An AI supported case study applying in vitro studies using the ONTOX toolbox: Protocol for probabilistic risk assessment of perfluoroctanoic acid (PFOA)

Summary

The project 'Ontology-driven and artificial intelligence-based repeated dose toxicity testing of chemicals for next generation risk assessment' (ONTOX) under the EU programme Horizon 2020 is running from 01.05.21 to 30.04.26 and is coordinated by Vrije Universiteit, Brussel, Belgium (project website, URL: ONTOX project). The vision of ONTOX is to provide a functional and sustainable solution for advancing human risk assessment of chemicals without the use of animals in line with the principles of 21st century toxicity testing and next generation risk assessment. ONTOX will perform a case study on probabilistic risk assessment (PRA) on the selected chemical perfluorooctanoic acid (PFOA). The exposure assessment will use already established methods from the newly published scoping review, "Accessible methods and tools to estimate chemical exposure in humans to support risk assessment: a systematic scoping review", or custom-made methods using R. The hazard characterisation will use some published methods as a starting point which will be adjusted and combined to fit this case study. The hazard identification/characterisation will use data from published literature and on in vitro data produced in ONTOX. The whole ONTOX toolbox will be used in this risk assessment, such as physiological based kinetic (PBK) models, quantitative in vivo in vitro extrapolation (QIVIVE), physiological maps (PMs) and boolean models , and a large transformer-based AI model. This is a protocol for the case study on PFOA, which will provide a proof-of-principle of PRA using in vitro studies.

Title:
An AI supported case study applying in vitro studies using the ONTOX toolbox: Protocol for a probabilistic risk assessment

---

Protocol for probabilistic risk assessment

Commissioner:
WP6 ONTOX

Commissioned date:
31.01.2025

Due date:
30.04.2026

Team:
Trine Husøy (team leader)
Camilla Svendsen
Maria Kalyva
Hubert Dirven
Michael Guy Diemar
Erwin L Roggen
Devon A. Barnes
Job Berkhout
Maria Corradi
Annika Drees
Ellen Fritsche
Domenico Gadeleta
Thomas Hartung
Ramiro Jover
Nynke Kramer

Eliska Kuchovska
Luiz Carlos Maia Ladeira
Thomas Luechtefeld
Amy Maerten
Alexandra Maertens
Rosalinde Masereeuw
Oddvar Myhre
Rita Ortega
Susana Proença
Alessandra Roncaglioni
Julen Sanz Serrano
Eva Serrano
Bernard Staumont
Marc Teunis
Anouk Verhoeven
Mathieu Vinken

Peer reviewer:
Monica Andreassen, Senior Advisor, NIPH

Approved by:
Cathrine Thomsen, Department Director, NIPH

Commission

This case study on probabilistic risk assessment is a part of the EU project “Ontology-driven and artificial intelligence-based repeated dose toxicity testing of chemicals for next generation risk assessment’ (ONTOX)” under the EU programme Horizon 2020, Grant Agreement 963845 – ONTOX.

Introduction and aim

The project 'Ontology-driven and artificial intelligence-based repeated dose toxicity testing of chemicals for next generation risk assessment' (ONTOX) under the EU programme Horizon 2020 is running from 01.05.21 to 30.04.26 and is coordinated by Vrije Universiteit, Brussel, Belgium (project website, URL: https://ontox-project.eu/project/). The vision of ONTOX is to provide a functional and sustainable solution for advancing human risk assessment of chemicals without the use of animals in line with the principles of 21st century toxicity testing and next generation risk assessment. Specifically, ONTOX will deliver a generic strategy to create innovative new approach methodologies (NAMs) to predict systemic repeated dose toxicity effects that, upon combination with tailored exposure assessment, will enable human risk assessment. The overview of the work package (WP) structure in ONTOX is given in figure 1.

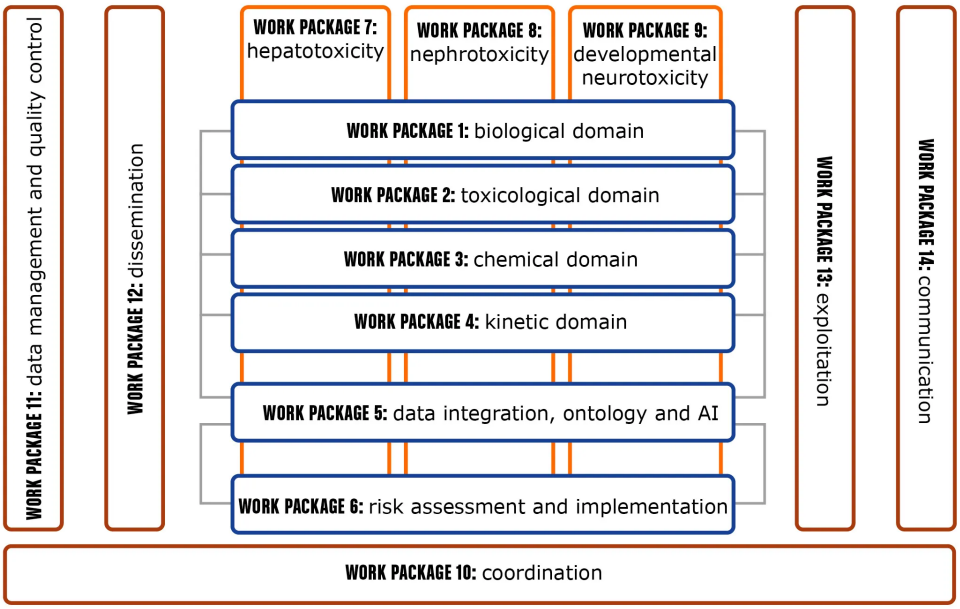

Figure 1: Overview of the work packages and their connection in ONTOX

Probabilistic risk assessment (PRA) incorporates variability and uncertainty into the risk assessment process. It provides a distribution of hazard, exposure or risk estimate, rather than a single point estimate. Probabilistic exposure assessment is more frequently used than probabilistic hazard assessment. However, in this case study, both the exposure and hazard assessment will be done in a probabilistic manner.

In ONTOX WP5, two distinct neural network models have been developed to facilitate chemical property prediction. The first model utilizes contrastive learning to derive property-specific similarity metrics for chemicals based on their molecular structure. This similarity-based approach enables the identification of compounds with comparable toxicological profiles.

The second model is a transformer-based sequence prediction system designed to infer multiple properties of a given compound. Unlike traditional methods, this model integrates both structural information and known properties to predict one or more additional characteristics. Specifically, it is trained to predict 3,797 boolean properties sourced from databases such as PubChem, ChEMBL, and the Integrated Chemistry Environment. These properties encompass a wide range of toxicological endpoints, including acute toxicity classifications (e.g., acute oral and dermal toxicity), receptor binding affinities, and high-level toxicological concerns such as nephrotoxicity, hepatotoxicity, developmental neurotoxicity, corrosivity, mutagenicity, and many more.

Furthermore, natural language processing (NLP) models developed within ONTOX contribute to automating the extraction of toxicological information from literature and other unstructured sources. This tooling supports case studies aimed at expanding Adverse Outcome Pathways (AOPs), Physiological Maps, and Physiologically Based Pharmacokinetic (PBPK) models. By leveraging AI-driven predictions, these models provide a probabilistic framework for risk assessment, making them valuable tools in modern toxicology.

Aim

To perform an AI-supported PRA for one well known environmental chemical using data from NAMs.

The objectives:

- Characterise the probability of hazard using in silico and vitro approaches
    - Estimate the probability of hazard using published methods from the literature and in vitro data from ONTOX and public sources
    - Explore the physiological maps and Adverse Outcome Pathways (AOPs) to quantify the hazard in humans
- Assess the probability of internal exposure in humans
    - Estimate the probability distribution for external exposure in a population
    - Estimate the probability distribution for external exposure in individuals using data from the EuroMix study
    - Estimate the probability distribution for internal exposure using PBK modelling and QIVIVE modelling
- Characterise the probability of risk from the probabilistic hazard assessment and the probabilistic exposure assessment.
- Identify obstacles in the PRA process and knowledge gaps in the methods and data.

Method

Overview over existing methods for PRA

The vision of ONTOX is to provide a sustainable solution for advancing human risk assessment of chemicals without the use of animals. The most relevant methods/tools for PRA are described in short below. None of these methods can directly be used for our purpose, however, we will use elements from these methods to develop an approach which is suitable for the purpose of the PRA in ONTOX.

The APROBA tool was first developed by EPA to quantify uncertainty (1) but was further developed by RIVM for use as a tool for PRA and subsequently called APROBA-Plus. The tool is based on toxicological data in animals and the input is benchmark dose calculations (BMDL and BMDU) and upper and lower bound of assessment factors. The tool also includes a probabilistic exposure assessment, which can be calculated by filling the input data in an Excel template. The tool is explained in detail in the paper by Chiu et al 2015 (2), where also the R script for the calculations is included. The tool was used by Pensirini et al 2022 for assessing the risk assessment of phthalate mixtures (3), where a schematic overview of the framework is included.

In 2024 a paper by Middleton et al. was published where a workflow for a PRA, including an approximate probabilistic hazard analysis, using in vitro data were described (4). The workflow included the use of physiologically based kinetic (PBK) models to translate external exposure to internal doses in humans. The hazard assessment was based on benchmark dose responses modeling of effects from in vitro studies to calculate a point of departure (POD). The calculated POD from several studies were summarized in a frequency distribution of the PODs, which were used in the risk assessment.

