Bioactivity: exposure ratios derived from a systemic NAM-toolbox distinguish between low and high-risk chemical-exposure scenarios

Sophie Cable[1], **Maria Teresa Baltazar[1]**, Fazila Bunglawala[1], Paul L. Carmichael[1], Leonardo Contreas[1], Matthew Philip Dent[1], Jade Houghton[1], Predrag Kukic[1], Sophie Malcomber[1], Beate Nicol[1], Katarzyna R Pryzbylak[1], Ans Punt[1], Georgia Reynolds[1], Joe Reynolds[1], Sharon Scott[1], Dawei Tang[1], Alistair M Middleton[1]. [1]Safety and Environmental Assurance Centra, Unilever, Colworth, MK44 1lQ

STEP 1: DEFINE TOOLBOX COMPONENTS AND PERFORM PROOF OF PRINCIPLE STUDY

A critical question for risk assessors and regulators is whether safety assessments based on non-animal data can be protective of human health. One important way of establishing scientific confidence in decision making using non-animal methods is through large scale data-driven projects across a broad range of chemistries and biology. Here we show the results of an evaluation activity of a core toolbox of *in vitro* assays and a risk assessment workflow for decision making using benchmark chemical exposure scenarios to interpret the performance of the toolbox and workflow.

The core components of this NAM-based NGRA workflow are:

- **Estimation of internal exposure** using different levels of input parameters to build the physiologically-based kinetic (PBK) models. Plasma Cmax values are estimated for every chemical-exposure scenario using either *in silico* only parameter estimates (L1), *in vitro* parameters from experimental data where available (L2), or calibrated model estimates using human clinical data (L3).
- **Estimation of a bioactivity point of departure (PoD)** was done across 3 different assays set ups consisting of the investigation of 63 specific protein targets (GPCRs, ion channels, enzymes etc.) as well as cellular stress mechanisms and effects on the transcriptome of 3 cell lines (HepG2, HepaRG, MCF7). Bayesian statistical models were built to analyse the cellular stress and transcriptomics data in a concentration-response manner and establish the most likely concentration at which an effect begins, thus determining a bioactivity platform PoD.
- **Calculation of a Bioactivity Exposure Ratio (BER)** combines inputs from the exposure and bioactivity assay modules, calculating the ratio between the plasma C_{max} estimates and the lowest platform PoD.

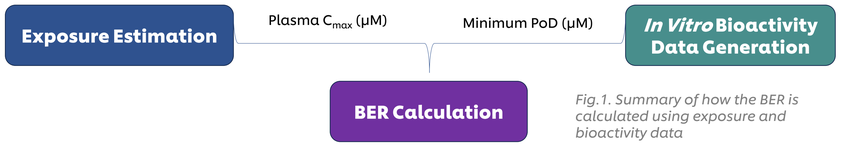


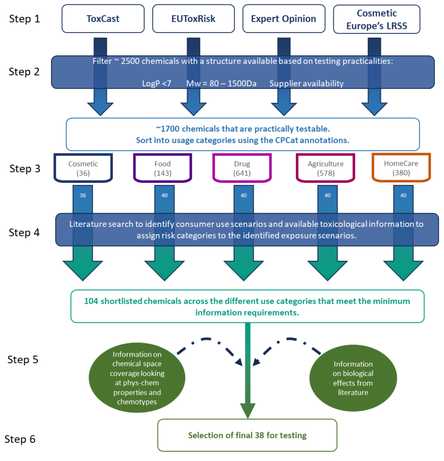
Fig.1. Summary of how the BER is calculated using exposure and bioactivity data

Calculation of a Bioactivity Exposure Ratio (BER) combines inputs from the exposure and bioactivity assay modules, calculating the ratio between the plasma Cmax estimates and the lowest platform PoD.

Conceptually a BER > 1 indicates a low risk of adverse effects in consumers if the following assumptions are true:
1. The *in vitro* measures of bioactivity provide appropriate biological coverage
2. There is confidence that the test systems are at least as sensitive to perturbation as human cells *in vivo*
3. The exposure estimate is conservative for the exposed population

The work from Middleton et al., 2022 identified BER threshold at which all exposure scenarios with a greater BER would be considered low risk, BER above 110, 11, 2.5, depending on the uncertainty associated with the exposure estimation (PBK L1, L2, L3)[1].

