NAM-Based Assessment Framework for Classification of Chemicals for Systemic Toxicity Potential

ASCCT-ESTIV Award Winners Series
February 25, 2026

Prachi Pradeep, PhD
Department of Chemical and Product Safety,
German Federal Institute for Risk Assessment (BfR),
Germany

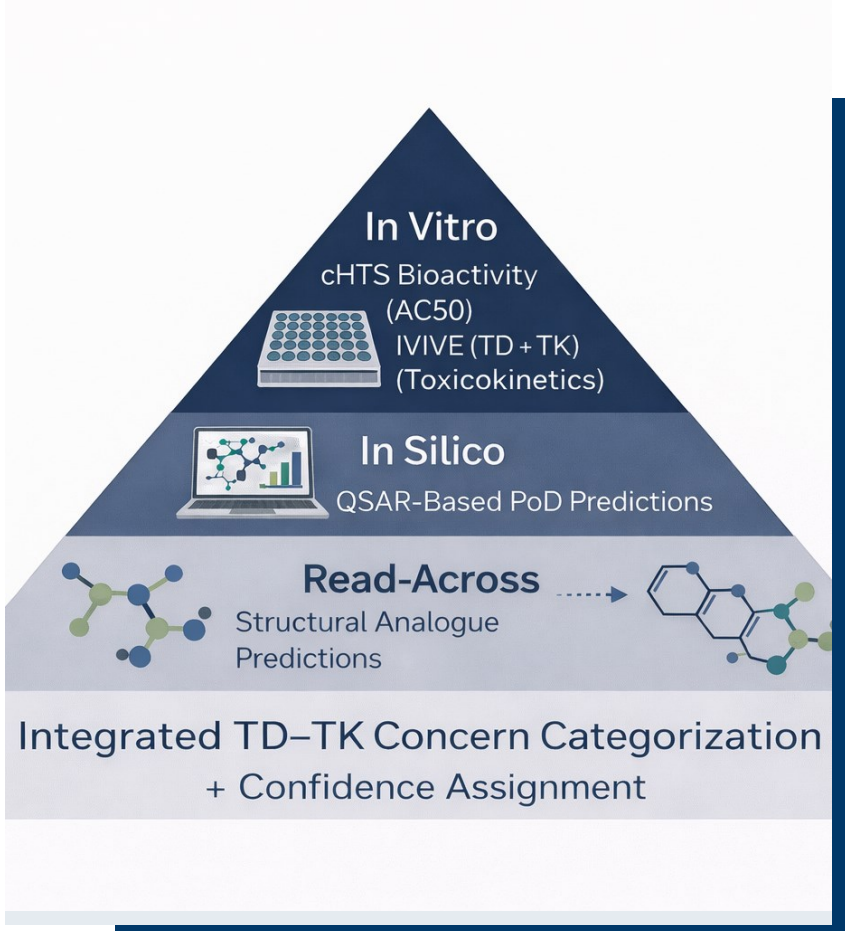


Image generated using ChatGPT (OpenAI), 2026.

The EPAA Designathon for Human Systemic Toxicity

The European Partnership for Alternative Approaches to Animal Testing

Established in 2005, EPAA is a collaboration between the European Union (EU) and industry stakeholders from 9 sectors

Vision: Replacement, reduction and refinement (3Rs) of animal use for meeting regulatory requirements through better and more predictive science

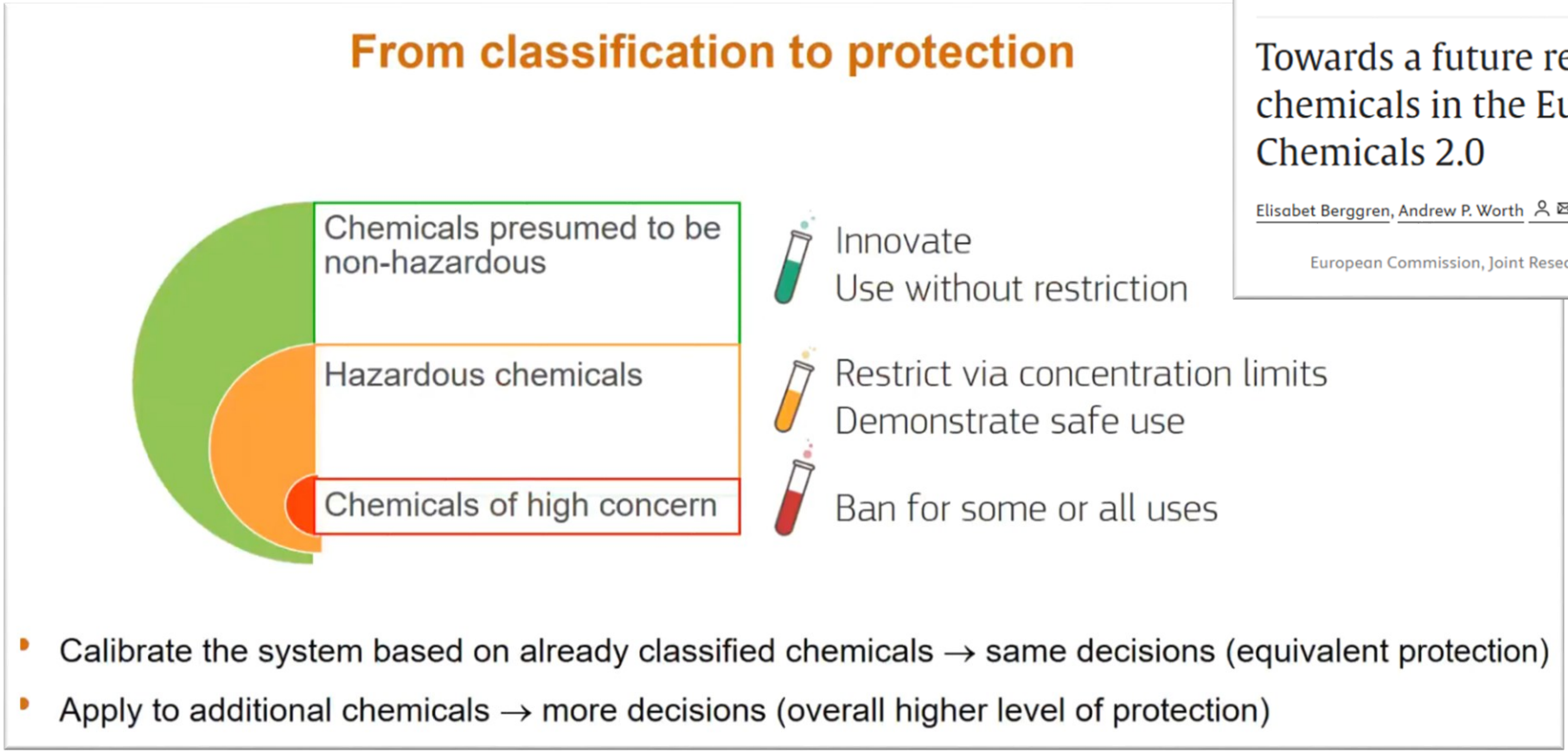
Goal: To assign a level of concern (high, medium and low) for systemic toxicity to each chemical based on "New Approach Methodology" (NAM) data related to their toxicodynamic and toxicokinetic properties

