Evaluation of *in vitro* non-animal methods for use in systemic safety decision making

SEAC | Unilever

Andrew White, Maria Baltazar, Joe Reynolds, Beate Nicol, Gopal Pawar, Fazila Bunglawala, Dawei Tang, Sharon Scott, Sophie Malcomber, Sarah Hatherell, Predrag Kukic, Matt Dent, Georgia Reynolds, Paul Carmichael, Alistair Middleton, Sophie Cable[1]

[1] Unilever's Safety and Environmental Assurance Centre, Colworth Science Park, MK44 1LQ

Pilot Study

CONCEPT AND OVERVIEW

The ASPA framework provides a tiered and modular framework for an evidence led non-animal method (NAM) based risk assessment. This development is being guided by case studies to evaluate the logic flows and potential gaps. A critical question arising at decision points in the workflow is whether the information is sufficient. At the end of Tier 1 this question tries to assess whether safety assessments based on non-animal data can "provide technically reliable information that is relevant to the understanding of human biology and health **protective** for the endpoint of concern" (Van der Zalm 2022). To date several attempts have been made to benchmark NAM based approaches against current *in vivo* data to assess their protectiveness as a surrogate for human protection and shown that for the majority of chemicals NAM based approaches have been protective (Paul Friedman et al. 2020; Reardon et al 2023; Zobl et al 2023). Here we present an approach that examines directly whether NAM-based assessments for systemic toxicity can be protective of human health without being overly conservative.

This workflow has been overlayed on the corresponding ASPA modules and follows a similar tiered approach. It consists of three main modules as outlined in Fig.1: Internal exposure (Plasma C_{max}) was estimated using different levels of input parameters to build the physiologically-based kinetic (PBK) models; *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 core assays consisting of binding to 63 specific protein targets (GPCRs, ion channels, enzymes etc.), assessment of cellular stress 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.

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.

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 cell *in vivo*
3. The exposure estimate is conservative for the exposed population

However there has been limited work up to this point to evaluate if this concept holds true in real cases. The results of this pilot study were used to define a threshold for benchmark chemicals at which the BER would be considered low risk.

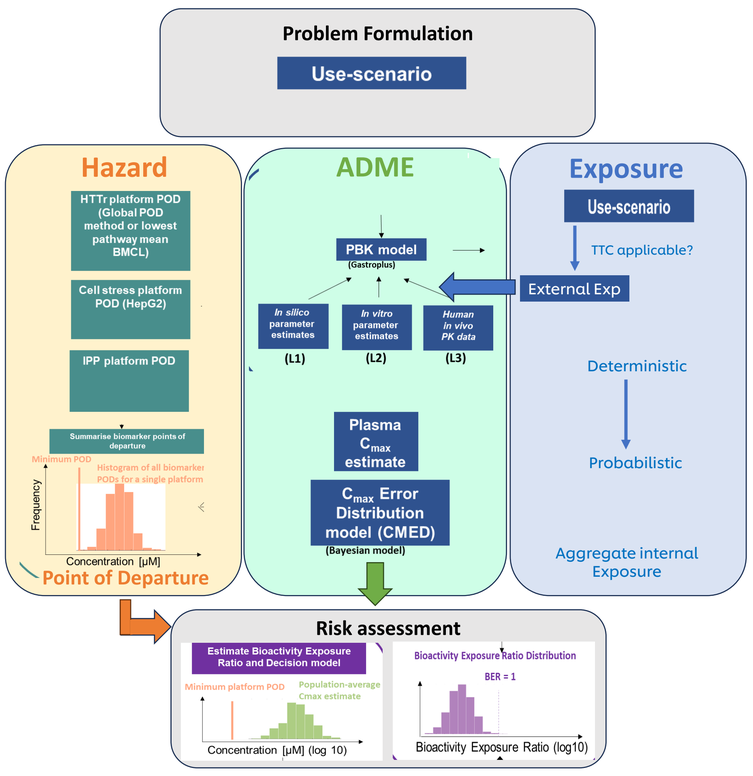

Fig.1. Proposed workflow for integration of exposure and bioactivity data for safety decision making overlayed on ASPA modules

RESULTS – non-animal toolbox 100% protective for high-risk chemical exposure scenarios

A pilot study was conducted using 10 chemicals and 24 benchmark exposure scenarios, with a risk classification defined for each chemical-exposure scenario. This work allowed for optimisation of the test systems and of the data analysis process, but also worked to define a method for conduction of a larger scale evaluation. A BER threshold was determined above which it is likely that the chemical exposure scenario is low risk.