In ONTOX, physiological maps are developed and providing invaluable insights into human physiological mechanisms and ultimately toxicological pathways. These maps allow for linkage to adverse outcome pathways (AOPs) relevant to specific toxicities in the liver, kidney and brain. In the PRA we will explore the use of these physiological maps in translating the hazard effect in vitro to human health effects. Below in Figure 2 is a schematic example of what the PRA workflow in ONTOX could look like. However, this is not necessarily a description of the final PRA workflow in ONTOX, but some of the elements are likely to be present.

A large transformer-based AI model generated in ONTOX will be used to get initial property-based predictions for hazard. This is a generative AI model, capable of providing compound-property conditioned predictions for chemicals, based on structure. It expands on the idea of QSAR and read-across structure activity relations (RASAR) (5). For steps in the hazard assessment where a probabilistic approach is not feasible, the reason for selecting a deterministic approach will be described in the risk assessment. The step will then be included in the list of data gaps and/or methodological challenges.

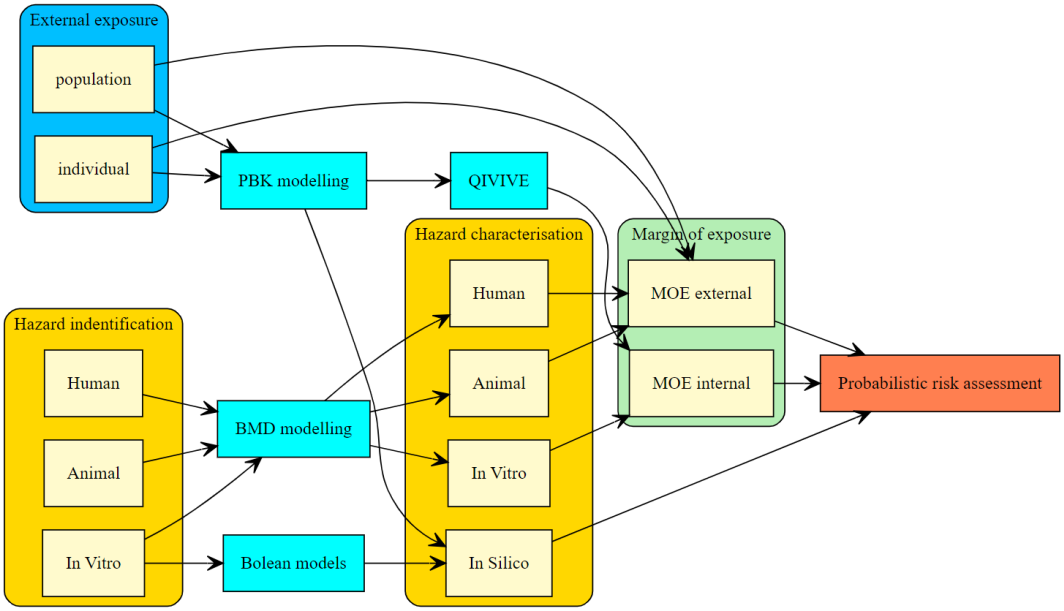

Figure 2. A schematic illustration of the planned probabilistic workflow in ONTOX

Justification for the selection of the chemical for the case study

Perfluoroctanoic acid (PFOA; CAS 335-67-1), a chemical that has been categorized as a "forever chemical" has been selected for this case study.

It was decided that it was only feasible to perform the case study on one chemical. The ONTOX chemical list, updated on 28.05, was the starting point for the selection of which chemical to use for the case study. All the chemicals in the ONTOX list cause adverse effects in the liver, kidney and/or brain.

In addition to being on the ONTOX chemical list, all the following criteria had to be met for the chemical to be considered a candidate for the case study, and to be included in a short list of relevant chemicals:

- The chemical is not a pharmaceutical or metal

And

- A PBK model is available or possible to create within the duration of the project

Chemicals that fulfilled these criteria were included in a shortlist. For these chemicals, the availability of the following data/information were collected:

- Data available to perform an exposure assessment
    - Biomonitoring data in EuroMix biomonitoring study
    - Summary data for a population
    - Concentration data in etc. food, consumer products, to make individual exposure assessment

- Hazard data
    - Available human data on toxicity of the chemical (RCT or observational studies)
    - Available animal data on the toxicity of the chemical

- Physiological map available for the biochemical pathway affected by the chemical in the context of ONTOX case studies
- Availability of a risk assessment, based on animal and human data

The shortlist was sent to work package (WP) and task leaders for input on which of the chemicals that are the most feasible for their work and to provide information if any of the chemicals are not within their applicability domain. The selection of PFOA was based on the input received, where in vitro data in human cells were available from ONTOX partners, and where most of the WPs could contribute into the case study.

Exposure assessment

Individual external exposure (WP6)

Individual external exposure will be studied in the EuroMix biomonitoring study, which was a part of the EU project "European Test and Risk Assessment Strategies for Mixtures" (EuroMix, 633172-2, https://www.euromixproject.eu/index.html), which is funded by the H2020 program. The study was performed between September 2016 and November 2017. For a detailed description of the study, see Husøy et al. 2019 (6) . For two 24-hour study periods separated by 2-3 weeks, adult volunteers (44 males and 100 females) in Norway kept detailed diaries on food consumption (type/brand, weight, time and packaging material) and the usage of PCPs (type/brand of product, time and number of applications). In parallel, 24 hours urine samples were collected and at the end of each study day blood samples were collected. Personal information such as body weight, height, age is available. The study was approved by the Regional Committees for Medical and Health Research Ethics (REK ID no 2015/1868) and all the participants provided their written informed consent. The study is anonymised and accepted for use in ONTOX by the Regional Committees for Medical and Health Research Ethics (REK ID no 2015/1868).

External exposure in a population (WP6)

A probabilistic external exposure assessment in a population will be performed based on input data from the literature and by using one of the tools described in our recent scoping review of methods for exposure modelling of environmental chemicals, such as the toolboxes MCRA, SHEDS, ConsExpo, USEtox, and PACEM, or custom made scripts in R. Input data can be food consumption and cosmetic use data from the EuroMix study, concentration data of the chemical in food, consumer product and air, and data such as food consumption, consumer use and inhalation . We will also use existing exposure estimates from the literature and opinions presented as summary data for a population. Based on these summary data a distribution of exposure can be calculated using R. All input data and scrips used in the exposure assessment will be openly published on GitHub.

PBK modelling and QIVIVE (WP4) – internal exposure

The first step in PBK modelling will be to evaluate how to translate the current PBK models for PFOA into the PK-Sim software or whether the model is coding in R will be used instead. Even in the case the PBK model is built in PK-Sim/MoBi it can be imported into R for further analysis.

To make probabilistic PBK simulations, both the population variability and parameter uncertainty needs to be accounted for. However, accounting for all parameter variability and uncertainty is both computational heavy and not necessarily informative because some parameters might have minor impact in the model. To make sure we account for the parameters (chemical and physiological) with highest impact for the PBK simulations, a sensitivity analysis will be performed. This sensitivity analysis should be performed observing concentrations of PFOA in plasma but also at the target sites (e.g. liver, kidney and fetus) under a relevant exposure scenario. The top 10 parameters having higher impact in the observed simulations will then be characterized either for their variability and/or uncertainty. There are different manners the variability and uncertainty of the parameter can be characterized:

- If the parameter is physiological (e.g. blood flow into a specific organ, protein content in serum) the distribution of values can be derived from medical literature or anatomy databases. Extracting these parameters from these sources can be supported by the Natural Language Processing models and workflows established in WP5.
- In case the parameter is the expression of a specific protein (e.g. transporter), there are databases of human variability that and also account for changes in age and pregnancy status.
- In case of a physicochemical parameter (e.g. logP, pKa), the uncertainty of the analytical method or QSAR used needs to be accounted. If there are different values reported for the parameter this variation is also accounted.
- If the parameter is an ADME rate (e.g. clearance from in vitro), derived from in vitro then, both the uncertainty of the in vitro methods and IVIVE procedure needs to be accounted for.

Based on the variability and/or uncertainty of the parameters a distribution of values is defined. Values of these distribution are pooled through a Markov Chain-Monte Carlo calculation and run for >1000 simulations.
This probabilistic PBK model can be evaluated with biomonitoring data and the distribution of the probabilistic parameters refined (e.g. Bayesian inference) and the model itself calibrated.
For this we need meetings and exchange of information between WP6 (NIPH) and WP4 (ESQlabs). It should also be noted that specific populations can be described (based on gender, ethnicity or diseased status) if needed.

Typically, QIVIVE is performed using reverse-dosimetry, but since ONTOX presents specific challenges (i.e. perform risk assessment for different toxicity mechanism at different target sites from a possible aggregate exposure), we consider that a forward dosimetry approach is more suitable. Briefly, we use a distribution of external concentrations from biomonitoring studies in the probabilistic PBK model to assess the resulting target sites concentrations in a form of a distribution. This distribution of internal concentrations in the different tissues resulting from real exposure, can then be compared to dose-response curves from in vitro and the overlap between the two evaluated (Figure 3). Before this comparison, the dose-response curves from in vitro studies need to be converted from nominal concentration to unbound and cellular-associated concentrations by using in silico models of the in vitro kinetics.