Proposal for a new classification matrix

		Activity (NAM-based toxicodynamics)		
		High	Medium	Low
Potential Systemic Availability (NAM-based toxicokinetics, based on ADME properties)	High	H	H	M
	Medium	H	M	L
	Low	M	L	L

BfR

From Classification to Protection




ELSEVIER
Regulatory Toxicology and Pharmacology
Volume 142, August 2023, 105431

Towards a future regulatory framework for chemicals in the European Union – Chemicals 2.0

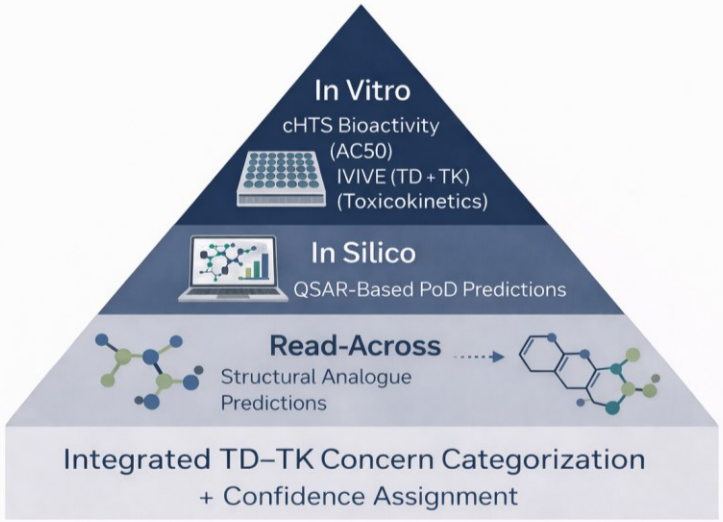
Elisabet Berggren, Andrew P. Worth

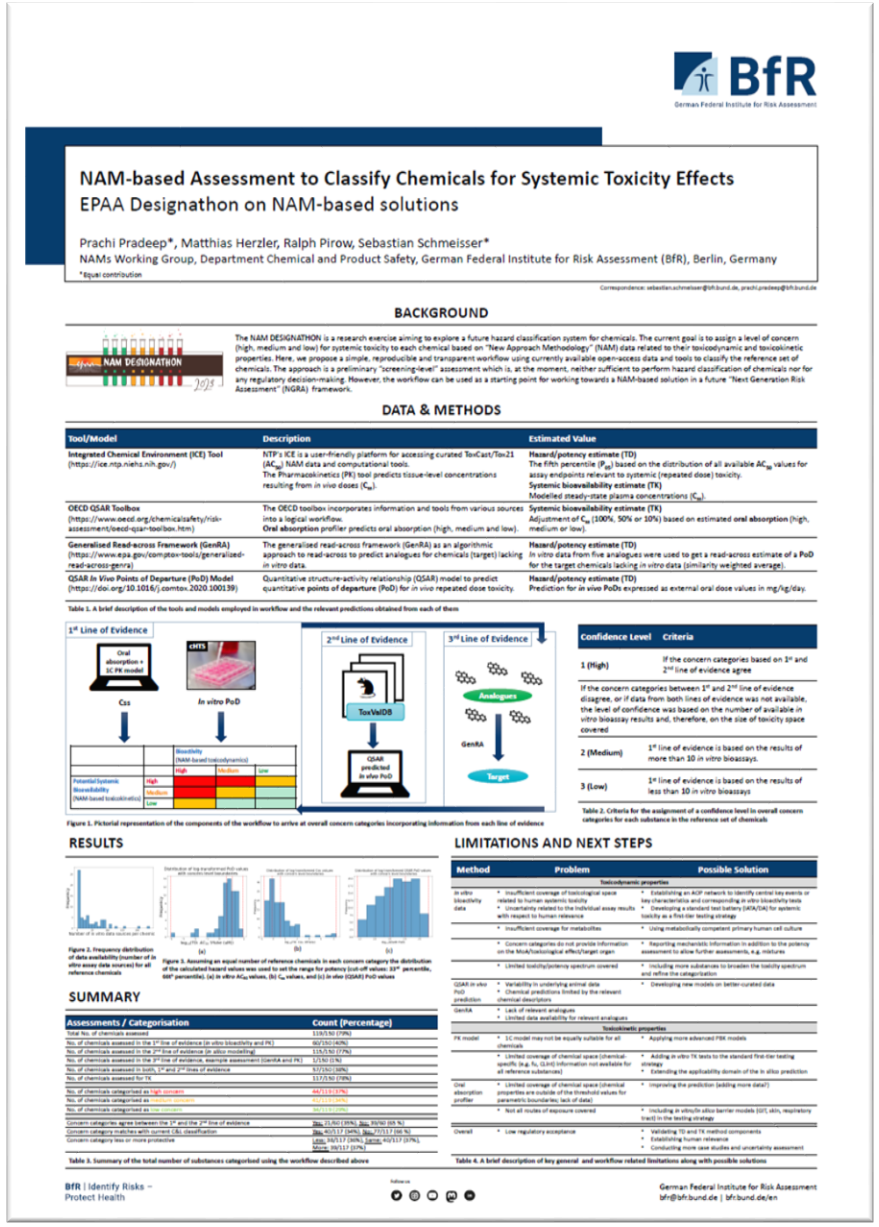
European Commission, Joint Research Centre (JRC), Ispra, Italy

NAM-based Assessment Workflow

Proposed Solution

- Simple and transparent stepwise workflow
- Open-access data and tools were used to classify the reference set of chemicals
- Integrating three lines of evidence
  - In vitro – cHTS Data
  - In silico – QSAR predictions
  - Data-gap Filling – Read-across predictions





