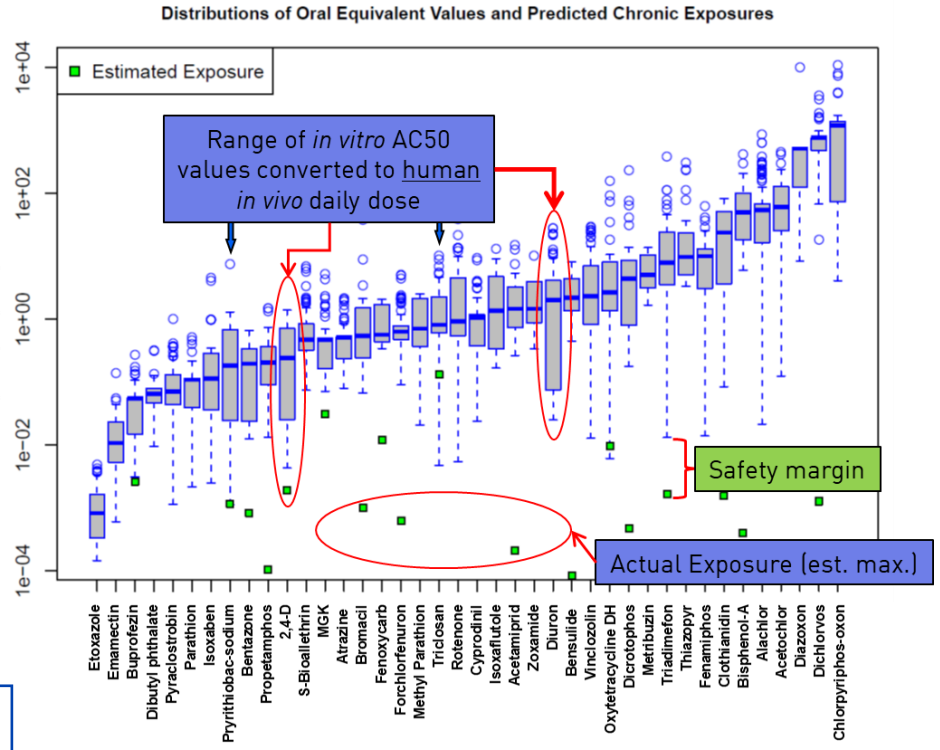
Evaluating a systemic safety toolbox for use in Next Generation Risk Assessment

Dr Alistair Middleton
Science leader in Computational Toxicology
Unilever Safety & Environmental Assurance Centre (SEAC)

Middleton et al (2022), *Tox Sci,* Volume 189, Issue 1, Pages 124-147

Safety without animal testing - Next Generation Risk Assessment (NGRA)

*NGRA is defined as **an exposure-led, hypothesis-driven** risk assessment approach that **integrates New Approach Methodologies (NAMs)** to assure **safety without the use of animal testing***

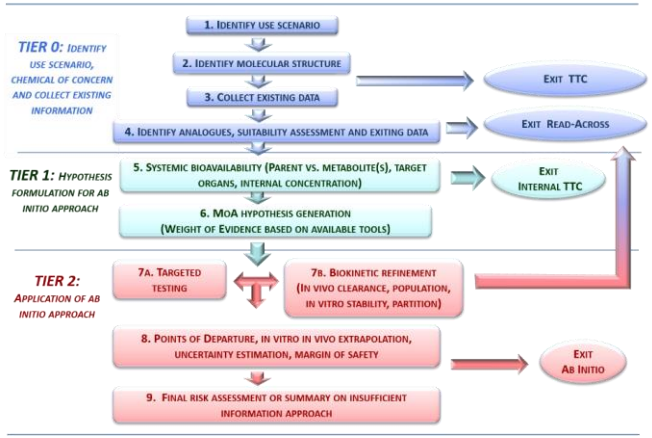


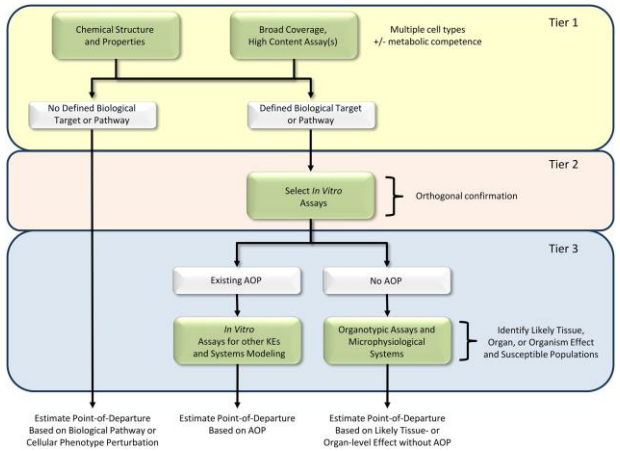
Slide from Dr Rusty Thomas, EPA, with thanks

Rotroff, *et al.* Tox.Sci 2010

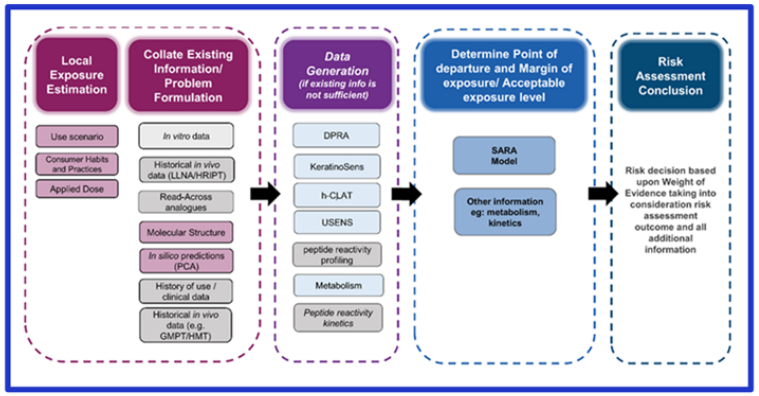
The hypothesis underpinning this type of NGRA is that **if there is no bioactivity observed at consumer-relevant concentrations, there can be no adverse health effects.**

Decision frameworks in NGRA



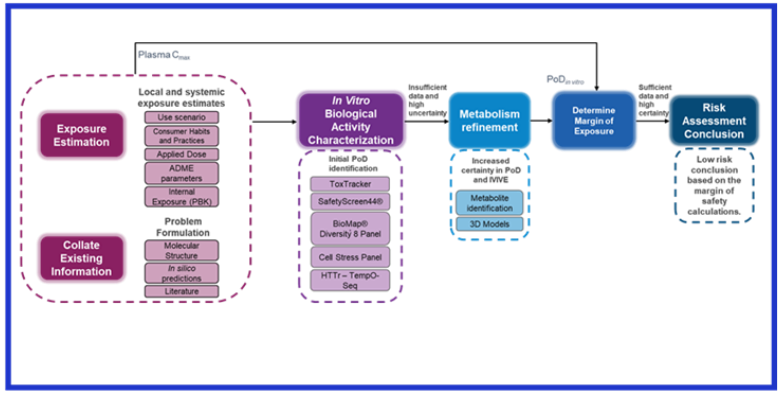


Skin Sensitisation



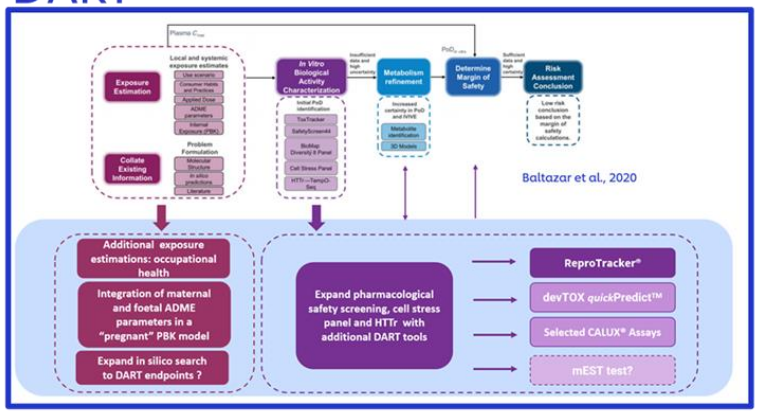
Reynolds et al (2021) Reg Tox Pharmacol, **127**, 105075

Systemic safety



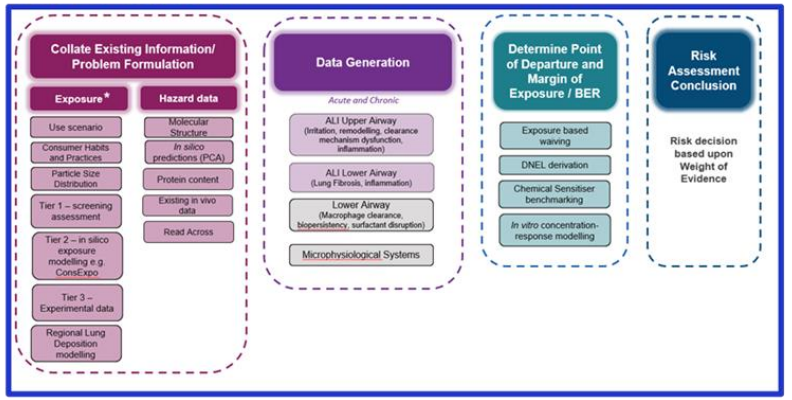
Baltazar *et al.*, (2020) *Tox Sci* , Volume 176, Issue 1, Pages 236–252