The benefit of this strategy is the forward dosimetry simulations only need to be performed once, and then can be compared with as many dose-response curves and target sites as desired. For the reverse dosimetry approach, for each benchmark concentration generated from NAMs, new simulations need to be performed which can take a considerable amount of time. In addition, in case different routes of exposure need to be accounted for (e.g. dermal and oral), reverse dosimetry leads to an identifiability issue with the respective routes of exposure doses.

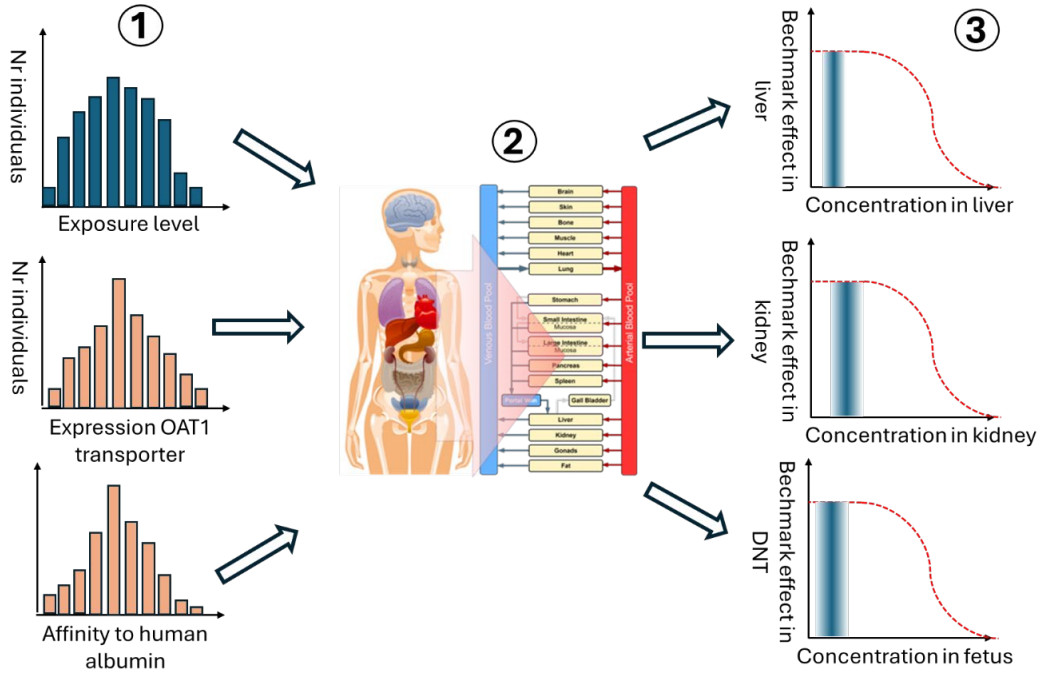

Figure 3. Forward dosimetry procedure to perform probabilistic risk assessment. In step 1 the variability and uncertainty of the real exposure and physiological and chemical parameters are characterized; in step 2, PBK models are used to perform probabilistic simulations using the distribution of parameters defined in step 1 and determine the internal concentrations in the target organs. -The resulting distributions of internal concentrations in each target organ are compared to dose-response data derived in vitro.

These strategies need to be discussed both with WP6 who will perform the final risk assessment but also all partners involved in performing hazard characterization (WP5-,7,8, and 9). It would be of added value if the discussion of the strategy would involve also stakeholders that perform risk assessment to understand what strategies are more suitable for their applications. The discussion of the QIVIVE strategy is also dependent on what is the ideal dose metric for the specific toxicity effect: Cmax, time-weighted AUC, or the QIVIVE is to be integrated with effect modelling (e.g. through use of quantitative AOPs (qAOP)).

Hazard identification

Identification of data from the literature

Literature searches will be performed to identify information useful for the hazard identification and characterisation. Although the PRA workflow will be developed for the use of new assessment methods (NAMs), we will also look for animal data and human data for validation purposes. A literature search on human and animal toxicity data was performed in the 2020 EFSA opinion on PFAS up to 16 August 2019 (7). We will search after human and animal data published after that date. The in vitro data search will be restricted to human cells.

Identification of human data (WP2)

Literature searches will be performed to identify human studies showing associations between exposure to PFOA and higher risk of one of the outcomes of ONTOX (e.g. female exposure to PFOA led to a higher risk of nonalcoholic fatty liver disease (NAFLD) than male exposure (OR =1.82, 95 % CI:1.01-3.26) (8).

A literature search will be conducted in PubMed/Ovid MEDLINE. Search strings will be designed to achieve a comprehensive search at the expense of a high number of false positives.

Main inclusion criteria: human case studies, clinical studies and patient-based meta-analysis providing indirect evidence (associations) between exposure to PFOA and a higher risk of hepatotoxicity (cholestasis and steatosis), nephrotoxicity (crystallopathy and tubular necrosis) and neurotoxicity (DNT and neural tube defects). No restrictions in time (year), language, or country of publication.

The list of identified articles will be uploaded to SysRev. A manual screen for inclusion/exclusion will be made by pairs of Reviewers, first for the abstract and then fulltext. Conflicts will be resolved by a third expert. After a manual screening of ≤5%, the AI-based auto-label tool will be used to include or exclude the remaining articles.

Full text PDFs will be retrieved and uploaded to a new SysRev project, and data of interest (exposure data, association figures, as well as clinical exploratory, biochemical, and histopathological information related to hazards) will be extracted in a tabulated form, either manually or by using the AI-based extraction tools of SysRev (WP5). The extracted data will be thoroughly curated manually.

Tabulated data (clinical info and association assessment) can be delivered in 4 months if the Sysrev tools can be applied to decrease the screening burden. Article inclusion

and data extraction are dependent on the auto-label tool and data extraction tools of Sysrev (WP5).

Identification of data from animal studies and in vitro (WP6)

An experienced research librarian will conduct the literature searches.
The literature searches will be conducted in the following bibliographic databases:

- Ovid MEDLINE®
- Embase
- ISI Web of Science
- Scopus
- Cochrane Database of Systematic Reviews
- Epistemonikos
- SciFinder (American Chemical Society)

The study selection will be based on the predefined eligibility criteria (Table 1). Reviewers will independently, in pairs of two, assess 1) titles and abstracts for relevance, and 2) full text articles against eligibility criteria. Disagreements will be resolved by consensus or by consulting a third author. To decrease the screening burden, machine learning will be used for the screening of titles and abstracts. The evidence retrieved from each bibliographic database will be imported and combined into the bibliographic reference management software SysRev (Built for data miners | Sysrev). Duplicates will be removed. Only one reviewer will screen those studies that have less than 20% likelihood of inclusion.

Table 1. Hazard eligibility criteria

Study design and population	Animal experimental studies (mammals and non-mammals) In vitro data using human cells
Exposure	PFOA (tested separately), oral exposure
Comparison	Placebo, no treatment, dose comparison
Outcomes	Studies addressing endpoints connected to hepatotoxicity (cholestasis and steatosis), nephrotoxicity (crystallopathy and tubular necrosis), and neurotoxicity (DNT and neural tube defects)
Publication year	No restriction
Country	No restriction

Language	English
Publication type	Peer-reviewed primary studies and systematic reviews

A publication qualifies as a systematic review if 1) a specific research question and the specific criteria used for selecting studies are described, 2) the authors have performed a systematic literature search, and 3) it includes a quality assessment of the selected studies (9).

Study appraisal

Risk of bias assessment will be performed on all included studies. The risk of bias evaluation of systematic reviews will be performed using the ROBIS tool (10) whereas the evaluation of human controlled studies and experimental animal studies will be performed according to the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration (OHAT handbook) (11).

Currently, no validated risk of bias tool for in vitro studies exists. However, a tool for assessing internal validity of in vitro studies, INVITES-IN is currently being developed and validated (12). INVITES-IN will be used for the risk of bias assessment of the included in vitro studies. Authors will independently, in pairs of two, perform the risk of bias evaluation.

Disagreements will be resolved by consensus or by consulting a third author. Studies assessed as being Tier 3 (having a high risk of bias) will not be included in the hazard characterization.

Data extraction

For studies assessed as being Tier 1 and 2, data for the dose response modelling will be collected. As well as other relevant information, e.g. species, number of animals, duration of exposure. The project group will jointly develop and test a data charting form. Authors will extract data from a sample of the included publications to ensure that the data extraction is consistently applied. Discrepancies will be resolved through discussions. If necessary, the data-charting form will be modified. Following calibration of the data extraction, one author will extract the data, and a second author will independently check the data extraction for accuracy and completeness. Disagreements will be resolved by consensus or by consulting a third author.

When sufficient data are not reported in a study, we will attempt to obtain them by contacting the corresponding author.

The final full paper list will also be handled for the extraction of mechanistic molecular interactions between chemical and biological entities (e.g., proteins, genes, structures)

and the chemical effects on biological systems by WP5. This information will be used to complement, refine, and validate the in silico models for probabilistic hazard characterization.