BfR
German Federal Institute for Risk Assessment

NAM-based Assessment to Classify Chemicals for Systemic Toxicity Effects
EPAA Designathon on NAM-based solutions

Prachi Pradeep*, Matthias Herzler, Ralph Pirow, Sebastian Schmeisser*
NAMs Working Group, Department Chemical and Product Safety, German Federal Institute for Risk Assessment (BfR), Berlin, Germany
*Equal contribution
Correspondence: sebastian.schmeisser@bfr.bund.de, prachi.pradeep@bfr.bund.de

BACKGROUND

The NAM DESIGNATHON is a research exercise aiming to explore a future hazard classification system for chemicals. The current goal is to assign a level of concern (high, medium and low) for systemic toxicity to each chemical based on "New Approach Methodology" (NAM) data related to their toxicodynamic and toxicokinetic properties. Here, we propose a simple, reproducible and transparent workflow using currently available open-access data and tools to classify the reference set of chemicals. The approach is a preliminary "screening-level" assessment which is, at the moment, neither sufficient to perform hazard classification of chemicals nor for any regulatory decision-making. However, the workflow can be used as a starting point for working towards a NAM-based solution in a future "Next Generation Risk Assessment" (NGRA) framework.

DATA & METHODS

Tool/Model	Description	Estimated Value
Integrated Chemical Environment (ICE) Tool (https://ice.ntp.niehs.nih.gov/)	NTP's ICE is a user-friendly platform for accessing curated ToxCast/Tox21 (AC_{50}) NAM data and computational tools. The Pharmacokinetics (PK) tool predicts tissue-level concentrations resulting from *in vivo* doses (C_{ss}).	Hazard/potency estimate (TD) The fifth percentile (P_{05}) based on the distribution of all available AC_{50} values for assay endpoints relevant to systemic (repeated dose) toxicity. Systemic bioavailability estimate (TK) Modelled steady-state plasma concentrations (C_{ss}).
OECD QSAR Toolbox (https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm)	The OECD toolbox incorporates information and tools from various sources into a logical workflow. Oral absorption profiler predicts oral absorption (high, medium and low).	Systemic bioavailability estimate (TK) Adjustment of C_{ss} (100%, 50% or 10%) based on estimated oral absorption (high, medium or low).
Generalised Read-across Framework (GenRA) (https://www.epa.gov/comptox-tools/generalized-read-across-genra)	The generalised read-across framework (GenRA) as an algorithmic approach to read-across to predict analogues for chemicals (target) lacking *in vitro* data.	Hazard/potency estimate (TD) *In vitro* data from five analogues were used to get a read-across estimate of a PoD for the target chemicals lacking *in vitro* data (similarity weighted average).
QSAR *In Vivo* Points of Departure (PoD) Model (https://doi.org/10.1016/j.comtox.2020.100139)	Quantitative structure-activity relationship (QSAR) model to predict quantitative points of departure (PoD) for *in vivo* repeated dose toxicity.	Hazard/potency estimate (TD) Prediction for *in vivo* PoDs expressed as external oral dose values in mg/kg/day.

Table 1. A brief description of the tools and models employed in workflow and the relevant predictions obtained from each of them

Figure 1. Pictorial representation of the components of the workflow to arrive at overall concern categories incorporating information from each line of evidence

Confidence Level	Criteria
1 (High)	If the concern categories based on 1st and 2nd line of evidence agree
If the concern categories between 1st and 2nd line of evidence disagree, or if data from both lines of evidence was not available, the level of confidence was based on the number of available *in vitro* bioassay results and, therefore, on the size of toxicity space covered	
2 (Medium)	1st line of evidence is based on the results of more than 10 *in vitro* bioassays.
3 (Low)	1st line of evidence is based on the results of less than 10 *in vitro* bioassays

Table 2. Criteria for the assignment of a confidence level in overall concern categories for each substance in the reference set of chemicals

RESULTS

Figure 2. Frequency distribution of data availability (number of *in vitro* assay data sources) for all reference chemicals

Figure 3. Assuming an equal number of reference chemicals in each concern category the distribution of the calculated hazard values was used to set the range for potency (cut-off values: 33rd percentile, 66th percentile). (a) *in vitro* AC_{50} values, (b) C_{ss} values, and (c) *in vivo* (QSAR) PoD values

SUMMARY

Assessments / Categorisation	Count [Percentage]
Total No. of chemicals assessed	116/192 (59%)
No. of chemicals assessed in the 1st line of evidence (*in vitro* bioactivity and PK)	92/192 (48%)
No. of chemicals assessed in the 2nd line of evidence (*in silico* modelling)	115/192 (77%)
No. of chemicals assessed in the 3rd line of evidence, example assessment (GenRA and PK)	17/192 (1%)
No. of chemicals assessed in both, 1st and 2nd lines of evidence	97/192 (50%)
No. of chemicals assessed for TK	117/192 (78%)
No. of chemicals categorised as high concern	44/113 (37%)
No. of chemicals categorised as medium concern	41/113 (36%)
No. of chemicals categorised as low concern	28/113 (25%)
Concern categories agree between the 1st and the 2nd line of evidence	Yes: 21/92 (39%), No: 39/92 (69 %)
Concern category matches with current C&L classification	Yes: 40/117 (34%), No: 77/117 (66 %)
Concern category less or more protective	Less: 39/117 (36%), Same: 40/117 (37%), More: 39/117 (33%)

Table 3. Summary of the total number of substances categorised using the workflow described above

LIMITATIONS AND NEXT STEPS

Method	Problem	Possible Solution
	Toxicodynamic properties	
in vitro bioactivity data	* Insufficient coverage of toxicological space related to human systemic toxicity * Uncertainty related to the individual assay results with respect to human relevance	* Establishing an AOP network to identify central key events or key characteristics and corresponding *in vitro* bioactivity tests * Developing a standard test battery (IATA/DA) for systemic toxicity as a first-tier testing strategy
	* Insufficient coverage for metabolites	* Using metabolically competent primary human cell culture
	* Concern categories do not provide information on the MoA/toxicological effect/target organ	* Reporting mechanistic information in addition to the potency assessment to allow further assessments, e.g. mixtures
	* Limited toxicity/potency spectrum covered	* Including more substances to broaden the toxicity spectrum and refine the categorisation
QSAR *in vivo* PoD prediction	* Variability in underlying animal data * Chemical predictions limited by the relevant chemical descriptors	* Developing new models on better-curated data
GenRA	* Lack of relevant analogues * Limited data availability for relevant analogues	
	Toxicokinetic properties	
PK model	* 1C model may not be equally suitable for all chemicals	* Applying more advanced PBK models
	* Limited coverage of chemical space (chemical-specific (e.g. f_{up} (1,04) information not available for all reference substances)	* Adding *in vitro* TK tests to the standard first-tier testing strategy * Extending the applicability domain of the *in silico* prediction
Oral absorption profiler	* Limited coverage of chemical space (chemical properties are outside of the threshold values for parametric boundaries; lack of data)	* Improving the prediction (adding more data?)
	* Not all routes of exposure covered	* Including *in vitro*/*in silico* barrier models (GIT, skin, respiratory tract) in the testing strategy
Overall	* Low regulatory acceptance	* Validating TD and TK method components * Establishing human relevance * Conducting more case studies and uncertainty assessment

