven. For data, there is a need to obtain chemical and biological information for reference chemicals. Key questions to be addressed for the development of a qAOP include: how to identify substances for regulatory action or further testing? What kind of criteria to apply to prioritise chemicals for investigation? How was the conclusion about a chemical reached? There also are many molecular properties and descriptors that may be obtained experimentally or estimated: which of them are informative to describe the chemical space of the qAOP? What type of chemical information should be included in the development of a qAOP? In addition, the incorporation of biological information represents the main strength of a qAOP. At present, the bioactivity of chemicals is measured e.g., binding and enzymatic assays for protein/receptor interaction and ligand-specific to target. Cellular stress responses of chemicals

are also measured, i.e., high-throughput transcriptomic analysis, and for general toxic effects (cytotoxicity). Tiered screening schemes can ease the selection of bioassays for obtaining appropriate data for modelling.

Increasing numbers of computational models are becoming available for the identification and prediction of MIEs and AOs. However, less is known about the intermediate KEs. Thus, there is an opportunity to consider these KEs when designing a strategy to develop a qAOP. At the same time, not all endpoints are well studied. As such, there is a need to decide how to choose and prioritise the endpoints as well as establishing definitively the MoAs associated with compounds. Omics analysis could be a solution in its role to generate hypotheses as well as to describe a MoA. However, an adverse effect is not monocausal, hence, there is the on-going debate of whether the AOs should be considered in isolation or, in the longer term, is it preferable to model systems? Modelling a number of endpoints simultaneously requires the development of networks that are able to cover the MoA of interest and reflect on the development and functions of cells, tissues and organs. This also opens up the possibility of considering other scenarios, such as population variability, species differences, health related conditions, age, single chemicals vs mixtures, i.e., chemical-chemical interactions, amongst others.

NAMs are increasingly exploited for data generation and target identification purposes, including for those data that are required for qAOP modelling. There are a number of NAM methods including advanced analytics, 2-D and 3-D *in vitro* models, organ-on-a-chip models etc. and the choice of method will have to be pragmatic based on suitability, the data required and, not least, cost. NAMs are not yet fully accepted as replacements to the traditional animal tests, especially in areas such as preclinical studies in drug discovery. However, it is anticipated that the transition from generating new animal data to a fully accepted NAMs pipeline will require integrated data analysis, e.g., NAMs in a WoE analysis.

A qAOP model should allow for the prediction of the tipping points (often known as Points of Departure, PoDs) that define subsequent downstream adverse effects at a range of biological complexity. This raises the intriguing possibility of being able to redefine how these causally inferred tipping points are expressed. For instance, PoDs are based on the activity and not (non-) specific effects. However, as we move into a probabilistic risk assessment framework, rather than defining a chemical as being hazardous or non-hazardous, or an exposure as being safe or not, the opportunity is to start to deal with probabilities of harm, leading to the discussion of what would be acceptable. Ideally, a qAOP should generalise the predictive knowledge to extend its life as much as possible and these techniques are ideally suited to support this analysis.

Overall, qAOPs can improve the identification of adverse effects in a more meaningful and optimised manner, support a new and better means of understanding the risk associated with exposure to a chemical, if the above challenges are addressed.

### 6.3.2. Validation of qAOPs

It is widely acknowledged that it is not possible to have a one-to-one replacement of complex *in vivo* toxicity tests with *in vitro* assays (Mahony et al. 2020). Instead, a full evolutionary replacement of animal tests that

involves integrative approaches, e.g., IATA, is envisaged to achieve such a goal. However, it remains challenging to integrate the variety of evidence, e.g., historical data, bioactivity, toxicity data from different sources given that, in addition, most NAMs are not validated for toxicity or hazard assessment. As part of the process of developing alternatives, there will also be a need to validate qAOPs, especially those that utilise available information to contribute to and strengthen the outcome of a model in an integrative manner.

The process of validating qAOPs remains an open question and will require acceptance at the international level (e.g., OECD). However, three overarching metrics can be proposed to validate a qAOP, including biological reasoning, predictive performance and the level of confidence. For instance, it is not sufficient to obtain statistics, such as conducting a sensitivity analysis, in the context of such complex and demanding models as qAOPs. An understanding of biological reasoning is required to explain the chosen endpoint and depict the mechanism as an AOP. Predictive performance analyses the tipping points that are inferred and, by comparison with exposure estimates, i.e., $C_{max}$, can show their relevance to humans. A further stumbling block that needs to be addressed, which is how to benchmark predictions and against what i.e., the gold standard is human data. With regard to this, human biomonitoring studies, can offer a perspective in this sense to demonstrate, for example, the (Q)IVIVE predictions. Notably, to make a qAOP model reliable and increase its adoption into practice, overall confidence informed by both the computational approach as well as the qualitative description of the AOP is needed. A schematic workflow for the assessment of qAOPs is proposed in Figure 6.1. This also allows for the quantification of uncertainty and captures the bias propagated through the modelling process.

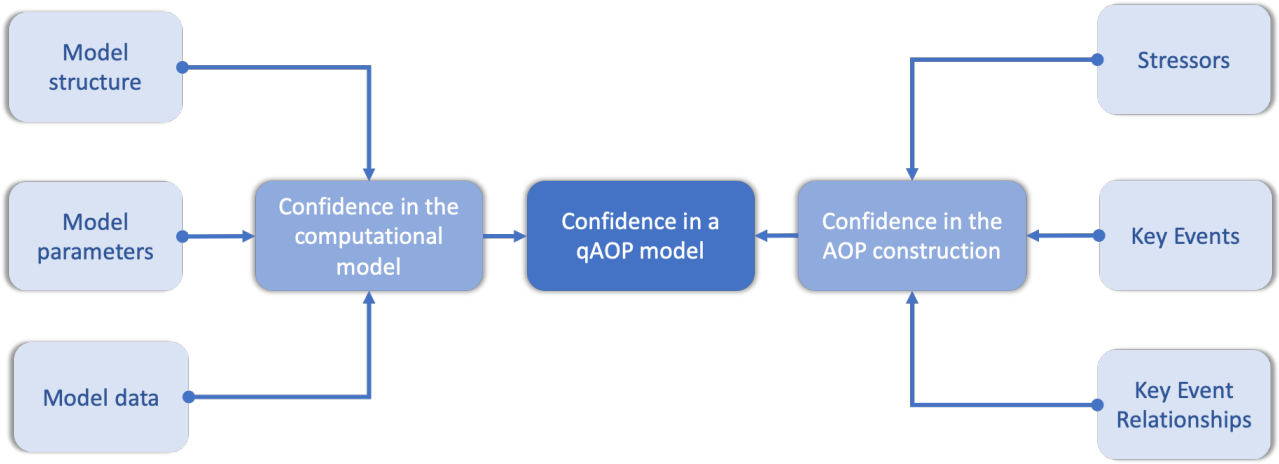


Figure 6.1. A potential framework for the assessment of the level of confidence of a qAOP model informed by both qualitative AOPs as well as computational approaches.

However, above any metrics that can be applied to evaluate its quality and trustworthiness, maybe the real validation of a qAOP rests in its usability in real life. The state-of-the-art models are not necessarily going to be those that make it to the implementation phase and thus have an impact on the toxicity assessment, e.g., improved decision-making. Real-world reliability, traceability and ease of maintenance are what will make the difference and turn qAOPs into practically applicable tools.

### 6.3.3. Reporting and implementation of qAOPs

Documenting and dissemination of qAOPs will be vital for their distribution, uptake and use. As part of this, a standardised reporting system can facilitate communication between model developers and reviewers/users. Additionally, such a reporting system can solve issues such as lack of harmonisation and consistency in data requirements between sectors and countries/regions and lack of coherent and transferable data resources. There are a number of reporting templates available for computational toxicology including those for QSAR, PBPK, read-across (OECD 2014b; OECD 2021; Schultz et al. 2015). However, since qAOPs are in their infancy and not as mature as these other computational toxicology models, little thought has yet been given to the protocols to collect, store, process and share data used for their development. Even though, as described throughout this thesis, there is no one-size-fits-all qAOP methodology, creating consistent reporting and documentation will increase the regulatory uptake of qAOPs. For example, decision trees, as a decision support tool can ease the formulation of appropriate protocols for qAOP model reporting and documentation, depending on the type of quantitative approach utilised. Open discussions with regulatory agencies such that their needs and expectations are met and to make the qAOPs flexible, represent another essential means to increase their adoption.

Implementation of qAOPs depends on their applications; these can include screening allowing for chemical prioritisation, enhancing the evidence informed by animal data, through to a complete replacement of testing by NAMs. Notably, it is acknowledged that Next Generation Risk Assessment (NGRA) is based on the integration of NAMs in a WoE assessment to make a safety decision based on non-animal data (Dent et al. 2018). A qAOP has the ability to contribute and increase the acceptance of NGRA through illustration with well-worked and successful case studies, e.g., Gilmour et al. (2020). Notably, the replacement of animal testing should be undertaken while achieving the same, or greater, level of protection of humans as afforded by the current techniques (Mahony et al. 2020). As such, we are at the start of a transition period from animal testing towards a complete non-animal CRA. However, revolutionary change will only occur when science is mature enough to explain biological processes in-depth and after regulators have acknowledged them. Thus, the hope around the qAOP makes it achievable given the current advances and aspirations (Cronin et al. 2021; Knight et al. 2021).

Hopefully, more and more qAOPs will become available that explore these possibilities. To achieve that and eliminate any perceived weaknesses, three potential directions are identified and proposed as next steps towards advancing the qAOP concept as presented below.

## 6.4. Potential next steps of advancing qAOPs

### 6.4.1. Showing qAOP as an iterative process

A qualitative AOP is a linear construction in terms of how it is represented structurally and how the knowledge is gathered to form the desired building blocks. However, whilst testing a linear AOP *in vitro*, in most cases linearity is not followed, e.g. as in Zgheib et al. (2019). This is because the order of the biological events and, especially, the associated biomarkers representative of a KE, may occur differently. Additionally, some compounds might not show activity against all selected biomarkers. There is also the possibility that a KE may be measured by several bioassays and, hence, it leads to a network of responses. These are several of the practical challenges that influence the final structure of a qAOP. At the same time, it is important to underline again the essentiality of wet experiments conducted with an AOP in mind to depict an endpoint causally. Data, knowledge and modelling are highly interdependent, as outlined throughout the thesis.

Given the likelihood of the lack of linearity in an AOP and the experimental measurement, it becomes challenging to decide on the parameters that would describe a biological system and the assessment question that a qAOP should model. Thus, development of a qAOP emerges as an iterative, cyclical process which proceeds until a sufficient level of detail is reached to provide predictions and robust mechanistic understanding for the decision regarding risk to be made. The formulation of a workflow, with a step-by-step description illustrated with examples on how to tackle this particularity about a qAOP, can increase the possibility for its development and application. The development of the workflow can be considered to be fundamental as guidance on how to approach the *ab initio* development of a qAOP.

### 6.4.2. Coupling qAOP with (Q)IVIVE

As Chapter 5 has proven, the available data to develop a qAOP are sparse, difficult to extract and integrate, in addition to other disadvantages including variability and reproducibility issues. Most of the data used at present for a qAOP and data that will be used for quantification purposes are, and will be derived, mostly from *in vitro* tests. Unfortunately, until recently *in vitro* studies conducted were not designed with AOPs, nor the acceptance of alternative methods to animal testing, in mind. Thus, to reiterate, a multidisciplinary effort to formulate and design informative experiments for the qAOP development pipeline is required.

In the view of the lack of data and the need for better experiments, there are several issues that will have to be addressed to progress to a successful outcome. Key to the use of NAMs and qAOPs is determining how their output is related to *in vivo*/clinical outcomes of relevance to CRA. Specifically, there is a need to better extrapolate PoDs from the commonly used static, short-term *in vitro* systems to the longer-term effects of *in vivo* exposure. Answering these types of practical issues will allow for qAOPs to be used to identify potentially harmful exposures to chemicals that can pose risks to human health. A significant part of the use of NAMs data within the qAOP framework will be the ability to extrapolate effects “up” from *in vitro* to *in vivo*. Thus, approaches that integrate qAOPs with (Q)IVIVE for an appropriate extrapolation will lead to better predictive

ability of potential adverse effects. This will enhance the regulatory acceptance of a qAOP to be routinely employed in CRA.

### 6.4.3. Integrating qAOP into a WoE analysis

A WoE analysis involves a variety of scientific data that should be organised and integrated in the best and intelligible way to provide, for instance, a complete uncertainty evaluation. There are further open items for discussion on how a qAOP could contribution to a WoE, given that it is intended to be a unifying framework. One school of thought is that the qAOP is an ideal structure to bring about integrated data analysis through WoE, with the possibility of providing an informed, evidence-based quantitative decision. Since uncertainties can be incorporated into the qAOP analysis, they will be able to provide a demonstration of the confidence that can be placed in the decisions. Thus, integration of the output of a qAOP within a WoE analysis will ground its role in regard to the assessment of the available supporting data.

Ideally, such an integrative framework should be organised as an evaluation tool sufficiently easy to be used by the AOP developers and any other interested party. In addition, it should be automated and provide a quantitative confidence score with sufficient details to appreciate the (in)consistencies as well as if, and where, additional efforts are needed. Development and exemplification of such a methodology will enhance the rigour, transparency and reproducibility of the quantitative analysis and shape chemical testing priorities leading to qualitatively improved decision-making.

## 6.5. Final thoughts: Putting the pieces together to allow the transition from endpoint-based to cause-and-effect Next Generation Risk Assessment

"*I would rather discover one causal relation than be king of Persia*" (Democritus, 430-380 BC)

The 21st Century has seen a shift from chemical risk assessment based on traditional animal tests identifying apical endpoints and concentrations that are "safe", to the prospect of exposure and mode of action led non-animal next generation risk assessment. There have been a number of drivers for this process and an undoubted catalyst, is the underpinning theme of this thesis, i.e., the AOP concept and philosophy. A decade on from the inception of AOPs, it is easy to see that Ankley et al. (2010) hit the sweet spot. In part, the world was in a mood for forward-looking and unifying frameworks, following the global financial crash of 2008; the vision of analysis of perturbing pathways, rather than finding adverse outcomes, had been presented to us as "Toxicity Testing in the 21st Century" (National Research Council 2007) and new technologies were becoming available that gave us an insight into those biochemical and physiological pathways. The AOP framework presented was elegant, fundamental and irrepressible! A decade later and the original linear, boxy AOP concept has been developed in several ways. With more data being generated from emerging technologies in chemical safety assessment, e.g., 2-D and 3-D *in vitro* models, batteries of *in vitro* tests, organ-on-a-chip, high-throughput and high-content screening technologies, as well as applied ML approaches, the opportunities for quantifying AOPs have arrived and more and more qAOP models should be alive due to the available resources. The qAOPs, firstly described by Villeneuve et al. (2014a), are considered a cornerstone to screen molecules and predict their points of departure from normal physiological pathways. However, as of mid-2020, there are only about 16 models for qAOPs publicly available (Spinu et al. 2020). The difficulties in developing a qAOP model are many and of diverse nature, including the lack of appropriate and heterogeneous data compared to the abundance in the qualitative mechanistic information and challenges in integrating multiscale omics data to identify and map causal effects induced by toxicants, to name a couple. Thus, what type of efforts should become the focus in advancing the concept of qAOP in predictive toxicology?

The modern approaches that are thought to be the basis of NGRA are exemplified by studies such as Baltazar et al. (2020). Subsequently, Knight et al. (2021) have identified the key elements that need to be brought together to support these new approaches in an unbiased way and Cronin et al. (2021) made a "call for action" that they be implemented. Whilst competent methodological approaches are being developed and demonstrated, a key element to their widespread use and application is trust. Trust is required at various levels, from the risk manager in industry, to the regulatory scientist in a government agency to the consumer. We also need to believe in and be able to demonstrate trust with the models themselves. As we move

towards an AI-driven society, understanding how we can trust will give stakeholders confidence and reduce the conspiracy theory laden “fake news”. In other words, if we can explain and justify our models, they will become more credible. An element of “disruptive thinking” is required to improve and increase acceptability of new methods (Mahony et al. 2020).

An essential criterion to increase trust in models is to demonstrate causality. In terms of the toxicology supporting CRA, causality is usually thought to be represented by the direct association with the mechanism of action producing the adverse outcome. Hence the popularity off the (q)AOP concept, which is firmly based around a mechanistic framework. Whilst this demonstration of causality is obvious to many in the area of toxicological risk assessment, it is worthwhile to consider the subject and meaning of causality in more detail, especially as it forms an integral part of many of our mathematical theorems, not least Bayesian theory.

The science of causality has been brought to light by Judea Pearl whom, in his recent book “*The Book of Why*” written together with Dana Mackenzie (Pearl and Mackenzie 2019), tells the “silent history of cause and effect”. Pearl and Mackenzie explain how the “Causal Revolution” is the new paradigm that is so much needed to progress computational modelling, especially in the context of AI. A reason for the need for the understanding of causal relationships is that traditional statistical thinking fails to address the real-world causal processes. Such processes imply the series of relationships essential for understanding mechanisms when circumstances change or not. With causality being the direct relationship of cause and effect, this surely being driven by evidence, often requiring painstaking effort and application to obtain, is it the neglected child of computational modelling and AI? At a time when there is a demand for computational modelling in toxicology and chemical safety assessment, possibly driven more by hope and hype than reality, it is surely time to step back and consider what causality means in the “new age” and why it will be crucial in a society requiring evidence, not words?

Thus, this final section of the thesis explores the key role of causality in developing models, as proposed by Judea Pearl, to underline similarities in addressing the problem formulation, and potentially facilitate the development, utilisation and acceptance of models of qAOP. As a consequence, it is intended that this section will support the acceptance of models at all levels from consumer to regulator, building on the concepts of causality from outside of toxicology to support our understanding of AOPs. The intention is to provide a stimulus to be “constructively disruptive” in our thinking, to improve understanding, acceptability and uptake, i.e., how we can use the “Causal Revolution” to bring about a “Toxicological Revolution” more rapidly.

Firstly, we must accept the fundamental concept that a qAOP model aims to identify causal relationships between a stressor, i.e., chemicals or genetic and environmental factors, and the toxicity endpoint/AO of regulatory (or other) interest. For example, Bal-Price et al. (2018b) compiled evidence showing that a brain

concentration of 20-30 nM of rotenone in rats leads to approximately 53% of inhibition of complex I, this, in turn, leads to a decrease in respiration rate of approximately 20-53% and, possibly more significantly, an approximately 20-60% decrease in ubiquitin proteasomal system activity which is involved in neuronal loss and motor impairment, the latter being responsible for Parkinsonian motor deficits (Figure 6.2.A). Hence, a causal effect is informed by both observational and experimental studies, herein, organised through the qualitative AOP framework. This allows the formulation of a series of assumptions and their quantitative translation as a qAOP model to simulate and predict the magnitude by which a downstream key event is altered, or perturbed, by a change in an upstream key event. Therefore, it is desirable that the AOP demonstrates causality in toxicology supplemented by computational modelling.

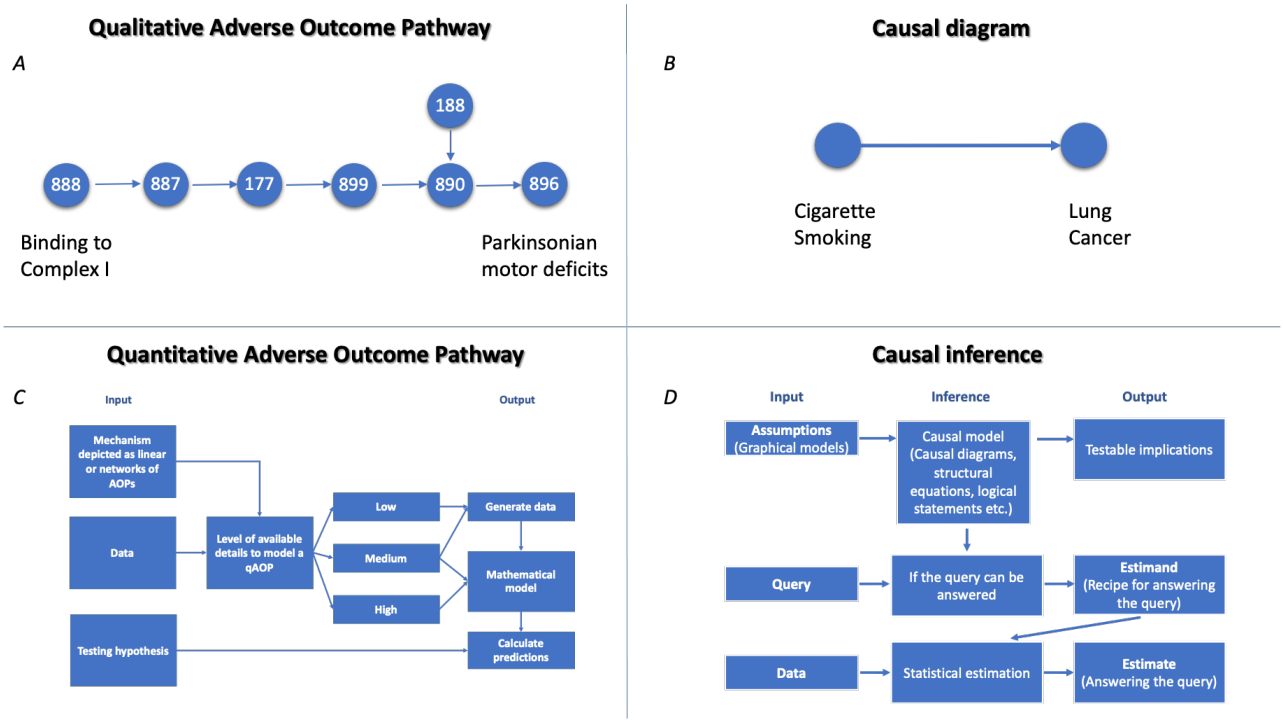


Figure 6.2. A. The AOP for Parkinsonian motor deficits taken as an example to underline one of the characteristics of a qAOP model, mainly understanding the cause and effect in the context of predictive toxicology (https://aopwiki.org/aops/3). Numbers represent the indices of the events in the OECD AOP-Wiki KB available at https://aopwiki.org/events/XXX, where xxx is the index in the node. B. A causal diagram representing the linkage between cigarette smoking and lung cancer. The debate about this cause-effect relationship led to several achievements including: (i) the establishment of randomised control trials methodology conducted by Doll and Hill to compare a treatment group (patients with diagnosed cancer) to a control group (healthy volunteers) (Doll and Hill 1954); (ii) Cornfield's inequality that described the hypothesis of the presence of a smoking gene that makes the difference of developing lung cancer and which has driven the sensitivity analysis methodology, and most importantly; (iii) Hill criteria that helped to summarise the evidence and which are now largely utilised. C. A putative scheme of a general process that a qAOP model implies. Depending on the available level of resources, an AOP can be used to generate data or model quantitatively to make predictions and test a hypothesis. D. The causal inference engine was proposed by Judea Pearl as described in the text and is taken from Pearl and Mackenzie (2019).