STEP 2: SELECT TEST CHEMICALS AND SET PERFORMANCE CRITERIA



Aims:
- **Avoid biasing** the evaluation through selection of only 'extreme' cases, e.g. fatally toxic chemicals and biologically inert chemicals
- Select chemicals covering a **broad range of chemistries and biology**
- Select chemicals with **exposure scenarios for which a risk classification could be assigned** using the available literature.

Fig.2. shows an overview of the chemical selection process, including several filtering steps to remove any chemicals that would be incompatible with the nature of the testing being conducted or for which there wasn't sufficient information available to define an exposure scenario with a defined risk classification.

The final selection of chemicals that met all the criteria included 9 chemicals primarily associated with cosmetic use, 21 primarily associated with medicinal use, 3 associated with food exposures, 5 agricultural chemicals and 1 primarily associated with occupational use.

PERFORMANCE CRITERIA

Calculate protectiveness and utility of toolbox

$$\text{Protectiveness} = \frac{H_U}{H_U + H_L} \qquad \text{Utility} = \frac{L_L}{L_L + L_U}$$

H_U - # of high risk exposures identified as uncertain risk
H_L - # of high risk exposures identified as low risk
L_U - # of low risk exposures identified as uncertain risk
L_L - # of low risk exposures identified as low risk

References
1. Middleton AM, et al (2022) 'Are Non-animal Systemic Safety Assessments Protective? A Toolbox and Workflow', *Toxicological Sciences*, Volume 189, Issue 1, p124–147
2. Cable S et al., (2024). Advancing systemic toxicity risk assessment: Evaluation of a NAM-based toolbox approach. *Toxicological Sciences*, Volume 204, Issue 1, p79–95
3. Dent, M. et al (2018) 'Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients', *Computational Toxicology*, Volume 7, p20-26

STEP 3: EVALUATE THE TOOLBOX

This toolbox and workflow is intended for use in quantitative early-tier risk assessment, where the primary goal is protectiveness: i.e. no classification of high-risk chemical exposure scenarios as low risk. It does this for over 90% of the benchmark chemical exposure scenarios

- There are a total of **8 different PoD types** generated by the systemic-safety toolbox: one associated with receptor profiling (IPP), one with cellular stress (CSP) and two for each of the three HTTr cell lines that were tested (one based on gene level changes and one on pathway level changes). Across the different chemicals tested in this work, **IPP gave the lowest PoD for 11 chemicals**, **CSP gave the lowest PoD for 5 chemicals** and **HTTr (gene level) gave the lowest PoD for 25 chemicals** (8 in HepaRG, 6 in HepG2, 11 in MCF-7).
- BERs were calculated using the lowest PoD across all bioactivity platforms tested and dividing them by the plasma Cmax estimates for each chemical exposure scenario. Fig.3 shows the resulting **BER plot when L2 PBK estimates are used and compared to the previously determined threshold of 11, giving a protectiveness and utility of 93% and 27% respectively[2].**
- **This is comparable to the performance of using traditional in vivo toxicology data for the risk assessment, as demonstrated in Fig.4. Where the NAM PoDs are more conservative than the in vivo PoDs in 22/24 cases (in vivo data were not found for all chemicals tested).**
- **Only the therapeutic doses of warfarin and occupational exposure to Trimellitic anhydride are misclassified as low risk** using this toolbox alone. However, the intended use is within a tiered and iterative framework encompassing all lines of evidence.
  - **Trimellitic anhydride is a known sensitiser**, and it is likely that in a risk assessment framework the risk posed by sensitisation via the inhalation route would limit the exposure below that which poses a systemic risk.
  - ***In vitro* data available for the activity of Warfarin at its target, VKORC1**, would change the risk assessment conclusion with a measured **IC50 giving a BER<<1**.
- It can reasonably be envisaged that **PBK models parameterised with *in vitro* data are the most likely future scenario for a novel risk assessment**, although the performance metrics improve as PBK models can be calibrated against human clinical data. Table 1 shows the resulting protectiveness and utility scores for the different PBK levels.