Table 4. A brief description of key general and workflow related limitations along with possible solutions

BfR | Identify Risks – Protect Health

German Federal Institute for Risk Assessment
bfr@bfr.bund.de | bfr.bund.de/en

BfR

Data Sources and Computational Tools

Tool|Model	Line of Evidence	Description	Estimated Quantity
Integrated Chemical Environment (ICE) Tool (https://ice.ntp.niehs.nih.gov/) A curated NTP/NICEATM resource integrating in vivo/in vitro data, chemical properties, predicted bioactivity, and computational tools (including ToxCast/Tox21) to support chemical safety assessment	In Vitro	**IVIVE Module:** In Vitro to In Vivo Extrapolation tool that translates in vitro assay activity concentrations into predicted in vivo equivalent doses using pharmacokinetic modeling	**Hazard/potency estimate (TD)** The fifth percentile (TD05) from available AC50 distribution for systemic (repeated-dose) endpoints
			Systemic bioavailability estimate (TK) Modelled steady-state plasma concentrations ($\mathbf{C_{ss}}$)
	Read-across	**Chemical Quest Module:** Similarity search tool using chemical structure fingerprints to identify similar chemicals with more data	**TD and TK** In vitro data from analogues were used to get a read-across TD and TK estimate for the target chemicals lacking in vitro data
QSAR In Vivo Points of Departure (PoD) Model (https://doi.org/10.1016/j.comtox.2020.100139) A structure-based QSAR model trained on in vivo repeat-dose toxicity data for ~3,500 chemicals (U.S. EPA's ToxValDB)	QSAR	QSAR model to predict quantitative PoDs (mg/kg/day) with confidence intervals to support hazard assessment in absence of experimental data	**TD** Prediction for **in vivo PoDs** expressed as external oral dose values in mg/kg/day.

Line of Evidence 1: In Vitro Bioactivity and IVIVE

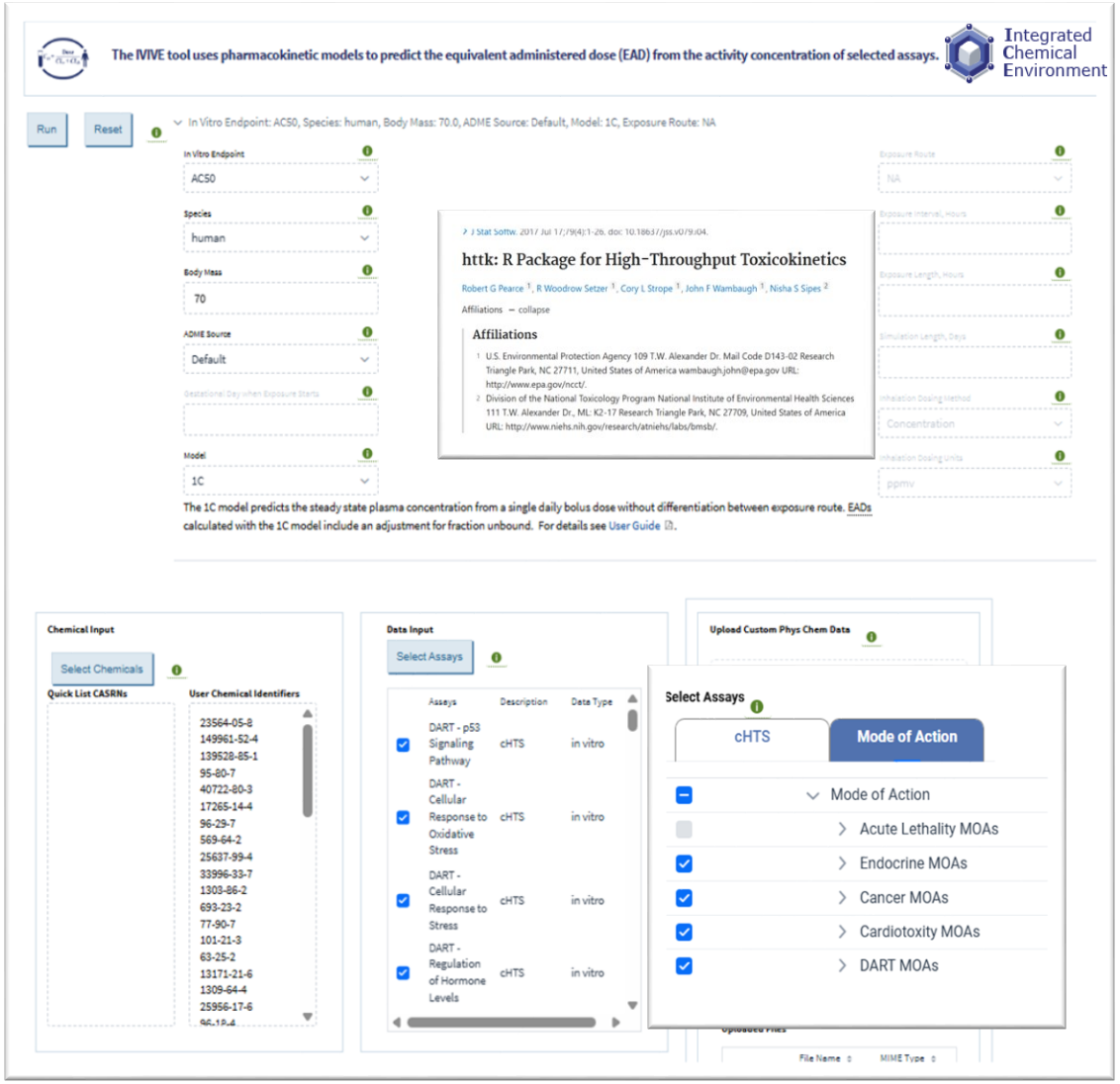
Purpose
Support hazard estimation using bioactivity and IVIVE data