To understand the fundamental principle of causality in an AOP, let us consider whether it is known or assumed, and if the AOP framework itself allows us to think in a causal way. To help us understand, the AOP can be placed in the context of established theory. For instance, Pearl refers to the "causal diagram" as the way to represent our scientific knowledge about a variable of interest, also termed as directed acyclic graphs. Thus, causal diagrams consist of variables (nodes) or quantities and arrows (edges) that indicate known or

suspected causal relationships between those variables (Figure 6.2.B). The author advocates the use of such diagrams due to the ease of drawing and comprehending them; because they can estimate all sorts of causal relationships: simple or complicated, deterministic or probabilistic, linear or nonlinear; and most importantly, they allow for the storage of information for future reference and application. Additionally, DAGs are models representing how we think the world/process of interest works. Once written, they help to find all testable implications, if the model we came up with is compatible with the data, and hence, the causality becomes a theory-driven approach. Furthermore, a causal diagram is a Bayesian network, a term coined by Judea Pearl in 1985. Each arrow implies a direct causal relation, or at least the possibility of one, in the direction of that arrow. Pearl confesses that "*[he] wanted Bayesian networks to operate like the neurons of a human brain; you touch one neuron, and the entire network responds by propagating the information to every other neuron in the system*" (Pearl and Mackenzie 2019). Furthermore, Bayesian networks are efficient in coping with contradictory and uncertain data that can be implemented on fast computer platforms and are understandable mathematically. Other advantages include transparency compared to other computational techniques, e.g., deep neural networks, allowing every step to be followed and the understanding of how and why each piece of evidence changed the network's beliefs; there is no need to intervene to tell it how to evaluate a new piece of data once the network is built up; updating can be done very quickly; and the network is integrative, which means that it reacts as a whole to any new information. However, not all Bayesian networks are causal as Pearl points out, "*while probabilities encode our beliefs about a static world, causality tells us whether and how probabilities change when the world changes, be it by the intervention or by act of imagination*" (Pearl and Mackenzie 2019). Pearl adds that "*the main differences between Bayesian networks and causal diagrams lie in how they are constructed and the uses to which they are put. A Bayesian network is literally nothing more than a compact representation of a huge probability table*" (Pearl and Mackenzie 2019). In other words, a Bayesian network can tell how likely one event is, given an observed one, while causal diagrams can answer interventional and counterfactual questions. Interventional questions ask what effect an intervention will produce on the observed variable. For example, does chemical X alone, or in combination with a chemical Y, induce adverse effects? Counterfactual questions ask "what if" referring to a process analysed retrospectively, "*we imagine a different scenario in order to change the circumstances being analysed*" Pearl explains (Pearl and Mackenzie 2019). For example, what if a person was exposed to chemical X, but not under the specified conditions, would the same adverse effect be observed? In other words, counterfactual questions allow us to assess situations that cannot be observed or measured in real-life, or in the past, due to ethical considerations or incapacity of performing such experiments; a change might not be possible to be forced in the past to see what could have happened otherwise. Also, Pearl comments that "*with Bayesian networks, we had taught machines to think in shades of grey, and this was an important step toward humanlike thinking. But we still couldn't teach machines to understand causes and effects*" (Pearl and Mackenzie 2019). Hence, the science focused on identifying patterns in data rather than understanding the reason for those patterns. Given this discussion, it becomes immediately obvious why an understanding of

Pearl's thoughts and concepts are relevant to the AOP framework. They provide a philosophical basis for the working of an AOP along with a strong justification of its quantification, especially through Bayesian networks and approaches. Additionally, the methodology of causality proposed by Pearl helps to answer both types of "why" questions: the straightforward one, when we seek to know the cause, and the one when we want to understand the mechanism itself.

The second reason for considering causality is that a qAOP model is hypothesis driven. AOPs give us the opportunity to test a hypothesis in order to examine the causal evidence for an adverse effect for human or ecological risk assessment. Mechanistic modelling, that can imply empirical dose-responses, Bayesian networks and systems biology, is considered most appropriate for such purposes. For example, a simplified AOP mechanistic model linking thiol oxidation to chronic kidney disease through oxidative stress and mitochondrial disruption was quantified, aiming to compare three quantitative approaches (Zgheib et al. 2019). Additionally, this qAOP model allowed for the evaluation of different levels of exposure of a chemical tested over time and the derivation of chemical-independent key event relationships by inversion of the empirical model applied. Therefore, a qAOP model serves as a tool to translate descriptive qualitative assumptions into the quantitative predictions of an AO in hazard and risk assessment as outlined by Figure 6.2.C.

Considering how we develop a hypothesis in more detail, we need to think about the data underlying or describing the relationships. Pearl asserts that "*data are profoundly dumb*" because they cannot tell us the "*why*" (Pearl and Mackenzie 2019). Therefore, as Pearl indicates, "*causal questions cannot be answered from data alone, it needs to formulate models that generate the data to understand their patterns*" (Pearl and Mackenzie 2019). Also, Pearl emphasises that "*data interpretation means hypothesising on how things operate in the real world*" (Pearl and Mackenzie 2019), and even though conclusions can be drawn with only partial information, and not necessarily knowing every causal relation between the variables of interest, a minimum causal hypothesis is always required; even missing data require a causal understanding. Importantly, Pearl makes the distinction of two types of data interpretation: deduction - reasoning from hypothesis to conclusion, and induction - reasoning from evidence to a hypothesis. Furthermore, Pearl refers to the methodology based on the Bayes' rule that consists of several steps: (1) formulate a hypothesis, (2) deduce a testable consequence of the hypothesis, (3) perform an experiment and collect evidence, and (4) update your belief in the hypothesis. Therefore, a qAOP does not require pattern recognition but rather can be placed in the context of measuring or obtaining the data for assays associated with key events, and hence, fits perfectly within the above methodology.

If we accept the arguments requiring consideration of causality in our models for toxicology, and especially qAOPs, then we need means to evaluate that causality. As we move to the holy grail of regulatory acceptance

of models and their predictions, evidence for the demonstration of causality within the model becomes paramount. Traditional approaches for evaluating causality in toxicology follow the criteria formulated by Austin Bradford Hill in 1965 when he attempted to summarise the arguments for the causal linkage of cigarette smoking to lung cancer (Figure 6.2.B) (Hill 1965). The so-called "modified Bradford Hill" criteria are used as a premise to act as a framework for semi-quantitative and quantitative weight-of-evidence qAOP models (Becker et al. 2017; OECD 2018b; Perkins et al. 2019b). Notwithstanding, Pearl describes the Hill criteria as being "*qualitative patterns of statistical trends*" and asserts that Hill himself called them "*viewpoints*" and not requirements. Pearl emphasises that the "*Hill's "viewpoints" are still useful as a description of how a discipline comes to accept a causal hypothesis, using a variety of evidence, but they came with no methodology to implement them. Each scientist just has to decide for him- or herself. But gut decisions can be wrong, especially if there are political pressures or monetary considerations*" (Pearl and Mackenzie 2019). For example, the consistency or strength of the association that comes with the Hill criteria by itself can prove nothing, "*if thirty studies each ignore the same confounder, all can easily be biased*" (Pearl and Mackenzie 2019). Put simply, a confounder represents a common cause between two variables. In the context of an AOP, we should think of confounders as being the modulating factors such as the age of a person, diet, genetic predispositions etc. Also, a confounder can help understand the difference between causal reasoning and causal inference, what we want to assess vs what we do actually assess using statistical methods. Causal inference cannot exist without confounders, while causal reasoning does not necessarily focus on confounding variables. Furthermore, Pearl asserts three levels of causation, named the "Ladder of Causation", referring to the human cognitive ability: seeing, doing, and imagining. Climbing on these three rungs, Pearl advocates that we can make the machines to think: "*a causal reasoning module will give the machines the ability to reflect on their mistakes, to pinpoint weaknesses in their software, to function as moral entities, and to converse naturally with humans about their own choices and intentions*" (Pearl and Mackenzie 2019). This describes the ideal world where we imagine AI to support and facilitate the chemical hazard and risk assessment and, hence, help with the complexity of the exposure to chemicals and the associated risks to human health.

Thus, we evaluate causality not only in qAOPs but also throughout toxicology, accepting the Hill criteria as a starting point, or framework, on which to hang our evidence. Understanding the limitations of the Hill criteria takes us a step closer to a more ideal approach; engaging with the opportunities given by the assessment of causality can offer us an insight into where to go next. Moreover, available means facilitate the start of assessing the causality by applying existing algorithms, or developing new ones based on formulated theories and principles, in investigating the causality in toxicology as summarised in Table 6.1.

Table 6.1. List of available algorithms to help to assess the causality.

| Package Name | Programming Language | URL[7] |
|---|---|---|
| **CausalLift** | Python | https://github.com/Minyus/causallift/ |
| **CausalML** | Python | https://github.com/uber/causalml |
| **DoWhy** | Python | https://github.com/Microsoft/dowhy |
| **EconML** | Python | https://github.com/Microsoft/EconML |
| **pylift** | Python | https://github.com/wayfair/pylift |
| **pymatch** | Python | https://github.com/benmiroglio/pymatch |
| **causaleffect** | R | https://cran.r-project.org/web/packages/causaleffect/ |
| **causalGAM** | R | https://cran.r-project.org/web/packages/CausalGAM/ |
| **dagitty** | R | https://cran.r-project.org/web/packages/dagitty/ |
| **ggdag** | R | https://cran.r-project.org/web/packages/ggdag/ |
| **mediation** | R | https://cran.r-project.org/web/packages/mediation/ |
| **pcalg** | R | https://cran.r-project.org/web/packages/pcalg/ |
| **uplift** | R | https://cran.r-project.org/web/packages/uplift/ |

Hence, the "causal inference engine" proposed by Pearl to handle causal reasoning aligns perfectly with the stages of the development of a qAOP model, which contrarily, relies heavily on the available resources today. Pearl presents his blueprint as a diagram similar to a decision tree in which inputs enter the inference engine and produce the outputs (Figure 6.2.D). It has three different kinds of inputs: assumptions given by the knowledge, queries and data. The first output is a yes/no decision to a query; if the answer is yes, an "estimand" is produced. The second output is a mathematical formula for generating the answer from hypothetical data. The third output is produced after the data are entered, which is an estimated answer, along with the statistical estimates of the uncertainty. Pearl adds "*this uncertainty reflects the limited size of the data set as well as possible measurement errors or missing data*" (Pearl and Mackenzie 2019). Rather than relying on the vague nature of the definition of the Hill criteria, is it time to think more about how we can better quantify uncertainty in AOPs and what these other theories tell us?

So, what have we learned? Modelling, in all its glory and infamy, is fundamental to the paradigm change dictated by 21$^{st}$ Century Toxicology. qAOPs are in their infancy, although like any infant – much is expected as they flourish and mature. We believe, at its heart, a model for a qAOP mainly implies two characteristics: namely the understanding of causal relationships and the testing of hypotheses. Hence, a qAOP model can be considered a causal model to predict the results of an action or an intervention. However, big data make us look for correlations and associations instead of causality. Additionally, most of the ML techniques applied, even though composed of independent and dependent variables, test/training and validation sets, do not infer causality of the included variables, rather they are trained to learn from experience. In the context of a qAOP model, causality is currently informed by the qualitative linear or network of AOPs that promote a structured format for understanding a stressor-induced mechanism of action assessed against the modified Hill criteria. Importantly, a qAOP model requires data as a mean to predict the adverse effects quantitatively. Unfortunately, the available resources do not fit within this framework. Thus, causal diagrams can dictate

[7]Last time accessed on March 16, 2021.

the production of data. For example, computational models can help retrieve observational data to construct causal relationships and consecutively can help design appropriate experiments to verify the assumptions. Causal models can help formulate and prioritise the assessment questions in a transparent manner, anticipate the potential consequences of a policy in decision-making, e.g., utilising counterfactual questions, identify emerging issues or questions, less evident to human expert judgement and avoid the bias in the decision-making process. Causal models can, and must, serve as a basis for the development of a robust, transparent, comprehensible and reproducible AI. More efforts are needed to shift towards understanding the causality rather than deriving data-driven models – for toxicology this means building models on mechanisms of action. Hence, we need causal assumptions in order to make reliable conclusions.

Something is certain, the causal revolution initiated by Judea Pearl is happening. It is not only emerging in epidemiology, sociology, and economics but also, in predictive toxicology, and it can contribute to the development of qAOP models.

The qAOP model has the power to computationally model the causal linkages, i.e., KERs, in a transparent manner, making use of available data and applying a range of methodologies. A qAOP model makes the data accessible, readable and (re)usable, hence, it has the capability to inform decision-making processes. Hopefully, qAOPs will one day be accepted as more than a screening tool. The combined use of PBPK, QSAR and qAOP models could, in principle, be used for a definitive chemical risk assessment.

# References


admetSAR v.2.0. In. http://lmmd.ecust.edu.cn/admetsar2/ Accessed 25.10.2020

Aguayo-Orozco A, Audouze K, Siggaard T, Barouki R, Brunak S, Taboureau O (2019a) sAOP: linking chemical stressors to adverse outcomes pathway networks. Bioinformatics 35(24):5391-5392 doi:10.1093/bioinformatics/btz570

Aguayo-Orozco A, Taboureau O, Brunak S (2019b) The use of systems biology in chemical risk assessment. Curr Opin Toxicol 15:48-54 doi:10.1016/j.cotox.2019.03.003

Allen TEH, Goodman JM, Gutsell S, Russell PJ (2016) A history of the molecular initiating event. Chem Res Toxicol 29(12):2060-2070 doi:10.1021/acs.chemrestox.6b00341

Allen TEH, Goodman JM, Gutsell S, Russell PJ (2020) Quantitative predictions for molecular initiating events using three-dimensional quantitative structure-activity relationships. Chem Res Toxicol 33(2):324-332 doi:10.1021/acs.chemrestox.9b00136

Amrhein V, Greenland S, McShane B (2019) Scientists rise up against statistical significance. Nature 567(7748):305-307 doi:10.1038/d41586-019-00857-9

Angrish MM, McQueen CA, Cohen-Hubal E, Bruno M, Ge Y, Chorley BN (2017) Editor's Highlight: Mechanistic toxicity tests based on an adverse outcome pathway network for hepatic steatosis. Toxicol Sci 159(1):159-169 doi:10.1093/toxsci/kfx121

Ankley GT, Bennett RS, Erickson RJ, et al. (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730-41 doi:10.1002/etc.34

Ankley GT, Edwards SW (2018) The adverse outcome pathway: A multifaceted framework supporting 21(st) century toxicology. Curr Opin Toxicol 9:1-7 doi:10.1016/j.cotox.2018.03.004

Aschner M, Ceccatelli S, Daneshian M, et al. (2017) Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use. ALTEX 34(1):49-74 doi:10.14573/altex.1604201

Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24(2):282–284 doi:10.1093/bioinformatics/btm554

Avila AM, Bebenek I, Bonzo JA, et al. (2020) An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs). Regul Toxicol Pharmacol 114:104662 doi:10.1016/j.yrtph.2020.104662

Bal-Price A, Crofton KM, Leist M, et al. (2015) International STakeholder NETwork (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. Arch Toxicol 89(2):269-87 doi:10.1007/s00204-015-1464-2

Bal-Price A, Hogberg HT, Crofton KM, et al. (2018a) Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. ALTEX 35(3):306-352 doi:10.14573/altex.1712081

Bal-Price A, Lein PJ, Keil KP, et al. (2017) Developing and applying the adverse outcome pathway concept for understanding and predicting neurotoxicity. Neurotoxicology 59:240-255 doi:10.1016/j.neuro.2016.05.010

Bal-Price A, Leist M, Schildknecht S, Tschudi-Monnet F, Paini A, Terron A (2018b) Adverse outcome pathway on inhibition of the mitochondrial complex I of nigro-striatal neurons leading to Parkinsonian motor deficits. OECD Series on Adverse Outcome Pathways, No 7, OECD Publishing, Paris doi:10.1787/b46c3c00-en

Bal-Price A, Meek MEB (2017) Adverse outcome pathways: application to enhance mechanistic understanding of neurotoxicity. Pharmacol Ther 179:84-95 doi:10.1016/j.pharmthera.2017.05.006

Bal-Price A, Pistollato F, Sachana M, Bopp SK, Munn S, Worth A (2018c) Strategies to improve the regulatory assessment of developmental neurotoxicity (DNT) using in vitro methods. Toxicol Appl Pharmacol 354:7-18 doi:10.1016/j.taap.2018.02.008

Baltazar MT, Cable S, Carmichael PL, et al. (2020) A Next-Generation Risk Assessment Case Study for Coumarin in Cosmetic Products. Toxicol Sci 176(1):236-252 doi:10.1093/toxsci/kfaa048

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101-13 doi:10.1038/nrg1272

Barrientos A, Moraes CT (1999) Titrating the effects of mitochondrial complex I impairment in the cell physiology. J Biol Chem 274(23):16188-97 doi:10.1074/jbc.274.23.16188
Bassett DS, Sporns O (2017) Network neuroscience. Nat Neurosci 20(3):353-364 doi:10.1038/nn.4502
Battistoni M, Di Renzo F, Menegola E, Bois F (2019) Quantitative AOP based teratogenicity prediction for mixtures of azole fungicides. Comput Toxicol(11):72-81 doi:10.1016/j.comtox.2019.03.004
Becker RA, Ankley GT, Edwards SW, et al. (2015) Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence. Regul Toxicol Pharmacol 72(3):514-37 doi:10.1016/j.yrtph.2015.04.004
Becker RA, Dellarco V, Seed J, et al. (2017) Quantitative weight of evidence to assess confidence in potential modes of action. Regul Toxicol Pharmacol 86:205-220 doi:10.1016/j.yrtph.2017.02.017
Behl M, Ryan K, Hsieh JH, et al. (2019) Screening for developmental neurotoxicity at the National Toxicology Program: The future Is here. Toxicol Sci 167(1):6-14 doi:10.1093/toxsci/kfy278
Beilmann M, Boonen H, Czich A, et al. (2018) Optimizing drug discovery by investigative toxicology: Current and future trends. ALTEX 36(2):289-313 doi:10.14573/altex.1808181
Bender A, Cortes-Ciriano I (2021) Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov Today doi:10.1016/j.drudis.2020.11.037
Bennett D, Bellinger DC, Birnbaum LS, et al. (2016) Project TENDR: Targeting environmental neuro-developmental risks. The TENDR Consensus Statement. Environ Health Perspect 124(7):A118-22 doi:10.1289/EHP358
Beronius A, Zilliacus J, Hanberg A, Luijten M, van der Voet H, van Klaveren J (2020) Methodology for health risk assessment of combined exposures to multiple chemicals. Food Chem Toxicol 143:111520 doi:10.1016/j.fct.2020.111520
Betarbet R, Sherer TB, Greenamyre JT (2005) Ubiquitin-proteasome system and Parkinson's diseases. Exp Neurol 191 Suppl 1:S17-27 doi:10.1016/j.expneurol.2004.08.021
Bishop C (2006) Pattern Recognition and Machine Learning. Springer-Verlag New York
Bloomingdale P, Housand C, Apgar JF, et al. (2017) Quantitative systems toxicology. Curr Opin Toxicol 4:79-87 doi:10.1016/j.cotox.2017.07.003
Bois FY (2013) Bayesian inference. In: Reisfeld B, Mayeno AN (eds) Computational toxicology: Volume II. Humana Press, Totowa, NJ, p 597-636
Bois FY, Jackson ET, Pekari K, Smith MT (1996) Population toxicokinetics of benzene. Environ Health Perspect 104(suppl 6):1405-1411 doi:10.1289/ehp.961041405
Bois FY, Jamei M, Clewell HJ (2010) PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278(3):256-267 doi:10.1016/j.tox.2010.06.007
Bove J, Perier C (2012) Neurotoxin-based models of Parkinson's disease. Neuroscience 211:51-76 doi:10.1016/j.neuroscience.2011.10.057
Brandes U, Robins G, McCranie ANN, Wasserman S (2013) What is network science? Network Science 1(1):1-15 doi:10.1017/nws.2013.2
Burgoon DL (2019) AOPXplorer. In. http://apps.cytoscape.org/apps/aopxplorer Accessed 8 May 2019
Burgoon LD, Angrish M, Garcia-Reyero N, Pollesch N, Zupanic A, Perkins E (2020) Predicting the probability that a chemical causes steatosis using adverse outcome pathway Bayesian networks (AOPBNs). Risk Anal 40(3):512-523 doi:10.1111/risa.13423
Carriger JF, Martin TM, Barron MG (2016) A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors. Aquat Toxicol 180:11-24 doi:10.1016/j.aquatox.2016.09.006
Carusi A, Davies MR, De Grandis G, et al. (2018) Harvesting the promise of AOPs: An assessment and recommendations. Sci Total Environ 628-629:1542-1556 doi:10.1016/j.scitotenv.2018.02.015
Carvaillo JC, Barouki R, Coumoul X, Audouze K (2019) Linking bisphenol S to adverse outcome pathways using a combined text mining and systems biology approach. Environ Health Perspect 127(4):47005 doi:10.1289/EHP4200
Chappell GA, Britt JK, Borghoff SJ (2020) Systematic assessment of mechanistic data for FDA-certified food colors and neurodevelopmental processes. Food Chem Toxicol 140:111310 doi:10.1016/j.fct.2020.111310