DART



Rajagopal et al (2022). Front. Toxicol., 07 March 2022

Inhalation



Unilever

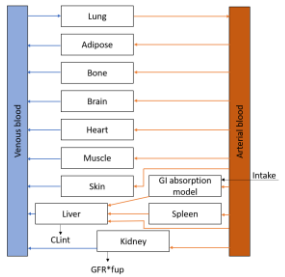
Decision frameworks in NGRA

Problem formulation – Tier 0

Initial BER estimate – Tier 1

Use case

Exposure route

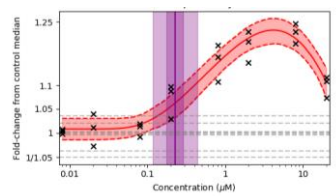


Exposure

Assay 1

Assay 2

Assay 3



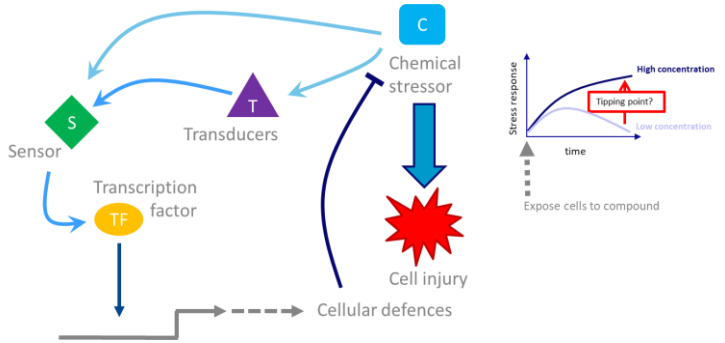
Hazard

- Internal exposure levels estimated using PBK models
- Set of assays covering broad range of biological effects & PODs.
- Key output: Bioactivity Exposure Ratio (**BER=POD/Exposure**)

Uncertain risk: exposure may trigger bioactivity

Small BER (i.e., Exposure close to POD)

BER refinement – Tier 2



- Refinement of internal exposure estimates:
  - Using clinical data
- Refined understanding of biological effects:
  - Repeat dose dynamics
  - Distinguishing bioactivity and adversity
  - Microphysiological systems

Low risk of exposure causing any bioactivity

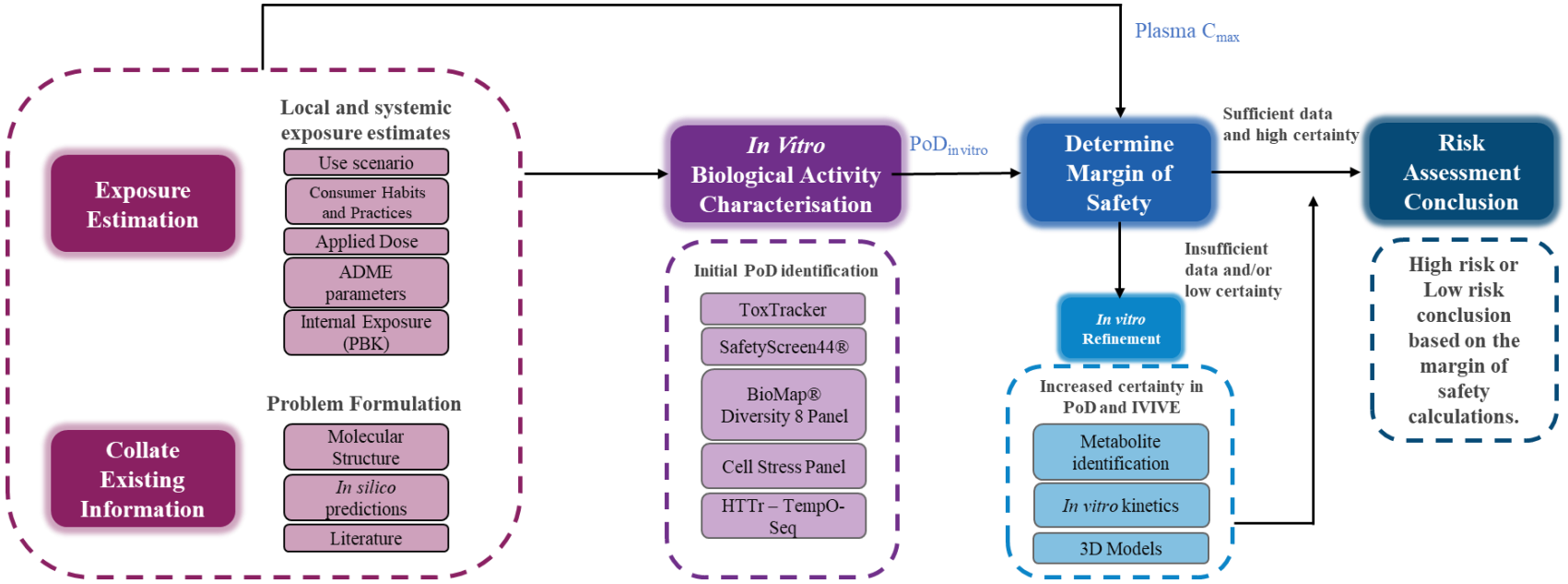
Large BER (i.e., Exposure << POD)

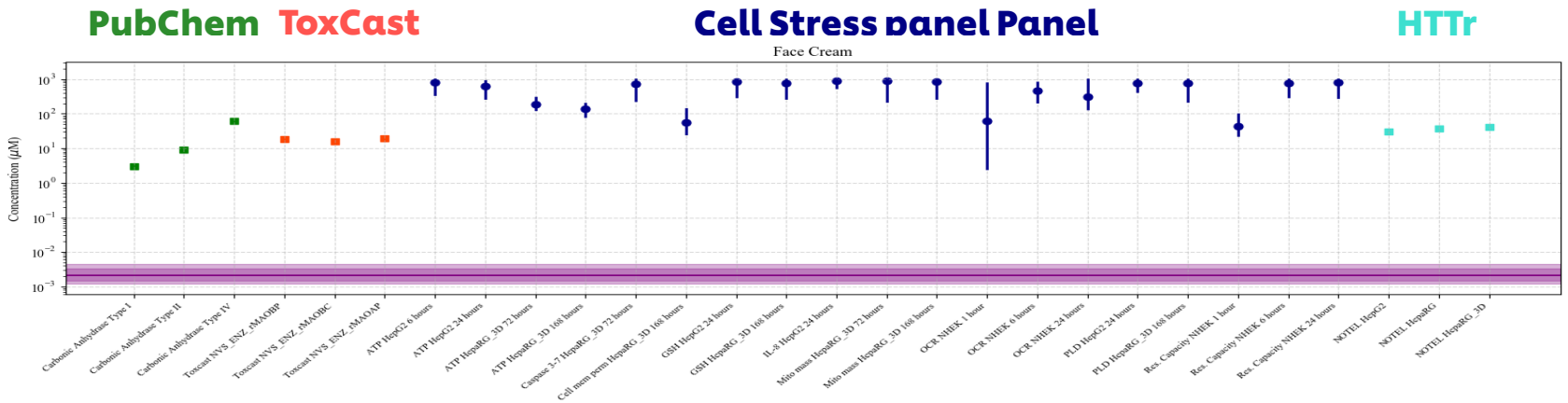
Hypothesis: If a chemical ingredient does not trigger any bioactivity at human relevant exposure levels, then there can be no adverse health effects.

Safety Decision

Gaining confidence in NAMs: first case study with coumarin

For coumarin, a safety assessment based on NAMs was at least as protective as the risk assessment based on traditional approaches

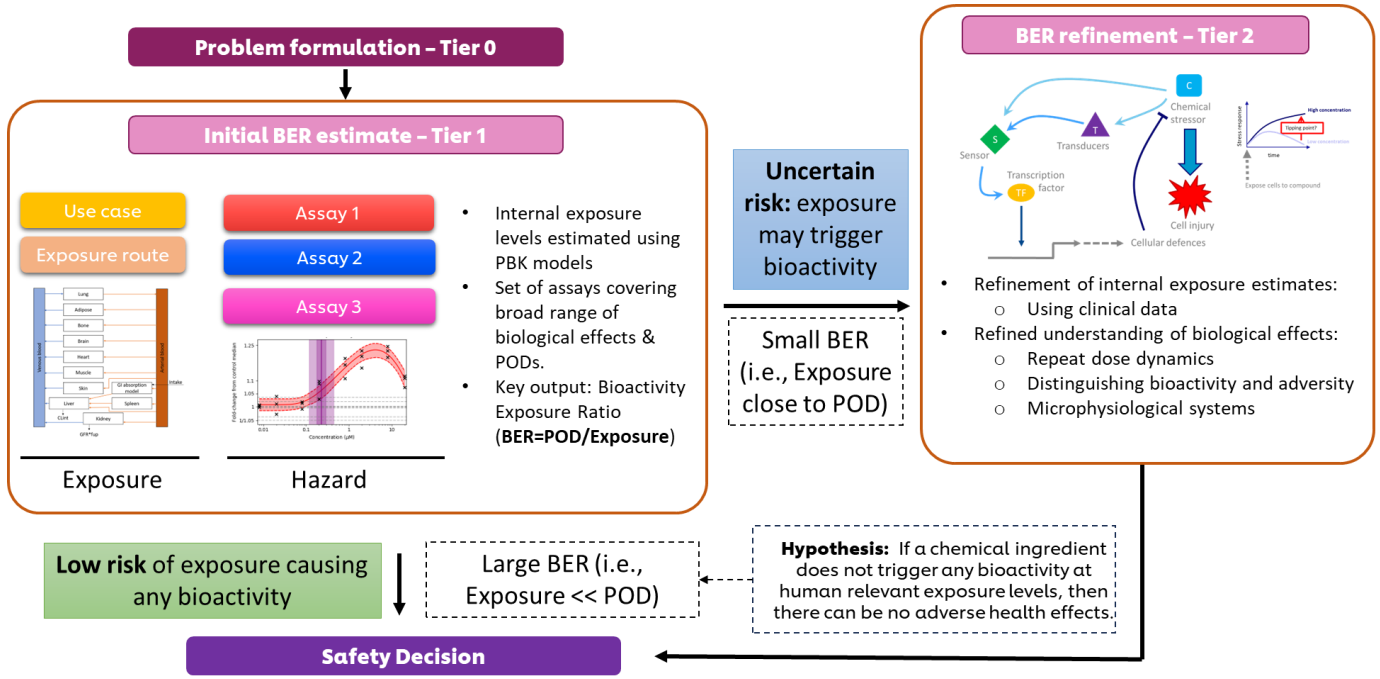




Baltazar *et al.*, (2020) *Tox Sci* Volume 176, Issue 1, 236–252

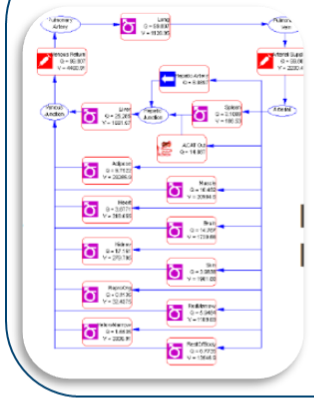
Building and evaluating a systemic safety toolbox

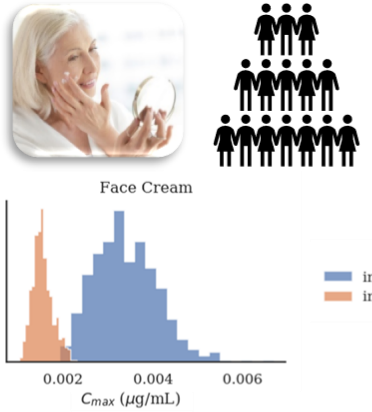
1. Focus the tools and workflows used to make decisions at Tier 1 of an NGRA framework. Includes both:
   - *In vitro* cell assays
   - Exposure models
2. Decisions that can be made with the toolbox are either that a given exposure level is **low risk**, or that the exposure scenario is of **uncertain risk.**
3. In principle, the toolbox will be used as part of a wider tiered assessment framework, which uses e.g., other data through **Tier 0**.