In silico predictions (WP3 and WP5)

It has been suggested that compounds of unknown hazards can be assigned to various levels of concern based on the number of activated molecular initiating events (MIEs) and the extent of their activation (13). In this regard, computational predictions will be generated for PFOA to assess its potential to bind to relevant protein targets (receptors, enzymes, transporters) linked to MIEs upstream in the adverse outcome pathways (AOPs) related to the project's adversities. Predictions will be made using three different types of computational models. The list of protein targets addressed by each model is provided in Table 2. In case experimental data for PFOA are available in the model's training sets, those data will be also provided.

Quantitative Structure-Activity Relationship (QSAR)

QSAR models will provide categorical predictions (active/inactive, based on a threshold of IC50 = 10,000 nM) about the capability of a PFOA to interact with protein targets, likely causing perturbations in the downstream AOPs. QSARs are based on machine learning techniques, including single-task and multi-task approaches, alongside different structural descriptions and statistical methods for artificial data balancing (e.g., SMOTE). The QSAR models includes an applicability domain assessment to identify compounds that fall within the QSAR chemical space and continuous probability values (ranging from 0.0 to 1.0), to differentiate between reliable and unreliable predictions. The QSAR models are ready for use, and predictions will be generated in a relatively short time without requiring external inputs.

Structure Activity Relationship (SAR)

SAR models will be used to establish qualitative relationships between the structure of PFOA and its activity by identifying structural alerts (SA)s, i.e. structural fragments in the structure of molecule that are the cause of the activation or suppression of a biological activity. SAs were extracted with the SARpy software (https://www.vegahub.eu/portfolio-item/sarpy/) and codified as SMARTS to classify chemicals for their capability to interact with protein targets associated to MIEs of project's adversities. SAR models will provide categorical predictions (active/inactive) for PFOA and will be associated with a continuous likelihood ratio (LR), a statistical parameter defining the precision of SA used to classify PFOA. Higher precision of the SA used indicates lower prediction uncertainty.

Similarity-Based Prediction in Hazard Identification

As part of ONTOX WP5's in silico hazard identification framework, a supervised-learning neural network model was developed using contrastive learning to derive property-specific similarity metrics for chemicals based on their molecular structure. The model is pretrained using a Variational Autoencoder (VAE), which learns a compressed representation of chemical structures before fine-tuning with contrastive learning to develop task-specific similarity functions. This approach makes it more effective than traditional similarity measures such as Tanimoto similarity on Morgan fingerprints. Early evaluations indicate that this method consistently outperforms conventional fingerprint-based approaches, providing more precise alignment between structurally related compounds and their toxicological properties. By leveraging deep learning to capture nuanced structural relationships, this model significantly enhances early hazard identification and facilitates efficient screening of large chemical libraries.

Property Transformer AI Model

The Property Transformer AI model is an encoder-decoder transformer designed for chemical property prediction. Unlike traditional Quantitative Structure-Activity Relationship (QSAR) models, which predict properties individually, this model formulates property prediction as a sequence-to-sequence problem. It tokenizes Self-Referencing Embedded Strings (SELFIES) of molecules alongside property-value pairs, allowing it to generate sequences of predicted properties for a given chemical.

This model is trained on the ChemHarmony dataset, which contains data on 118 million substances, enabling it to learn complex relationships between molecular structure and known properties (https://arxiv.org/pdf/2408.17320 and biobricks-ai/chemharmony: integrated chemical-property-values from many source databases.) . By incorporating causal masking and learned embeddings of property tokens, the Property Transformer can leverage existing chemical knowledge to make conditional predictions—for example, estimating the probability of acute inhalation toxicity given knowledge of a chemical's acute oral toxicity. This capability allows for transfer learning across chemical properties, improving prediction accuracy and enabling hypothesis generation and counterfactual analysis (e.g., "If this chemical were acidic, would it be more likely to be corrosive?").

By integrating both structural and property-based information, the Property Transformer represents a major advancement in AI-driven chemical hazard prediction, offering a powerful and flexible tool for toxicological assessments.

Docking

The DockTox docking server (https://chemopredictionsuite.com/DockTox) will be used to simulate the binding of the PFOA with 23 proteins associated to MIEs of the

liver, brain, and kidney toxicity. The server provides a series of parameters describing the interaction between chemicals and proteins (e.g. binding affinity, number of ligand-protein interactions) which will be used to refine and confirm predictions generated by QSAR and SAR models.
When multiple predictions are available for the same protein target from different models, an integration strategy will be applied to reduce uncertainty and provide a unique consensus output. Reliability measures (e.g., prediction probability, likelihood ratio) will be considered to favour predictions with higher reliability during the integration.
Additionally, computational predictions for relevant physicochemical and toxicokinetic properties—provided by QSAR models identified from the literature —will be generated and delivered to WP4 as input for PBK simulations. Properties available for predictions are boiling and melting points, octanol/water partition coefficient and distribution rate at pH =7, Henry's law constant, water solubility, vapor pressure, pKa, Caco2 permeability, fraction unbound to plasma protein, skin permeation, blood brain barrier permeability (categorical), bioavailability (F30%, categorical), human intestinal absorption (categorical, higher or lower than 30%), P-glycoprotein inhibition and substrate activity (categorical).

Table 2: Details on the output provided and the MIE covered by each type of computational model

Model	Output	MIE proteins
Quantitative Structure Activity Relationshp (QSAR)	Activity prediction (categorical, active/inactive) Applicability domain (categorical, in / out) Prediction probability (continuous, 0.0-1.0)	*PXR, PPAR(α, γ), MitoTox, AHR, LXR, NrF2, BSEP, MRP(2,3,4), OATP1B(1,3), PgP, ACHE, VGSC, NMDA, TTR, THR(α, ß), TPO, NIS, BMP, CYP26, Histone deacetilase, WNT, SMO, FGFR(1-4), COX1, OAT1*
Structure Activity Relationship (SAR) based on structural alerts (SA)s	Activity prediction (categorical, active/inactive) Likelyhood ratio (continuous, 0.0 to infinite)	*PXR, PPAR(α, γ), MitoTox, AHR, LXR, NrF2, BSEP, MRP(2,3,4), OATP1B(1,3), PgP, ACHE, VGSC, NMDA, TTR, THR(α, ß), TPO, NIS, BMP, CYP26, Histone Deacetilase , WNT, SMO, FGFR(1-4), COX1, OAT1*
Docking	Binding parameters (e.g. binding affinity, ligand-protein interactions) (continuous)	*PXR, PPAR(α, γ), AHR, LXR(α, ß), BSEP, OATP1B1, PgP, ACHE, TTR, THR(α, ß), BMP, Histone Deacetilase (2, 4, 6, 7, 8), TXNRD1, COX1, POLG1, ACE*

Hepatotoxicity (WP7)

The battery of in vitro assays detecting molecular initiating events (MIEs) and key events (KEs) in the adverse outcome pathway (AOP) networks for chemical-induced steatosis and cholestasis in the liver will be used (Table 3) (14;15). These will be assessed at the transcriptional and functional levels. The in vitro cell model of differentiated human liver hepatoma HepaRGTM will be used. The highest non-cytotoxic concentration and two 10-fold serial dilutions of the chemicals will be tested. In case of no cytotoxicity, the highest soluble concentration will be used as the test concentration. Data will be delivered by the end of February 2025. This delivery is independent of input/data from other work packages/tasks. No incorporation of artificial intelligence is expected from this work package; all in vitro data will be integrated using the CRStats tool as explained below (see 5.3).

Table 3. Battery of *in vitro* assays addressing molecular initiating events and key events relevant to chemical-induced steatosis and cholestasis in the liver.

Disease	Key Event
Steatosis	Intrahepatic lipid accumulation
	Fatty acid uptake
	Mitochondrial beta-oxidation
	Very low-density lipoprotein secretion
Cholestasis	Transporter changes
	Bile canalicular changes
Steatosis & cholestasis	Oxidative stress
	Endoplasmic reticulum stress
	Mitochondrial impairment

Steatosis

Cultured cells will be exposed daily to a range of five concentrations of PFOA for 72 hours, covering levels reported in human plasma. This range of concentrations will include the CC10 value (defined as the concentration of chemical needed to induce 10% of cytotoxicity) and four ten-fold serial dilutions. The following processes related to chemical-induced liver steatosis will be assessed at the transcriptional level: peroxisome proliferator-activated receptor alpha and gamma (PPARα, PPARγ), de novo lipogenesis (stearoyl-coenzyme A desaturase-1 (SCD1) and diacylglycerol o-acyltransferase 2 (DGAT2)), fatty acid uptake (cluster of differentiation 36 (CD36)), lipid secretion (apolipoprotein B (APOB)), and fatty acid β-oxidation (carnitine palmitoyltransferase 1 (CPT1) and peroxisomal acyl-coenzyme A oxidase 1 (ACOX1)) will be assessed by real-time quantitative polymerase chain reaction assay (RT-qPCR). Changes in the KEs intrahepatic lipid accumulation, fatty acid uptake and mitochondrial β-oxidation will be

assessed using the fluorescent probes 4,4-difluoro-1,3,5,7,8-pentamethyl-4-bora-3a,4a-diaza-s-indacene (BODIPY 493/503), 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-hexadecanoic acid (BODIPY™ FL C16) and a fluorescent coumarin dye bound to nonanoic acid (FAOBlue), respectively. Cultured cells will be incubated with the probe and analyzed by spectrophotometry. Changes in the KE very low-density lipoprotein (VLDL) secretion will be assessed by quantifying APOB100 in the extracellular medium via enzyme-linked immunosorbent assay (ELISA).