Chauhan V, Said Z, Daka J, et al. (2019) Is there a role for the adverse outcome pathway framework to support radiation protection? Int J Radiat Biol 95(2):225-232 doi:10.1080/09553002.2019.1532617
Chou AP, Li S, Fitzmaurice AG, Bronstein JM (2010) Mechanisms of rotenone-induced proteasome inhibition. Neurotoxicology 31(4):367-72 doi:10.1016/j.neuro.2010.04.006
Chu VR (2018) Assessing the Effects of Chemical Mixtures using a Bayesian Network-Relative Risk Model (BNRRM) Integrating Adverse Outcome Pathways (AOPs) in Four Watersheds. WWU Graduate School Collection 699 doi:https://cedar.wwu.edu/wwuet/699
Clippinger AJ, Allen D, Behrsing H, et al. (2018) Pathway-based predictive approaches for non-animal assessment of acute inhalation toxicity. Toxicol in Vitro 52:131-145 doi:10.1016/j.tiv.2018.06.009
Coady K, Browne P, Embry M, et al. (2019) When are adverse outcome pathways and associated assays “fit for purpose" for regulatory decision-making and management of chemicals? Integr Environ Assess Manag 15(4):633-647 doi:10.1002/ieam.4153
Collier ZA, Gust KA, Gonzalez-Morales B, et al. (2016) A weight of evidence assessment approach for adverse outcome pathways. Regul Toxicol Pharmacol 75:46-57 doi:10.1016/j.yrtph.2015.12.014
Conolly RB, Ankley GT, Cheng W, et al. (2017) Quantitative adverse outcome pathways and their application to predictive toxicology. Environ Sci Technol 51(8):4661-4672 doi:10.1021/acs.est.6b06230
Crofton KM, Mundy WR, Shafer TJ (2012) Developmental neurotoxicity testing: a path forward. Congenit Anom (Kyoto) 52(3):140-6 doi:10.1111/j.1741-4520.2012.00377.x
Cronin M, Doe J, Pereira M, Willett C (2021) Re: A call for action on the development and implementation of new methodologies for safety assessment of chemical-based products in the EU – A short communication. Regul Toxicol Pharmacol 122:104911 doi:10.1016/j.yrtph.2021.104911
Cronin MTD, Richarz A-N (2017) Relationship between adverse outcome pathways and chemistry-based in silico models to predict toxicity. Appl In Vitro Toxicol 3(4):286-297 doi:10.1089/aivt.2017.0021
Cytoscape (2018) NetworkAnalyzer Online Help. In. https://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/ Accessed 8 May 2019
Dent M, Amaral RT, Da Silva PA, et al. (2018) Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Comput Toxicol 7:20-26 doi:10.1016/j.comtox.2018.06.001
Dexter DT, Jenner P (2013) Parkinson disease: from pathology to molecular disease mechanisms. Free Radic Biol Med 62:132-144 doi:10.1016/j.freeradbiomed.2013.01.018
Dias V, Junn E, Mouradian MM (2013) The role of oxidative stress in Parkinson's disease. J Parkinsons Dis 3(4):461-91 doi:10.3233/JPD-130230
Doering JA, Villeneuve DL, Poole ST, et al. (2019) Quantitative response-response relationships linking aromatase inhibition to decreased fecundity are conserved across three fishes with asynchronous oocyte development. Environ Sci Technol 53(17):10470-10478 doi:10.1021/acs.est.9b02606
Doering JA, Wiseman S, Giesy JP, Hecker M (2018) A cross-species quantitative adverse outcome pathway for activation of the aryl hydrocarbon receptor leading to early life stage mortality in birds and fishes. Environ Sci Technol 52(13):7524-7533 doi:10.1021/acs.est.8b01438
Doll R, Hill AB (1954) The mortality of doctors in relation to their smoking habits; a preliminary report. Br Med J 1(4877):1451-5 doi:10.1136/bmj.1.4877.1451
Dreier DA, Mello DF, Meyer JN, Martyniuk CJ (2019) Linking mitochondrial dysfunction to organismal and population health in the context of environmental pollutants: Progress and considerations for mitochondrial adverse outcome pathways. Environ Toxicol Chem 38(8):1625-1634 doi:10.1002/etc.4453
EC (2019) EU Chemicals Policy 2030. Building on the past, moving to the future. In. https://euchemicalspolicy2030.teamwork.fr/docs/report.pdf
Edwards SW, Tan YM, Villeneuve DL, Meek ME, McQueen CA (2016) Adverse outcome pathways-organizing toxicological information to improve decision making. J Pharmacol Exp Ther 356(1):170-81 doi:10.1124/jpet.115.228239
Esteves AR, Arduino DM, Silva DF, Oliveira CR, Cardoso SM (2011) Mitochondrial dysfunction: The road to alpha-synuclein oligomerization in PD. Parkinsons Dis 2011:693761 doi:10.4061/2011/693761
Foran CM, Rycroft T, Keisler J, Perkins EJ, Linkov I, Garcia-Reyero N (2019) A modular approach for assembly of quantitative adverse outcome pathways. ALTEX 36(3):353-362 doi:10.14573/altex.1810181

Frank CL, Brown JP, Wallace K, Mundy WR, 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. Toxicol Sci 160(1):121-135 doi:10.1093/toxsci/kfx169

Fritsche E, Grandjean P, Crofton KM, et al. (2018) Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes. Toxicol Appl Pharmacol 354:3-6 doi:10.1016/j.taap.2018.02.004

Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 141(1):2-18 doi:10.1037/a0024338

Fujita KA, Ostaszewski M, Matsuoka Y, et al. (2014) Integrating pathways of Parkinson's disease in a molecular interaction map. Mol Neurobiol 49(1):88-102 doi:10.1007/s12035-013-8489-4

Furxhi I, Murphy F, Poland CA, Sheehan B, Mullins M, Mantecca P (2019) Application of Bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics. Nanotoxicology 13(6):827-848 doi:10.1080/17435390.2019.1595206

Gao Y, Kang L, Zhang Y, Feng J, Zhu L (2019) Toxicokinetic and toxicodynamic (TK-TD) modeling to study oxidative stress-dependent toxicity of heavy metals in zebrafish. Chemosphere 220:774-782 doi:10.1016/j.chemosphere.2018.12.197

Garcia-Reyero N, Murphy CA (2018) A systems biology approach to advancing adverse outcome pathways for risk assessment. Springer International Publishing

Gautier F, Tourneix F, Assaf Vandecasteele H, van Vliet E, Bury D, Alepee N (2020) Read-across can increase confidence in the Next Generation Risk Assessment for skin sensitisation: A case study with resorcinol. Regul Toxicol Pharmacol 117:104755 doi:10.1016/j.yrtph.2020.104755

Gelman A (2006) Multilevel (hierarchical) modeling: What it can and cannot do. Technometrics 48(3):432-435 doi:10.1198/004017005000000661

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis, Third Edition. Taylor & Francis

Gelman A, Hill J, Vehtari A (2020) Regression and other stories. Cambridge University Press

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statist Sci 7(4):457-472 doi:10.1214/ss/1177011136

Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. Nature 521(7553):452-9 doi:10.1038/nature14541

Gilmour N, Kern PS, Alepee N, et al. (2020) Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. Regul Toxicol Pharmacol 116:104721 doi:10.1016/j.yrtph.2020.104721

Glass CK, Saijo K, Winner B, Marchetto MC, Gage FH (2010) Mechanisms underlying inflammation in neurodegeneration. Cell 140(6):918-34 doi:10.1016/j.cell.2010.02.016

Grandjean P, Landrigan PJ (2006) Developmental neurotoxicity of industrial chemicals. Lancet 368(9553):2167-78 doi:10.1016/S0140-6736(06)69665-7

Greim H, Snyder R (2018) Toxicology and risk assessment: A comprehensive introduction, 2 edn. Wiley

Grivennikova VG, Maklashina EO, Gavrikova EV, Vinogradov AD (1997) Interaction of the mitochondrial NADH-ubiquinone reductase with rotenone as related to the enzyme active/inactive transition. Biochim Biophys Acta 1319(2-3):223-32 doi:10.1016/s0005-2728(96)00163-6

Groh KJ, Carvalho RN, Chipman JK, et al. (2015) Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere 120:764-777 doi:10.1016/j.chemosphere.2014.09.068

Groh KJ, Tollefsen KE (2015) The Challenge: Adverse outcome pathways in research and regulation-Current status and future perspectives. Environ Toxicol Chem 34(9):1935 doi:10.1002/etc.3042

Gust KA, Collier ZA, Mayo ML, Stanley JK, Gong P, Chappell MA (2016) Limitations of toxicity characterization in life cycle assessment: Can adverse outcome pathways provide a new foundation? Integr Environ Assess Manag 12(3):580-90 doi:10.1002/ieam.1708

Hack CE, Haber LT, Maier A, et al. (2010) A Bayesian network model for biomarker-based dose response. Risk Anal 30(7):1037-51 doi:10.1111/j.1539-6924.2010.01413.x

Hardy A, Benford D, Halldorsson T, et al. (2017) Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15(8):e04971 doi:10.2903/j.efsa.2017.4971

Harrill JA, Freudenrich T, Wallace K, Ball K, Shafer TJ, Mundy WR (2018) Testing for developmental neurotoxicity using a battery of in vitro assays for key cellular events in neurodevelopment. Toxicol Appl Pharmacol 354:24-39 doi:10.1016/j.taap.2018.04.001

Hartung T, FitzGerald RE, Jennings P, et al. (2017) Systems Toxicology: Real World Applications and Opportunities. Chem Res Toxicol 30(4):870-882 doi:10.1021/acs.chemrestox.7b00003

Hassan I, El-Masri H, Kosian PA, Ford J, Degitz SJ, Gilbert ME (2017) Neurodevelopment and thyroid hormone synthesis inhibition in the rat: Quantitative understanding within the adverse outcome pathway framework. Toxicol Sci 160(1):57-73 doi:10.1093/toxsci/kfx163

Hecker M, LaLone CA (2019) Adverse outcome pathways: Moving from a scientific concept to an internationally accepted framework. Environ Toxicol Chem 38(6):1152-1163 doi:10.1002/etc.4385

Hespanhol L, Vallio CS, Costa LM, Saragiotto BT (2019) Understanding and interpreting confidence and credible intervals around effect estimates. Braz J Phys Ther 23(4):290-301 doi:10.1016/j.bjpt.2018.12.006

Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine. J R Soc Med 58(5):295-300 doi:10.1177/003591576505800503

Hoffmann S, de Vries RBM, Stephens ML, et al. (2017) A primer on systematic reviews in toxicology. Arch Toxicol 91(7):2551-2575 doi:10.1007/s00204-017-1980-3

Hothorn LA, Ralph P (2020) Use compatibility intervals in regulatory toxicology. Regul Toxicol Pharmacol 116:104720 doi:10.1016/j.yrtph.2020.104720

Howdeshell KL, Hotchkiss AK, Gray LE, Jr. (2017) Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. Int J Hyg Environ Health 220(2 Pt A):179-188 doi:10.1016/j.ijheh.2016.11.007

Jameson RR, Seidler FJ, Qiao D, Slotkin TA (2006) Adverse neurodevelopmental effects of dexamethasone modeled in PC12 cells: identifying the critical stages and concentration thresholds for the targeting of cell acquisition, differentiation and viability. Neuropsychopharmacology 31(8):1647-58 doi:10.1038/sj.npp.1300967

Jannicke Moe S, Madsen AL, Connors KA, et al. (2020a) Development of a hybrid Bayesian network model for predicting acute fish toxicity using multiple lines of evidence. Environ Model Softw 126:104655 doi:10.1016/j.envsoft.2020.104655

Jannicke Moe S, Wolf R, Xie L, Landis WG, Kotamaki N, Tollefsen KE (2020b) Quantification of an Adverse Outcome Pathway Network by Bayesian Regression and Bayesian Network Modeling. Integr Environ Assess Manag doi:10.1002/ieam.4348

Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M (2015) Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol 89(12):2355-83 doi:10.1007/s00204-015-1634-2

Jeong J, Choi J (2019) Adverse outcome pathways potentially related to hazard identification of microplastics based on toxicity mechanisms. Chemosphere 231:249-255 doi:10.1016/j.chemosphere.2019.05.003

Jeong J, Garcia-Reyero N, Burgoon L, et al. (2019) Development of adverse outcome pathway for PPARgamma antagonism leading to pulmonary fibrosis and chemical selection for its validation: ToxCast database and a deep learning artificial neural network model-based approach. Chem Res Toxicol 32(6):1212-1222 doi:10.1021/acs.chemrestox.9b00040

Jeong J, Song T, Chatterjee N, Choi I, Cha YK, Choi J (2018) Developing adverse outcome pathways on silver nanoparticle-induced reproductive toxicity via oxidative stress in the nematode Caenorhabditis elegans using a Bayesian network model. Nanotoxicology 12(10):1182-1197 doi:10.1080/17435390.2018.1529835

Kaplan L, Chow BW, Gu C (2020) Neuronal regulation of the blood-brain barrier and neurovascular coupling. Nat Rev Neurosci 21(8):416-432 doi:10.1038/s41583-020-0322-2

Kleinstreuer NC, Sullivan K, Allen D, et al. (2016) Adverse outcome pathways: From research to regulation scientific workshop report. Regul Toxicol Pharmacol 76:39-50 doi:10.1016/j.yrtph.2016.01.007

Knapen D, Angrish MM, Fortin MC, et al. (2018) Adverse outcome pathway networks I: Development and applications. Environ Toxicol Chem 37(6):1723-1733 doi:10.1002/etc.4125

Knapen D, Vergauwen L, Villeneuve DL, Ankley GT (2015) The potential of AOP networks for reproductive and developmental toxicity assay development. Reprod Toxicol 56:52-5 doi:10.1016/j.reprotox.2015.04.003

Knight DJ, Deluyker H, Chaudhry Q, Vidal JM, de Boer A (2021) A call for action on the development and implementation of new methodologies for safety assessment of chemical-based products in the EU - A short communication. Regul Toxicol Pharmacol 119:104837 doi:10.1016/j.yrtph.2020.104837

KNIME RDKit Nodes for KNIME (trusted extension) v.3. In. https://www.knime.com/rdkit Accessed 25.10.2020

Krewski D, Andersen ME, Tyshenko MG, et al. (2020) Toxicity testing in the 21st century: progress in the past decade and future perspectives. Arch Toxicol 94(1):1-58 doi:10.1007/s00204-019-02613-4

Kruschke JK (2013) Bayesian estimation supersedes the t test. J Exp Psychol Gen 142(2):573-603 doi:10.1037/a0029146

Labelle C, Marinier A, Lemieux S (2019) Enhancing the drug discovery process: Bayesian inference for the analysis and comparison of dose–response experiments. Bioinformatics 35(14):i464-i473 doi:10.1093/bioinformatics/btz335

LaLone CA, Ankley GT, Belanger SE, et al. (2017a) Advancing the adverse outcome pathway framework-An international horizon scanning approach. Environ Toxicol Chem 36(6):1411-1421 doi:10.1002/etc.3805

LaLone CA, Villeneuve DL, Wu-Smart J, et al. (2017b) Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. Sci Total Environ 584-585:751-775 doi:10.1016/j.scitotenv.2017.01.113

Laroche C, Annys E, Bender H, et al. (2019) Finding synergies for the 3Rs - Repeated Dose Toxicity testing: Report from an EPAA Partners' Forum. Regul Toxicol Pharmacol 108:104470 doi:10.1016/j.yrtph.2019.104470

Lassiter TL, MacKillop EA, Ryde IT, Seidler FJ, Slotkin TA (2009) Is fipronil safer than chlorpyrifos? Comparative developmental neurotoxicity modeled in PC12 cells. Brain Res Bull 78(6):313-22 doi:10.1016/j.brainresbull.2008.09.020

Lazic SE, Mellor JR, Ashby MC, Munafo MR (2020) A Bayesian predictive approach for dealing with pseudoreplication. Sci Rep 10(1):2366 doi:10.1038/s41598-020-59384-7

Leist M, Ghallab A, Graepel R, et al. (2017) Adverse outcome pathways: opportunities, limitations and open questions. Arch Toxicol 91(11):3477-3505 doi:10.1007/s00204-017-2045-3

Li J, Settivari R, LeBaron MJ, Marty MS (2019) An industry perspective: A streamlined screening strategy using alternative models for chemical assessment of developmental neurotoxicity. Neurotoxicology 73:17-30 doi:10.1016/j.neuro.2019.02.010

Li Z, Ivanov AA, Su R, et al. (2017) The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. Nat Commun 8:14356 doi:10.1038/ncomms14356

Lin MT, Beal MF (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443(7113):787-795 doi:10.1038/nature05292

Liu H, Wang L, Lv M, et al. (2014) AlzPlatform: an Alzheimer's disease domain-specific chemogenomics knowledgebase for polypharmacology and target identification research. J Chem Inf Model 54(4):1050-60 doi:10.1021/ci500004h

Lopachin RM (2004) The changing view of acrylamide neurotoxicity. Neurotoxicology 25(4):617-30 doi:10.1016/j.neuro.2004.01.004

Lopachin RM, Decaprio AP (2005) Protein adduct formation as a molecular mechanism in neurotoxicity. Toxicol Sci 86(2):214-25 doi:10.1093/toxsci/kfi197

Ma XL, Chen C, Yang J (2005) Predictive model of blood-brain barrier penetration of organic compounds. Acta Pharmacol Sin 26(4):500-12 doi:10.1111/j.1745-7254.2005.00068.x

MacKay C, Davies M, Summerfield V, Maxwell G (2013) From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. ALTEX 30(4):473-86 doi:10.14573/altex.2013.4.473

Macmillan DS, Chilton ML (2019) A defined approach for predicting skin sensitisation hazard and potency based on the guided integration of in silico, in chemico and in vitro data using exclusion criteria. Regul Toxicol Pharmacol 101:35-47 doi:10.1016/j.yrtph.2018.11.001

Mahony C, Ashton RS, Birk B, et al. (2020) New ideas for non-animal approaches to predict repeated-dose systemic toxicity: Report from an EPAA Blue Sky Workshop. Regul Toxicol Pharmacol 114:104668 doi:10.1016/j.yrtph.2020.104668

Margiotta-Casaluci L, Owen SF, Huerta B, et al. (2016) Internal exposure dynamics drive the adverse outcome pathways of synthetic glucocorticoids in fish. Sci Rep 6:21978 doi:10.1038/srep21978

McElreath R (2016) Statistical rethinking. Chapman and Hall/CRC, New York

Meek ME, Boobis A, Cote I, et al. (2014a) New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 34(1):1-18 doi:10.1002/jat.2949

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

Meek MEB (2017) AOPs in hazard characterization for human health. Curr Opin Toxicol 3:80-86 doi:10.1016/j.cotox.2017.06.002

Mellor CL, Steinmetz FP, Cronin MT (2016) Using molecular initiating events to develop a structural alert based screening workflow for nuclear receptor ligands associated with hepatic steatosis. Chem Res Toxicol 29(2):203-12 doi:10.1021/acs.chemrestox.5b00480

Mora Lagares L, Minovski N, Novic M (2019) Multiclass classifier for P-glycoprotein substrates, inhibitors, and non-active compounds. Molecules 24(10) doi:10.3390/molecules24102006

Muller EB, Lin S, Nisbet RM (2015) Quantitative adverse outcome pathway analysis of hatching in zebrafish with CuO nanoparticles. Environ Sci Technol 49(19):11817-24 doi:10.1021/acs.est.5b01837

Mundy WR, Padilla S, Breier JM, et al. (2015) Expanding the test set: Chemicals with potential to disrupt mammalian brain development. Neurotoxicol Teratol 52:25-35 doi:10.1016/j.ntt.2015.10.001

Mustieles V, D'Cruz SC, Couderq S, et al. (2020) Bisphenol A and its analogues: A comprehensive review to identify and prioritize effect biomarkers for human biomonitoring. Environ Int 144:105811 doi:10.1016/j.envint.2020.105811

Myatt GJ, Ahlberg E, Akahori Y, et al. (2018) In silico toxicology protocols. Regul Toxicol Pharmacol 96:1-17 doi:10.1016/j.yrtph.2018.04.014

National Research Council (2007) Toxicity Testing in the 21st Century: A Vision and a Strategy. doi:10.17226/11970

Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR (2007) A primer on learning in Bayesian networks for computational biology. PLoS Comput Biol 3(8):e129 doi:10.1371/journal.pcbi.0030129

Netzwerkerin (2018) Eccentricity and closeness. In. https://sites.google.com/site/netzwerkerin/home/closeness Accessed 8 May 2019

OECD (2006) Test No. 221: Lemna sp. growth inhibition test. OECD Guidelines for the testing of chemicals, Section 2 In. https://doi.org/10.1787/9789264016194-en

OECD (2007) Test No. 426: Developmental Neurotoxicity Study. OECD Publishing, Paris

OECD (2014a) The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. In: OECD Series on Testing and Assessment. OECD Publishing, Paris. https://www.oecd-ilibrary.org/content/publication/9789264221444-en

OECD (2014b) Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models,

OECD (2016) Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA). Series on Testing & Assessment No 260. (ENV/JM/MONO(2016)67). In. https://one.oecd.org/document/ENV/JM/MONO(2016)67/en/pdf

OECD (2017) Revised guidance document on developing and assessing adverse outcome pathways (ENV/JM/MONO(2013)6). In. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en

OECD (2018a) Test No. 443: Extended One-Generation Reproductive Toxicity Study. OECD Publishing, Paris

OECD (2018b) Users' handbook supplement to the guidance document for developing and assessing adverse outcome pathways. OECD Series on adverse outcome pathways No 1. Series on testing & assessment No. 233. (ENV/JM/MONO(2016)12). In. https://doi.org/10.1787/5jlv1m9d1g32-en

OECD (2020) Test No. 326: Identification and characterisation of Parkinsonian hazard liability of deguelin by an AOP-based testing and read across approach (ENV/JM/MONO(2020)22). In. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2020)22&docLanguage=en

OECD (2021) OECD Guidance Document on the Characterisation, Validation and Reporting of PBK Models for Regulatory Purposes. GD ENV/CBC/MONO(2021)1. Series on Testing and Assessment No. 331. In. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-CBC-MONO(2021)1%20&doclanguage=en

Online BBB Predictor v.0.9. In. https://www.cbligand.org/BBB/ Accessed 25.10.2020

Paparella M, Bennekou SH, 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. Reprod Toxicol 96:327-336 doi:10.1016/j.reprotox.2020.08.002

Patlewicz G, Simon TW, Rowlands JC, Budinsky RA, Becker RA (2015) Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul Toxicol Pharmacol 71(3):463-77 doi:10.1016/j.yrtph.2015.02.011