The key NAMs in our NGRA approach

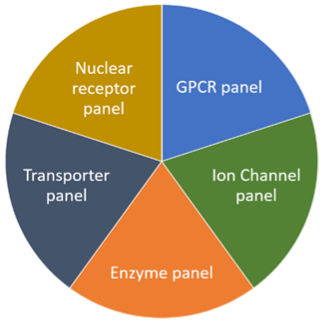
PBK Modelling





Toxicology in Vitro (2020), **63**, 104746

In vitro pharmacological profiling

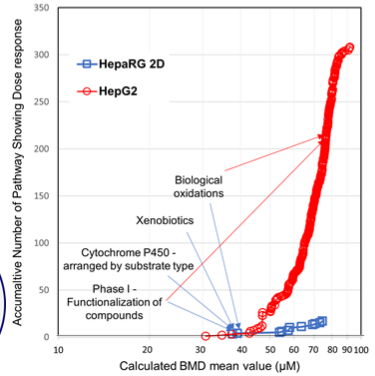


eurofins | Cerep

Transcriptomics

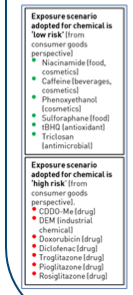
- Use of full human gene panel ~ 21k
- 24 hrs exposure
- 7 concentrations
- 3 cell lines HepG2/ HepaRG/ MCF7
- 3D HepaRG spheroid

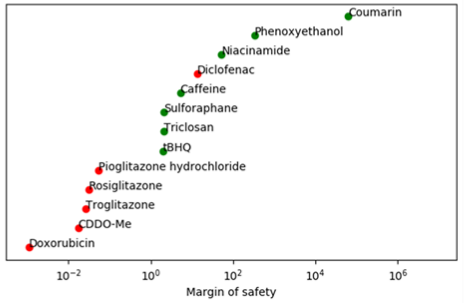
BIFROST
BMDexpress 2



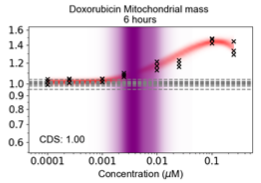
Cellular Stress Pathways

13 chemicals, 36 Biomarkers; 3 Timepoints; 8 Concentrations; ~10 Stress Pathways





cyprotex AN EVOTEC COMPANY



Toxicol Sci (2020), **176**, 11-33

Key aims: 1) select *in vitro* assays that can cover both specific and non-specific mechanisms of toxicity, and 2) can be used to detect early perturbations that may lead to relevant toxicity effects, before the onset of adversity.

Unilever

How do we build scientific confidence in the systemic safety toolbox?

1. Determine whether the toolbox is fit for purpose
   - Can the toolbox be used to make safety decisions that are protective of human health?
   - Do the various assays and cell types used in the toolbox provide sufficient biological coverage?
   - Are the PBK exposure estimates sufficiently accurate?
2. When evaluating the toolbox, take into account all relevant safety data in assessing the approach:
   - Where available/possible, take into account human safety data.
   - Consider both chronic and acute exposure scenarios
   - Ensure we are protective for a broad range of systemic toxicities.
3. Identify an appropriate safety decision model
   - For example, setting a threshold value on the bioactivity exposure ratio.

How do we build scientific confidence in a systemic safety toolbox?

1. Determine whether the toolbox is fit for purpose
   - Can the toolbox be used to make safety decisions that are protective of human health?
   - Do the various assays and cell types used in the toolbox provide sufficient biological coverage?
   - Are the PBK exposure estimates sufficiently accurate?
2. When evaluating the toolbox, take into account all relevant safety data in assessing the approach:
   - Where available/possible, take into account human safety data.
   - Consider both chronic and acute exposure scenarios
   - Ensure we are protective for a broad range of systemic toxicities.
3. Identify an appropriate safety decision model
   - For example, setting a threshold value on the bioactivity exposure ratio.

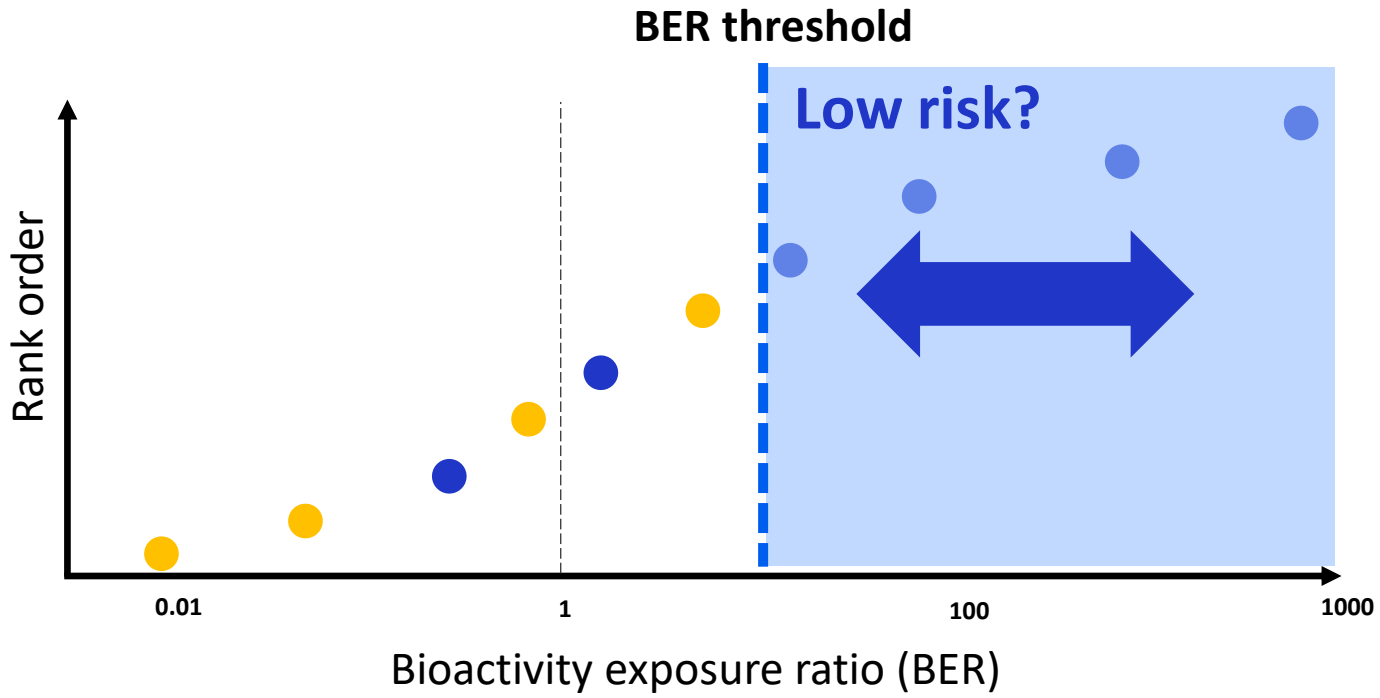
How do we build scientific confidence in a systemic safety toolbox?

1. Determine whether the toolbox is fit for purpose
   - Can the toolbox be used to make safety decisions that are protective of human health?
   - Do the various assays and cell types used in the toolbox provide sufficient biological coverage?
   - Are the PBK exposure estimates sufficiently accurate?
2. When evaluating the toolbox, take into account all relevant safety data in assessing the approach:
   - Where available/possible, take into account human safety data.
   - Consider both chronic and acute exposure scenarios
   - Ensure we are protective for a broad range of systemic toxicities
3. Identify an appropriate safety decision model
   - For example, setting a threshold value on the bioactivity exposure ratio.

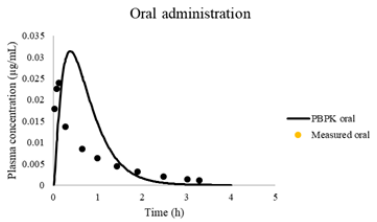
How do we build scientific confidence in a systemic safety toolbox?

Chemical exposures scenarios

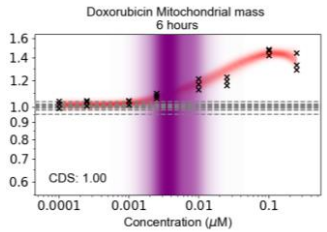
- ‘Low’ risk (from consumer goods perspective) – e.g. foods, cosmetics
- ‘High’ risk (from consumer goods perspective) – e.g. drugs



Define typical use-case scenarios benchmark chemical-exposures; Mixture of High and low risk



PBK models of systemic exposure



In-vitro cell assays, estimate PoDs

Calculate the bioactivity exposure ratio

Unilever

Overall evaluation strategy

Step 1 (pilot study)*

- Define what the toolbox contains (which NAMs) and the workflow through which they should be used.
- Define process of how the toolbox will be evaluated, and the metrics that will be used to determine its 'performance'
- Explore using a small set of chemicals and exposure scenarios (11 chemicals, 25 exposure scenarios)
- Define **prototype decision model** for determining the BER threshold.