Fig.2. shows the results of this pilot study with high risk exposure scenarios coloured in yellow and low risk exposure scenarios coloured in blue. As expected there is some overlap in the BERs calculated for both high and low risk scenarios but a threshold could be set based on the different inputs above which the likelihood of a scenario being low risk was > 95%. At PBK level 2 (*in vitro* parameter inputs) this threshold is BER > 11, where all exposure scenarios are low risk from a consumer perspective; the thresholds for L1 and L3 are BER > 110 and BER > 2.5 respectively.

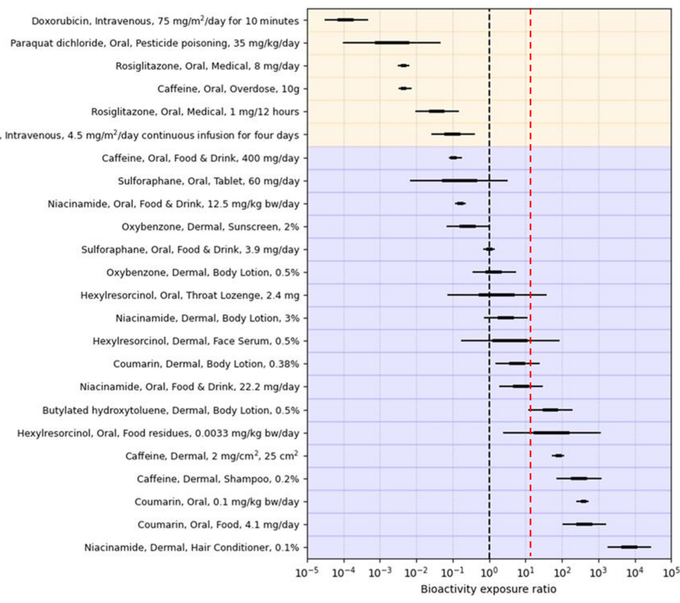

Fig.2. Calculated BER values for 24 chemical exposure scenarios as determined using the modules and workflow shown in Fig.1. High risk chemical exposure scenarios are shown in yellow, low risk chemical exposure scenarios are shown in blue. The bars represent the 95% confidence interval of the calculated BER when considering uncertainty in the exposure estimate. The red dotted is at BER = 11, the black dotted line shows a BER = 1 to visualise the conceptual approach to interpreting the BER values in the context of benchmark chemical exposure scenarios.

Full Evaluation

SELECTION OF TEST CHEMICALS

Aims:

- To avoid biasing the evaluation through selection of only 'extreme' cases, e.g. highly toxic chemicals and biologically inert chemicals
- To select chemicals covering a broad range of chemistries and biology
- To select chemicals with exposure scenarios for which a risk classification for human safety could be assigned using the available literature.

Fig.3. 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.

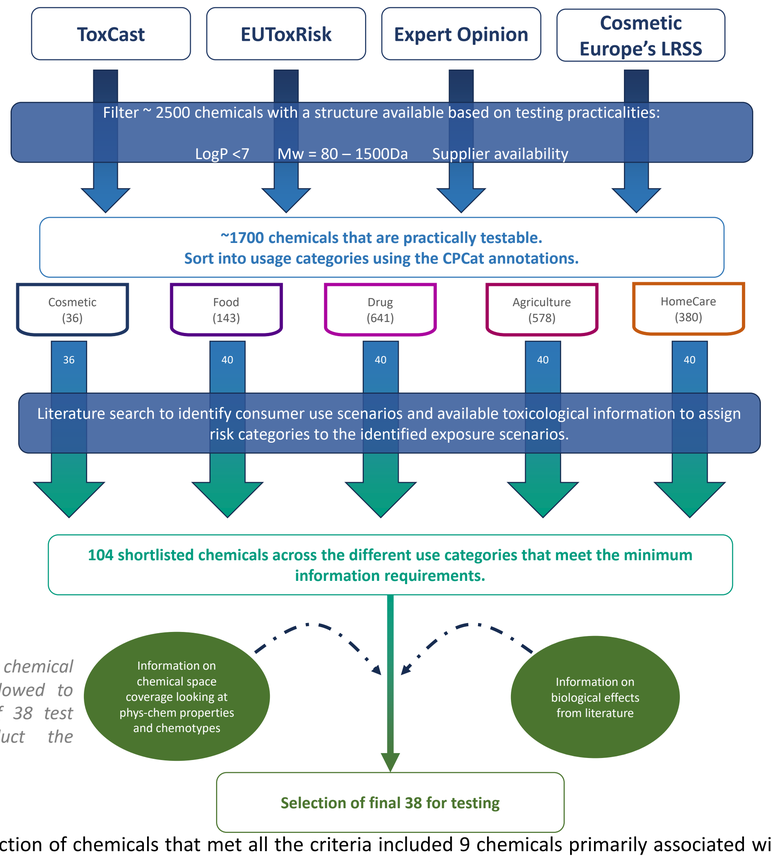

Fig.3 Schematic of the chemical selection process followed to determine the list of 38 test chemicals to conduct the evaluation on.