Cholestasis

Cultured cells will be exposed to a range of five concentrations of PFOA for 1-6 hours, covering levels reported in human plasma. This range of concentrations will include the CC10 value and four ten-fold serial dilutions. The transcriptional changes of the following drug transporters related to chemical-induced liver cholestasis will be assessed by RT-qPCR: bile salt export pump (BSEP), multidrug resistance-associated protein 2 and 3 (MRP2, MRP3), sodium/bile acid cotransporter (NTCP), organo anion transporter family 1B1 and 1B3 (OATP1B1, OATP1B3). At the functional level, the MIEs transporter changes and bile canalicular changes will be addressed. The activity of the former will be assessed focusing on the basolateral (i.e. sinusoidal) and apical (i.e. bile canalicular) membranes using probe-based fluorescent assays. Cultured cells will be exposed to the concentration range of the chemical for 1 hour. Fluorescence changes in basolateral efflux will be quantified by spectrophotometry using 5(6)-carboxyfluorescein diacetate (CDFDA), and basolateral uptake using cholyl-lys-fluorescein (CLF) in response to chemical exposure. Bile canalicular efflux will be assessed by fluorescence microscopy and quantified using CDFDA. Bile canalicular changes will be assessed by phase contrast imaging after exposing cultured cells to the concentration range of the chemical for 6 hours. Bile canalicular area and counts will be quantified to assess bile canalicular dilatation and disruption, respectively.

Steatosis & cholestasis

Cultured cells will be daily exposed to the concentration range of the chemical for a period of 1 hour. General KEs related to the adverse effects of both chemical-induced liver steatosis and cholestasis will be assessed by functional assays. Oxidative stress will be measured by identification of reactive oxygen species by the fluorescent probe 2'-7'-dichlorodihydrofluorescein diacetate (DCFH-DA). Endoplasmic reticulum stress will be quantified by protein aggregation with thioflavin T. Mitochondrial impairment will be assessed by mitochondrial membrane potential quantification using the fluorescent probe Rhodamine123. All three outcomes will be measured by spectrophotometry.

Nephrotoxicity (WP8)

WP8 will assess nephrotoxic potential by PFOA using a variety of assays (16). These include metabolic activity for general cytotoxicity and IC50 determination, lactate dehydrogenase (LDH) release for membrane integrity, N-acetylglucosaminidase (NAG) release for proximal tubule-specific toxicity, mitochondrial membrane potential and reactive oxygen species (ROS) production for mitochondria-specific toxicity. Furthermore, inflammatory cytokine release will be assessed for TNF-α, IL-1β, IL-18, IL-8 and IL-6 to evaluate the inflammatory response induced by PFOA exposure on proximal tubule epithelial cells. In addition, gene expression profiles will be assessed for the common urinary biomarkers: kidney injury molecule-1 (Kim-1), NGAL and cystatin C, and the cell defensive mediators hemeoxygenase 1 (HMOX1) and clusterin. This analysis will provide detailed insights into the nephrotoxic effects induced by PFOA on proximal tubule epithelial cells, the predominant cell type for toxicant-induced nephrotoxicity. We hypothesize that PFAO will significantly alter cell viability, inflammatory responses, and ROS production, particularly by affecting the mitochondria. By assessing this range of biomarkers, the cellular mechanisms of PFAO-induced nephrotoxicity may be unraveled, which will provide robust data for PRA.

The data can be delivered in the beginning of 2025.

Developmental neurotoxicity (WP9)

First, a battery of in vitro assays detecting molecular initiating events (MIEs) and key events (KEs) in the adverse outcome pathway (AOP) network for chemically induced cognitive function defects will be used to functionally assess neurotoxicity in the developing brain in vitro systems (Table 4) (17;18). Second, a computational model of neural tube closure will be used to assess chemically induced neural tube closure defects in the developing brain (19).

Table 4. The battery of *in vitro* assays addressing molecular initiating events and key events relevant to cognitive function defects in the developing brain. *assay optimization ongoing

In vitro system	Test method (assay)	Key Event
Primary hNPCs (3D)	NPC1	NPC proliferation
	NPC2a	radial glia migration
	NPC2b	neuronal migration
	NPC2c	oligodendrocyte migration
	NPC3	neuronal differentiation
	NPC4	neurite growth
	NPC5	oligodendrocyte differentiation
	LDH	cell death
	CTB	cell viability
Primary hNPCs (3D)	NPC6*	Oligodendrocyte maturation

Neucyte cells (2D)	hNNF*	Neural network formation
hiPSC-derived hiNPCs (2D)	hiNPC1	hiNPC proliferation
	hiNPC2	hiNPC differentiation
	hiNPC3	hiNPC maturation
	hiNPC4	hiNPC synaptogenesis
	LDH	cell death
	CTB	cell viability

Cognitive function defects

Different in vitro systems are used to cover a broad selection of KEs. First, primary human neural progenitor cells (hNPCs) isolated from cortices of gestational week 16–19 fetuses were used as 3D neurospheres under proliferative and differentiative conditions and represent the fetal developmental stage. The hNPC assays are part of the OECD-supported DNT IVB (20) and are carried out at IUF. Second, adherent human induced pluripotent stem cell-derived neural progenitor cells (hiNPCs) representing the embryonic stage in its proliferative state were matured under differentiation conditions to form synapses characteristic of the 3rd trimester-developing brain. These assays are carried out at NIPH. The PFOA concentrations tested (Table 5) were based on levels found in the blood of Scandinavian populations (21). PFOA was dissolved in DMSO. Data obtained from hNPC experiments are analyzed and ready to be forwarded upon demand. These data do not show any DNT-specific effects or cytotoxicity caused by PFOA on both proliferating and differentiating hNPCs. Moreover, samples of PFOA in the in vitro medium with and without cells were collected for kinetic analysis and shipped to WP4 (Nynke Kramer, Sylvia Adam). The analysis of the results is probably still ongoing and may be followed up upon with Nynke Kramer. Experiments with hiNPCs have been partly carried out and will be delivered by the end of the year 2024. These experiments haven't shown so far any DNT-specific effects either.

WP9 offers in its in vitro battery additional assays that measure the key neurodevelopmental endpoint of human neural network formation (hNNF) and oligodendrocyte maturation (NPC6). The NPC6 assay is being optimized and thus it is currently not planned to use it for PFOA testing. The hNNF assay is currently employing the commercial NeuCyte in vitro system causing the assay to be very expensive to perform. We are currently working on its optimization using a different in-lab-made cell line in the framework of the PARC project. There are no indications that PFOA is a DNT-specific chemical in our in vitro systems based on the above-mentioned results and published results of the rat neural network formation assay (22). We hypothesize that this might be due to the involvement of microglia in PFOA's mode of action that are not present in our in vitro systems. Due to these reasons, we do not plan to test PFOA in the hNNF assay. We are considering testing it in the optimized version of the assay but this might be possible only by the end of 2025. Finally, the key neurodevelopmental process of proliferation is assessed by the NPC1 assay with no observed effects induced by PFOA, thus it is not planned to test PFOA using the hiNPC1 assay as well.

The delivery of these data is independent of input/data from other work packages/tasks. No incorporation of artificial intelligence is planned in this cognitive function defects case study. The analysis of the probability of hazard is applied to WP7-9 in the same way using the CRStats software and is described in section 5.3.

Table 5. Concentrations of PFOA used in WP9 in vitro experiments corresponding to PFOA levels found in the blood of Scandinavian populations. *tested only in hiNPCs; **Δ** tested only in hNPCs.

Compound	Exposure concentrations [nM] - times human blood levels						
	0.1xΔ	**0.5x**	**1x**	**10x***	**100x**	**500x**	**1000x**
PFOA	0.4209	2.105	4.209	42.094	420.94	2104.7	4209.4

Neural tube closure

To predict the probability of neural tube closure defects (NTDs) caused by chemical exposure, a dynamic systems model of neural tube closure was built in the CompuCell3D software modeling environment (19). The computational model is agent-based and represents pivotal human-relevant physiological morphogenetic signals and cellular dynamics underlying neural fold elevation, bending and fusion. These dynamics include various cellular behaviors such as proliferation, differentiation, adhesion, migration, apoptosis, contraction, and secretion, driven by a gene regulatory network derived from a physiological map (23). We tested and calibrated the model by introducing perturbations to the gene expression levels in the gene regulatory network based on data from existing literature, including mouse gene knockout studies and human clinical data of patients suffering from NTDs. Our model recapitulated the effects of these perturbations in both incidence and degree of defect and provided a percentage probability (the chance) that these perturbations would result in a neural tube defect. We estimate that the first chemical perturbations will be tested in the fall of 2024, thus, the results of PFOA perturbations could be delivered by the end of the year. The expected outcome is the probability of the closure of the neural tube dependent on the PFOA concentration.

This work is not dependent on input from other work packages or tasks. The computational model does not use AI for its predictions. The results of the computational model will be validated using the in vitro assay UKN2 developed in the laboratory of Marcel Leist (Risk-Hunt3R collaboration) and the Zebrafish Embryo Toxicity Test developed at RIVM. The UKN2 assay is part of the OECD-supported DNT-IVB (20) and measures the migration and viability of neural crest cells. PFOA has already been tested in this assay and no effects were observed up to 20 µM of PFOA (24).