Step 2 (extended evaluation)

- Evaluate the toolbox using ~40 chemicals with ~100 exposure scenarios based on the toolbox established in the pilot study.
- Use learnings from the toolbox evaluation to refine the toolbox in terms of NAM composition and the decision model.

Unilever

*Middleton et al (2022), *Tox Sci,* Volume 189, Issue 1, Pages 124-147

Overall evaluation strategy

Step 1 (pilot study)*

- Define what the toolbox contains (which NAMs) and the workflow through which they should be used.
- Define process of how the toolbox will be evaluated, and the metrics that will be used to determine it's 'performance'
- Explore using a small set of chemicals and exposure scenarios (11 chemicals, 25 exposure scenarios)
- Define **prototype decision model** for determining the BER threshold.

Step 2 (extended evaluation)

- Evaluate the toolbox using ~40 chemicals with ~100 exposure scenarios based on the toolbox established in the pilot study.
- Use learnings from the toolbox evaluation to refine the toolbox in terms of NAM composition and the decision model.

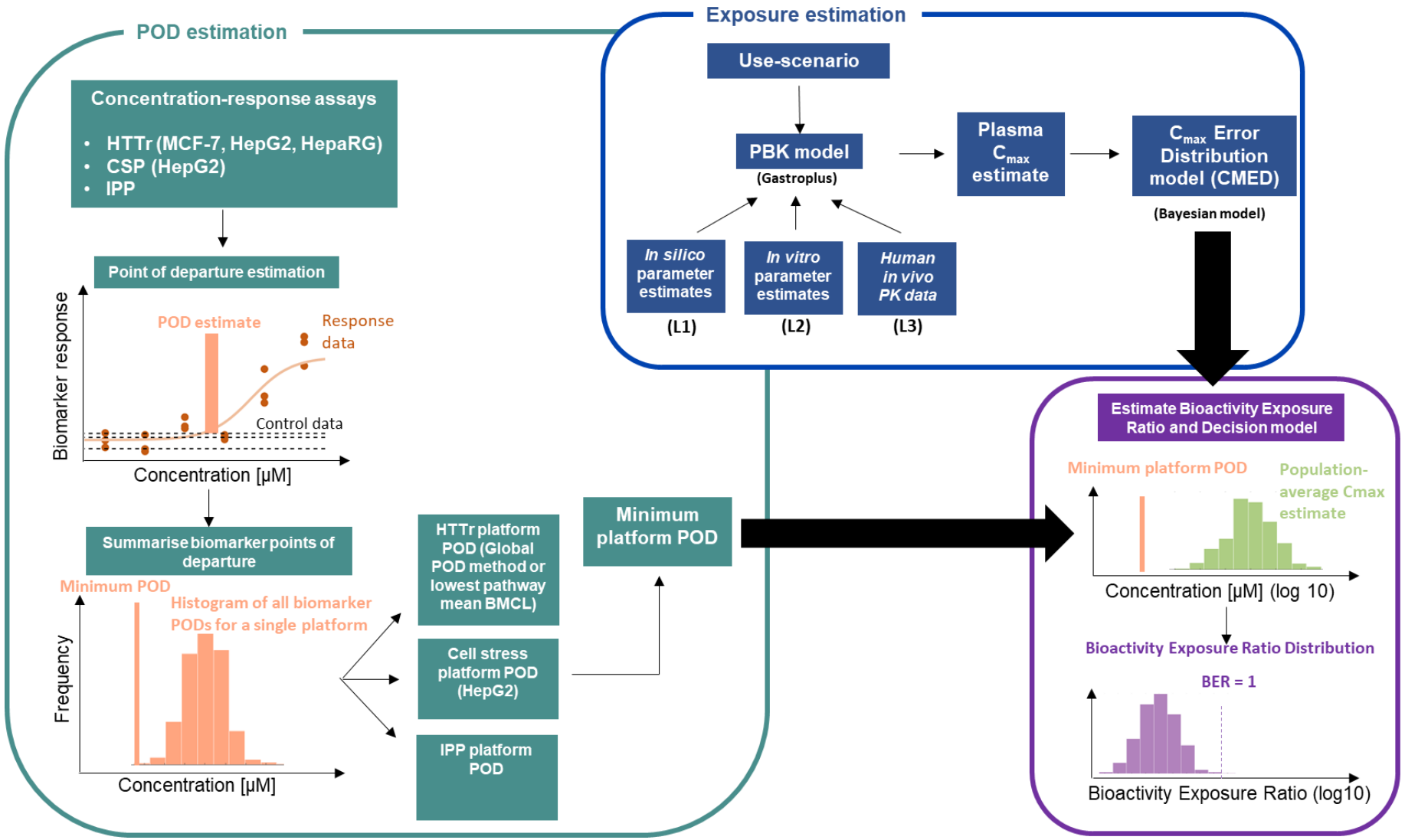
*Middleton et al (2022), *Tox Sci,* Volume 189, Issue 1, Pages 124-147

Stage 1: defining the benchmark chemical exposure scenarios

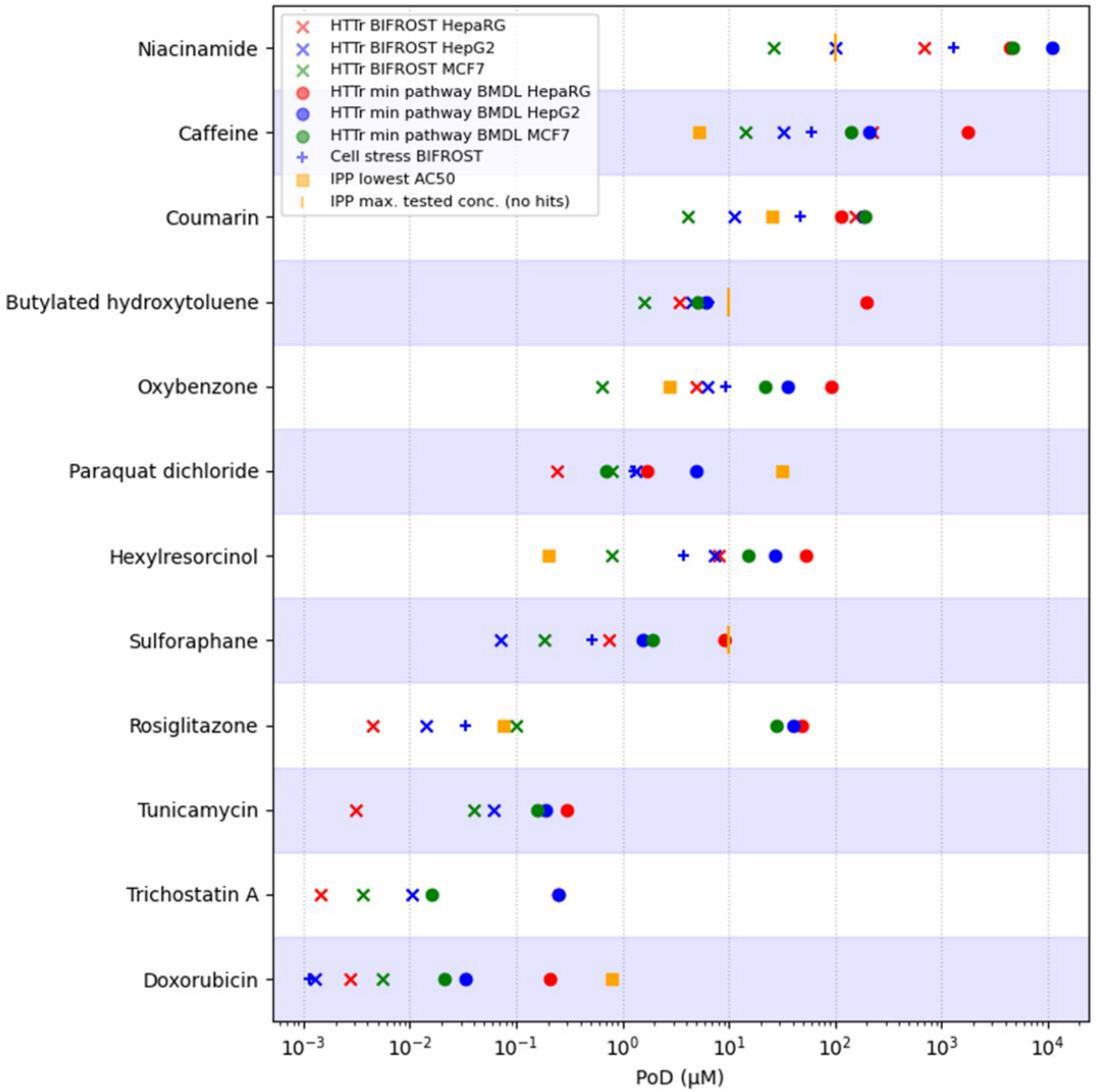
Chemical	Exposure scenario	Risk classification
Oxybenzone	**2 scenarios:** 0.5%; 2% sunscreen	Low risk
Caffeine	**2 scenarios:** 0.2% shampoo & coffee oral consumption 50 mg	Low risk
Caffeine	10g – fatal case reports	High risk
Coumarin	**3 scenarios:** 4 mg/d oral consumption; 1.6% body lotion (dermal); TDI 0.1 mg/kg oral	Low risk
Hexylresorcinol	**3 scenarios:** Food residues (3.3 ug/kg); 0.4% face cream; throat lozenge 2.4 mg	Low risk
BHT	Body lotion 0.5%	Low risk
Sulforaphane	**2 scenarios:** Tablet 60 mg/day; food 4.1-9.2 mg/day	Low risk
Niacinamide	**4 scenarios:** oral 12.5-22 mg/kg; dermal 3% body lotion and 0.1 % hair condition	Low risk
Doxorubicin	75 mg/m2 IV bolus 10 min; 21 days cycles; 8 cycles	High risk
Rosiglitazone	8 mg oral tablet	High risk
Valproic Acid (VPA)	**2 scenarios:** oral tablet 1000 mg & > 60 mg/kg	High risk
Paraquat	Accidental ingestion 35 mg/kg	High risk