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. A key question of using low tier, broad screening approaches, such as those that comprise this toolbox, is whether they provide enough coverage to be used for ab initio non-animal risk assessments. One way to look at assessing the coverage provided by this workflow is through mapping the diversity of the chemical and biological space provided by the choice of test chemicals.

The coverage of the chemical space was investigated through characterising the structural diversity of the test chemicals by chemotyping and comparing to the chemotypes present in structures annotated for cosmetic use in the CPCat database. This showed a very similar spread of chemotypes across the reference cosmetic chemicals and our test chemicals.

This was then also visualised in Fig.4, representing each chemical using RDKit[2] descriptors and the UMAP[3] technique. Given the limited number of test chemicals in this evaluation the structural coverage appears to be fairly even across a representative cosmetics chemical structural space.

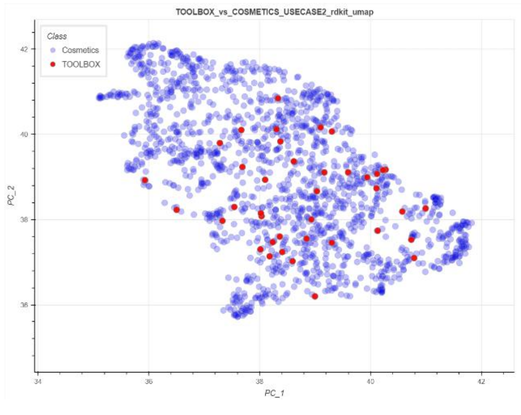

Fig.4. Visualisation of the chemical structural space covered by CPCat cosmetics (blue) and the test chemicals used in this evaluation (red).

RESULTS – non-animal toolbox 98% protective for high-risk chemical exposure scenarios

PROTECTIVENESS:

Does the toolbox workflow identify all high-risk exposure scenarios as requiring further evaluation, i.e. BER < predefined L1/L2/L3 threshold ?

Of our test chemical exposure scenarios, **98% of the 46 classified as high risk from the literature would also be classified as high risk using the toolbox and would require further (higher-tier) evaluation to refine and progress the indicated exposure scenario.**

Potential considerations for where a lack of protectiveness could occur:

The chemical has a specific mode of action not picked up in our test systems:

- Warfarin exposure was correctly identified as requiring further (higher-tier) evaluation, but the BER was very close to the threshold. Warfarin specifically interacts with VKORC1 which is not present in any of the test systems that make up this toolbox. Literature data is available for warfarin in this assay which, if integrated into the workflow, would give a BER of 0.088

- The Cmax estimate calculated at L2 is an underestimate of the in vivo exposure
  - The current L2 definition does not specify which parameters need to be derived experimentally, key parameters could be *in silico* and this might not be reflected in the error calculated under the assumption of an L2 prediction
  - The chemical might rely on active transport to enter cells, which isn't reflected in the PBK model without specific information. This is the case for Digoxin where the L2 prediction underestimates the L3 value by more than 50 times due a lack of consideration of transporters.

UTILITY:

Does the toolbox workflow identify all low-risk exposure scenarios as low risk, i.e. BER > predefined L1/L2/L3 threshold ?

Of our test chemical exposure scenarios, **8 of the 24 classified as low risk from the literature would be classified as low risk using this approach. This gives the current toolbox a utility of 33%.**

Future avenues of exploration to address reasons for lack of utility:

- The exposure estimate is a significant overestimate of the likely *in vivo* exposure and more data would be needed to refine this.
  - E.g. not all dermal exposure scenarios had good quality dermal penetration data available and so a default of 100% was assumed.
- The concentration-response analysis method used is overly sensitive and does not correct for all false positives
  - This is likely to be the case for examples like panthenol where the BER is being driven by a small number of genes with low level responses.
- The test systems are broadly conservative and require interpretation in the context of the full weight of evidence risk assessment/IATA, which has not been considered in this early tier evaluation. At this stage, if required, the assessment could progress in a tiered and iterative way in line with the ICCR principles[4], generating data in higher tier models or working to address remaining sources of uncertainty.

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. Landrum, G. 'RDKit Documentation, *https://www.rdkit.org/docs/*
3. UMAP Documentation, https://umap-learn.readthedocs.io/
4. 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

seac.unilever.com