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

SEAC | Unilever

Alistair Middleton, Maria Baltazar, Joe Reynolds, Beate Nicol, Gopal Pawar, Fazila Bunglawala, Dawei Tang, Sharon Scott, Sophie Malcomber, Sarah Hatherell, Predrag Kukic, Matt Dent, Andrew White, Georgia Reynolds, Paul Carmichael, Sophie Cable[1]
[1] Unilever's Safety and Environmental Assurance Centre, Colworth Science Park, MK44 1LQ

CONCEPT AND OVERVIEW

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.

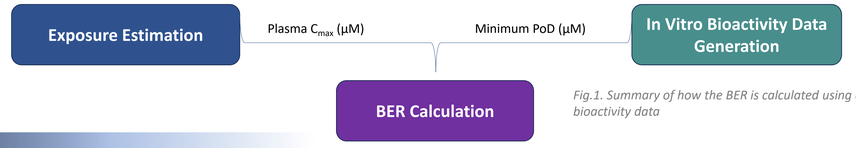

Fig.1. Summary of how the BER is calculated using exposure and bioactivity data

METHODS

- A pilot study was conducted using 10 chemicals with 24 identified exposure scenarios classified as high or low risk from a systemic toxicity perspective. Data were generated in the 3 core bioactivity platforms and PBK models built to estimate plasma Cmax levels.
- The pilot study was used to optimise the experimental design of the assays (e.g. plate layouts, choice of control chemicals) and to determine BER thresholds above which a confident low risk decision could be made. Analogous to the margin of safety approach where it is typically considered that a MoS > 100 represents low risk.
- For the full evaluation 38 chemicals were selected with the aim to avoid bias from selection of only highly toxic or inert chemicals, to cover a broad range of chemistries and biologies, and chemicals with definable exposure scenarios that could be classified as high or low risk using traditional toxicological approaches.
- Data were generated across the same bioactivity platforms as the pilot study and the performance assessed in terms of protectiveness and utility.
- PoDs from in vivo studies (traditional PoDs) were also identified for a subset of the evaluation chemicals and compared to the NAM PoDs. These included subchronic/chronic NOAELs, and in the case of Ketoconazole an adjusted 28-day NOAEL. The same activity was performed in that a minimum PoD was taken and compared to the external exposure estimates (in mg/kg bw/day) and the same performance metrics derived.

PILOT STUDY Results

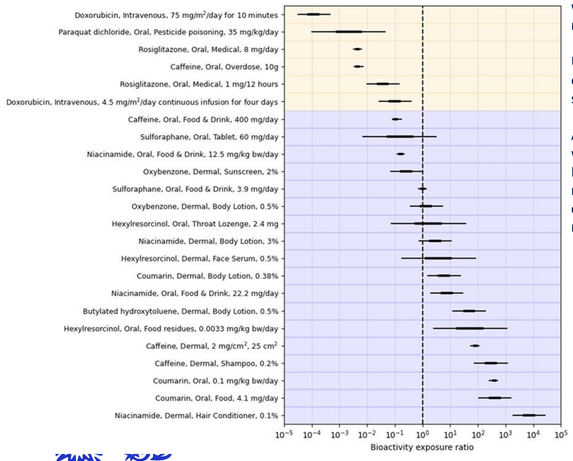

Fig.2. Calculated BER values for 24 chemical exposure scenarios as determined using the modules and workflow shown in Fig.1. High risk chemical exposure scenarios are shown in yellow, low risk chemical exposure scenarios are shown in blue. The bars represent the 95% confidence interval of the calculated BER when considering uncertainty in the exposure estimate. The 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.

Separation of the high and low risk scenarios was achieved, which was to be expected given the high potency and relatively inert nature of the chemicals and use scenarios identified at this stage.

Uncertainty in the plasma C_{max} estimates meant that the range for each BER plotted lead to some overlap for both high and low risk scenarios.

A threshold was set based on the different PBK inputs above which the likelihood of a scenario being low risk was > 95%. At PBK level 2 (in vitro parameter inputs) a BER threshold of 11 meant that all high risk exposure scenarios had a smaller BER calculated and all exposure scenarios with a value > 11 were low risk from a consumer perspective.

PBK Input	Threshold BER required for exposure scenario to be identified as low risk	Probability of overturning
In silico only (Level 1)	110	0.1
At least one in vitro parameter (Level 2)	11	0.1
Calibrated to clinical data (Level 3)	2.9	N/A

Table 1. BER Thresholds for different PBK levels derived from the pilot study. Corrected from Middleton et al., 2022

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. 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

FULL EVALUATION RESULTS

- 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.
- 2 of these were inhalation exposures, 7 IV exposures, 11 dermal exposures and 44 oral.
- For 1,2-Octanediol, Ethylzingerone, Panthenol and Cypermethrin, no in vitro ADME data were available, and therefore only L1 PBK models could be built.
- **The toolbox and workflow are 93% protective for high risk chemical exposure scenarios.** This increases to 98% protective when clinical data are used to calibrate the PBK models where available.