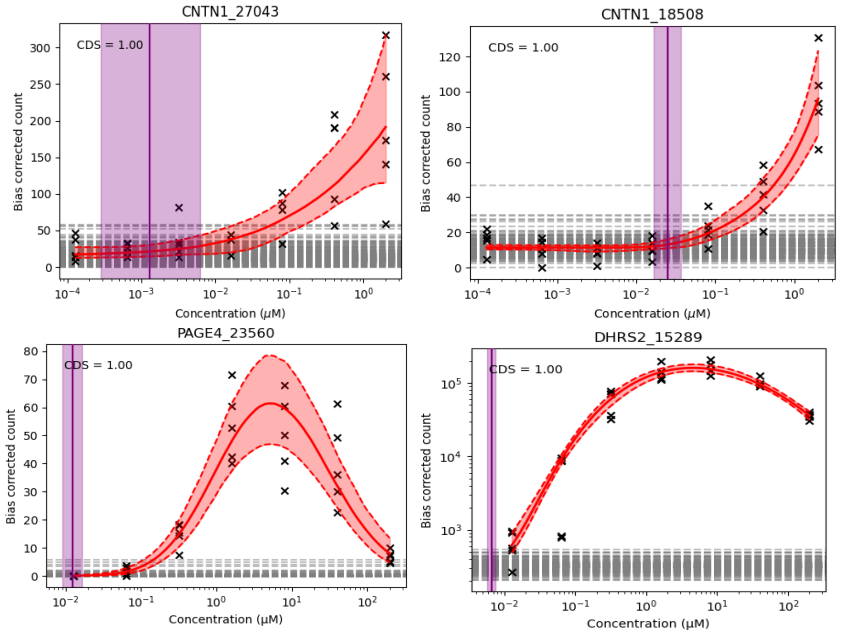
Unilever

The systemic toolbox workflow to estimating a BER



POD estimation



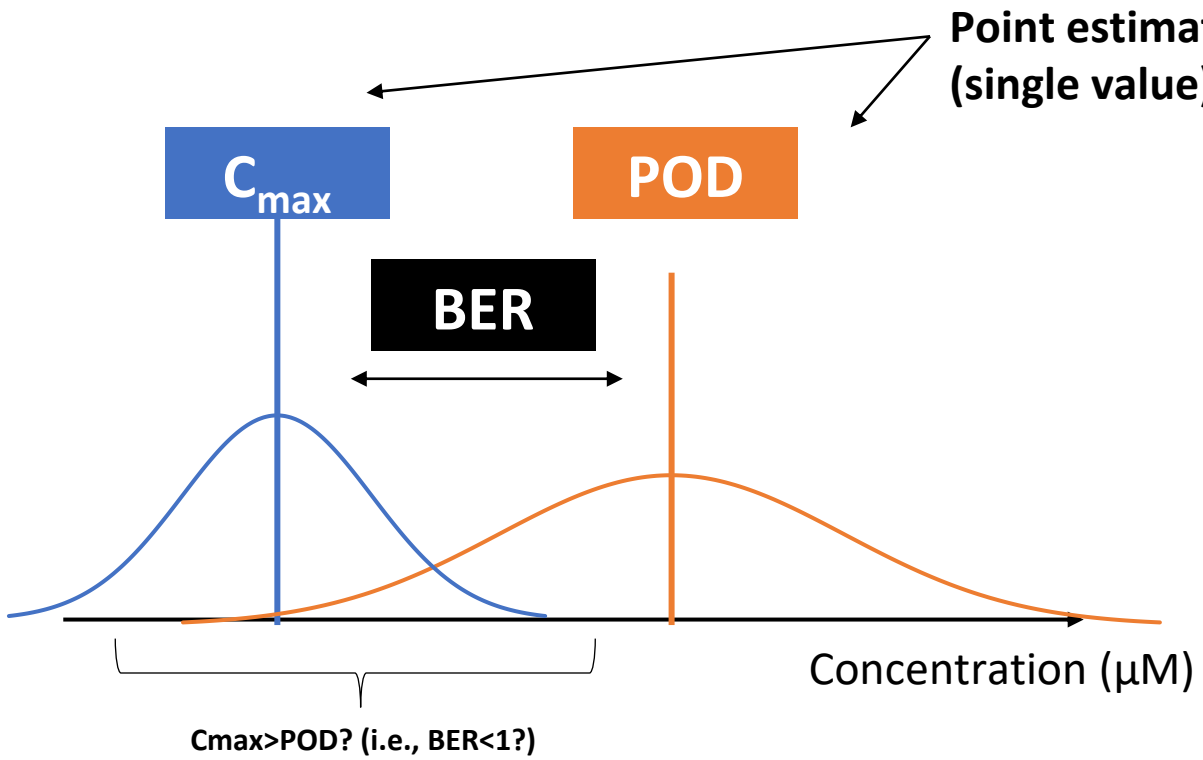


- For 8/10 of compounds tested in the pilot study, HTTr provided the most conservative (lowest POD) when basing the POD on individual genes.
- Pathway based PODs from the HTTr data were typically much higher.

Uncertainty quantification and decision making

Why do we care about quantifying uncertainty?

- Using the point estimates, Cmax appears to be below the POD.
- The true values of both metrics are subject to uncertainty.
- These uncertainties can be captured in terms of distributions.
- The distributions show the range of plausible values for the Cmax and POD.
- Quantifying uncertainty in quantities like Cmax and the POD can be helpful to determine when a safety decision can be made with confidence, or when more refinement is needed.

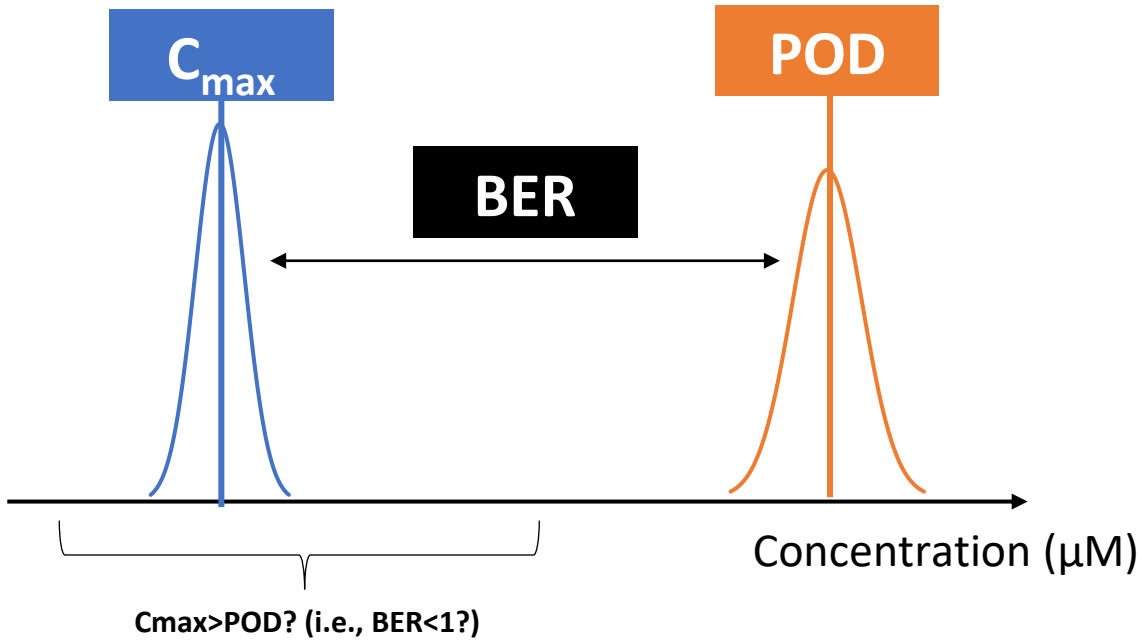


$$\text{Prob}(\text{BER} > 1)=?$$

Uncertainty quantification and decision making

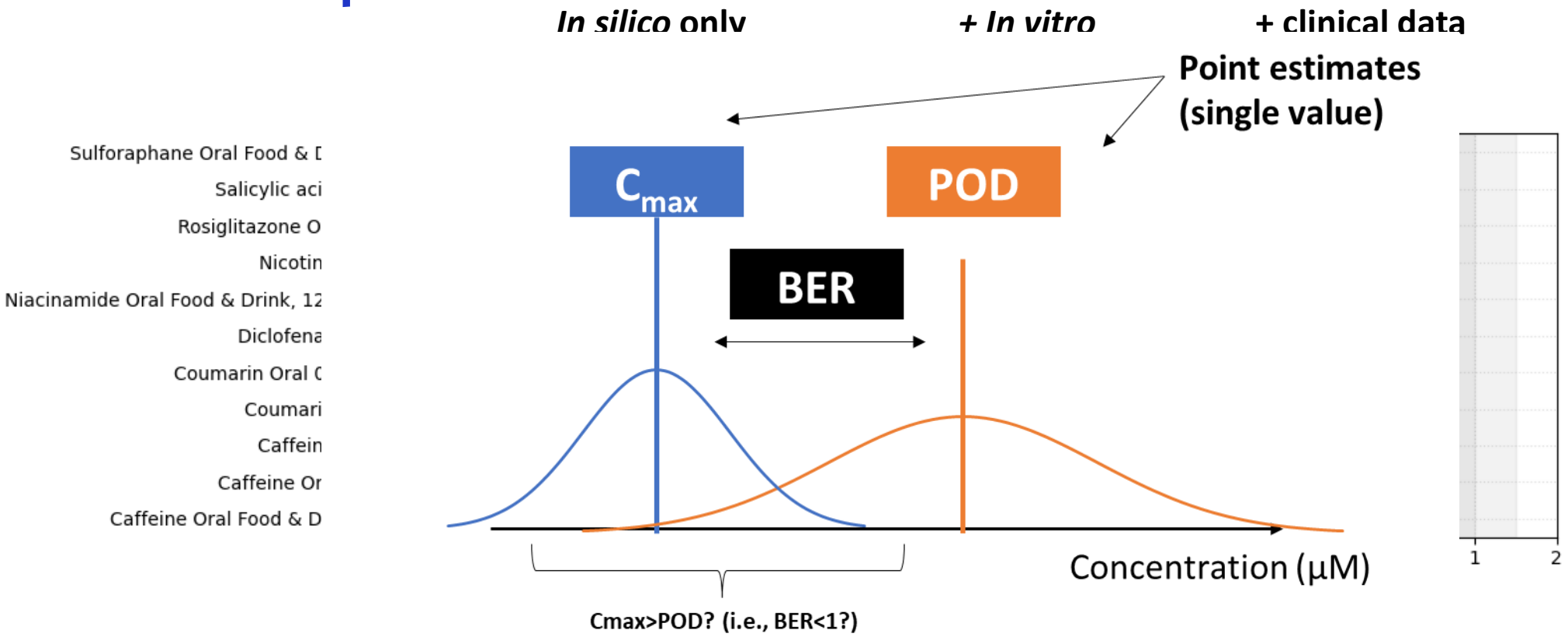
Why do we care about quantifying uncertainty?