Toxicodynamics (TD)
Value: In vitro bioactivity data, AC_{50}
Source: Tox21 and ToxCast data
Focus: Assays relevant to systemic (repeated-dose) toxicity aggregated by mode of action
Data handling:

- Compiled all available AC_{50} values
- Calculated 5th percentile (TD05) of distribution as a conservative Point of Departure (PoD) for hazard estimation

Toxicokinetics (TK)
Value: Steady-state plasma concentration (Css)
Assumptions:

- Model: 1-compartment (1C) population-based PK model from httk R package (implemented in ICE)
- Dose: 1 mg/kg/day, oral
- Species: Human
- Body Weight: 70.0 kg
- ADME Source: Default

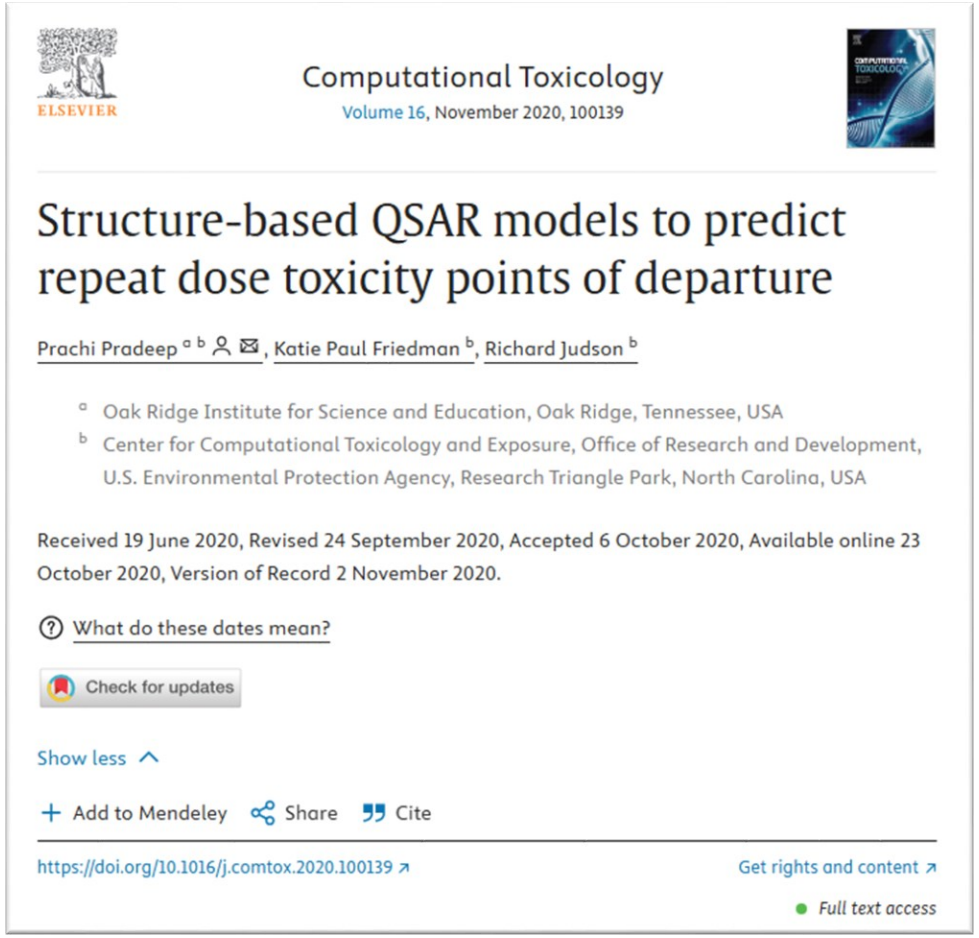


Line of Evidence 2: QSAR-based In Silico Predictions

Purpose
Support the assessment with an additional line of evidence

Model Basis
- Approach: In silico QSAR models for repeated-dose toxicity
- Training data: Large in vivo legacy dataset for ~3500 chemicals from US EPAs Toxicity Value database (ToxValDB)
- Output: Predicted in vivo PoDs (external oral dose, mg/kg/day) and a confidence interval

Conceptual Basis
- Predicted PoDs reflect both toxicodynamics (TD) and toxicokinetics (TK)
- Directly used for concern category assignment



ELSEVIER

Computational Toxicology
Volume 16, November 2020, 100139

Structure-based QSAR models to predict repeat dose toxicity points of departure

Prachi Pradeep [a b], Katie Paul Friedman [b], Richard Judson [b]

[a] Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA
[b] Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

Received 19 June 2020, Revised 24 September 2020, Accepted 6 October 2020, Available online 23 October 2020, Version of Record 2 November 2020.

What do these dates mean?