Hazard characterisation

In this section we describe the hazard characterization based on the animal data first. Not because the animal data is more important, but because the RA methods are most developed for the use of animal data. The approach described for animal data will then be applied for human and in vitro data. The main goal, however, is to make use of the in vitro data in PRA, by adapting the approach from animal data.

Animal

The APROBA-Plus tool (see Method section) will be adapted for characterizing the toxicological effects in the animal data. We will use the PROAST tool to calculate a distribution of benchmark doses (BMDs) (25), based on selection of benchmark dose responses (BMR) for the included health effects. The selection of the BMR distribution will be based on the guidance for selection of BMRs in the Guidance on the use of the benchmark dose approach in risk assessment, published by EFSA in 2022 (26). The guidance state that "the BMR should reflect the dose where an effect becomes adverse and, therefore, depends on the nature of the endpoint selected (including apical and non-apical endpoints) and the relation to the BMD" and "Ideally, the BMR is set numerically so that it reflects the onset of a human-relevant adverse effect, meaning that a response above the BMR is considered adverse.". The BMR is crucial for the BMD calculation, and methods for setting the BMR have been described in the EFSA guidance (26). In line with the PRA principle, the BMR could be a distribution calculated based on different experiments which would include the variability in the BMR value. The BMR distribution will then be used to calculate a distribution of BMDs for each health effect for use in the risk assessment.
In addition, a distribution of assessment factors will be included for calculation of the probabilistic hazard in the APROBA-Plus tool, to get the final distribution of hazard for each health effect.

Human

The same principles as used for the animal data will be used for the human data to calculate the distribution of hazard in APROBA-Plus tool. However, we expect the number of human studies to be limited. The assessment factors will be reduced since the uncertainty from extrapolation of hazard between animal and humans will not be needed. The PROAST tool will be used to calculate a distribution of BMDs.

In vitro

The dose-response results obtained in WP7, WP8, and WP9 in vitro experiments are used to calculate the probability of hazard using the benchmark response (BMR) and benchmark concentrations (BMC) approach. The BMR is specific for every assay endpoint and marks the extent of response considered to be statistically significant and toxicologically meaningful. For example, BMRs of the WP9 hNPCs assays are calculated based on the coefficient of variation (CV) of the lowest compound concentration times 1.5. The CV is established based on the entire dataset of a study. BMCs are defined as the point of the concentration-response curve that intersects with the BMR. An alternative method for establishing the BMR for the different in vitro studies were suggested by US-EPA in 2012, where one standard deviation (1SD) was used for BMR. The distribution of these BMRs will be used in the dose response modelling from the in vitro studies, and the resulting hazard distributions for the non-apical endpoints. We will also investigate whether our calculations can be organized in an adapted workflow as described in the paper by Middleton, Reynolds (4).

Several software's are available for BMD calculations. The CRStats software available at https://iuf-duesseldorf.shinyapps.io/crstatslive4/ has previously been used by partners in ONTOX working with in vitro studies. CRStats uses BMRs and BMCs are calculated using an R software analysis pipeline (27) according to Blum, Masjosthusmann (24) in CRStats statistical software. An alternative software is the PROAST software also run as an R package or as a web tool (25). Both tools provide BMD dose accompanied by lower and upper 95% confidence intervals (BMCL, BMCU). The CRStats software selects the best-fitting model (e.g. linear, exponential, non-monotonic, etc.) for the in vitro data, while the PROAST tool use model averaging using Bayesian methods. Both the CRStats software and the PROAST software will be tested and the most appropriate tool for the in vitro results will be selected.

Physiological maps workflow

Establishment of the physiological maps (if you need to create one)

The next steps will require the use of a physiological map (PM). If you don't have a map that is suitable for your probabilistic risk assessment, you will find here a proposed workflow and references to create one. The workflow to build PMs is described in Staumont et al., in prep. The main steps include data collection, data processing, graphical representation, map visualization, sharing, review and update (inspired from Mazein, Acencio (28)). Data is retrieved from literature and pathway databases (e.g. Reactome, Wikipathways) with the help of Sysrev data mining tool, and integrated into a diagram using the CellDesigner editor (29). Data collection and processing can be assisted by AI/natural language processing -based tools such as INDRA (30). The PM is then uploaded to the Molecular Interaction Networks Visualization (MINERVA) platform, a

navigable tool (31), enabling better visualization and facilitating feedback from domain experts.

Molecular target identification on physiological maps

The identification of possible chemical-protein interactions can be performed directly using the MINERVA platform (https://minerva.pages.uni.lu/doc/). MINERVA has specific instances to host PMs, these are the places where PMs can be visualized. The MINERVA instance for the ONTOX project, containing the liver, kidney, and brain PMs, is located at: https://ontox.elixir-luxembourg.org/minerva/ . If one or more targets are present as elements of the PM, they can be highlighted by MINERVA for further investigation by the user. This can be achieved manually, using the MINERVA graphical interface, or programmatically, using the minervar. This last is an R package containing the relevant functions of MINERVA that can be reused in an R script. For more details about minervar please see:
https://onlinelibrary.wiley.com/doi/epdf/10.1002/pro.4565 and
https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2023.1101505/full .

Another useful tool is BioBricks (https://biobricks.ai/) that can also be used to interrogate several databases and extract the information in a standard format. A documentation about the installation of BioBricks and its use (with examples) is available at https://docs.biobricks.ai/ . Being the BioBricks output format from different databases standardized, these data can be easily integrated, providing more insights about a chemical and its biological interaction(s).

Relevant databases for studying chemical-biological properties interactions supported by the MINERVA platform are:

- ChEMBL
- DrugBank
- CTD

It is important to recall that the graphical interface of MINERVA uses only these databases and cannot interrogate other interaction databases that can be of interest for the user. However other databases can be interrogated by the package minervar. This action requires you to write a script, as minervar has no graphical interface as previously mentioned. BioBricks is also useful for the same purpose. Biobricks, MINERVA and minervar are actually under continuous development and improvement.
A possible workflow can be to extract all biological entities from an existing PM (or newly created PM) using the minervar package (as it can do that automatically for all PM entities) and then interrogate specific databases for possible interactions.

Chemical-target interaction characterization

Using the list of targets, the interactions between the chemical and each of the targets can be modelled using computational methods (e.g. QSAR) (Ortega-Vallbona et al., in

prep., WP3). The interactions are then quantified and characterized (input needed from WP3 (section “In silico predictions (WP3)”)).

Modelling

Map to Model

Task:	map conversion into dynamic model
Input:	CellDesigner SBML file
Output:	SBML-qual format model
Estimated timeframe:	few minutes

Physiological maps can be used to build large-scale Boolean models reflecting the complexity of the biology of the system under investigation. For that, an automated approach for converting a map, or part of it, can be used. CaSQ (CellDesigner as SBML-qual) (32) is a Python-based tool that converts maps represented using the Systems Biology Graphical Notation (33) process description language (https://sbgn.github.io/specifications) to a SMBL-qual format that suits different simulation platforms. In this stage, it is important to prepare the map in a way that represents the phenotypes under investigation and alternative phenotypic endpoints, for example, pathological and physiological mechanisms.

After preparing the map in a single CellDesigner SBML file, the conversion into a Boolean model in SBML-qual format can be done using CaSQ. Map preparation can take from a few days to a few weeks, depending on the scale of the map. Conversion using CaSQ usually takes no more than a few minutes.

Model format conversion

Task:	model format conversion
Input:	model file
Output:	model file in a different format
Estimated timeframe:	few minutes

The SBML-qual format is suitable for some analysis and platforms/software, as described on Table 1 from Hemedan, Niarakis (34). Conversion to other formats can be done using an in-house script by Ladeira and Hemedan, as shown in the Table 6 below

Model format converter: https://github.com/luiz-ladeira/m_format_converter/tree/main

Table 6. adapted from bioLQM documentation – https://github.com/colomoto/bioLQM).

type	import	export	format	description
B	yes	yes	bnet	bnet functions format
B	yes	yes	booleannet	Alternative functions format
B	yes	yes	boolsim	boolsim format
B	no	yes	cnet	cnet functions format
B	yes	yes	boolfunctions	Raw functions format
M	no	yes	ginml	GINML format
M	no	yes	gna	GNA (non-xml) format
B	no	yes	bnd	MaBoSS format
M	yes	yes	mnet	Multi-valued functions format
M	no	yes	an	Pint format
M	yes	yes	sbml	SBML-qual v1.0 format
M	yes	yes	tt	Truth table format *

Type: 'b'/'B'/'M' Boolean/Booleanized/Multivalued
* Truth table format can reach huge file size (many Gbs)

Model verification

Task:	model verification
Input:	model
Output:	reviewed model
Estimated timeframe:	up to several days

The trajectory analysis using MaBoSS (in a vanilla model, without parameterization) can be used to check the relative behavior of biological entities in the model, still not considering the quantitative aspects of the dynamics. The behavior of the model must reflect what is expected, according to what is described in the scientific literature. This

verification process can take several days, depending on the scale of the model, the familiarity of the person who is performing this step with the map and the biology of the system. All relationships should be verified and corrected as needed. Updating the map based on the model verification process might be done to correctly represent the biological processes.