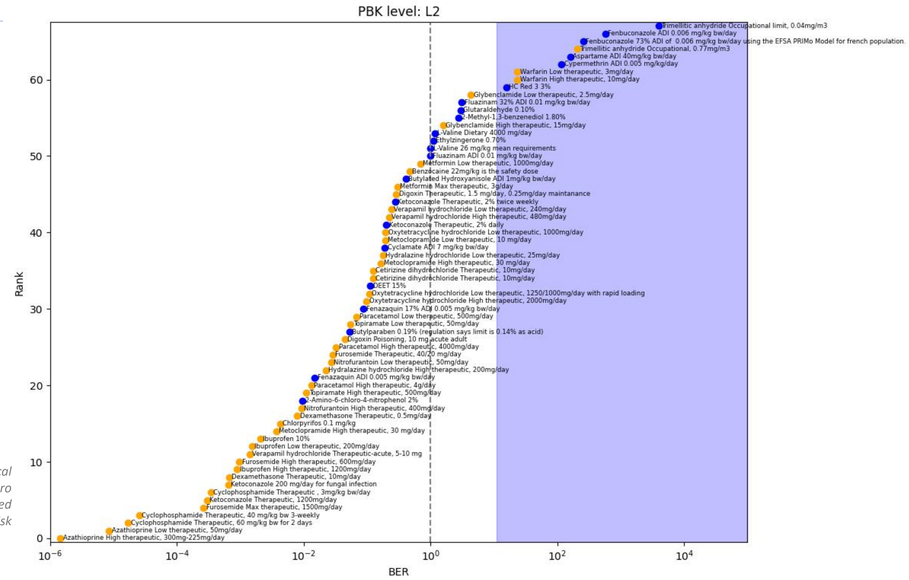

Fig.3. Plot showing the BER values determined for all chemical exposure scenarios using PBK models parameterised using in vitro data and the lowest PoD across all the bioactivity platforms tested (orange dots – high risk exposure scenarios, blue dots – low risk exposure scenarios).

PROTECTIVENESS:

Does the toolbox workflow identify all high-risk exposure scenarios as not low risk, i.e. BER < 11?

Of our test chemical exposure scenarios, **93% of the high risk 36 benchmark scenarios would also be classified as not low risk (i.e., uncertain risk) using this approach.**

Potential reasons for lack of protectiveness:

- The chemical has a specific mode of action not picked up in our test systems:
  - Warfarin specifically interacts with VKORC1 which is not present in any of the test systems that make up this toolbox. Literature data on the binding of warfarin is available in the literature, 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 > 11?

Of our test chemical exposure scenarios, **5 of the 21 classified as low risk from the literature would be classified as low risk using this approach. This gives the current toolbox a utility of 24%.**

Potential 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.

COMPARISON TO USE OF TRADITIONAL DATA IN EARLY TIER ASSESSMENT

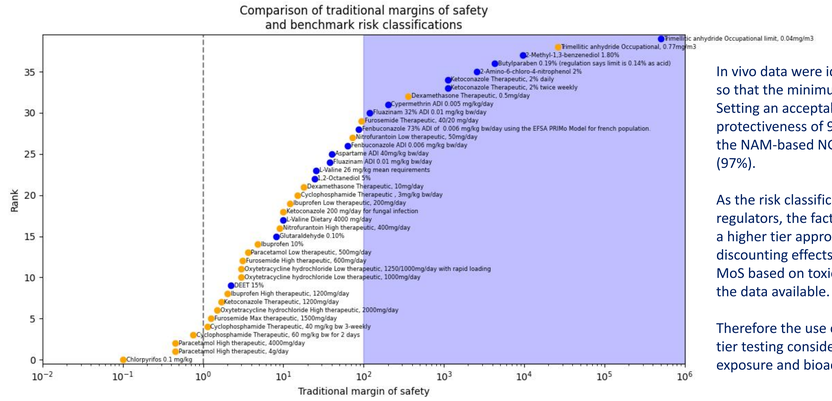

Fig.3. Plot showing the MoS values comparing the external applied doses (mg/kg bw/day) to the minimum traditional NOAEL for chemicals where appropriate in vivo data were available. Blue dots represent low risk benchmark scenarios, orange dots represent high risk benchmark exposure scenarios, the black dotted line is plotted at MoS =1 and the blue shaded region represents the area where the MoS > 100.

In vivo data were identified for 25 of the test chemicals. The data were extracted from Toxval and curated so that the minimum NOAEL of the repeat dose studies was taken forwards.
Setting an acceptable threshold at an MoS >100, using the in vivo data in an early tier manner gives a protectiveness of 91% (21 out of 23) and a utility of 47% (8 out of 17). For the same subset of chemicals, the NAM-based NGRA toolbox gives comparable performance metrics albeit with a higher protectiveness (97%).

As the risk classifications are set based on the traditional data and the opinions of authorities or regulators, the fact that so many of the low risk benchmarks fall below the threshold demonstrates that a higher tier approach to the available data is often required. In these cases that could be through discounting effects seen that were not relevant for humans, adjusting the acceptable threshold for an MoS based on toxicokinetic or toxicodynamic information or taking a weight of evidence approach to all the data available.

Therefore the use of NAMs in early tier assessment should be considered in the same light, and higher tier testing considered to improve the utility of a safety decision making framework based on in vitro exposure and bioactivity data.

seac.unilever.com