Check for updates

Show less

Add to Mendeley Share Cite

https://doi.org/10.1016/j.comtox.2020.100139

Get rights and content

Full text access

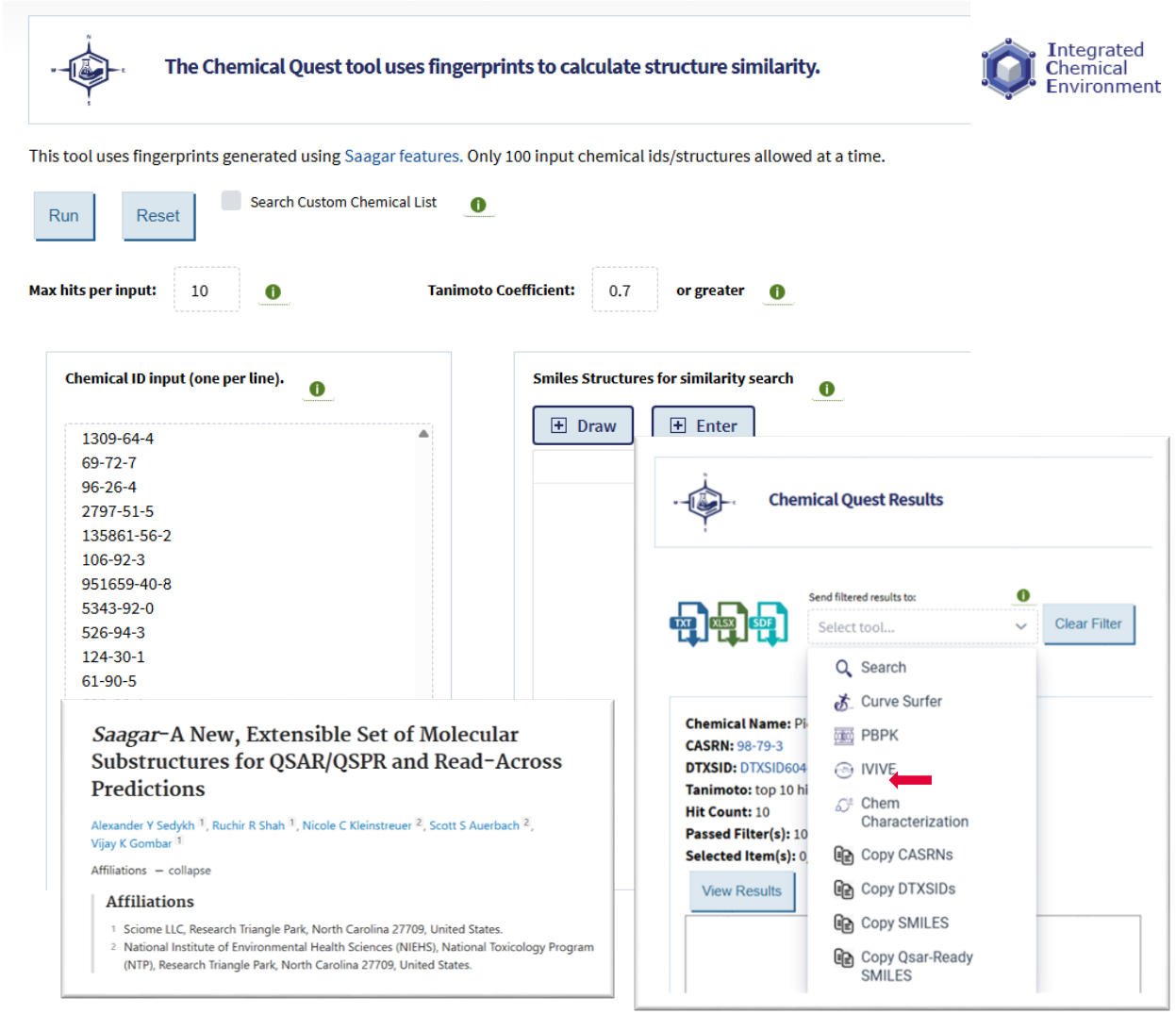
Line of Evidence 3: Read-across for Data Gap Filling

Purpose
Address data gaps for chemicals lacking relevant in vitro AC50 data

Analog Identification
- ICE - Chemical Quest Tool
- Structural similarity search using Saagar fingerprints
- Similarity criterion: Tanimoto coefficient ≥ 0.7

Data Retrieval & Application
- Identified analogs routed to IVIVE functionality for in vitro data retrieval

Data-gap Filling
- Mean AC50 and Css values used for read-across to fill target chemical data gaps for TD and TK assessment



TD–TK Concern Level Categorization

Assumptions

- Log10-transformed TD, TK, and QSAR PoD values
- Equal number of reference chemicals per concern category

Proposal for a new classification matrix

		Activity (NAM-based toxicodynamics)		
		High	Medium	Low
Potential Systemic Availability (NAM-based toxicokinetics, based on ADME properties)	High	H	H	M
	Medium	H	M	L
	Low	M	L	L

Concern Categorization

- **In Vitro**

TD → AC50 5th percentile
TK → Css (95th percentile)
Log10 → tertiles → High/Medium/Low

- **QSAR**

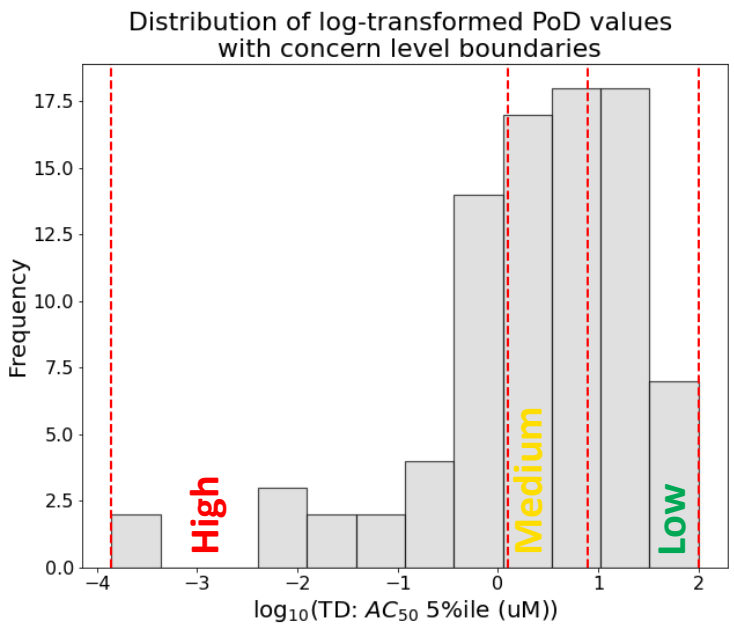
Predicted in vivo PoD (mg/kg/day)
Log10 → tertiles → High/Medium/Low