Pearl J, Mackenzie D (2019) The Book of Why. The New Science of Cause and Effect. Penguin Random House UK

Pease CK, Gentry RP (2016) Systematic review in chemical risk assessment - A chemical industry perspective. Environ Int 92-93:574-7 doi:10.1016/j.envint.2015.12.007

Perkins EJ, Antczak P, Burgoon L, et al. (2015) Adverse outcome pathways for regulatory applications: Examination of four case studies with different degrees of completeness and scientific confidence. Toxicol Sci 148(1):14-25 doi:10.1093/toxsci/kfv181

Perkins EJ, Ashauer R, Burgoon L, et al. (2019a) Building and applying quantitative adverse outcome pathway models for chemical hazard and risk assessment. Environ Toxicol Chem 38(9):1850-1865 doi:10.1002/etc.4505

Perkins EJ, Gayen K, Shoemaker JE, et al. (2019b) Chemical hazard prediction and hypothesis testing using quantitative adverse outcome pathways. ALTEX 36(1):91-102 doi:10.14573/altex.1808241

Pittman ME, Edwards SW, Ives C, Mortensen HM (2018) AOP-DB: A database resource for the exploration of Adverse Outcome Pathways through integrated association networks. Toxicol Appl Pharmacol 343:71-83 doi:10.1016/j.taap.2018.02.006

Poewe W, Seppi K, Tanner CM, et al. (2017) Parkinson disease. Nat Rev Dis Primers 3:17013 doi:10.1038/nrdp.2017.13

Pollesch NL, Villeneuve DL, O'Brien JM (2019) Extracting and benchmarking emerging adverse outcome pathway knowledge. Toxicol Sci 168(2):349-364 doi:10.1093/toxsci/kfz006

PreADMET ADME Prediction v.2.0. In. https://preadmet.bmdrc.kr/adme-prediction/ Accessed 25.10.2020

Priyadarshi A, Khuder SA, Schaub EA, Shrivastava S (2000) A meta-analysis of Parkinson's disease and exposure to pesticides. Neurotoxicology 21(4):435-40

Punt A, Bouwmeester H, Blaauboer BJ, et al. (2020) New approach methodologies (NAMs) for human-relevant biokinetics predictions. Meeting the paradigm shift in toxicology towards an animal-free chemical risk assessment. ALTEX 37(4):607-622 doi:10.14573/altex.2003242

Ramsay RR, Singer TP (1992) Relation of superoxide generation and lipid peroxidation to the inhibition of NADH-Q oxidoreductase by rotenone, piericidin A, and MPP+. Biochem Biophys Res Commun 189(1):47-52 doi:10.1016/0006-291x(92)91523-s

Roper C, Tanguay RL (2020) Chapter 40 - Tox21 and adverse outcome pathways. In: Pope CN, Liu J (eds) An Introduction to Interdisciplinary Toxicology. Academic Press, p 559-568

Roth N, Sandstrom J, Wilks MF (2020) A case study applying pathway-oriented thinking to problem formulation for planning a systematic review. Environ Int 140:105768 doi:10.1016/j.envint.2020.105768

Rowland MA, Mayo ML, Perkins EJ, Garcia-Reyero N (2019) Stochastically modeling multiscale stationary biological processes. PLoS One 14(12):e0226687 doi:10.1371/journal.pone.0226687

Rugard M, Coumoul X, Carvaillo JC, Barouki R, Audouze K (2020) Deciphering adverse outcome pathway network linked to bisphenol F using text mining and systems toxicology approaches. Toxicol Sci 173(1):32-40 doi:10.1093/toxsci/kfz214

Sachana M, Bal-Price A, Crofton KM, et al. (2019) International regulatory and scientific effort for improved developmental neurotoxicity testing. Toxicol Sci 167(1):45-57 doi:10.1093/toxsci/kfy211

Salvatier J, Wiecki T, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science(2:e55) doi:10.7717/peerj-cs.55

Sandstrom von Tobel J, Zoia D, Althaus J, et al. (2014) Immediate and delayed effects of subchronic Paraquat exposure during an early differentiation stage in 3D-rat brain cell cultures. Toxicol Lett 230(2):188-97 doi:10.1016/j.toxlet.2014.02.001

Saravanan KS, Sindhu KM, Mohanakumar KP (2005) Acute intranigral infusion of rotenone in rats causes progressive biochemical lesions in the striatum similar to Parkinson's disease. Brain Res 1049(2):147-55 doi:10.1016/j.brainres.2005.04.051

Schmidt BZ, Lehmann M, Gutbier S, et al. (2017) In vitro acute and developmental neurotoxicity screening: an overview of cellular platforms and high-throughput technical possibilities. Arch Toxicol 91(1):1-33 doi:10.1007/s00204-016-1805-9

Schultz IR, Watanabe KH (2018) The development of quantitative AOPs. In: Garcia-Reyero N, Murphy CA (eds) A systems biology approach to advancing adverse outcome pathways for risk assessment. Springer International Publishing, Cham, p 263-280

Schultz TW, Amcoff P, Berggren E, et al. (2015) A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol 72(3):586-601 doi:10.1016/j.yrtph.2015.05.016

Segura-Aguilar J, Kostrzewa RM (2015) Neurotoxin mechanisms and processes relevant to Parkinson's disease: an update. Neurotox Res 27(3):328-54 doi:10.1007/s12640-015-9519-y

Semenova E, Williams DP, Afzal AM, Lazic SE (2020) A Bayesian neural network for toxicity prediction. Comput Toxicol 16:100133 doi:10.1016/j.comtox.2020.100133

Shao K, Shapiro AJ (2018) A Web-Based System for Bayesian Benchmark Dose Estimation. Environ Health Perspect 126(1):017002 doi:10.1289/EHP1289

Slenter DN, Kutmon M, Hanspers K, et al. (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res 46(D1):D661-D667 doi:10.1093/nar/gkx1064

Song Y, Xie L, Lee Y, Tollefsen KE (2020) De novo development of a quantitative adverse outcome pathway (qAOP) network for Ultraviolet B (UVB) radiation using targeted laboratory tests and automated data mining. Environ Sci Technol 54(20):13147-13156 doi:10.1021/acs.est.0c03794

Spinu N, Bal-Price A, Cronin MTD, Enoch SJ, Madden JC, Worth AP (2019) Development and analysis of an adverse outcome pathway network for human neurotoxicity. Arch Toxicol 93(10):2759-2772 doi:10.1007/s00204-019-02551-1

Spinu N, Cronin MTD, Enoch SJ, Madden JC, Worth AP (2020) Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. Arch Toxicol 94(5):1497-1510 doi:10.1007/s00204-020-02774-7

Stephens ML, Betts K, Beck NB, et al. (2016) The emergence of systematic review in toxicology. Toxicol Sci 152(1):10-6 doi:10.1093/toxsci/kfw059

Sturla SJ, Boobis AR, FitzGerald RE, et al. (2014) Systems toxicology: from basic research to risk assessment. Chem Res Toxicol 27(3):314-29 doi:10.1021/tx400410s

Terron A, Bal-Price A, Paini A, et al. (2018) An adverse outcome pathway for Parkinsonian motor deficits associated with mitochondrial complex I inhibition. Arch Toxicol 92(1):41-82 doi:10.1007/s00204-017-2133-4

Tschudi-Monnet F, FitzGerald R (2018) Adverse outcome pathway on chronic binding of antagonist to N-methyl-D-aspartate receptors during brain development leading to neurodegeneration with impairment in learning and memory in aging. OECD Series on Adverse Outcome Pathways, No 8, OECD Publishing, Paris doi:10.1787/95f569ad-en

UN (2015) Transforming our world: The 2030 Agenda for sustainable development. A/RES/70/1. In. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

UNEP (2019) Global Chemicals Outlook II. From Legacies to Innovative Solutions: Implementing the 2030 Agenda for Sustainable Development. In. http://hdl.handle.net/20.500.11822/27651

USEPA (1998) Health effects guidelines OPPTS 870.6300 Developmental Neurotoxicity Study.
van Leeuwen CJ, Vermeire TG (2007) Risk assessment of chemicals: an introduction, 2 edn. Springer Netherlands
Villeneuve DL, Angrish MM, Fortin MC, et al. (2018a) Adverse outcome pathway networks II: Network analytics. Environ Toxicol Chem 37(6):1734-1748 doi:10.1002/etc.4124
Villeneuve DL, Crump D, Garcia-Reyero N, et al. (2014a) Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142(2):312-20 doi:10.1093/toxsci/kfu199
Villeneuve DL, Crump D, Garcia-Reyero N, et al. (2014b) Adverse outcome pathway development II: best practices. Toxicol Sci 142(2):321-30 doi:10.1093/toxsci/kfu200
Villeneuve DL, Landesmann B, Allavena P, et al. (2018b) Representing the process of inflammation as key events in adverse outcome pathways. Toxicol Sci 163(2):346-352 doi:10.1093/toxsci/kfy047
Vinken M, Knapen D, Vergauwen L, Hengstler JG, Angrish M, Whelan M (2017) Adverse outcome pathways: a concise introduction for toxicologists. Arch Toxicol 91(11):3697-3707 doi:10.1007/s00204-017-2020-z
Whaley P, Halsall C, Agerstrand M, et al. (2016) Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. Environ Int 92-93:556-64 doi:10.1016/j.envint.2015.11.002
Whitton PS (2007) Inflammation as a causative factor in the aetiology of Parkinson's disease. Br J Pharmacol 150(8):963-76 doi:10.1038/sj.bjp.0707167
Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018 doi:10.1038/sdata.2016.18
Williams DP, Lazic SE, Foster AJ, Semenova E, Morgan P (2020) Predicting drug-induced liver injury with Bayesian machine learning. Chem Res Toxicol 33(1):239-248 doi:10.1021/acs.chemrestox.9b00264
Wittwehr C, Aladjov H, Ankley G, et al. (2017) How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. Toxicol Sci 155(2):326-336 doi:10.1093/toxsci/kfw207
Wittwehr C, Blomstedt P, Gosling JP, et al. (2020) Artificial intelligence for chemical risk assessment. Comput Toxicol 13:100114 doi:10.1016/j.comtox.2019.100114
Worth AP, Patlewicz G (2016) Integrated approaches to testing and assessment. In: Eskes C, Whelan M (eds) Validation of alternative methods for toxicity testing. Springer International Publishing, p 317-342
Yang H, Lou C, Sun L, et al. (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067-1069 doi:10.1093/bioinformatics/bty707
Yozzo KL, McGee SP, Volz DC (2013) Adverse outcome pathways during zebrafish embryogenesis: a case study with paraoxon. Aquat Toxicol 126:346-54 doi:10.1016/j.aquatox.2012.09.008
Zaltieri M, Longhena F, Pizzi M, Missale C, Spano P, Bellucci A (2015) Mitochondrial Dysfunction and alpha-Synuclein Synaptic Pathology in Parkinson's Disease: Who's on First? Parkinsons Dis 2015:108029 doi:10.1155/2015/108029
Zgheib E, Gao W, Limonciel A, et al. (2019) Application of three approaches for quantitative AOP development to renal toxicity. Comput Toxicol(11):1-13 doi:10.1016/j.comtox.2019.02.001
Zuang V, Dura, A., Asturiol Bofill, D., Batista Leite, S., Berggren, E., Bernasconi, C., Bopp, S., Bowe, G., Campia, I., Carpi, D., Casati, S., Coecke, S., Cole, T., Corvi, R., Deceuninck, P., Fortaner Torrent, S., Franco, A., Gribaldo, L., Grignard, E., Halder, M., Holloway, M., Kienzler, A., Macko, P., Madia, F., Milcamps, A., Morath, S., Munn, S., Paini, A., Palosaari, T., Pistollato, F., Price, A., Prieto Peraita, M., Sund, J., Viegas Barroso, J., Wittwehr, C., Worth, A. and Whelan, M. (2019) EURL ECVAM Status Report on the Development, Validation and Regulatory Acceptance of Alternative Methods and Approaches. doi:10.2760/25602

# Appendices

## Appendix I. Overview of the 23 definitions for the qAOP concept

The definitions were obtained from the literature (as referred to in Chapter 2); listed in chronological order, the common features of a qAOP model derived from the highlighted key words, and the type of qAOP (quantitative WoE, probabilistic or mechanistic) implied by the definition.

| Common features | Type of qAOP | Extracted qAOP definition |
|---|---|---|
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Additional considerations | Mechanistic | *"For maximum application across levels of biological organization, AOPs should encompass **test endpoints or nodes** that can be quantitatively related to demographic traits of the model **population**. Models that describe **lower-level** mechanistic detail be coupled with those being used for modeling population processes and necessitates **empirically** derived **quantitative relationships** between the individual endpoint and inputs to the demographic rates of the population models. Establishing these relationships often requires simultaneous bottom-up and top-down **integration** of the physiology of the organism with the natural history of the population."* (Kramer et al. 2011) |
| Quantitative approaches<br>Additional considerations | Mechanistic | *"Ideally, causality across AOPs is approached not only in a qualitative, but also in a **quantitative** way relating **exposure** to the adverse outcome… The situation is relatively straightforward if the extent to which a MIE or KE is altered and is known to be sufficient to trigger the final AO. It is assumed that the AO occurs only after a biologically meaningful **overall threshold** has been passed." (Bal-Price et al. 2015)* |
| Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Mechanistic | *"The use of AOPs in a **full risk assessment** would require a **quantitative** description of the links between **suborganismal** changes, ecologically relevant outcomes in **individuals** and **population**-level responses. Moreover, a complete risk assessment would need to be based not only on the simplified **toxicodynamic** sequence of events depicted in the AOP itself, but also take into account **chemical**-(e.g., external and internal exposure) and **situation**-specific (e.g., outcomes for a specific field population) aspects. **Bioavailability** and **toxicokinetic** processes require specific attention if the AOP is to be used for any quantitative assessments with **regulatory relevance**. Furthermore, to enable AOP application for quantitative risk assessment, the linkages between KEs and AOs need to be defined quantitatively. To establish a quantitative AOP, thresholds for upstream MIE or KE to trigger the downstream KEs or AOs need to be defined, taking into account the potential **modifying factors** as well as site-specific contexts to the fullest extent possible."* (Groh et al. 2015) |

| | | |
|---|---|---|
| Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Mechanistic | *"A further academic perspective for future AOP research, with a strong link to **regulatory applications**, is to **quantitatively** describe the **causal links** of the AOP chain to facilitate the prediction of AOs based on the analysis of molecular initiating events or key events and the **extrapolation** between models and species. For a full quantitative approach, established **toxicokinetic** and **toxicodynamic** models could be applied to effectively integrate **exposure** and toxicokinetic information."* (Groh and Tollefsen 2015) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Regulatory applicability | Mechanistic | *"The need for predictive power has been a strong impetus to develop adverse outcome pathway (AOP) analyses, which describe networks of causally linked events at different levels of biological organization, and to develop **general quantitative methods** that summarize toxicant impact in process-based toxicity measures. **Integration** of these two developments into **quantitative AOP approaches** promises to yield powerful **predictive tools** for **ecological risk assessment**. These approaches, in which toxicity metrics relate to chemical, biological and ecological processes, eliminate or reduce the dependence of toxicity assessments on experimental design, choices of endpoint, and species of organism and chemical compound. Furthermore, quantitative AOPs open the way to using results from **(semi)automated high-throughput** and **high content screening tests** to anticipate the impact of toxicants on processes at ecologically relevant levels of biological organization. Typically, those rapid and cost-efficient screening tests record molecular, cellular or individual responses to toxicant exposure in a **dose–response manner** in order to **rank the hazard** of a group of compounds. **Data** from those screening tests could also be analyzed within a quantitative AOP framework."* (Muller et al. 2015) |
| Quantitative approaches | Probabilistic and mechanistic | *"In an idealized case, an AOP would include a description of all key events, delineation of methods which can be used to measure each key event, descriptions of each key event relationship (KER), and **quantitative models** for each KER to permit **statistical prediction** of a downstream key event from an upstream key event. If all of this information were available, quantitative predictions of the adverse outcome (AO) could be made from an upstream key event."* (Patlewicz et al. 2015) |
| Quantitative approaches<br>Additional considerations | Mechanistic | *"Stronger scientific confidence in KE relationships are required to enable AOPs to be used to understand fully pathway homology **across species**. They permit development of **quantitative models** for **extrapolating** difficult to measure outcomes such as **population** level effects from KEs."* (Perkins et al. 2015) |

| | | |
|---|---|---|
| Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Quantitative weight of evidence and mechanistic | *"Semiquantitative and quantitative AOPs (qAOPs) have been targeted for development to serve as **predictive models** in the quantitative toolbox for human and environmental health **impact assessment**. To enable transition of AOPs to **life cycle impact assessment** (LCIA), AOPs must become more quantitative in nature, with specific emphasis on establishing **quantitative dose–response relationships** between MIEs and various KEs to the adverse outcome of concern. Therefore, to facilitate LCIA, the underlying AOP framework must be quantitative and able to predict dose–response relationships between activation of MIEs or other key events to adverse outcomes that are directly involved in fitness (i.e., **effects** that impair the ability of individuals to survive or reproduce). A quantitative AOP should provide numerically based assessments to attempt to **quantify weight of evidence** (WoE)."* (Gust et al. 2016) |
| Regulatory applicability | Mechanistic | *"To provide information useful for **chemical risk assessment**, AOPs need to have some quantitative information relevant to important KERs. Beyond characterizing the pathway, it is important to understand the **dose** that activates the pathway, and if it is **relevant** to human or ecological exposure scenarios."* (Kleinstreuer et al. 2016) |
| Mechanistic knowledge and associated data | Mechanistic | *"It is important to understand the key event relationships (KERs) and to **provide relevant information** or, even better, **quantitative data** supporting KERs, especially between the early KEs."* (Bal-Price et al. 2017) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Probabilistic and mechanistic | *"The term quantitative AOP (qAOP) refers to a loosely defined, but relatively advanced stage in the progression of AOP development and description. At this stage, **quantitative understanding** of the relationships underlying transition from one KE to the next, as well as **critical factors that can modulate** those relationships, are sufficiently well-defined to allow **quantitative prediction** of the **probability** or **severity** of the AO occurring for a given activation of the MIE. **Information** concerning the quantitative understanding of what defines the transition from one KE in an AOP to the next is thus captured and included (where possible) in the KE relationship descriptions. That quantitative understanding may take **many forms**, depending on the extent of the **available, relevant data**. In the case of a relatively limited data set containing little or no or dose–response and time-course information, the relationship between adjacent KEs may be as simple as a **linear regression equation** linking an upstream with an immediately downstream KE. With richer data sets, reflecting fuller dose-response and time-* |

| | | |
|---|---|---|
| | | *course designs, the quantitative understanding may be encoded into sophisticated **biologically based models** that simulate complex, nonlinear, dynamics that can result from **feedback loops**, **adaptive** and **compensatory responses**, **stochastic influences**, **interactions** with other pathways, and/or influences of external or internal **modulating factors**. Whatever form they take, quantitative understanding of the KE relationships encompassed in an AOP description can facilitate a broader spectrum of **applications**. Consequently, there is interest in developing the quantitative understanding and description of AOPs to the extent that **regulatory needs** warrant and resources allow."* (Conolly et al. 2017) |
| Regulatory applicability | Quantitative weight of evidence and mechanistic | *"KERs facilitate inference or extrapolation based on the premise that if the upstream KE is altered to a sufficient degree, predictable changes (qualitative or quantitative) can be expected in the downstream event in the hypothesized AOP. For empirical support, qualitative consideration of the extent of supporting data or **WOE** for hypothesized AOPs takes into account "patterns" of quantitative relationships for KERs (i.e., the extent to which **temporal and dose response** patterns align with what would be anticipated, for essential key events in an AOP). This differs from quantitation of the KERs, addressing essentially how much change in KEup is needed to evoke some unit of change in KEdown as a basis for developing **predictive response models**."* (Meek 2017) |
| Mechanistic knowledge and associated data<br>Quantitative approaches | Mechanistic | *"Currently there are few, or no, examples of **(Q)SAR** or **QAAR** models for Key Event Relationships, although some **in silico models** for Key Event Relationships are becoming available, especially in the form of quantitative AOPs (qAOPs). Quantitative models for the MIE, as well as for Key Events and Key Event Relationships require a more **complete data set** with information from a greater number of **compounds** covering a range of **activity** and **properties**." (Cronin and Richarz 2017)* |
| Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Mechanistic | ***"Computational predictive modelling** can be applied to quantitatively describe the sequences of key events (KEs) and their relationships (KE relationships, KER) and the biologic processes of pathogenesis that comprise an AOP. The **reliability** of predictions from these models is improved with greater **understanding of the biologic foundation** of the AOP. Predictions also are improved by complete appreciation of the **quantitative determinants** from upstream biologic perturbations to subsequent downstream overt organismal effects comprising the AO. When **mathematical** descriptions of these relationships are biologically driven **equations**, a mechanistic* |

| | | |
|---|---|---|
| | | *quantitative AOP (qAOP) is derived."* (Hassan et al. 2017) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Regulatory applicability | Probabilistic and mechanistic | *"Information captured in the "quantitative understanding of the linkage" section of the KER descriptions within the AOP framework provides the foundation for addressing this desire for quantification. Quantitative AOPs can be described in various ways, ranging from* ***expert judgment-based scoring****, requiring limited information, where elements of the AOP are weighted using expert opinion, to more* ***probabilistic approaches****, where statistical relationships exist between the MIE/KE and the adverse outcome, to* ***mechanistic approaches****. The more mechanistic approaches employ* ***mathematical models*** *or relationships of MIE, KE, and KER (e.g., response–response relationships between KERs) to quantitatively* ***predict the risk*** *of an adverse effect given specified* ***initial conditions*** *(e.g., a set of exposure conditions)."* (LaLone et al. 2017a) |
| Quantitative approaches | Mechanistic | *"Whereas an AOP description lays out the sign posts within a biological system that indicate progression toward an AO, computational models can quantitatively* ***simulate*** *the* ***dynamics*** *of the complex biology at multiple scales that dictate dose–response and time–course behaviors and* ***define the conditions*** *under which perturbation of early KEs in the pathway will ultimately lead to the AO, or not."* (Wittwehr et al. 2017) |
| Quantitative approaches<br>Additional considerations | Mechanistic | *"It is possible to assemble quantitative AOPs (qAOPs) that consider* ***quantitative relationships*** *between KEs, including* ***feedback*** *models* ***designed*** *to reflect system regulation, to predict* ***AOs****."* (Ankley and Edwards 2018) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Additional considerations | Mechanistic | *"The* ***integration*** *of all* ***information*** *will also lead to the development of quantitative AOPs that can be used for* ***dose-response analyses****, and iteratively, inform* ***refinements*** *of the next generation of mechanistic IATAs."* (Clippinger et al. 2018) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Probabilistic and mechanistic | *"The most advanced developments in AOPs, known as quantitative AOPs (qAOPs), have potential utility to* ***quantitative ecological risk assessments****. A qAOP describes* ***quantitative response–response relationships*** *linking the molecular initiating event and adverse outcome to enable quantitative prediction of the* ***probability*** *of occurrence or* ***severity*** *of an adverse outcome for a given magnitude of chemical interaction with a molecular initiating event. Depending upon the extent of* ***mechanistic understanding*** *and the needs in terms of* ***regulatory application****, a qAOP could be as simple as a* ***linear regression*** *that quantitatively links the molecular initiating event to the adverse outcome, or as complex as a consecutive series of* ***nonlinear models*** *which* |