- Using the point estimates, Cmax appears to be below the POD.
- The true values of both metrics are subject to uncertainty.
- These uncertainties can be captured in terms of distributions.
- The distributions show the range of plausible values for the Cmax and POD.
- Quantifying uncertainty in quantities like Cmax and the POD can be helpful to determine when a safety decision can be made with confidence, or when more refinement is needed.



$$\text{Prob}(\text{BER} > 1)=?$$

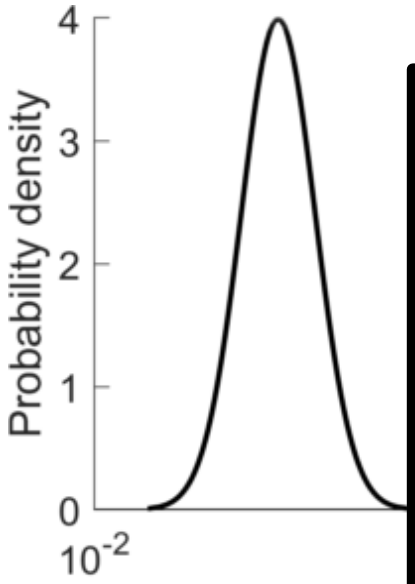
Quantifying PBK model accuracy and uncertainty for different chemical exposure scenarios



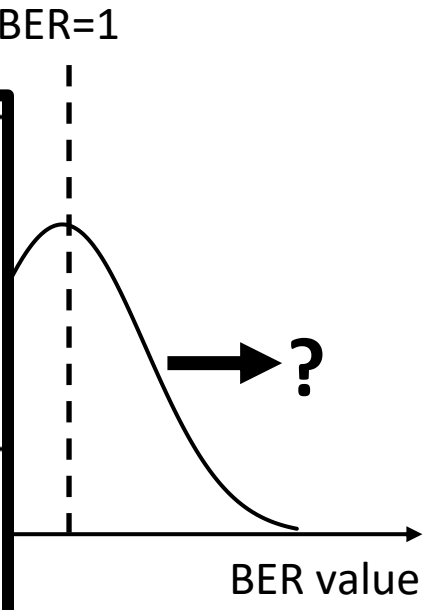
- The accuracy of PBK model Cmax estimates can be quantified by comparing the predicted Cmax value to measured values for different clinical datasets.
- The Cmax Error Distribution (CMED) model was developed using these data to quantify the uncertainty in a PBK Cmax prediction novel substance or exposure scenario, depending on how the PBK model had been parameterised.

Estimating the Bioactivity Exposure Ratio distribution

$\text{Prob}(\text{BER} > 1) \sim 0$

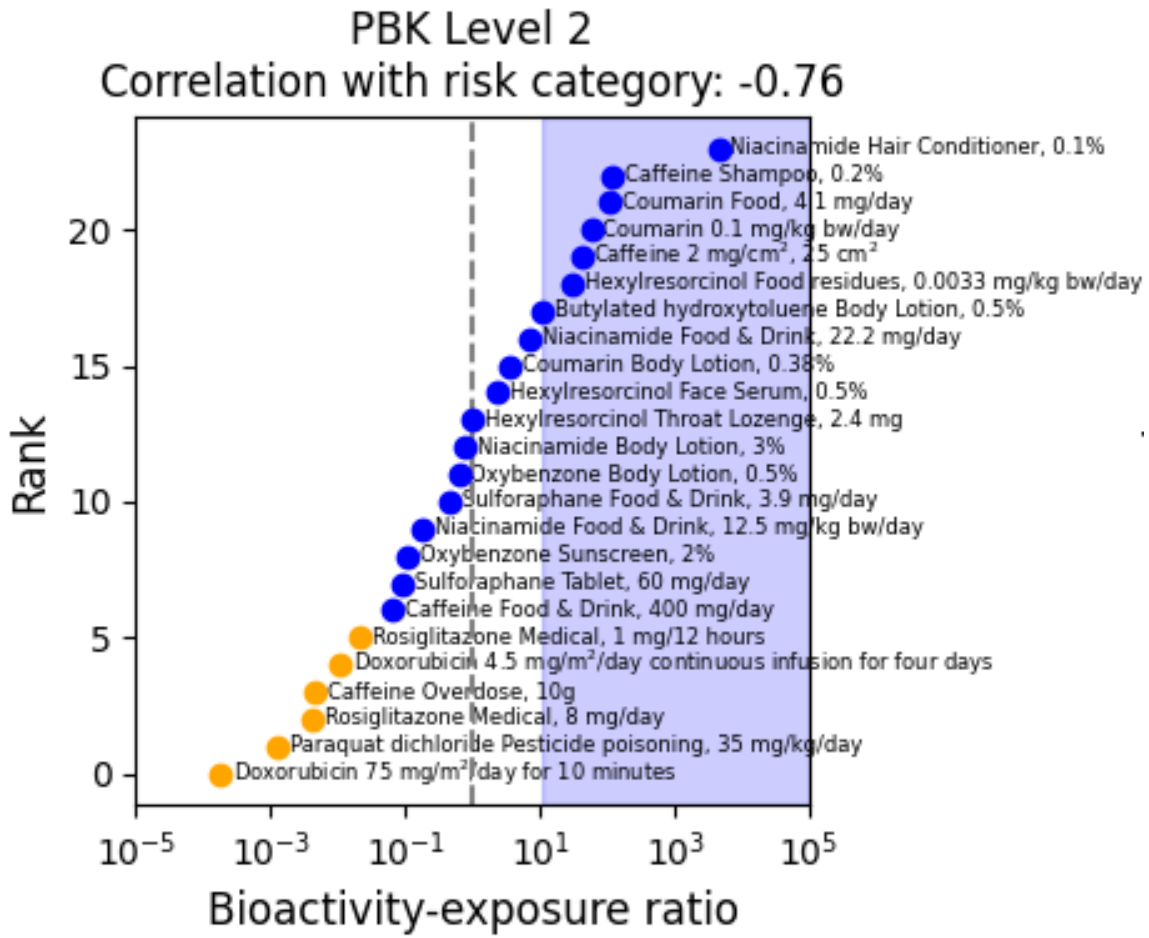


PBK Level	Threshold BER Required for Exposure to Be Identified as Low Risk	Confidence Threshold ($p_{threshold}$) Required for Exposure Scenario to Be Identified as Low Risk
1	110	.98
2	11	.97
3	2.5	.95



- The distribution representing uncertainty C_{max} estimate can be combined with the minimum PODs to form a single BER distribution. (Currently this distribution does not take into account POD uncertainty).
- The minimum POD was selected in order to ensure safety decisions are sufficiently conservative.

Systemic safety toolbox pilot study results: 100% protective for all PBK levels



Blue: low risk chemical-exposure scenario

Yellow: high risk chemical-exposure scenario

Exposure scenarios within the **blue shaded region** are identified as **low risk**

Across the various PBK parameterisation levels:

- **100% protective** (i.e., 100% of all high-risk exposure scenarios were correctly identified as not low risk)
- **Up to 69% utility** (i.e., 69% of all low-risk exposures were correctly identified as low risk).

Overall evaluation strategy

Step 1 (pilot study)*

- Define what the toolbox contains (which NAMs) and the workflow through which they should be used.
- Define process of how the toolbox will be evaluated, and the metrics that will be used to determine its 'performance'
- Explore using a small set of chemicals and exposure scenarios (11 chemicals, 25 exposure scenarios)
- Define **prototype decision model** for determining the BER threshold.

Step 2 (extended evaluation)

- Evaluate the toolbox using ~40 chemicals with ~100 exposure scenarios based on the toolbox established in the pilot study.
- Use learnings from the toolbox evaluation to refine the toolbox in terms of NAM composition and the decision model.

