10.10.2024

www.ideaproject.info

Outlook: How to tailor considerations of uncertainty into risk assessment

Andreas Natsch,
Givaudan

Givaudan
Human by nature

Outlook - Considerations on uncertainty

1. The scope of the problem
2. Uncertainty of LLNA data
3. Uncertainty of predictions by regression models and GARD-DRP
4. Comparing Uncertainty of LLNA data and PoD predictions
5. Options to address uncertainty in NAM-based QRA
    1. Default assessment factor based on overall uncertainty
    2. Refine assessment factor based un uncertainty assessment for related chemicals

The scope of the problem

- **The quantitative models based on NAM to predict a point of departure come with their own uncertainty:**
  - A) **Experimental uncertainty** – due to the intrinsic experimental variability. This – depending on the experimental model - can be lower as compared to the animal tests (e.g. a kDPRA has a low run-to-run variability)
    B) **Mechanistic uncertainty** – the models are surrogates, reflecting and correlating to key events, but they are by definition proxy measurements

At the same time there is also **uncertainty in the in vivo models** used in classical risk assessment:

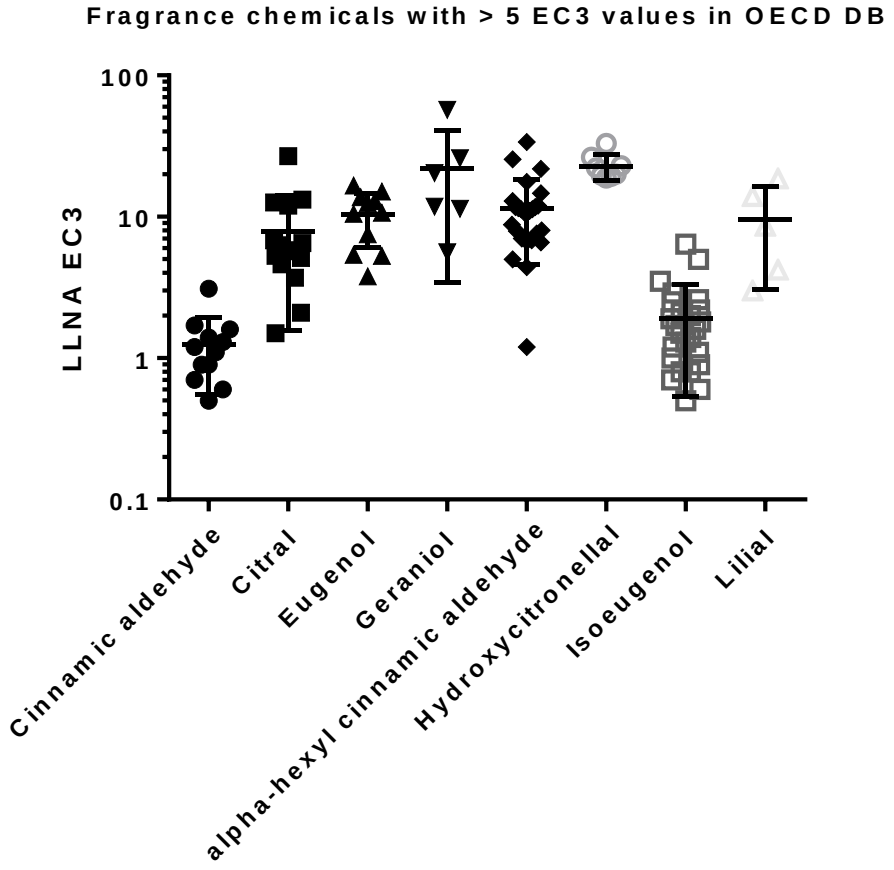
- For the LLNA there is both **experimental variability** – a single LLNA value does not fully represent the 'true' EC3 values which would be obtained from averaging over multiple studies (see below)
- For the LLNA we also have a **mechanistic uncertainty** – while cell proliferation is a close proxy for sensitization induction (arguably a closer proxy as compared to the *in vitro* endpoints), it still is a proxy