| | | |
|---|---|---|
| | | *describe responses at several levels of biological organization and **simulate** associated internal and external **modifying factors**."* (Doering et al. 2018) |
| Mechanistic knowledge and associated data<br>Quantitative approaches | Mechanistic | *"A quantitative adverse outcome pathway (qAOP) is a **mathematical/computational model** that represents the **dynamic** processes linking a molecular initiating event with an adverse outcome. A unique feature that distinguishes a qAOP from other biologically based mathematical models is the prediction of key events that are part of the qualitative adverse outcome pathway and are **measurable experimentally**."*<br>(Schultz and Watanabe 2018) |
| Additional considerations | Mechanistic | *"Quantitative AOPs will help answer what **level of in vitro perturbation** should be used as a point of departure (PoD) for quantitative in vitro to in vivo extrapolations (QIVIVE)."* (Beilmann et al. 2018) |
| Quantitative approaches | Probabilistic and mechanistic | *"For each pair of KEs, a quantitative KE relationship (KER) can be derived as a **response-response function** or a **conditional probability matrix** describing the anticipated change in a KE based on the response of the prior KE. This transfer of response across KERs can be used to assemble a quantitative AOP*." (Foran et al. 2019) |
| Mechanistic knowledge and associated data<br>Quantitative approaches<br>Regulatory applicability<br>Additional considerations | Probabilistic and mechanistic | *"Quantitatively, a KER may be defined in terms of **regressions** between KEs response-response relationships or dose-dependent transitions. They may take the form of **simple mathematical equations** or sophisticated **biologically based computational models** that consider other **modulating factors**, such as compensatory responses, or interactions with other biological or environmental variables. Depending on the level and nature of **empirical data available**, there is a continuum of AOPs from purely descriptive qualitative AOPs to qAOP models with detailed response-response relationships that allow one to infer the **magnitude** or **probability** of an AO. Here, we define a full qAOP model to be any mathematical construct that models the dose response or response-response relationships of all KERs described in an AOP, a partial qAOP as a construct that models the dose/response-response relationships of more than one KER, and a quantitative KER as a construct that models a single dose/response-response relationship. qAOP models support explicit **incorporation** of complex relationships, such as **feedback loops**, **thresholds**, and **signaling cascades** that are generally embedded in the KE or KER of descriptive AOPs. Models incorporating complex biological relationships can create predictions with greater biological fidelity to **support hazard and risk assessment** than models with simplified assumptions."* (Perkins et al. 2019a) |

# Appendix II. Supplemental results of Chapter 4

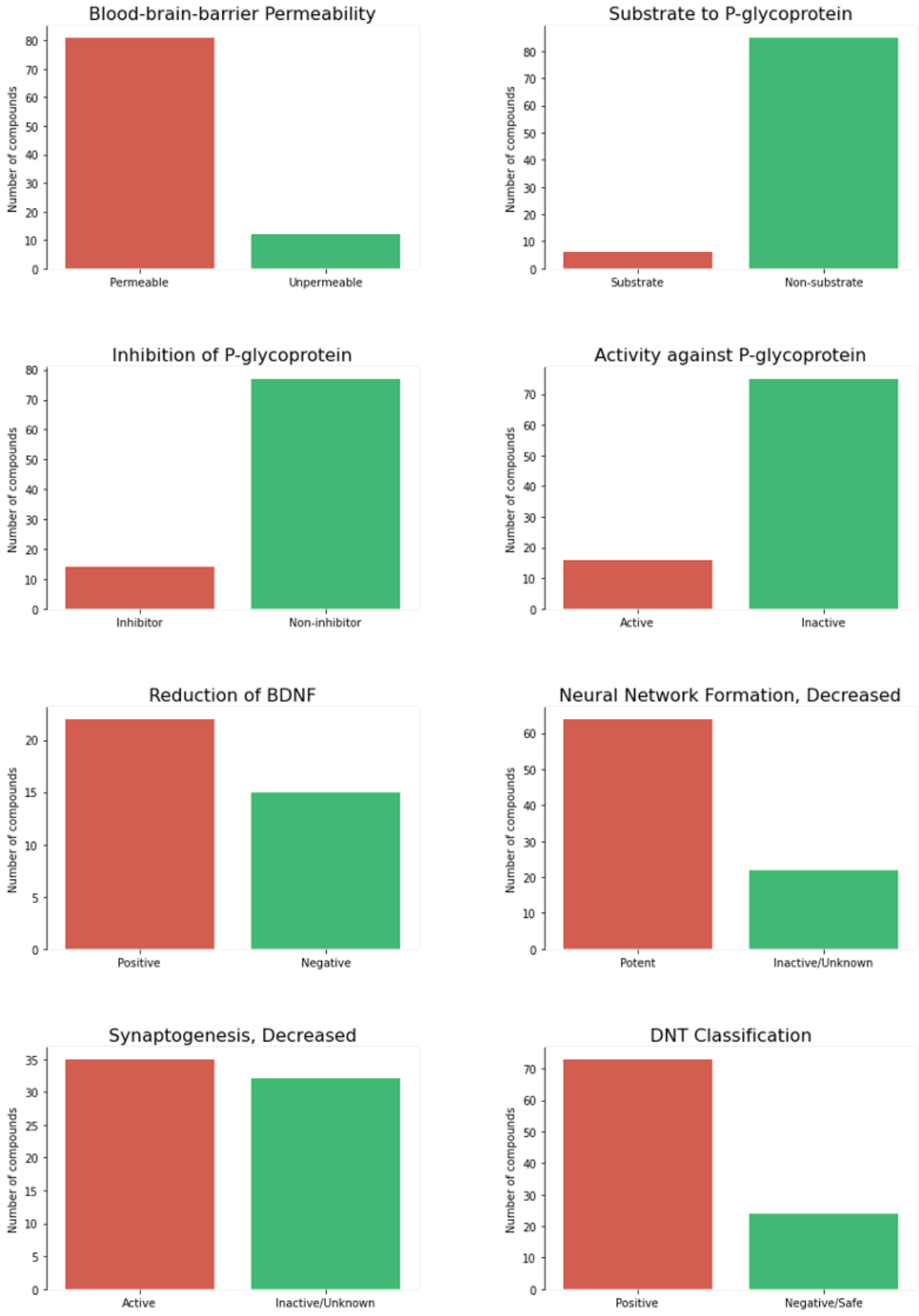


Figure S1. An overview of the distribution of the categorical type of variables that the machine-readable dataset contained to be analysed by the proposed Bayesian hierarchical model.

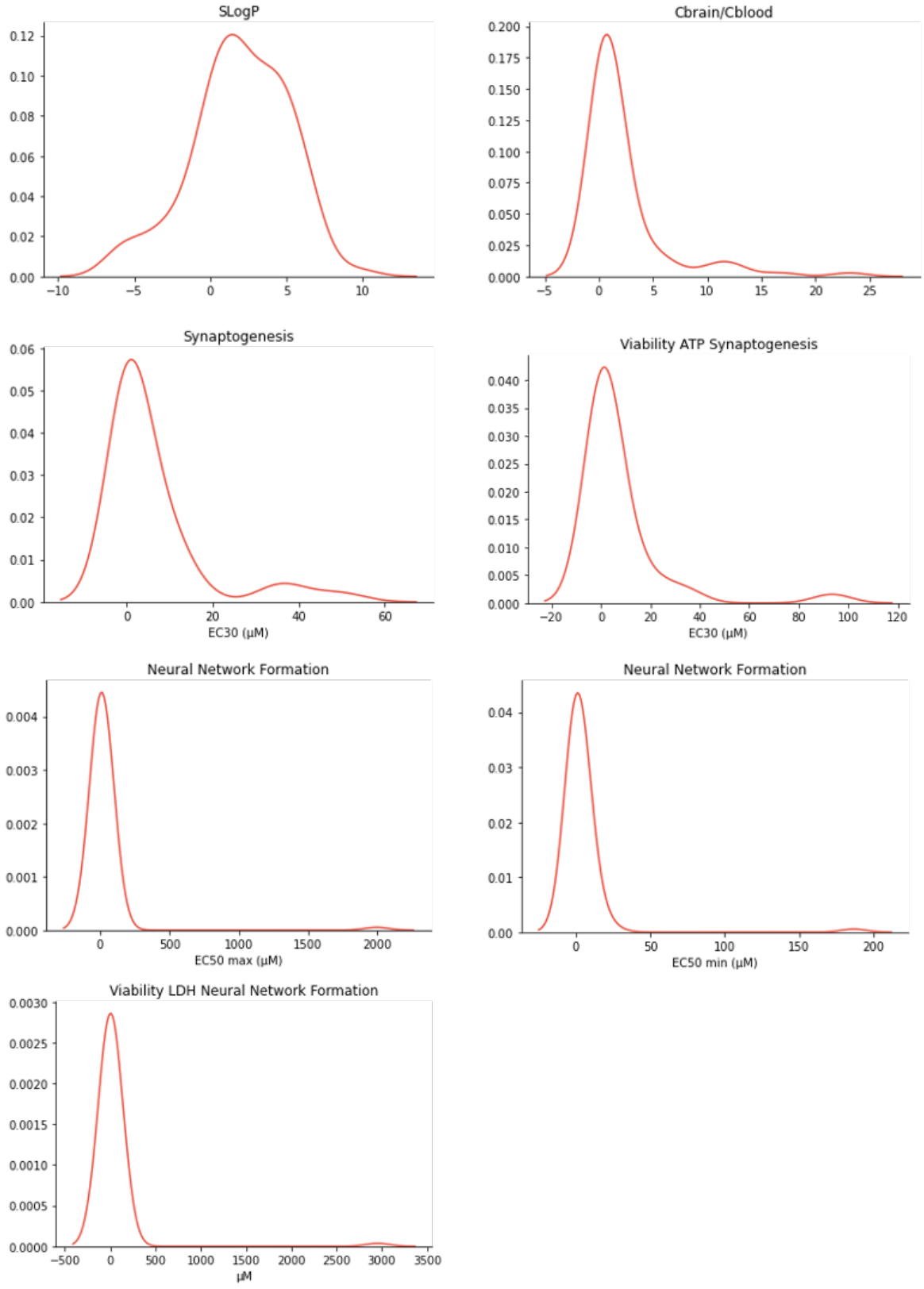


Figure S2. An overview of the distribution of the continuous type of variables that the machine-readable dataset contained to be analysed by the proposed Bayesian hierarchical model.

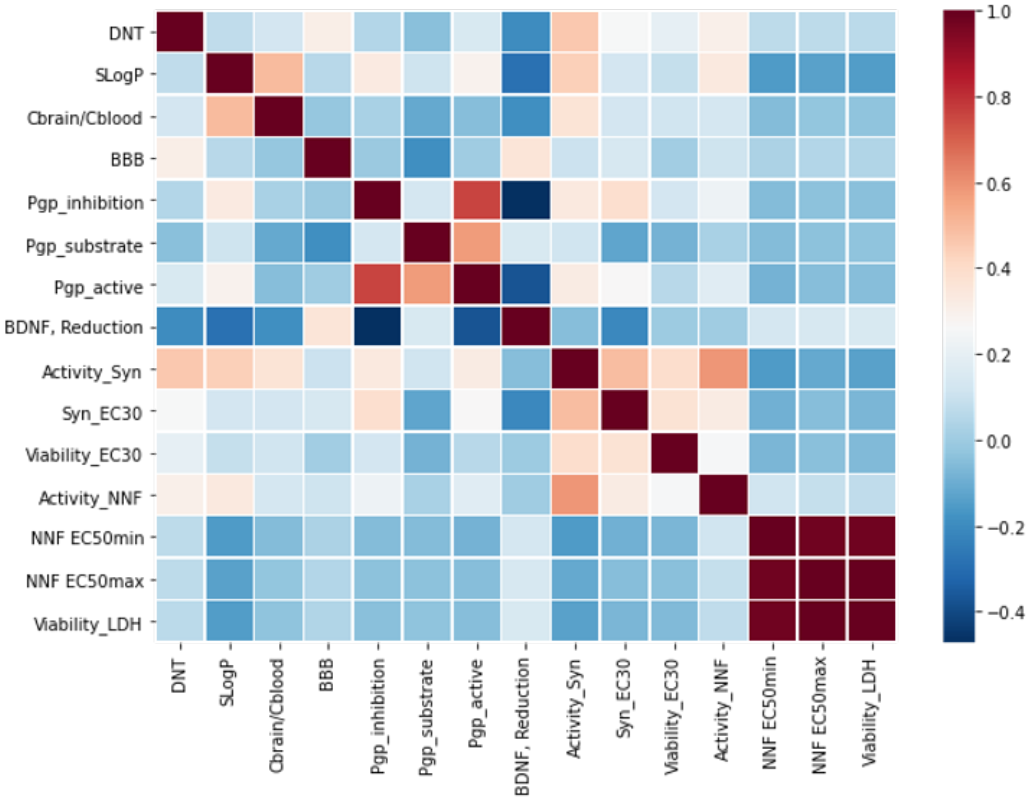


Figure S3. A visual representation of the correlation matrix between both categorical and continuous variables of the machine-readable dataset. It was generated based on the Pearson correlation coefficient. A value of one shows a total positive linear correlation.

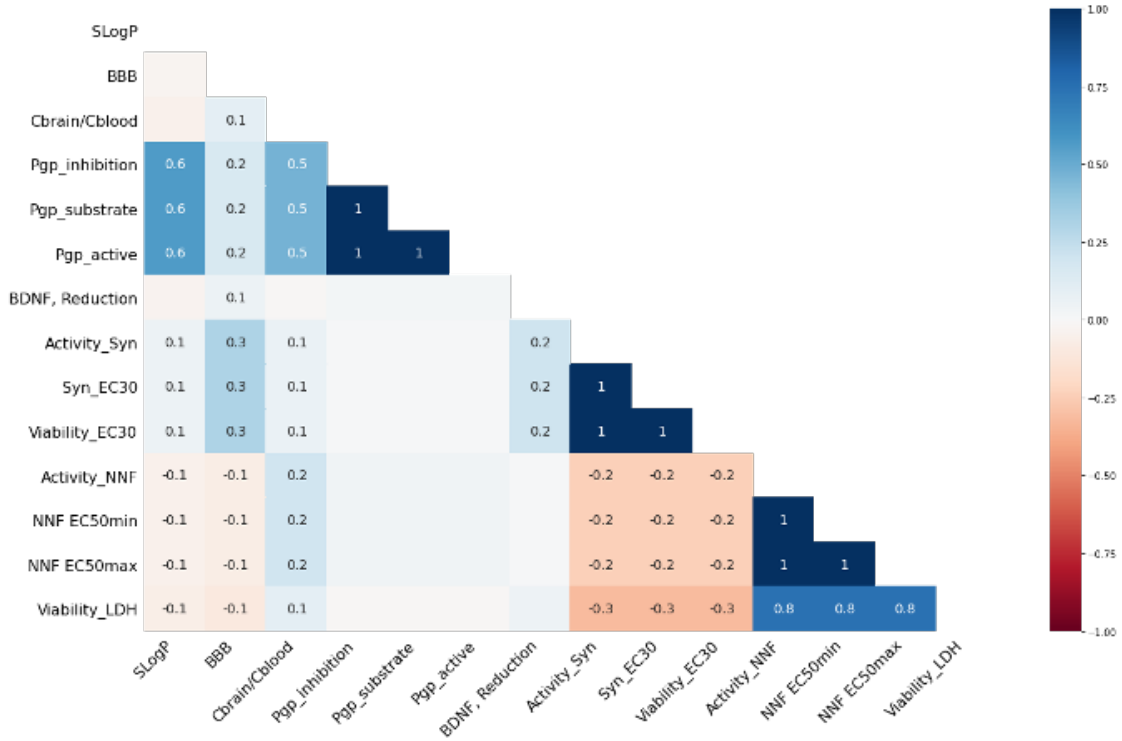


Figure S4. A visual representation of the correlation matrix of the missing values for all variables of the machine-readable dataset. It shows how strongly the presence or absence of one variable affects the presence of another variable. It was calculated using the Missingno v.0.4.2 Python package.

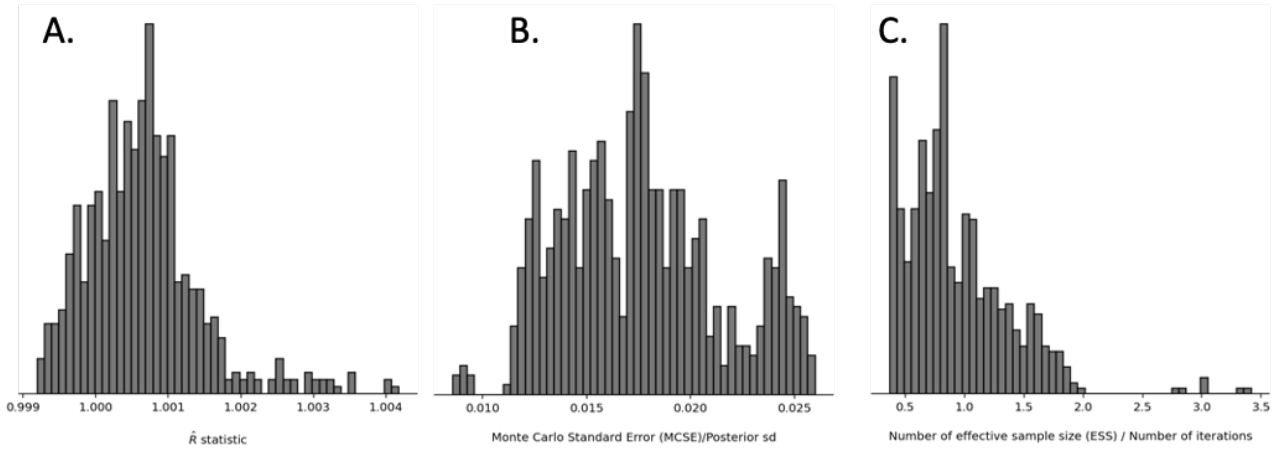


Figure S5. A visual summary of the statistical parameters that shows how well the samples from the inference step has been converged. A. The R-hat statistics, also known as Gelman-Rubin convergence diagnostic, indicates whether there is a lack of convergence of the posterior sampling. Herein, it did not exceed the threshold of 1.005. B. The Monte Carlo standard error is given by the posterior standard deviation divided by the square root of the number of the effective samples. The smaller it is, the closer the posterior mean is expected to be to the actual value. Herein, it did not exceed 10% of the posterior standard deviation. C. The number of the effective sample size should exceed the actual number of the samples. Herein, it is larger than 10% of the number of samples.

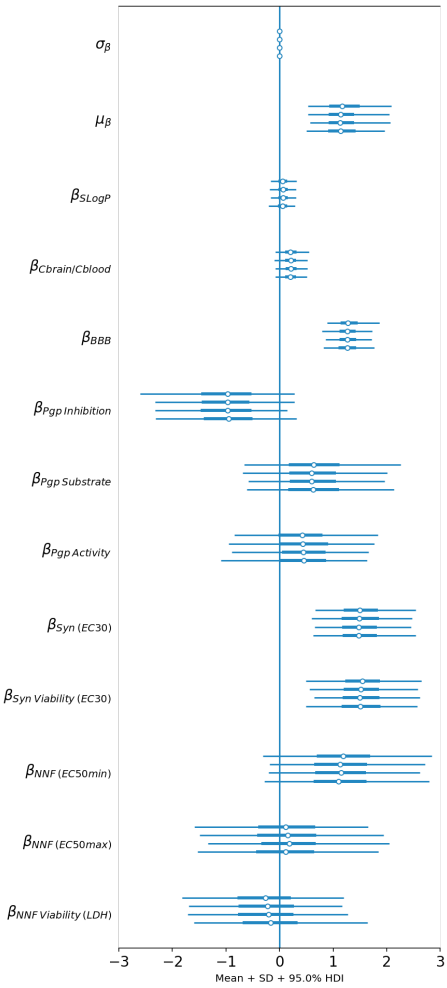


Figure S6. A forest plot of the statistical summary of the hyperpriors and priors of the inference sampling. It shows the results for each of the parameter computed on data. The dot represents the mean, the thicker line is the standard deviation, and the thin line represents the 95% of the Bayesian credible interval, also known as the highest density interval (HDI), of each independent chain out of four chains run in total.