Unilever

*Middleton et al (2022), *Tox Sci,* Volume 189, Issue 1, Pages 124-147

Overall evaluation strategy

Step 1 (pilot study)*

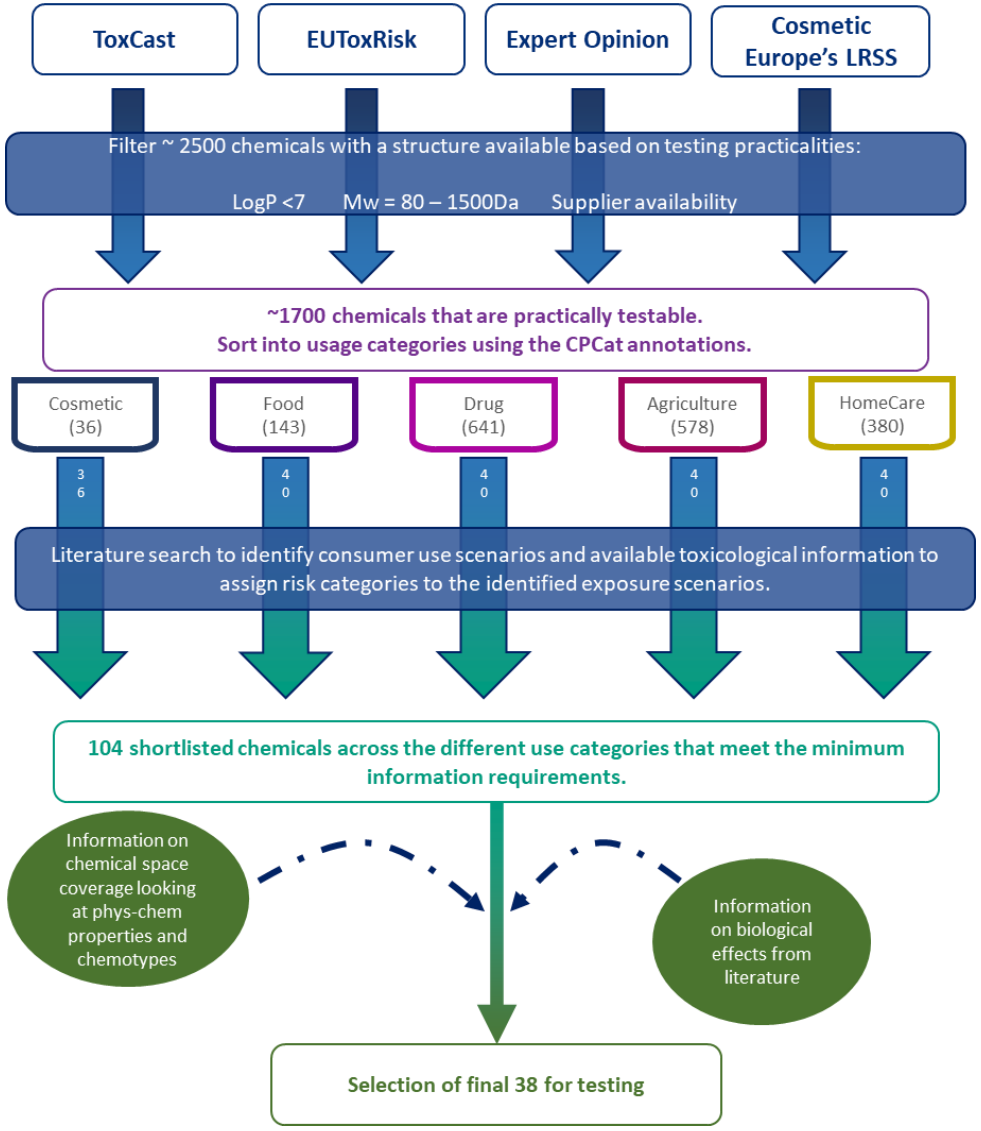
- Define what the toolbox contains (which NAMs) and the workflow through which they should be used.
- Define process of how the toolbox will be evaluated, and the metrics that will be used to determine it's 'performance'
- Explore using a small set of chemicals and exposure scenarios (11 chemicals, 25 exposure scenarios)
- Define **prototype decision model** for determining the BER threshold.

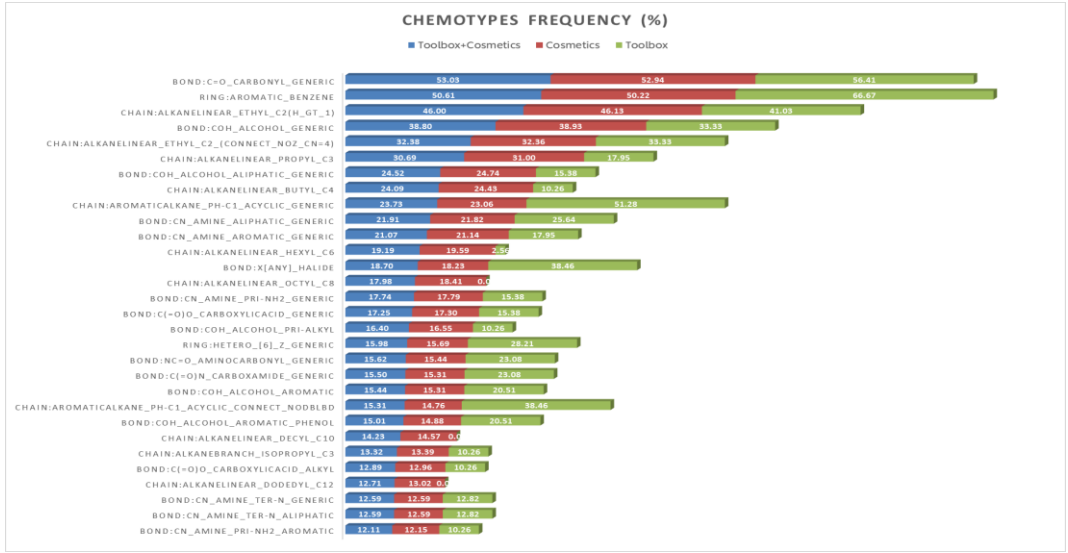
Step 2 (extended evaluation)

- Evaluate the toolbox using ~40 chemicals with ~100 exposure scenarios based on the toolbox established in the pilot study.
- Use learnings from the toolbox evaluation to refine the toolbox in terms of NAM composition and the decision model.

*Middleton et al (2022), *Tox Sci,* Volume 189, Issue 1, Pages 124-147

Expanding the set of benchmark chemical exposure scenarios

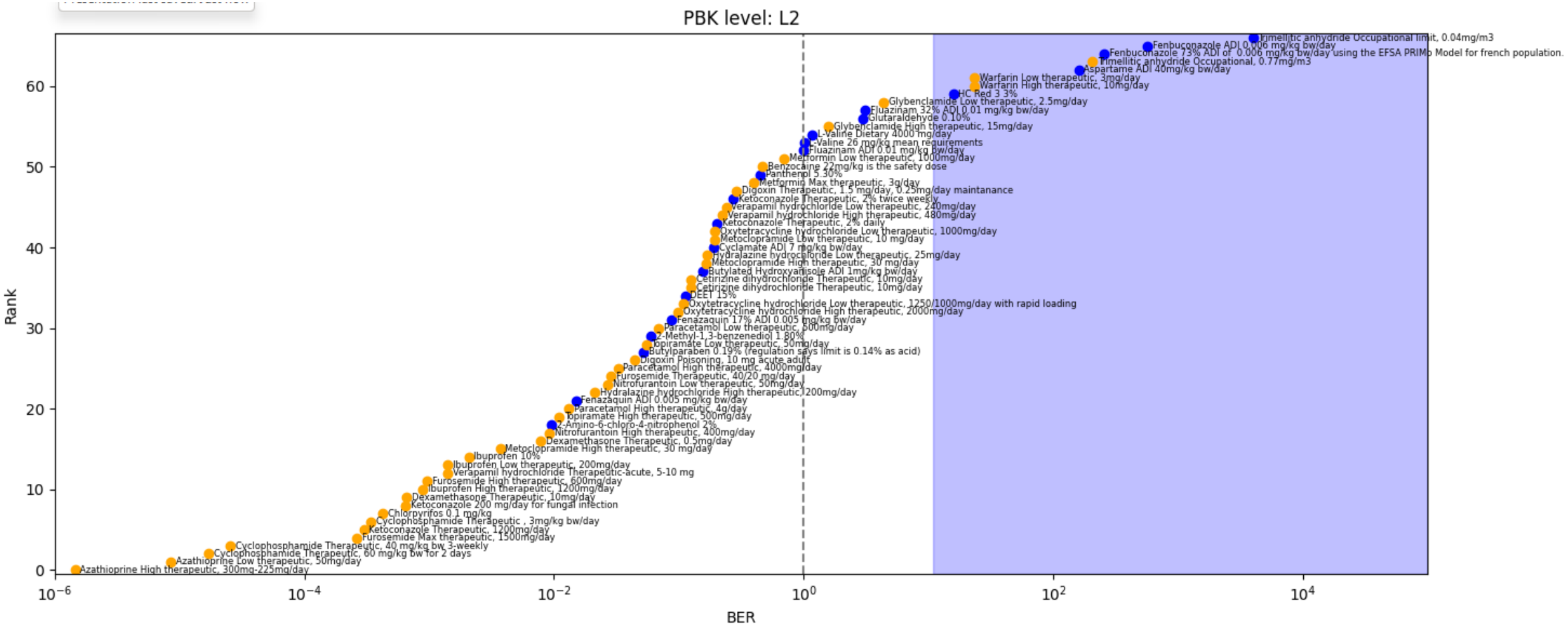




- Manual chemical or exposure scenario selection may result in strong biases.
- Therefore, for the extended evaluation, benchmarks were selected using a semi-randomised processed.
- The final set of benchmarks represented a wide range of different potencies, chemotypes and potential toxicity mechanisms.