Model reduction

Task:	model reduction
Input:	model
Output:	reduced model
Estimated timeframe:	up to several days

Model reduction is done to improve computing efficiency for highly demanding simulations and analyses, like the attractor search, for example. In some cases, the software or script requires a limited size of the model, like the BoolNet exhaustive attractor search, which requires a model with up to 29 nodes. There are options that could fit the requirements of the analysis and handle larger models. However, this is something that requires a case-by-case investigation.

A model reduction process consists of removing from the model nodes and edges (connections), which are not key to the dynamic process representation. At the end, the model should still reproduce the biology of the system, guaranteeing its reliability in generating consistent outcomes. Many reduction methods are available in the scientific literature; for example, Saadatpour, Albert (35). Other model reduction methods can be found in Wang, Saadatpour (36).

Model parameterization

Task:	model parameterization
Input:	gene, protein, and/or phenotype activity levels; Omics data (transcriptomics or proteomics), functional activity levels, and other different data in a fit-for-purpose manner.
Output:	a MaBoSS configuration file
Estimated timeframe:	up to several days

Parameterization of the model consists of setting initial values for the nodes. In the case of stochastic simulations, continuous values can be used to give different weights to

nodes on the network. These values can be calculated using different reasoning and data sources. Most commonly, differentially expressed gene fold changes are used to parameterize large-scale Boolean networks. Other omics data and parameters from functional assays can also be considered.

Chemical target activities, predicted by WP3, can be used to parameterize target nodes to reflect their activities upon chemical interaction. These quantitative values would be provided by WP3 based on their chemical-target interactions, converting a list of targets generated from querying databases for chemical-target interactions.

The values should be normalized to a scale from zero to one, to fit the requirements of the configuration file for Boolean simulations using MaBoSS (37). Parameterization is done in a case-by-case manner, depending on data availability and the interpretation of the available data.

Applications of fully parameterized probabilistic Boolean models can be found in Hemedan, Schneider (38).

Probabilistic characterization of hazards

Task:	probabilistic characterization of hazards
Input:	model and parameters
Output:	**trajectory:** quantitative prediction of the activity of a node. **sensitivity:** correlation values of nodes that influence the development of a (phenotype) node activity. **attractor search:** truth table, with the state of each node for each endpoint condition.
Estimated timeframe:	**trajectory:** few days **sensitivity:** few days **attractor search:** up to several days - depending on model size and complexity and computational power available.

To characterize hazards at the molecular level in a probabilistic manner, one can rely on probabilistic Boolean modelling (34). Hazard may be represented as a phenotype node at the end of a pathway, in which the disruption of a pathway leads to the occurrence of that phenotype (toxicity endpoint). From a physiological map, these pathological endpoints need to be added. Comparing a physiological map with an adverse outcome pathway can be useful for identifying related phenotypes, and enriching the map with them for modelling. Gene- or protein-rich pathways allow a richer parameterization of the model and, therefore, a more refined prediction.

Among the possibilities of analysis and simulation that can be done, three are key for a more comprehensive characterization of the hazard and the disruption in the system.

Trajectory analysis

The main objective of this analysis in the PRA process is to predict the chance (in percentage) of activation of a given toxicological endpoint, represented by a phenotype in the Boolean network. A final quantitative value will be taken from the point when the simulation reaches a stable state. This value reflects the activity level of the node, after simulation from an initial point which reflects the initial parameters of the model. Different parameterization is required to reflect possible variability of the initial point, in order to consider a global chance, with its deviations, for the activity of that node, reflecting a populational level for the chance of toxicity. The main output is the quantitative prediction of the activity of a node in the network. For this, one might use MaBoSS (37), PhysiBoSS (39), CellCollective (40), or others. Validation of the outcomes using new in vitro or legacy data from the literature might be needed to correctly interpret the outcome and increase confidence in the predictions.

A secondary outcome is a graph illustrating the relative activities of different key nodes. This graph displays how node activities are related to each other and can help in describing a temporal aspect of toxicity development.

Sensitivity analysis

A sensitivity analysis gives the critical biological entities present on the network that play an important role in leading the system to a specific level of activity of the node (toxicological phenotype) under investigation. This analysis helps in understanding the mechanisms leading to toxicity and describing which nodes have more weight in leading to such a toxicological endpoint. The main output are the correlation values of nodes that influence the development of a (phenotype) node activity. Different tools can be used, such as CellCollective (40), RMut - https://github.com/csclab/RMut (41) , BoolNet (42), and others.

Attractors search

Attractors represent the long-term behavior of the system. Different initial conditions may lead the system to different attractors. They represent all possible system organizations leading to a specific phenotype condition (36). This could help in understanding and describing the mechanistic aspect of the toxicity development, by calculating all possibilities of the system's evolution towards a steady state. Attractor search usually demands higher computational power for processing and can take up to many days. A potential solution to this problem is to first simplify the networks before moving on to the dynamic analysis (see model reduction). The main output of this analysis is a truth table, with the state of each node for each endpoint condition. This truth table can be translated into graphs, or graphical overlays to be visualized on the map of Boolean network. Tools for this analysis are available, as CellCollective (40), GinSim (43), BoolNet (42), and other, as described in Table 1 of Hemedan, Niarakis (34).

Other roles for PMs in supporting mechanistic risk assessment

- AOP development and refinement:

By providing mechanistic details about the physiology of the organ-specific processes, PMs also help to identify gaps in AOPs and to improve them. In particular, the MIE and KE contained in AOPs (which can also be standardized using SBGN representation) can be linked and integrated with the biochemical processes of the PMs. Such integration allows for AOP development and refinement, and therefore hazard identification and characterization (also support for WP7,8,9; section 4.2)

- Data visualization

The MINERVA platform, on which PMs are visualized, allows for the visualization of data. In particular, omics data can be overlayed on top of the PMs. These visualization properties will help domain experts (WP7,8,9) to identify key biological pathways or mechanisms, and support the development or improvement of in vitro experiments. This will contribute to hazard identification and characterization (section 4.2)

- AI-related tasks in WP5 (through WikiPathways and Biobricks)

Several biological pathways are represented in one PM. For each PM, some pathways (or group of pathways) are also included in smaller modules called submaps (i.e. fragments of the main map). Submaps will be included in the WikiPathways database which will allow the exploitation of their data using Biobricks (as WikiPathways is one of the bricks in Biobricks).

Gaps and limitations (merge with item 7)

It is important to mention that the steps described here are easily applicable to a single perturbation on network nodes, reflecting an acute toxicity development. Chronic toxicity and repeated-dose toxicity can be predicted, but not without using more complex modeling approaches.

It has already been demonstrated that chronic and repeated-dose biological perturbations (therapies) can be modeled and predicted in a probabilistic manner using these technologies (39)
. The same framework can be adapted to toxicological endpoints. However, adaptations and using multiple modeling formalisms might be needed. For example, using PhysiBoSS, one can combine ODE models describing the chemical interaction with each one of its targets, in which the output of these models is used to parameterize the target nodes on a large boolean model and simulate the general behavior of the system for specific doses and different exposure schemes. This allows for covering not only acute toxicity predictions but also chronic, repeated-dose, and recovery simulation predictions for different doses. The challenge is that an ODE model is needed for each chemical-target interaction. The timing for developing such a model will vary depending on the number of targets and the information available.

Risk characterisation

The margin of exposure (MOE) approach will be used for the PRA. The comparison between the distribution of exposure and distribution of hazard will be used to sample MOEs, which will give a probability distribution of possible MOEs.

For the PRA based on the in vitro data, the hazard distribution will be compared with the internal exposure after calculations using the PBPK models and QIVIVE, since the in vitro data will measure hazard at the internal cellular levels in humans.

This probability of MOEs will be used to draw conclusions of the PRA of PFOA based on the in vitro, animal and human data, depending on the data availability. A comparison of the output based on the in vitro data and animal/human data will be described.

Uncertainty, data gaps and deviation from the protocol

This case study is a first attempt to use in vitro studies to assess PRA of PFOA. Therefore, deviation from the protocol will be expected, as this is groundbreaking work. A list of the deviations from the protocol will be described accompanied with a scientific reasoning for the deviation.

In addition, a list of uncertainties and data gaps will be included to inform future work to improve these methods.

Per review of the project protocol

The project protocol was reviewed by Monica Andreassen, Senior Advisor at NIPH.