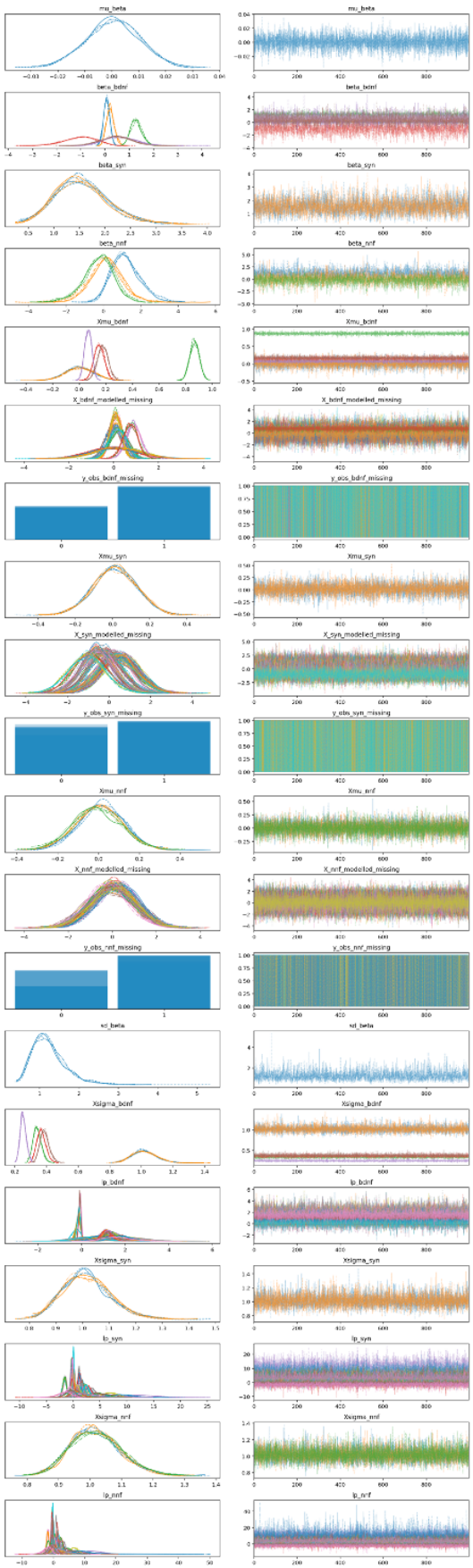


Figure S7. Distributions of the sampled values for each parameter of the Bayesian hierarchical model. On the left plot, the estimated probability distribution for each independent trace is shown. On the right plot, the actual sampling followed through the distribution is shown. The colours indicate the one set of the predictors on which the parameters have been computed.

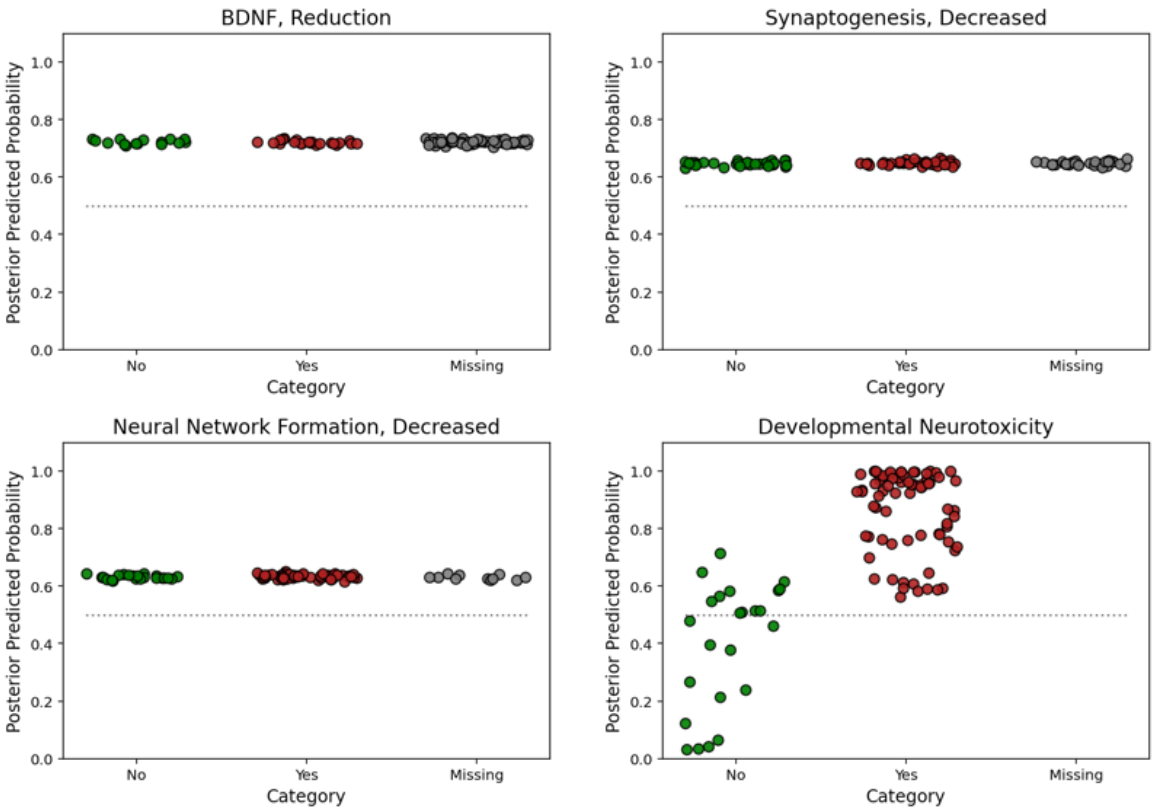


Figure S8. Results of the predictions on the outcomes for which the Bayesian hierarchical model has been screened.

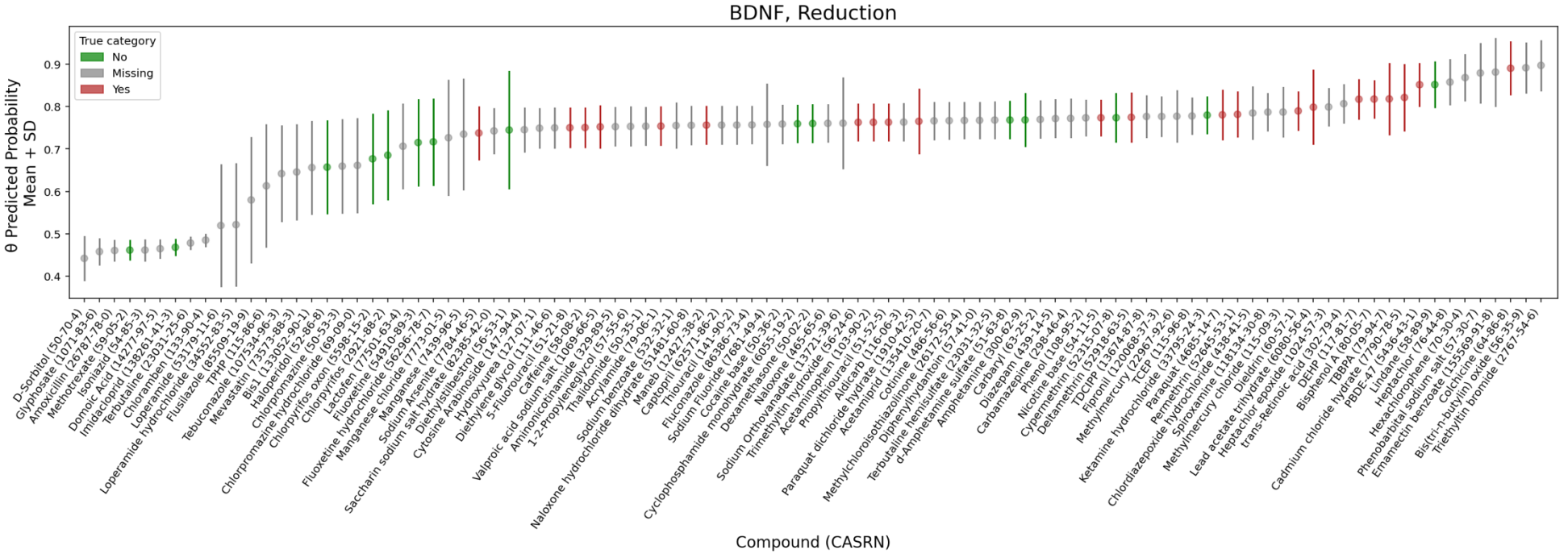


Figure S9. The predicted probabilities of the $\theta$ likelihood function colour-coded based on the correct category of compounds being associated with inducing or not the reduction of brain-derived neurotrophic factor (BDNF).

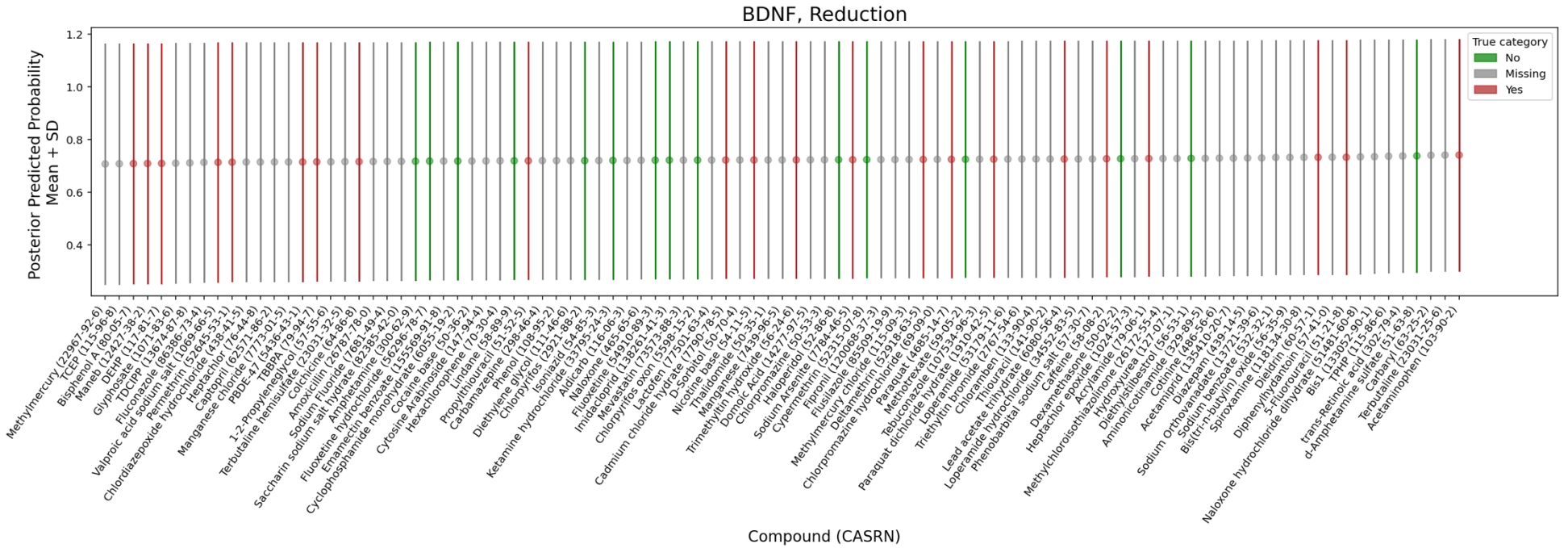


Figure S10. The posterior predicted probabilities calculated on the observed variables defined based on a binary classification resultant of the literature review. It shows ordered compounds for potency for inducing the reduction of brain-derived neurotrophic factor (BDNF).

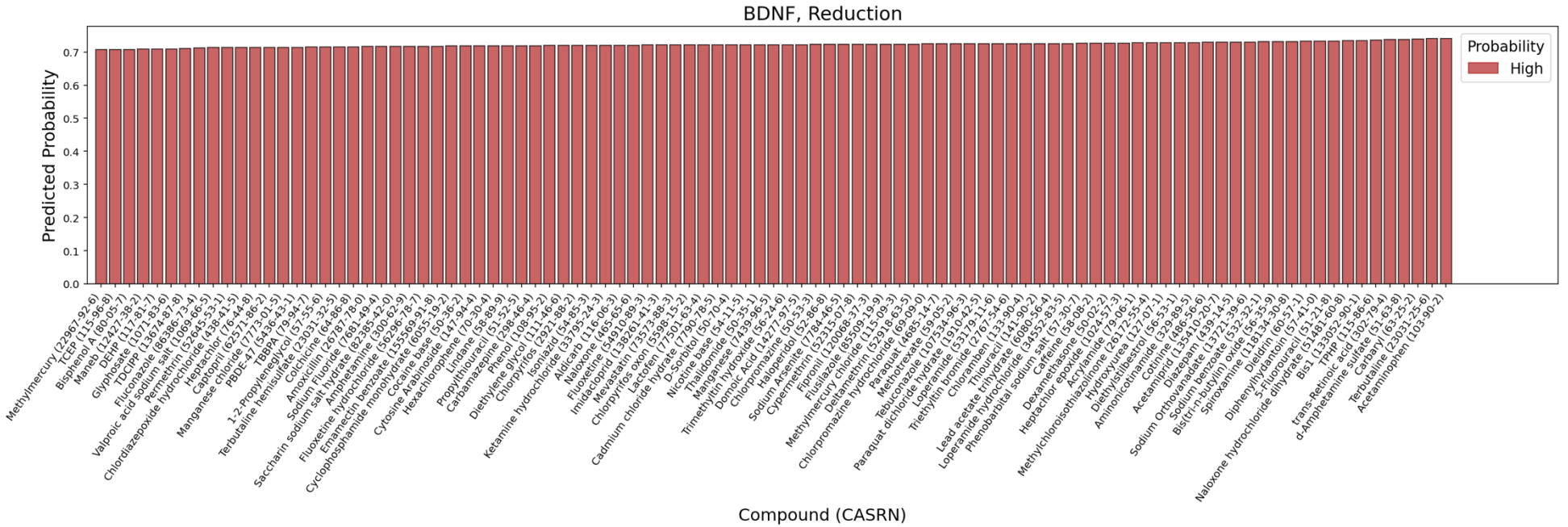


Figure S11. The predicted probabilities colour-coded based on two thresholds estimated from the results set to group the compounds for their high, medium and low probability of inducing the KE for potential future purposes, such as additional investigation and prioritisation testing schemes. Herein, all the compounds have been predicted with a high level of probability.

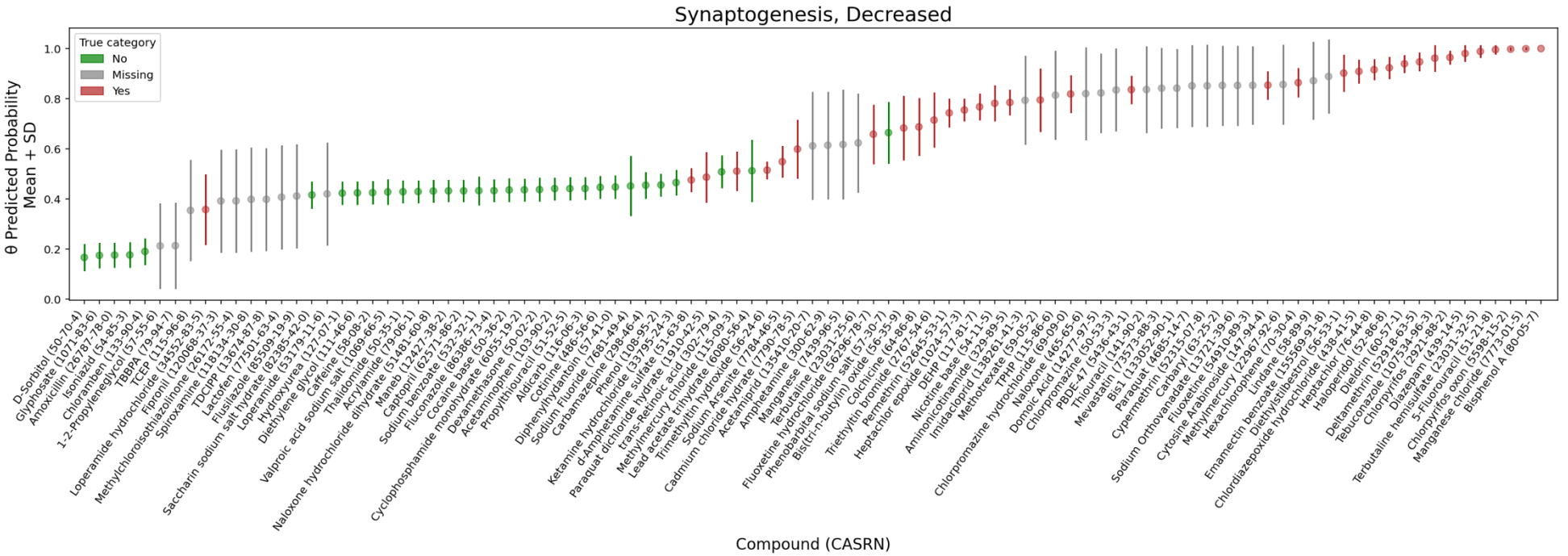


Figure S12. The predicted probabilities of the $\theta$ likelihood function colour-coded based on the true category of compounds being associated with inducing or not the decrease of synaptogenesis.

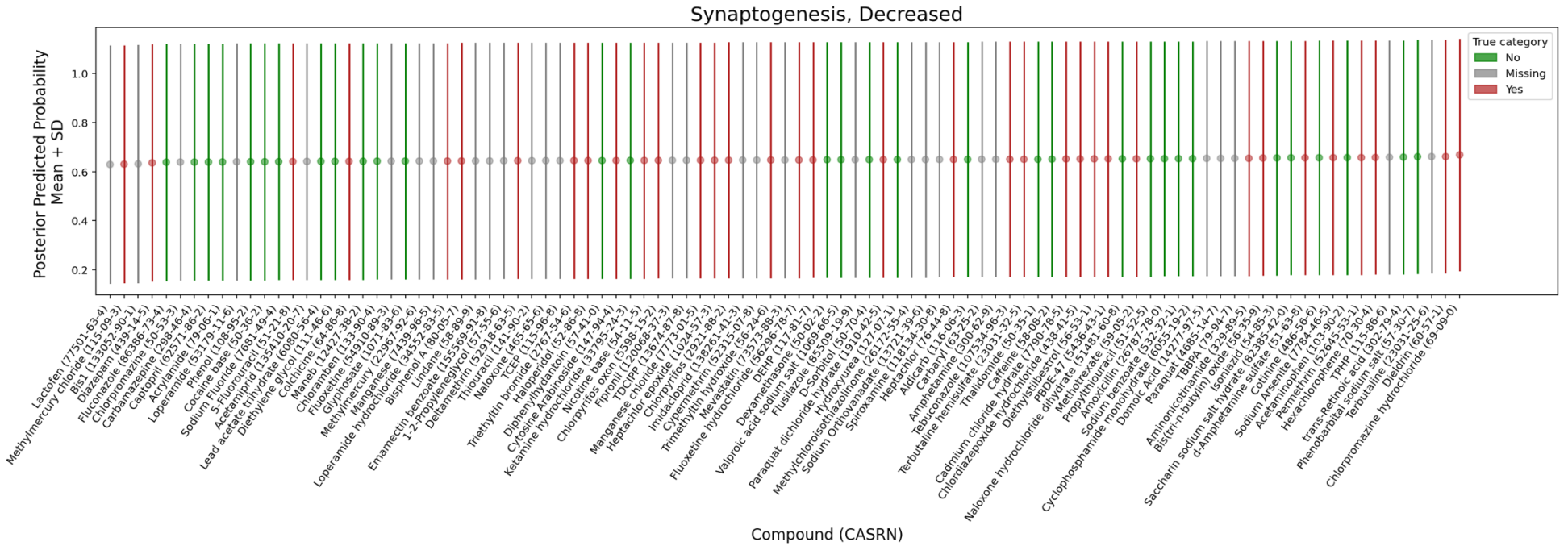


Figure S13. The posterior predicted probabilities calculated on the observed variables defined based on the classification proposed by the *in vitro* study. It shows ordered compounds for potency for inducing the decrease of synaptogenesis.

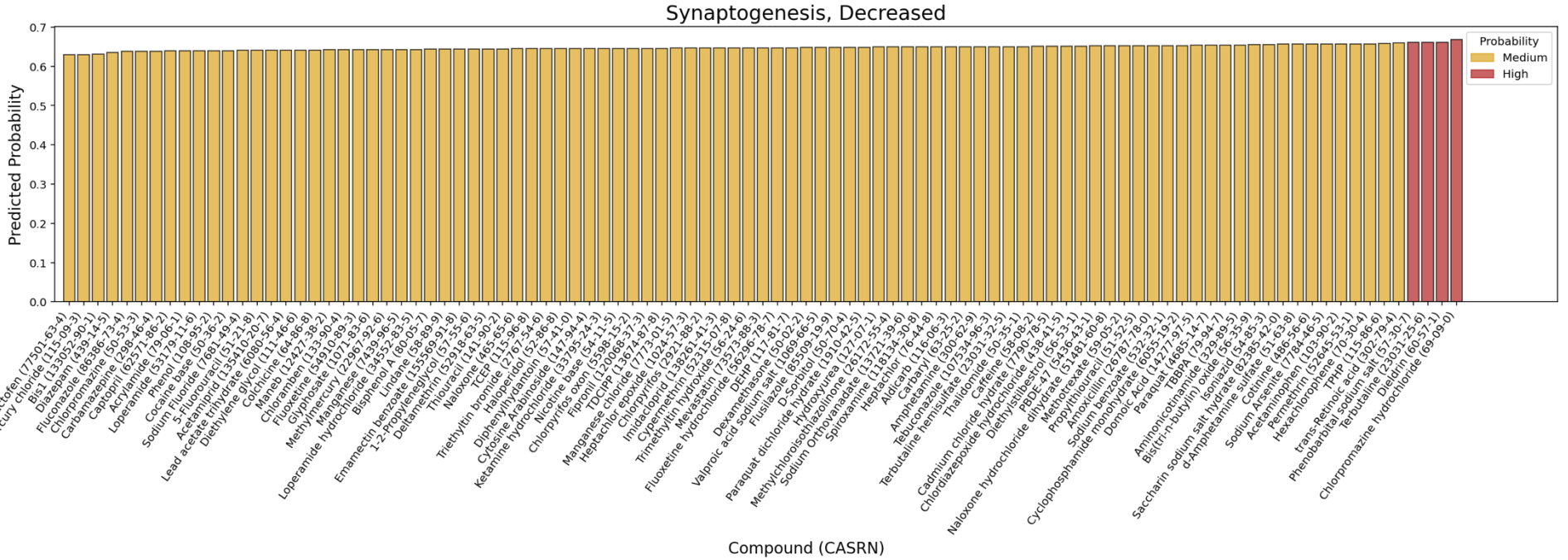


Figure S14. The predicted probabilities colour-coded based on two thresholds estimated from the results set to group the compounds for their high, medium and low probability of inducing the KE for potential future purposes such as additional investigation and prioritisation testing schemes. Herein, none of the compounds has been predicted with a low level of probability.