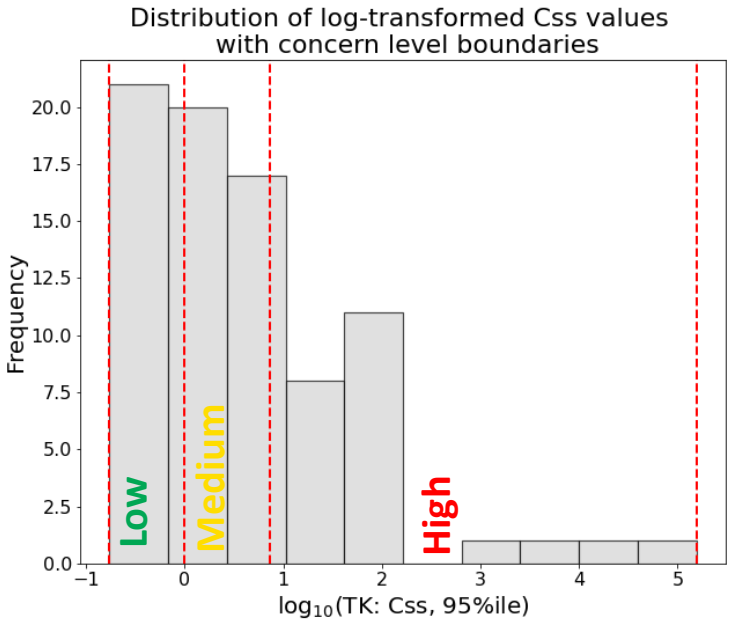
- **Read-across**

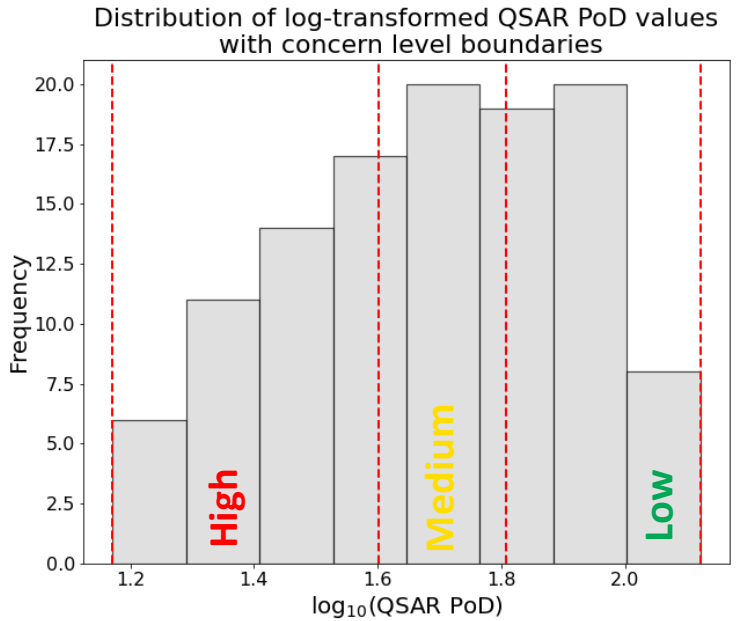
TD → Analog-based TD05
TK → Analog-based Css (95th percentile)
Log10 → tertiles → High/Medium/Low

Concern Category Assignment Framework

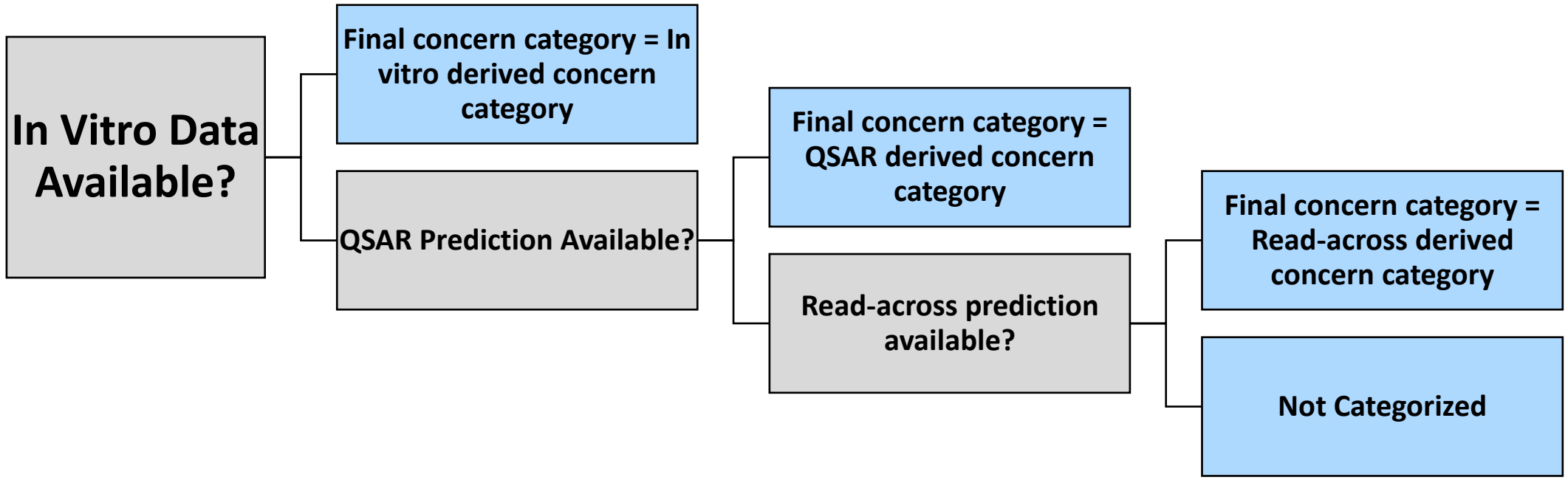
Assuming an equal number of reference chemicals in each concern category, the distribution of the calculated hazard values was used to set the range for potency (cut-off values: 33rd percentile, 66th percentile) for each level