References

1. EPA USEPA. Risk Assessment Forum White Paper: Probabilistic Risk Assessment Methods and Case Studies. EPA 2014;(004).
2. Chiu WA, Slob W. A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. Environ Health Perspect 2015;123(12):1241-54. DOI: 10.1289/ehp.1409385
3. Penserini L, Cantoni B, Vries D, Turolla A, Smeets P, Bokkers BGH, et al. Quantitative chemical risk assessment for mixtures: Application to alkylphenol mixtures and phthalate mixtures in tap and bottled water. Environ Int 2022;165:107294. DOI: 10.1016/j.envint.2022.107294
4. Middleton AM, Reynolds J, Cable S, Baltazar MT, Li H, Bevan S, et al. Are Non-animal Systemic Safety Assessments Protective? A Toolbox and Workflow. Toxicol Sci 2022;189(1):124-47. DOI: 10.1093/toxsci/kfac068
5. Luechtefeld T, Marsh D, Rowlands C, Hartung T. Machine Learning of Toxicological Big Data Enables Read-Across Structure Activity Relationships (RASAR) Outperforming Animal Test Reproducibility. Toxicol Sci 2018;165(1):198-212. DOI: 10.1093/toxsci/kfy152
6. Husøy T, Andreassen M, Hjertholm H, Carlsen MH, Norberg N, Sprong C, et al. The Norwegian biomonitoring study from the EU project EuroMix: Levels of phenols and phthalates in 24-hour urine samples and exposure sources from food and personal care products. Environ Int 2019;132:105103. DOI: 10.1016/j.envint.2019.105103
7. EFSA. Risk to human health related to the presence of perfluoroalkyl substances in food. EFSA Journal 2020;18.
8. Liu G, Zhang B, Hu Y, Rood J, Liang L, Qi L, et al. Associations of Perfluoroalkyl substances with blood lipids and Apolipoproteins in lipoprotein subspecies: the POUNDS-lost study. Environ Health 2020;19(1):5. DOI: 10.1186/s12940-020-0561-8
9. Shaheen N, Shaheen A, Ramadan A, Hefnawy MT, Ramadan A, Ibrahim IA, et al. Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability. Front Res Metr Anal 2023;8:1268045. DOI: 10.3389/frma.2023.1268045
10. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225-34. DOI: 10.1016/j.jclinepi.2015.06.005
11. NTP. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. 2019.
12. Svendsen C, Whaley P, Vist GE, Husoy T, Beronius A, Consiglio ED, et al. Protocol for designing INVITES-IN, a tool for assessing the internal validity of in vitro studies. Evid Based Toxicol 2023;1(1):1-15. DOI: 10.1080/2833373x.2023.2232415
13. Leist M, Ghallab A, Graepel R, Marchan R, Hassan R, Bennekou SH, et al. Adverse outcome pathways: opportunities, limitations and open questions. Arch Toxicol 2017;91(11):3477-505. DOI: 10.1007/s00204-017-2045-3
14. van Ertvelde J, Verhoeven A, Maerten A, Cooreman A, Santos Rodrigues BD, Sanz-Serrano J, et al. Optimization of an adverse outcome pathway network on chemical-induced cholestasis using an artificial intelligence-assisted data collection and confidence level quantification approach. J Biomed Inform 2023;145:104465. DOI: 10.1016/j.jbi.2023.104465
15. Verhoeven A, van Ertvelde J, Boeckmans J, Gatzios A, Jover R, Lindeman B, et al. A quantitative weight-of-evidence method for confidence assessment of adverse

outcome pathway networks: A case study on chemical-induced liver steatosis. Toxicology 2024;505:153814. DOI: 10.1016/j.tox.2024.153814
16. Barnes DA, Firman JW, Belfield SJ, Cronin MTD, Vinken M, Janssen MJ, et al. Development of an adverse outcome pathway network for nephrotoxicity. Arch Toxicol 2024;98(3):929-42. DOI: 10.1007/s00204-023-03637-7
17. Koch K, Bartmann K, Hartmann J, Kapr J, Klose J, Kuchovska E, et al. Scientific Validation of Human Neurosphere Assays for Developmental Neurotoxicity Evaluation. Front Toxicol 2022;4:816370. DOI: 10.3389/ftox.2022.816370
18. Davidsen N, Lauvas AJ, Myhre O, Ropstad E, Carpi D, Gyves EM, et al. Exposure to human relevant mixtures of halogenated persistent organic pollutants (POPs) alters neurodevelopmental processes in human neural stem cells undergoing differentiation. Reprod Toxicol 2021;100:17-34. DOI: 10.1016/j.reprotox.2020.12.013
19. Berkhout, J.A.; Piersma, A.H.; Belmonte, J.; Legler, J.; Spencer, R.M.; Knudsen, T.B.; Heusinkveld, H. J. A computational dynamic systems model for in silico prediction of neural tube closure defects. Toxicology 2024.
20. OECD. Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery. 2023.
21. Berntsen HF, Berg V, Thomsen C, Ropstad E, Zimmer KE. The design of an environmentally relevant mixture of persistent organic pollutants for use in in vivo and in vitro studies. J Toxicol Environ Health A 2017;80(16-18):1002-16. DOI: 10.1080/15287394.2017.1354439
22. Carstens KE, Freudenrich T, Wallace K, Choo S, Carpenter A, Smeltz M, et al. Evaluation of Per- and Polyfluoroalkyl Substances (PFAS) In Vitro Toxicity Testing for Developmental Neurotoxicity. Chem Res Toxicol 2023;36(3):402-19. DOI: 10.1021/acs.chemrestox.2c00344
23. Heusinkveld HJ, Staal YCM, Baker NC, Daston G, Knudsen TB, Piersma A. An ontology for developmental processes and toxicities of neural tube closure. Reprod Toxicol 2021;99:160-7. DOI: 10.1016/j.reprotox.2020.09.002
24. Blum J, Masjosthusmann S, Bartmann K, Bendt F, Dolde X, Donmez A, et al. Establishment of a human cell-based in vitro battery to assess developmental neurotoxicity hazard of chemicals. Chemosphere 2023;311(Pt 2):137035. DOI: 10.1016/j.chemosphere.2022.137035
25. RIVM. PROAST. 2012.
26. EFSA. Guidance on the use of the benchmark dose approach in risk assessment. EFSA Journal 2022;20.
27. Kessel HE, Masjosthusmann S, Bartmann K, Blum J, Donmez A, Forster N, et al. The impact of biostatistics on hazard characterization using in vitro developmental neurotoxicity assays. ALTEX 2023;40(4):619-34. DOI: 10.14573/altex.2210171
28. Mazein A, Acencio ML, Balaur I, Rougny A, Welter D, Niarakis A, et al. A guide for developing comprehensive systems biology maps of disease mechanisms: planning, construction and maintenance. Front Bioinform 2023;3:1197310. DOI: 10.3389/fbinf.2023.1197310
29. Funahashi A, Jouraku A, Matsuoka Y, Kitano H. Integration of CellDesigner and SABIO-RK. In Silico Biol 2007;7(2 Suppl):S81-90.
30. Bachman JA, Gyori BM, Sorger PK. Automated assembly of molecular mechanisms at scale from text mining and curated databases. Mol Syst Biol 2023;19(5):e11325. DOI: 10.15252/msb.202211325
31. Hoksza D, Gawron P, Ostaszewski M, Hasenauer J, Schneider R. Closing the gap between formats for storing layout information in systems biology. Brief Bioinform 2020;21(4):1249-60. DOI: 10.1093/bib/bbz067
32. Aghamiri SS, Singh V, Naldi A, Helikar T, Soliman S, Niarakis A. Automated inference of Boolean models from molecular interaction maps using CaSQ. Bioinformatics 2020;36(16):4473-82. DOI: 10.1093/bioinformatics/btaa484

33. Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nat Biotechnol 2009;27(8):735-41. DOI: 10.1038/nbt.1558
34. Hemedan AA, Niarakis A, Schneider R, Ostaszewski M. Boolean modelling as a logic-based dynamic approach in systems medicine. Comput Struct Biotechnol J 2022;20:3161-72. DOI: 10.1016/j.csbj.2022.06.035
35. Saadatpour A, Albert R, Reluga TC. A Reduction Method for Boolean Network Models Proven to Conserve Attractors. SIAM J Appl Dyn Syst 2013;12(4):1997-2011. DOI: 10.1137/13090537X
36. Wang RS, Saadatpour A, Albert R. Boolean modeling in systems biology: an overview of methodology and applications. Phys Biol 2012;9(5):055001. DOI: 10.1088/1478-3975/9/5/055001
37. Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics 2017;33(14):2226-8. DOI: 10.1093/bioinformatics/btx123
38. Hemedan AA, Schneider R, Ostaszewski M. Applications of Boolean modeling to study the dynamics of a complex disease and therapeutics responses. Front Bioinform 2023;3:1189723. DOI: 10.3389/fbinf.2023.1189723
39. Letort G, Montagud A, Stoll G, Heiland R, Barillot E, Macklin P, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics 2019;35(7):1188-96. DOI: 10.1093/bioinformatics/bty766
40. Helikar T, Kowal B, McClenathan S, Bruckner M, Rowley T, Madrahimov A, et al. The Cell Collective: toward an open and collaborative approach to systems biology. BMC Syst Biol 2012;6:96. DOI: 10.1186/1752-0509-6-96
41. Koschutzki D, Schwobbermeyer H, Schreiber F. Ranking of network elements based on functional substructures. J Theor Biol 2007;248(3):471-9. DOI: 10.1016/j.jtbi.2007.05.038
42. Mussel C, Hopfensitz M, Kestler HA. BoolNet--an R package for generation, reconstruction and analysis of Boolean networks. Bioinformatics 2010;26(10):1378-80. DOI: 10.1093/bioinformatics/btq124
43. Gonzalez AG, Naldi A, Sanchez L, Thieffry D, Chaouiya C. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. Biosystems 2006;84(2):91-100. DOI: 10.1016/j.biosystems.2005.10.003

Time plan

Commissioned date: 31.01.2025
Due date: 30.04.2026