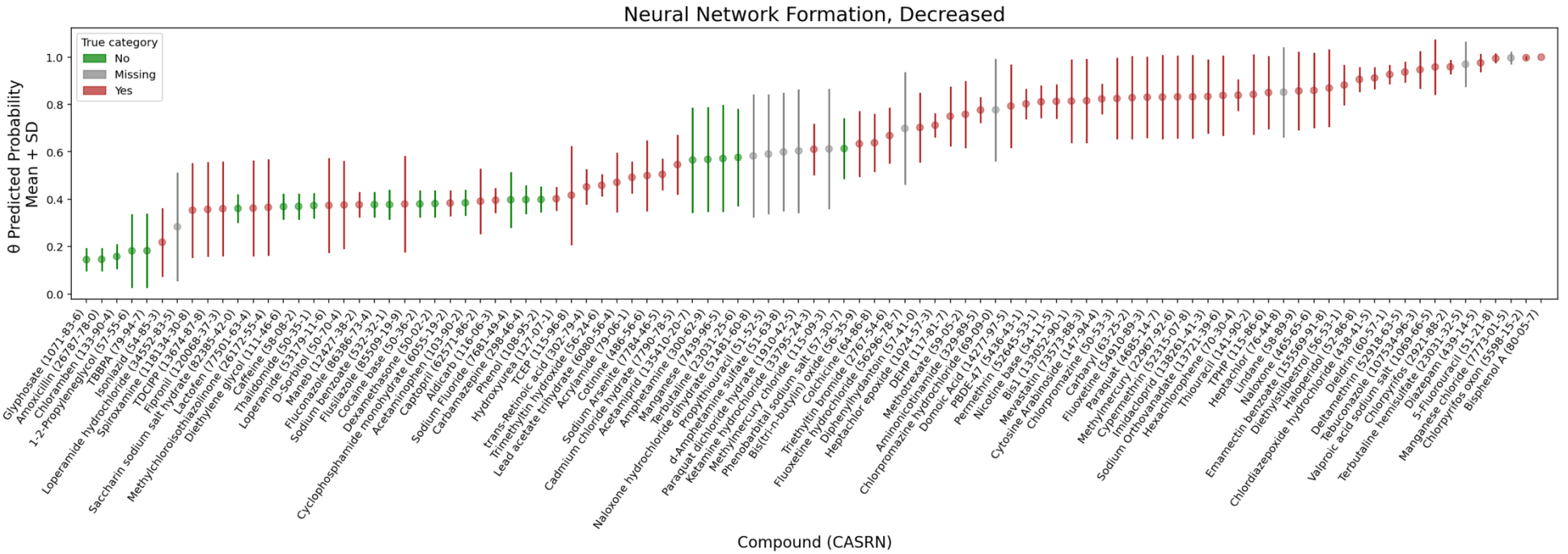


Figure S15. The predicted probabilities of the $\theta$ likelihood function colour-coded based on the true category of compounds being associated with inducing or not the decrease of neural network formation.

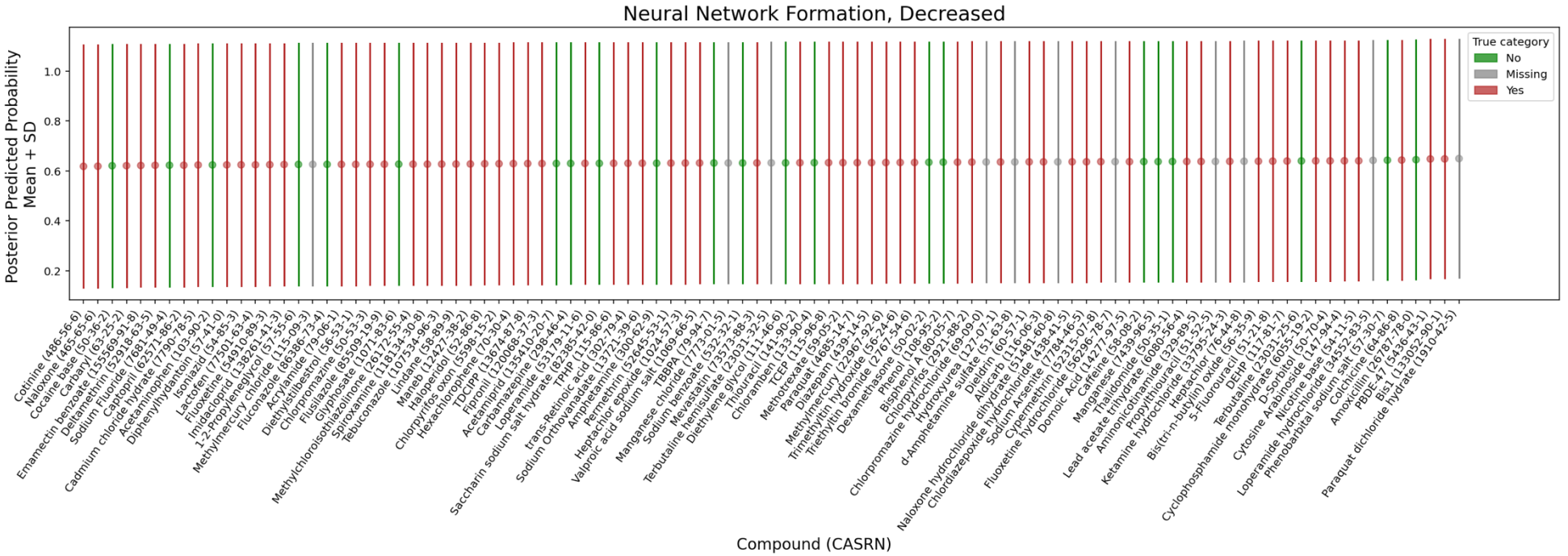


Figure S16. The posterior predicted probabilities calculated on the observed variables defined based on the classification proposed by the *in vitro* study. It shows ordered compounds for potency for inducing the decrease of neural network formation.

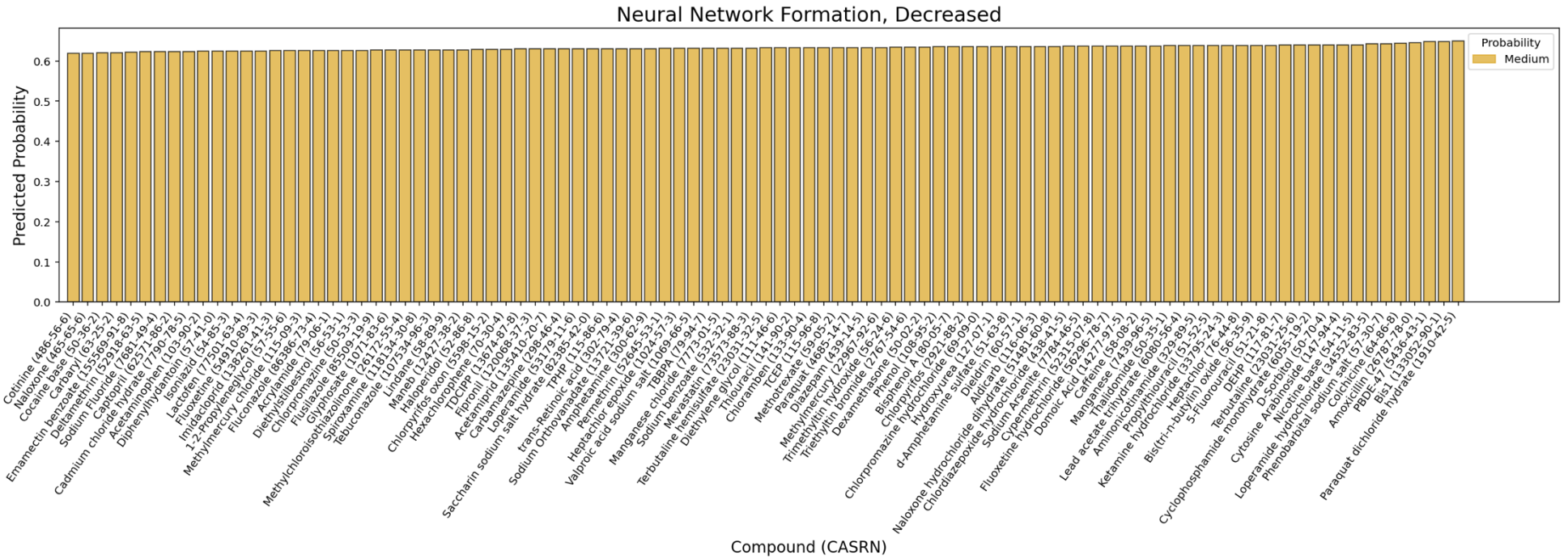


Figure S17. The predicted probabilities colour-coded based on two thresholds estimated from the results set to group the compounds for their high, medium and low probability of inducing the KE for potential future purposes such as additional investigation and prioritisation testing schemes. Herein, all the compounds have been predicted with a medium level of probability.

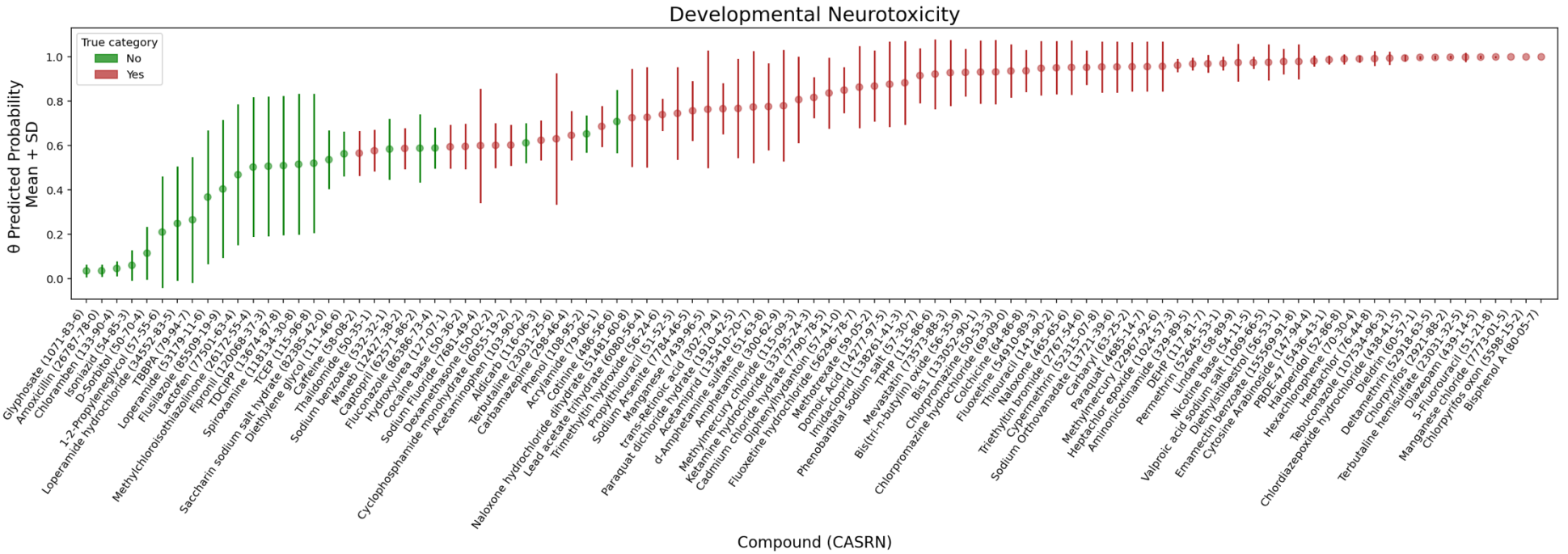


Figure S18. The predicted probabilities of the $\theta$ likelihood function colour-coded based on the correct category of the compounds being associated with inducing, or not, DNT effects.

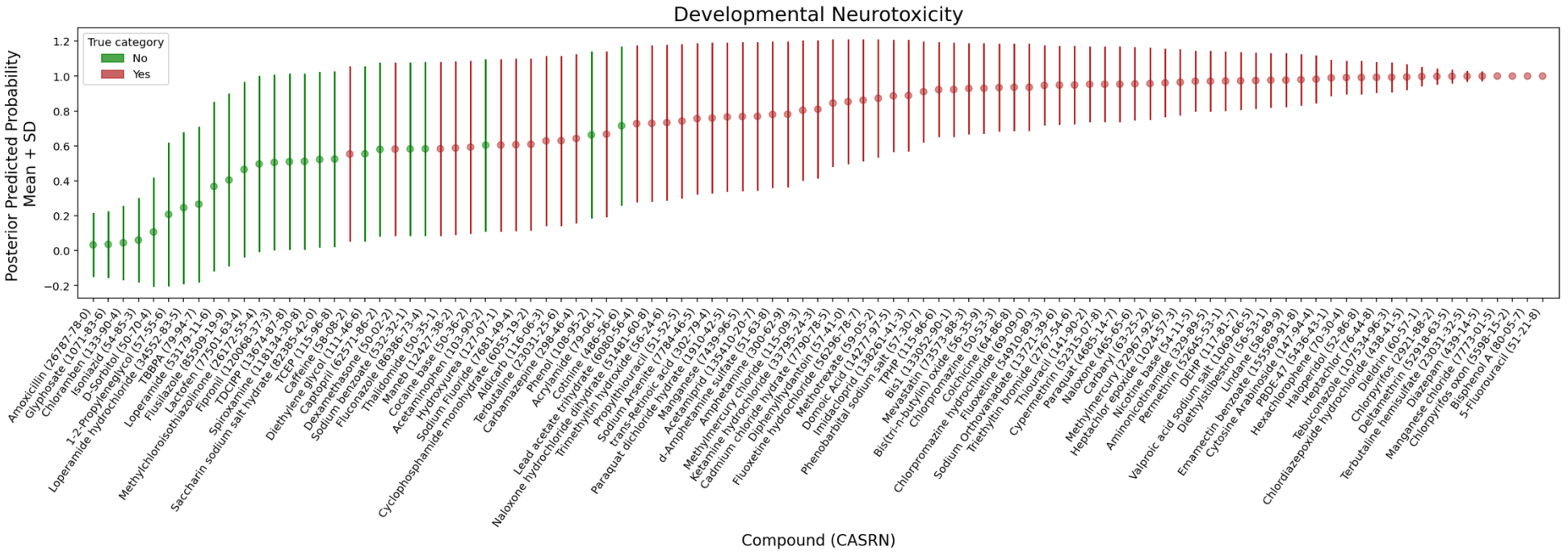


Figure S19. The posterior predicted probabilities calculated on the observed variables defined based on *in vivo* studies. Compounds were ordered for potency for DNT.

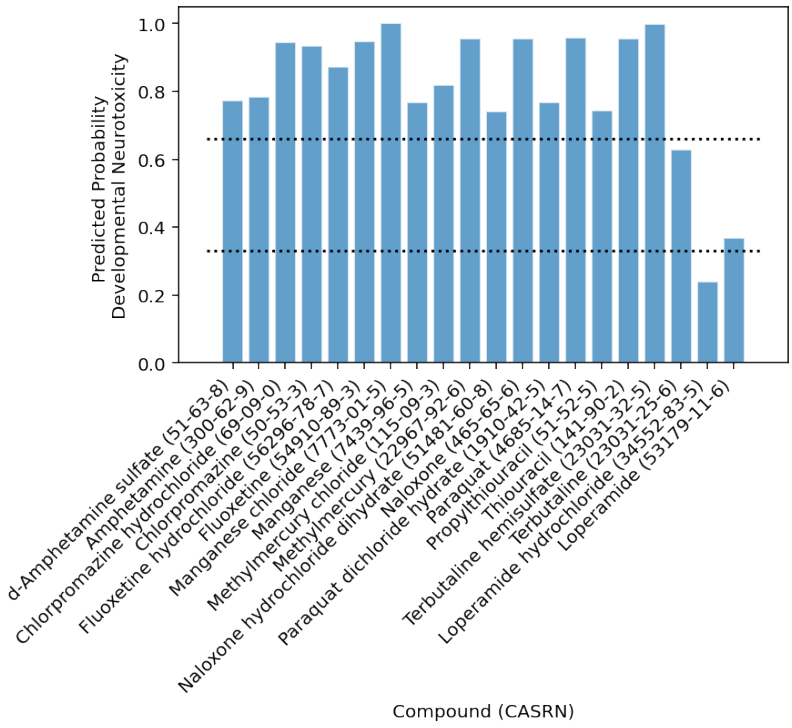


Figure S20. Predictions for the ten compounds tested under different CASRN and that were kept separate throughout the modelling. The dotted lines represent the thresholds set to categorise the compounds into low, medium, and high level of probability of inducing DNT.

# Appendix III. Supplemental results of Chapter 5

Table S1. A summary of the resultant fitted response-response functions per each time point and KER that link three KEs, i.e., inhibition of complex I, mitochondrial dysfunction and impaired of proteostasis using *drc* R package v. 3.0-1.

| KER | KER Name | Time point | Best-fit model | Mathematical equation |
|---|---|---|---|---|
| **1** | Complex I, Inhibition -> ROS production, Increase | 4 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 76.9 + \frac{613.3 - 76.9}{1 + \exp(-2.1(\log(x) - \log(514.4)))}$ |
| **2** | Complex I, Inhibition -> Mitochondrial Membrane Permeability, Decrease | 4 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 84.8 + \frac{112.2 - 84.8}{1 + \exp(11.2(\log(x) - \log(90.3)))}$ |
| **3** | ROS production, Increase -> Mitochondrial Membrane Permeability, Decrease | 4 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 14.3 + \frac{130.9 - 14.3}{1 + \exp(3.5(\log(x) - \log(125.8)))}$ |
| **1** | ROS production, Increase -> Lipid peroxidation, Increase | 24 h | Three-parameter log-logistic (LL.3) | $f(x,(b,d,e) = \frac{102.3}{1 + \exp(-6.7(log(x) - log(84.5)))}$ |
| **2** | Lipid peroxidation, Increase -> Mitochondrial Membrane Permeability, Decrease | 24 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 78.5 + \frac{109.7 - 78.5}{1 + \exp(7.2(\log(x) - \log(70.7)))}$ |
| **3** | ROS production, Increase -> Mitochondrial Membrane Permeability, Decrease | 24 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = -559.9 + \frac{105.7 + 559.9}{1 + \exp(7.9(\log(x) - \log(193.9)))}$ |
| **1** | ROS production, Increase -> Lipid peroxidation, Increase | 48 h | Three-parameter log-logistic (LL.3) | $f(x,(b,d,e) = \frac{162.6}{1 + \exp(-8.7(log(x) - log(116.3)))}$ |
| **2** | Lipid peroxidation, Increase -> Mitochondrial Membrane Permeability, Decrease | 48 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 59.9 + \frac{105.2 - 59.9}{1 + \exp(12.4(\log(x) - \log(123.9)))}$ |
| **3** | ROS production, Increase -> Mitochondrial Membrane Permeability, Decrease | 48 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 67.9 + \frac{101.5 - 67.9}{1 + \exp(277.4(\log(x) - \log(136.8)))}$ |
| **4** | ROS production, Increase -> 20S Proteasome Activity, Decrease | 48 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 124.8 + \frac{145.1 - 124.8}{1 + \exp(264.6(\log(x) - \log(90.1)))}$ |
| **5** | Mitochondrial Membrane Permeability, Decrease -> 20S Proteasome Activity, Decrease | 48 h | Four-parameter log-logistic (LL.4) | $f(x,(b,c,d,e)) = 63.4 + \frac{91.4 - 63.4}{1 + \exp(-57.9(\log(x) - \log(85.1)))}$ |

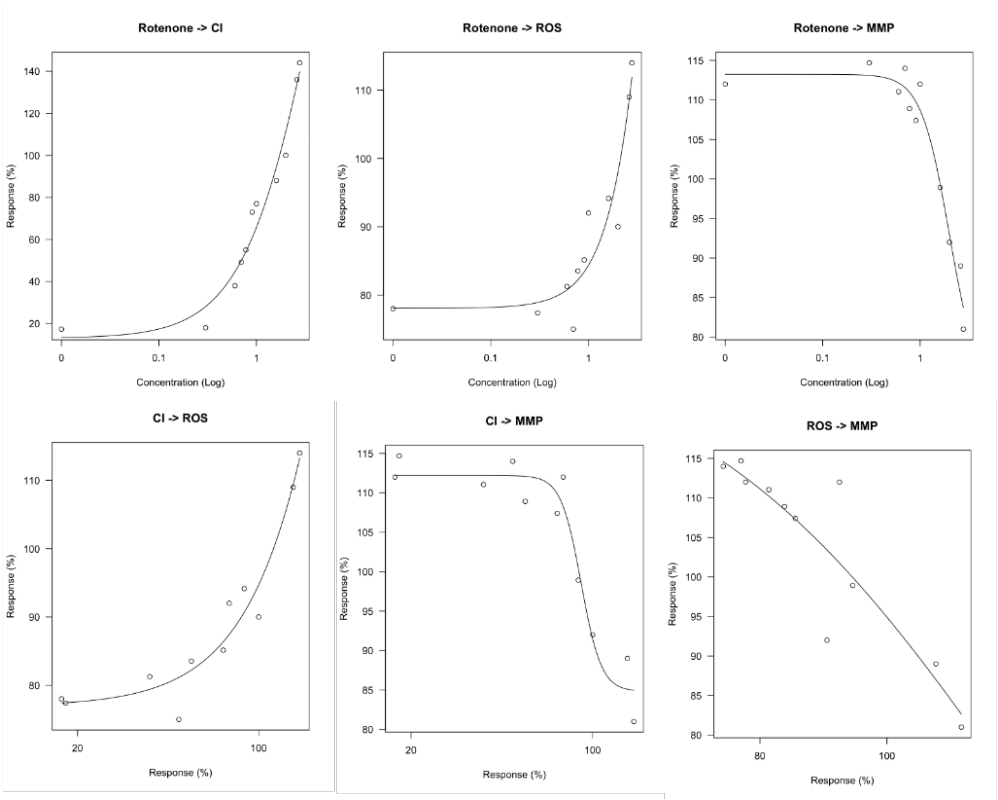


Figure S1. The fitted curves to the causal relations of both KEs, i.e., inhibition of complex I leading to mitochondrial dysfunction, and KERs for data measured after four hours conducted in *drc* R package after the imputation of the missing data using *mice* R package. CI, Complex I, Inhibition; ROS, Production of reactive oxygen species, Increase; MMP, Mitochondrial Membrane Permeability, Decrease.