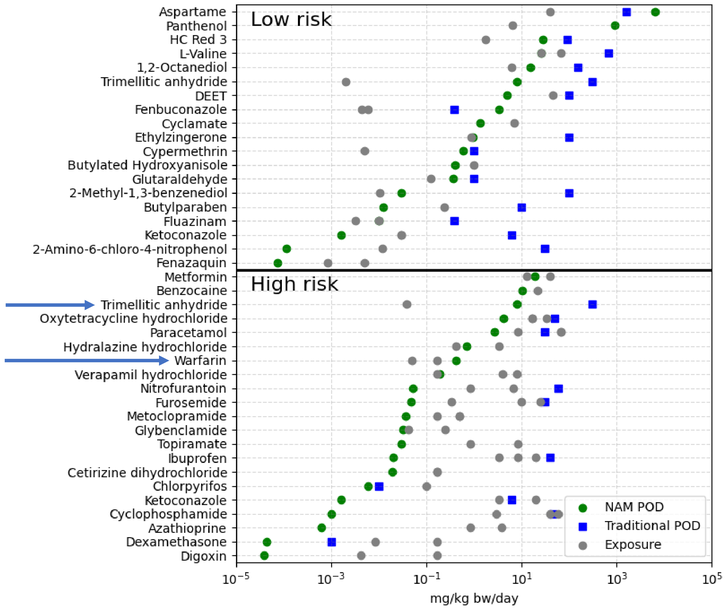


Fig. 4 Summary plot of the external exposure estimates with the converted minimum NAM PoDs and traditional PoDs, separated by the risk classification of the corresponding exposure scenarios. Traditional PoDs are only reported for the 25 chemicals where data were available. Blue arrows highlight the two examples where the Cmax calculated for a high risk chemical exposure scenario is below the NAM PoD

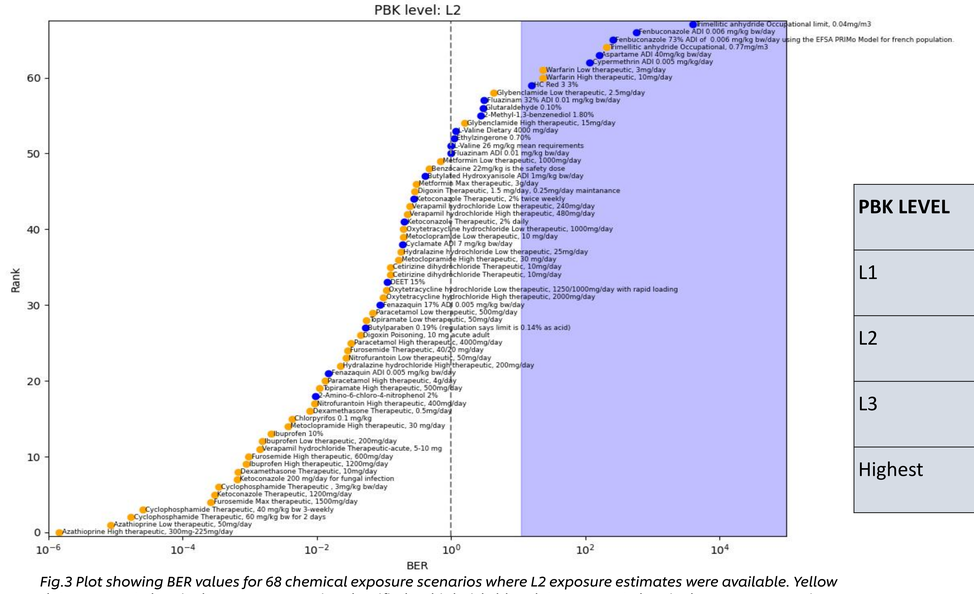


Fig.3 Plot showing BER values for 68 chemical exposure scenarios where L2 exposure estimates were available. Yellow dots represent chemical exposure scenarios classified as high risk, blue dots represent chemical exposure scenarios classified as low risk. The blue shaded region represents a BER of 11 or above, with a BER > 11 being the previously determined threshold for a low risk decision at L2. The black dashed line represents BER = 1.

PBK LEVEL	PROTECTIVENESS	UTILITY
L1	93% (43 out of 46)	8% (2 out of 24)
L2	93% (43 out of 46)	27% (6 out of 22)
L3	98% (40 out of 41)	0% (0 out of 3)
Highest	96% (44 out of 46)	29% (7 out of 24)

CONCLUSIONS

- A NAM-based toolbox can be used to make decisions that **are protective of human health in at least 93% of cases**, despite not predicting the mode of action.
- **The current proposed toolbox is intended to sit within a tiered risk assessment framework** and does not differentiate bioactivity from adversity at this stage. The observed low utility could be addressed by the incorporation of further testing or more detailed interpretation of the Tier 0 and Tier 1 results.
- **More chemicals** should be tested to build the reference database from 38 chemicals and 70 benchmark exposure scenarios to **increase confidence** in the applicability of this approach.

seac.unilever.com