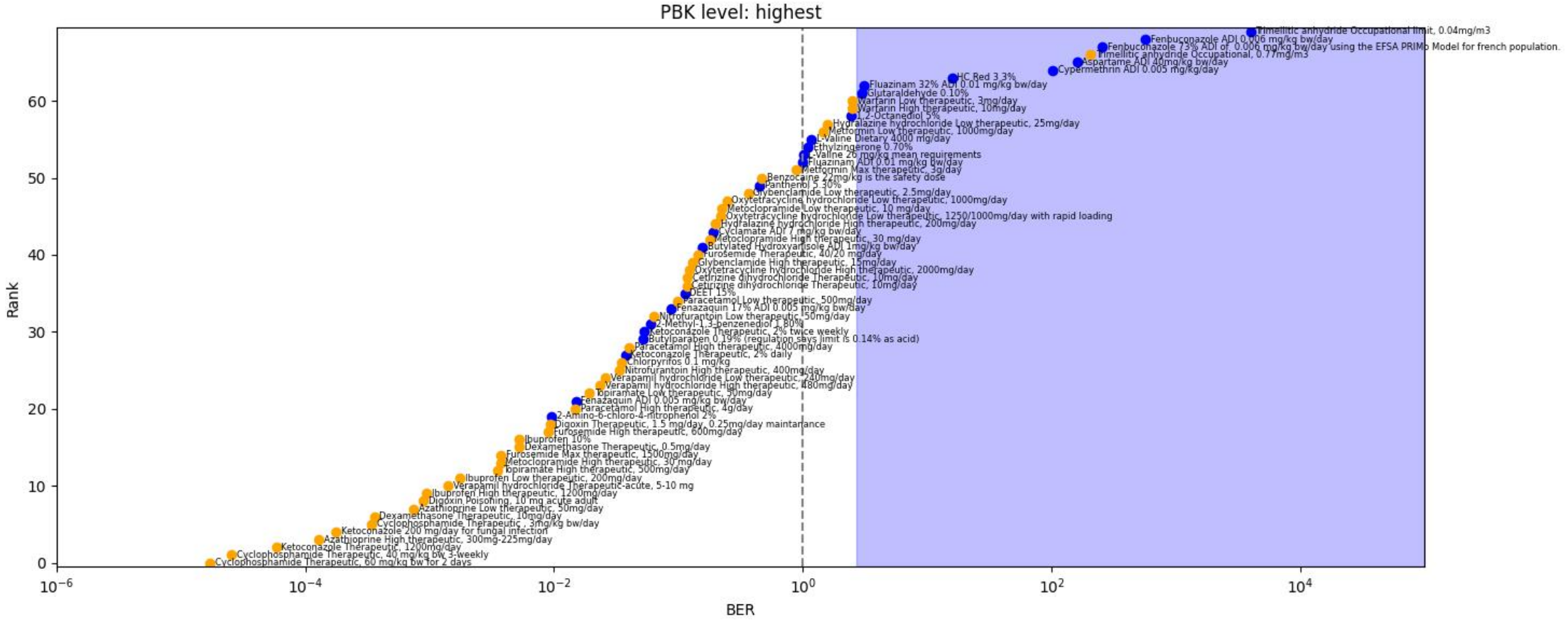
Unilever

Toolbox performance: PBK L2 exposure estimates



Protectiveness: 93% (43 out of 46)
Utility: 24% (5 out of 21)
Balanced accuracy: 59%

Toolbox performance: Highest available PBK level

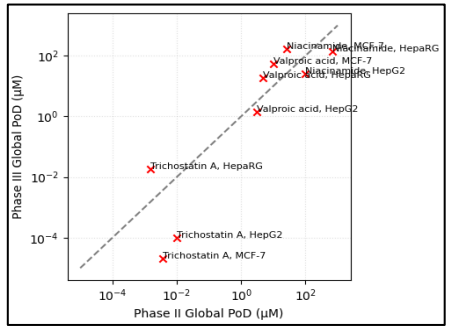


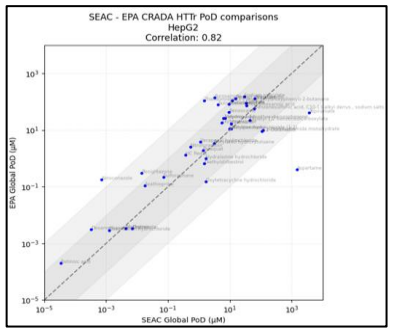
Protectiveness: 98% (45 out of 46)
Utility: 33% (8 out of 24)
Balanced accuracy: 66%

Discussion and next steps

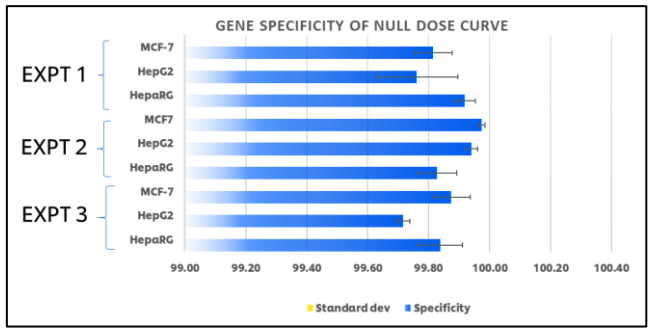
- We have now extended the evaluation to 38 chemicals and 70 exposure scenarios. Protective for 93-98% of scenarios (depending on PBK level).
- **Unilever-EPA CRADA**: Generating data for 10 cell lines, using high-throughput transcriptomics and phenotypic profiling.
- We are continuing to further establishing scientific confidence through a range of activities.

POD reproducibility

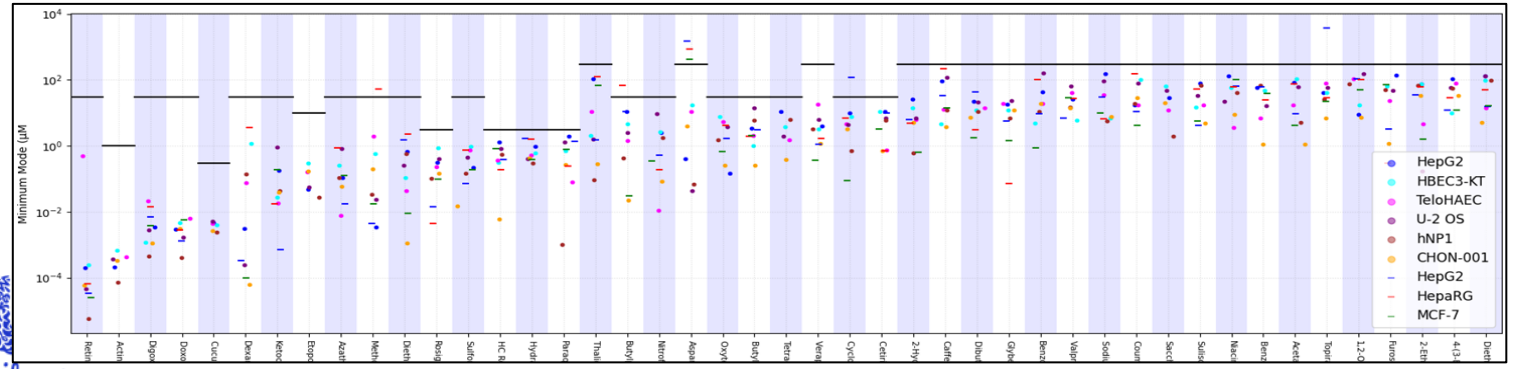


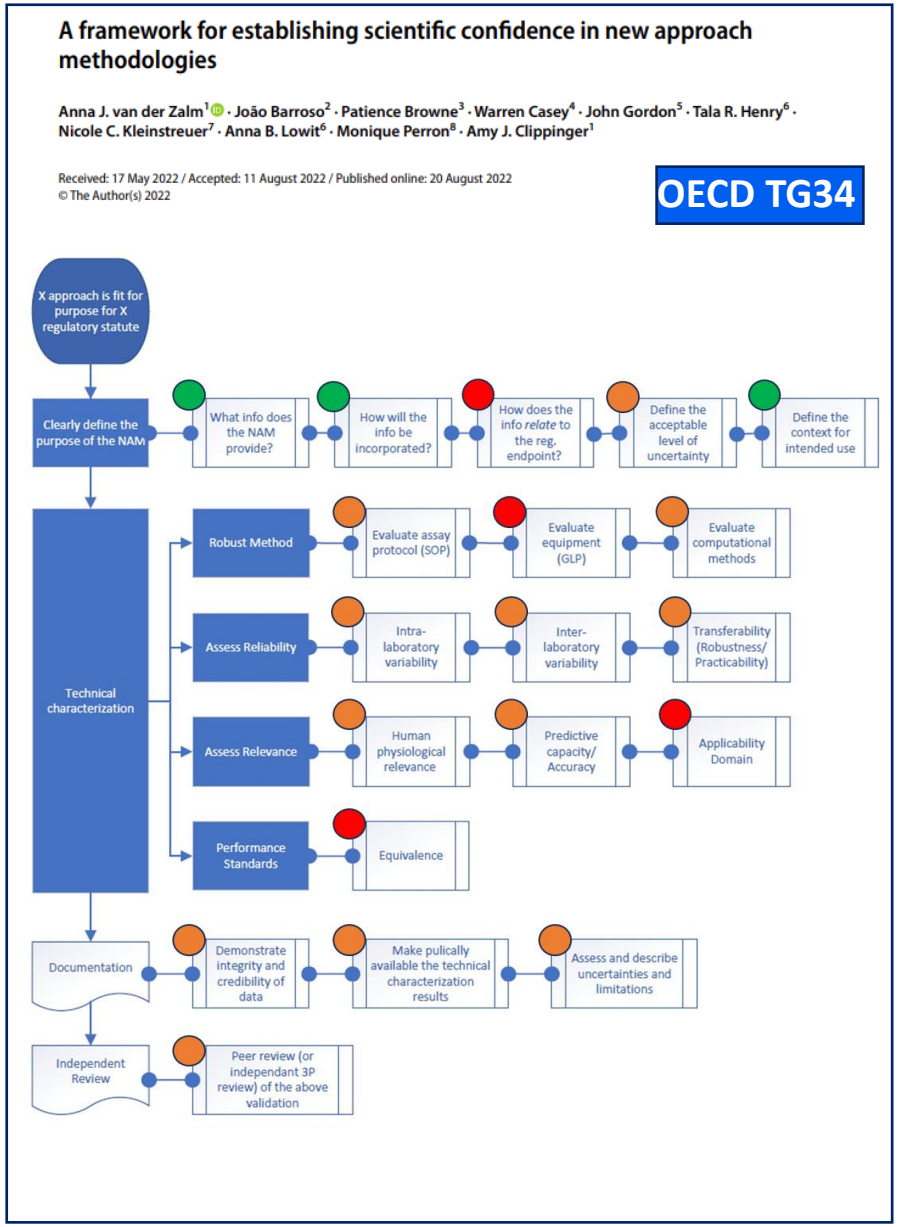


Evaluating different POD methods (sensitivity, specificity etc)



Cell line selection and POD diversity





Acknowledgements

Unilever: Maria Baltazar, Sophie Cable , Joe Reynolds, Georgia Reynolds, Beate Nicol, Sharon Scott, Sophie Malcomber, Annabel Rigarlsford, Katarzyna Przybylak, Predrag Kukic, Dawei Tang, Matthew Dent, Andrew White, Paul Carmichael, Sarah Hatherell, Richard Cubberley, Carl Westmoreland

US-EPA: Richard Judson, Josh Harrell, Logan Everett, Imran Shah, Joseph Bundy, Laura Taylor, Jacob Fredenburg

Thank You

seac.unilever.com

Unilever