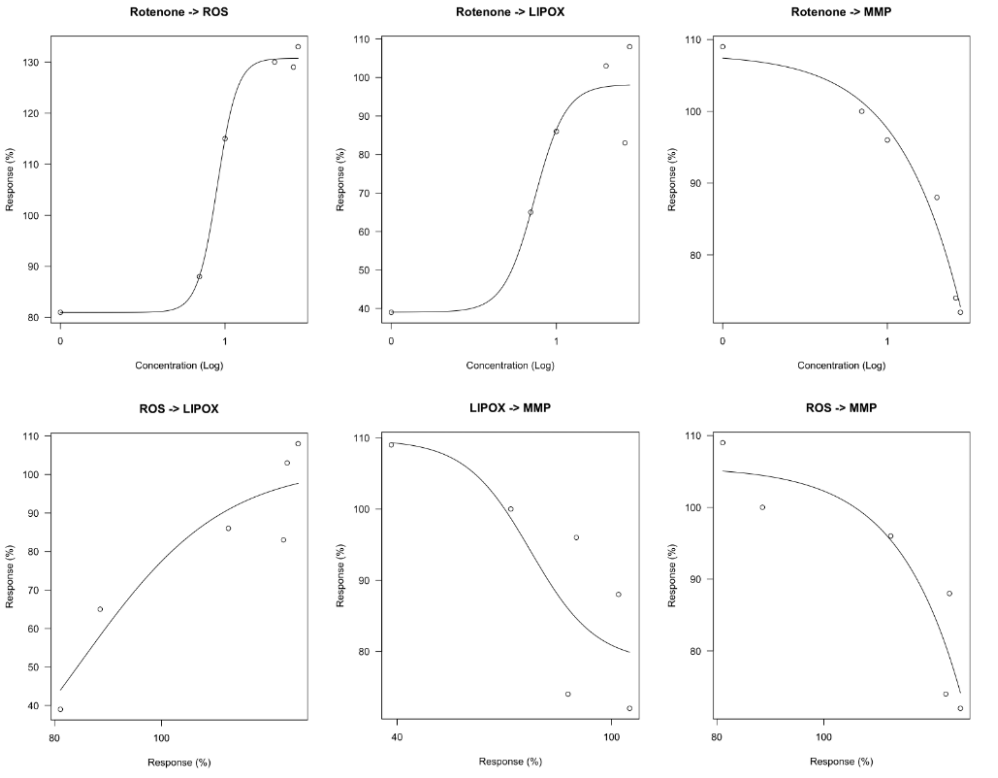


Figure S2. The fitted curves to the causal relations of the KE mitochondrial dysfunction, and KERs that encompass the characteristic biomarkers for data measured after 24 hours conducted in *drc* R package after the imputation of the missing data using *mice* R package. ROS, Production of reactive oxygen species, Increase; LIPOX, Lipid peroxidation, Increase; MMP, Mitochondrial Membrane Permeability, Decrease.

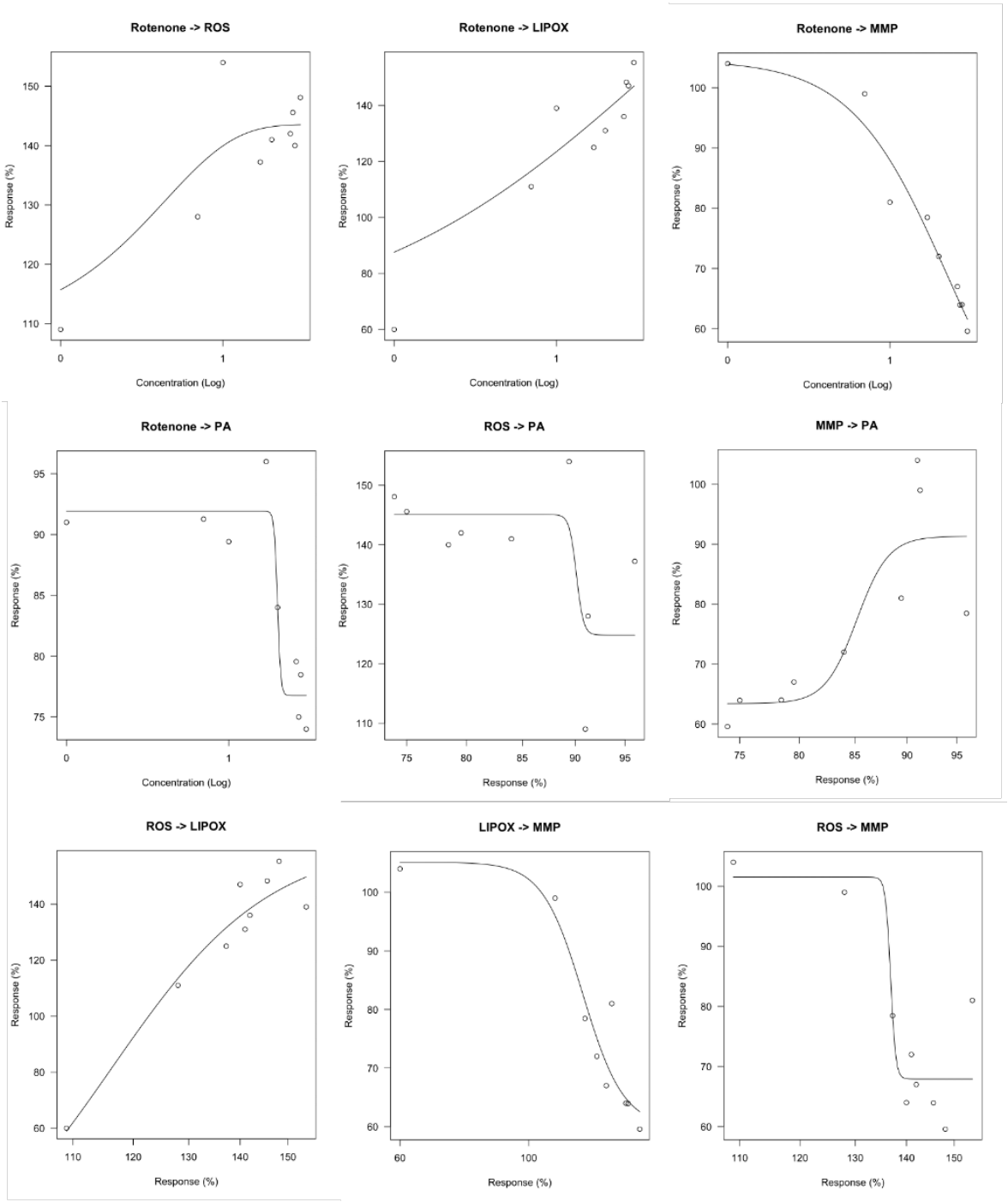


Figure S3. The fitted curves to the causal relations of both KEs, i.e., mitochondrial dysfunction leading to impaired protestasis, and KERs for data measured after 48 hours conducted in *drc* R package after the imputation of the missing data using *mice* R package. ROS, Production of reactive oxygen species, Increase; LIPOX, Lipid peroxidation, Increase; MMP, Mitochondrial Membrane Permeability, Decrease; PA, 20S Proteasome Activity, Decrease.

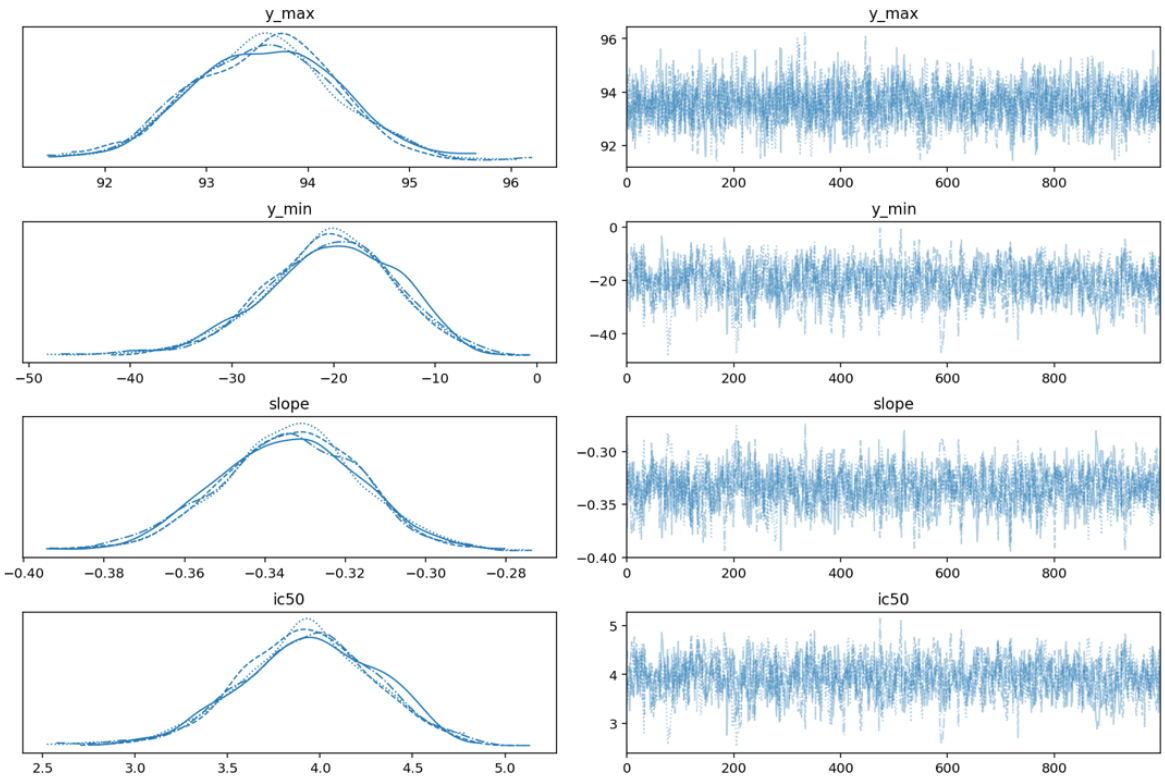


Figure S4. Distribution of the inferences of the parameters that describe the log-logistic regression of concentration-responses for the KE inhibition of complex I by rotenone at 4h.

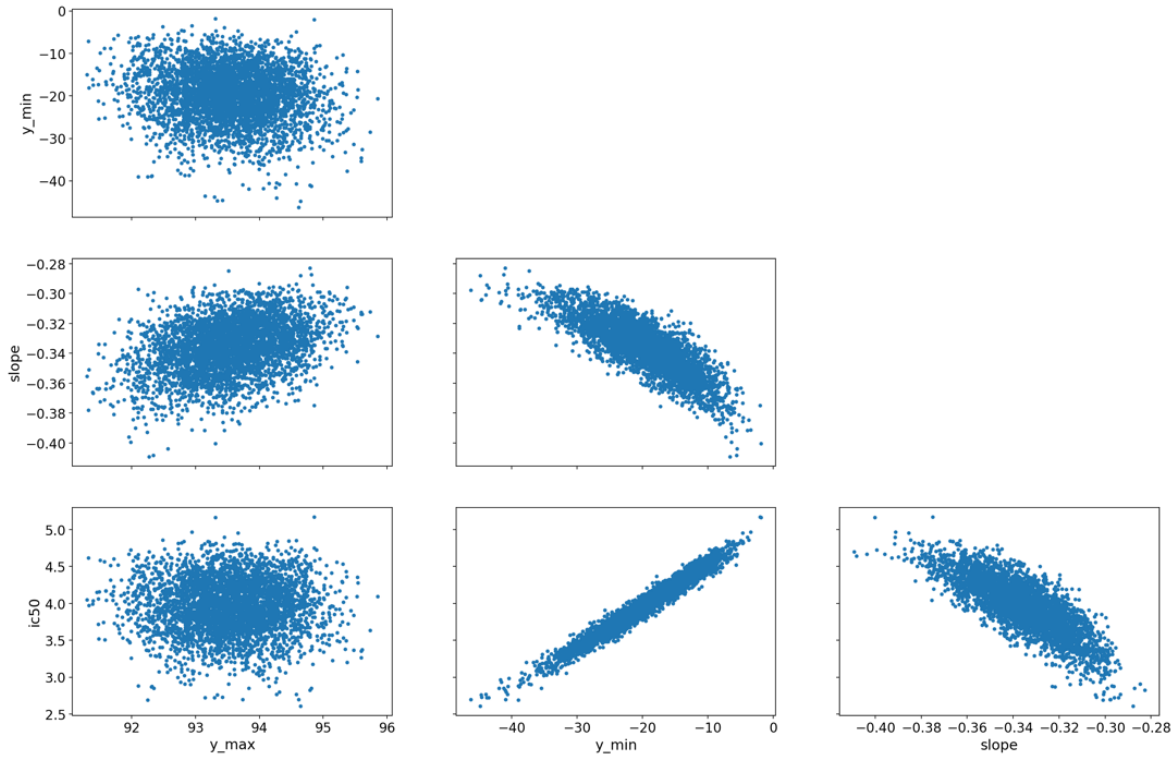


Figure S5. Pairplot of the sampled parameters that describe the pairwise relationships between the parameters of the log-logistic regression of concentration-responses for the KE inhibition of complex I by rotenone at 4h.

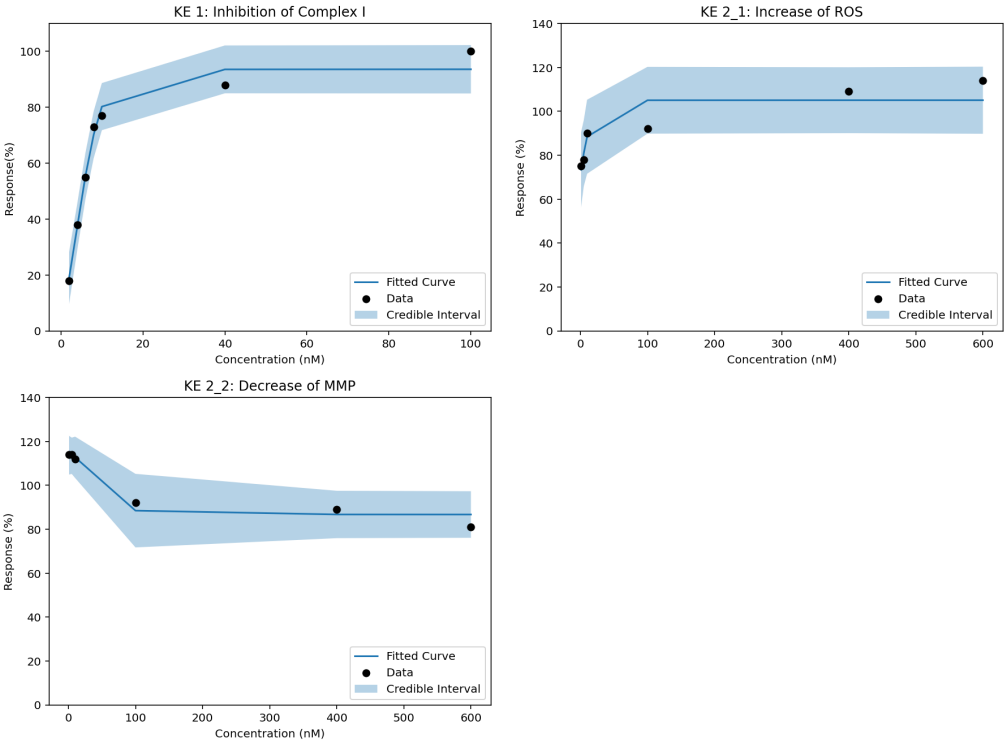


Figure S6. Fitted log-logistic regression to the observed data and associated credible interval to describe concentration-responses of each individual KE induced by rotenone at 4h.

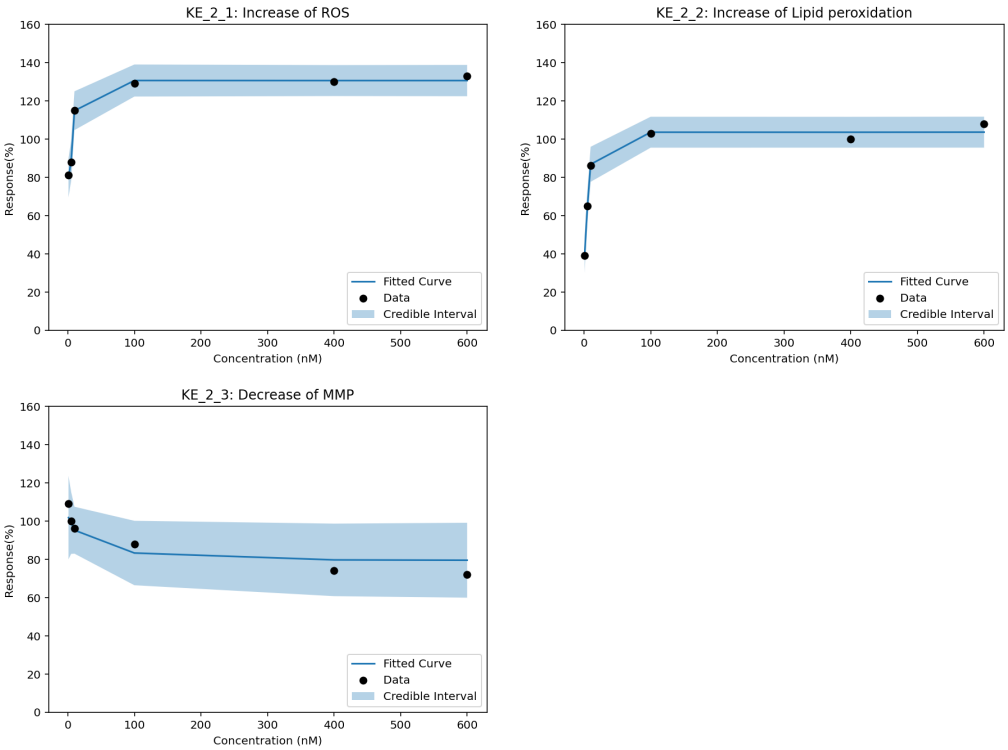


Figure S7. Fitted log-logistic regression to the observed data and associated credible interval to describe concentration-responses of each individual KE induced by rotenone at 24h.

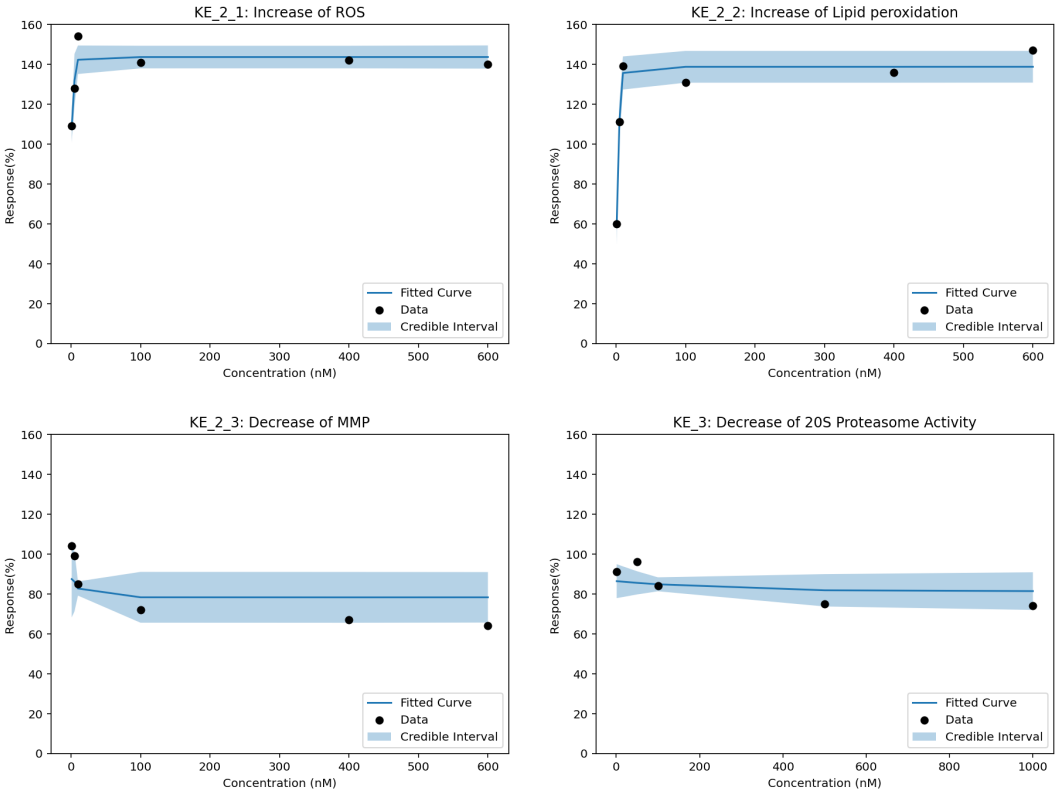


Figure S8. Fitted log-logistic regression to the observed data and associated credible interval to describe concentration-responses of each individual KE induced by rotenone at 48h.

## Appendix IV. List of published articles of the research presented in this thesis

1. Spinu, N., Cronin, M.T.D., Enoch, S.J. *et al.* Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Arch Toxicol* **94,** 1497–1510 (2020). https://doi.org/10.1007/s00204-020-02774-7
2. Spinu, N., Bal-Price, A., Cronin, M.T.D. *et al.* Development and analysis of an adverse outcome pathway network for human neurotoxicity. *Arch Toxicol* **93,** 2759–2772 (2019). https://doi.org/10.1007/s00204-019-02551-1

## Appendix V. List of recorded presentations about the research conducted in this thesis

1. ASCCT and ESTIV Webinar, Development and Use of Adverse Outcome Pathway Network to Support Assessment of Organ Level Effects, available at https://youtu.be/8kXlpJYDlik.
2. PyMCon2020 Conference, Estimating the Causal Network of Developmental Neurotoxicants Using PyMC3, available at https://youtu.be/2nfcwZCLDAE.
3. in3 Project Online Open Days, Modelling of quantitative Adverse Outcome Pathways: Progress report (Part I), available at https://youtu.be/3HhV1VRYgR8.
4. in3 Project Online Open Days Modelling of quantitative Adverse Outcome Pathways: Progress report (Part II), available at https://youtu.be/dsbo6eno3_g.

## Appendix VI. The permalink of the associated GitHub repository that contains digital supplementary materials of the present thesis

https://github.com/nicospinu/phd_thesis