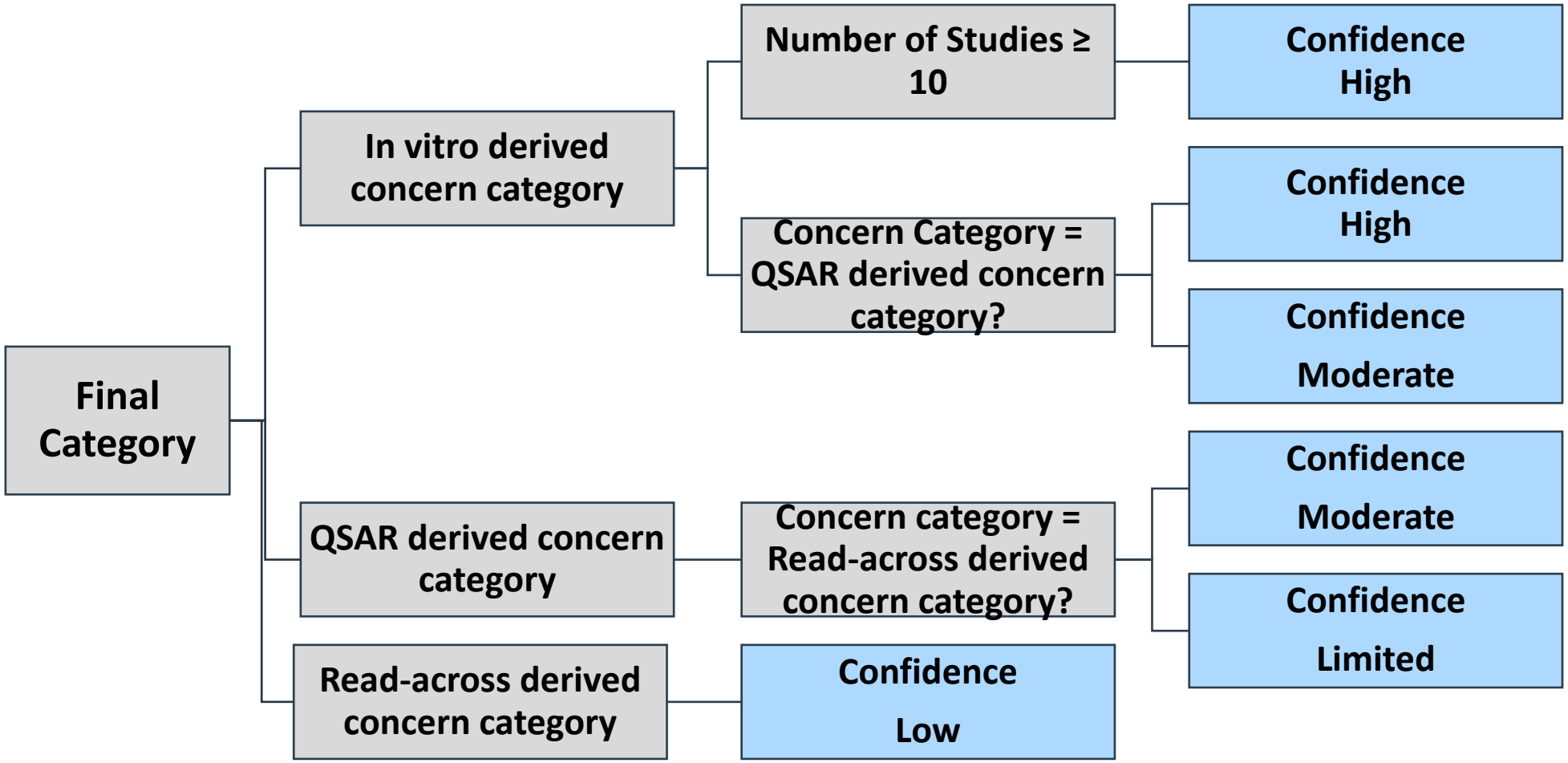




Concern Category Integration Framework



Confidence Assignment Framework



Concern Category Distribution and Concordance Analysis

Assessments	Count (Percentage)
Total number of chemicals assessed	123/150 (82%)
No. of chemicals assessed in the 1st line of evidence (In vitro)	83/150 (55%)
No. of chemicals assessed in the 2nd line of evidence (QSAR)	115/150 (77%)
No. of chemicals assessed in the 3rd line of evidence (Read-across)	13/67 (19%)
No. of chemicals assessed in both, 1st and 2nd lines of evidence	75/150 (50%)
Concern categories agree between the 1st and the 2nd line of evidence	28/76 (37%)
No. of chemicals categorized as high concern	46/123 (37%)
No. of chemicals categorized as medium concern	40/123 (33%)
No. of chemicals categorized as low concern	37/123 (30%)
Concern category matches with current C&L classification	39/123 (32%)
Concern category less or more protective	Less: 37/123 (30%) More: 43/123 (35%)

Summary

- The proposed framework provides a structured, NAM-based screening/categorization approach that enables systematic prioritization of chemicals

- It integrates in vitro bioactivity, IVIVE-derived data, QSAR predictions, and read-across within a transparent decision logic and confidence framework

However,

- Not all reference chemicals could be categorized due to limitations in availability of relevant data

- While there are technical limitations (regarding tools) that can be addressed in the future, the relevance of the data needs to be demonstrated

- Improvements are needed for various aspects both on the TD and TK data/tools

Limitations for NAM-Based Regulatory Screening

Method	Problem	Possible Solution
	Toxicodynamic properties	
In vitro bioactivity data	• Insufficient coverage of toxicological space related to human systemic toxicity • Uncertainty related to the individual assay results with respect to human relevance	• Establishing an AOP network to identify central key events or key characteristics and corresponding *in vitro* bioactivity tests • Developing a standard test battery (IATA/DA) for systemic toxicity as a first-tier testing strategy
	Insufficient coverage for metabolites	• Using metabolically competent primary human cell culture
	Concern categories do not provide information on the MoA/toxicological effect/target organ	• Reporting mechanistic information in addition to the potency assessment to allow further assessments, e.g. mixtures
	Limited toxicity/potency spectrum covered	• Including more substances to broaden the toxicity spectrum and refine the categorization
QSAR *in vivo* PoD prediction	• Variability in underlying animal data • Chemical predictions limited by the relevant chemical descriptors	• Developing new models on better-curated data
Read-across	• Lack of relevant analogues • Limited data availability for relevant analogues	

Limitations for NAM-Based Regulatory Screening

Method	Problem	Possible Solution
	Toxicokinetic properties	
PK model	• 1C model may not be equally suitable for all chemicals	• Applying more advanced PBK models
	• Limited coverage of chemical space (chemical-specific (e.g. fu, CLint) information not available for all reference substances)	• Adding *in vitro* TK tests to the standard first-tier testing strategy • Extending the applicability domain of the in silico prediction
Oral absorption profiler	• Limited coverage of chemical space (chemical properties are outside of the threshold values for parametric boundaries; lack of data)	• Improving the prediction (adding more data?)
	• Not all routes of exposure covered	• Including *in vitro/in silico* barrier models (GIT, skin, respiratory tract) in the testing strategy
Overall	• Low regulatory acceptance	• Validating TD and TK method components • Establishing human relevance • Conducting more case studies and uncertainty assessment

THANK YOU

Prachi Pradeep, PhD
prachi.pradeep@bfr.bund.de

ACKNOWLEDGMENTS

Sebastian Schmeisser, PhD
Ralph Pirow, PhD
Matthias Herzler, PhD

Department of Chemical and Product Safety
German Federal Institute for Risk Assessment (BfR)
Berlin, Germany

Verbraucherschutz zum Mitnehmen

BfR2GO – das Wissenschaftsmagazin des BfR

bfr.bund.de/de/wissenschaftsmagazin_bfr2go.html

Folgen Sie uns

- @bfrde | @bfren | @Bf3R_centre
- @bfrde
- youtube.com/@bfr_bund
- social.bund.de/@bfr
- linkedin.com/company/bundesinstitut-f-r-risikobewertung
- soundcloud.com/risikobewertung