- For human data, variability comes from the intrinsic population variability and the fact that a HRIPT panel samples only a random part of this variability. Additional uncertainty is due to different protocols used over time and the expert reading of the reactions.
- For human data we also have the statistical uncertainty of sampling a rare event: A HRIPT study finding 2 positives in a panel of 100 may upon repetition come with 0 or 4 positives just by the stochastic nature of sampling rare events within a limited panel

- For QRA based on NAM we have to quantify and implement the uncertainty to have similar protection as compared to classical assessment
- However: only addressing uncertainty of NAM and not comparing it to uncertainty of in vivo data would not be an equal level playing field

Uncertainty of LLNA data

- For the LLNA we have many historical repeat data, which allow to quantify the experimental variability
- OECD created a curated LLNA database of LLNA values meeting OECD test guideline requirements, 17 materials have at least 5 repeat LLNA values, and eight of these are fragrance chemicals



We can measure prediction **accuracy of an individual LLNA EC3** value to predict the consolidated median (MLLP) as a measure to predict the 'true' EC3 by calculating the geometric or median absolute fold change between the MLLP and any individual experimental value:

Median (n=16)[1)]: 1.5-fold

Geometric mean (n=16): 1.7-fold

Median fragrance chemicals (n=8): 1.4-fold

Geometric mean fragrance chemicals (n=17): 1.6-fold

This gives us a measure of the prediction accuracy for a single LLNA value to predict the 'best guess' LLNA – i.e. the uncertainty just coming from experimental LLNA uncertainty – *without considerations how good our proxy is / mechanistic uncertainty*

Givaudan

1) One chemical with pos and neg values excluded

Uncertainty of predictions by regression models: All chemicals – LLNA

- The prediction **accuracy for the regression models predicting the EC3** value have a geometric mean misprediction of 3.2-fold and a median of 2.5-fold. This is higher as compared to the variability of the LLNA to predict the 'best guess' LLNA value (as expected) (n = 188)

Similar predictivity of the models based on the following:

- kDPRA and KeratinoSens
- kDPRA and h-CLAT
- kDPRA, KeratinoSens and h-CLAT

Model	Input parameters	Fold-misprediction[1] (Geomean)	Fold-misprediction (Median)	Chemicals > 5 – fold underpre-dicted[2] n, (%)	Chemicals > 10 – fold under-predicted n, (%)	Chemicals > 5 – fold over-predicted[2] n, (%)	Chemicals > 10 – fold over-predicted n, (%)
EQ1	kDPRA, KS	3.3	2.5	33 (18%)	20 (11%)	16 (9%)	7 (4%)
EQ4	kDPRA, h-CLAT	3.2	2.4	30 (16%)	17 (9%)	16 (9%)	7 (4%)
EQ5	kDPRA, KS, h-CLAT	3.1	2.3	35 (19%)	17 (9%)	18 (10%)	6 (3%)
EQ6	KS, h-CLAT	3.5	2.6	33 (18%)	19 (10%)	19 (11%)	8 (4%)

[1] The ratio between the higher and the lower values of the measured and predicted EC3 value. Predicted EC3 > 100% were set to 100%.
[2] Under-predicted chemicals: those for which the measured LLNA EC3 is < than the predicted EC3; over-predicted chemicals: Those with measured LLNA EC3 > than the predicted value.

Uncertainty of predictions by regression models: All chemicals - PV

- The prediction **accuracy for the regression models predicting the PV** have a geometric mean misprediction of 3.5-fold and a median of 2.6-fold.

- Overall models trained on PV values and trained on LLNA data perform similarly
- N = 139

Model	Input parameters	Fold-misprediction (Geomean) vs. LLNA	Fold-misprediction (Geomean) vs PV	Fold-misprediction (Median) vs. LLNA	Fold-misprediction (Median) vs PV
EQ5 (trained on LLNA)	kDPRA, KS, h-CLAT	3.1	3.6	2.3	2.5
EQ5d (trained on PV)	kDPRA, KS, h-CLAT	3.4	3.5	2.6	2.6
EQ5e (trained on PV, excluding cytotoxicity input)	kDPRA, KS, h-CLAT	3.3	3.5	2.8	2.7

[1] The ratio between the higher and the lower values of the measured and predicted EC3 value. Predicted EC3 > 100% were set to 100%.
[2] Under-predicted chemicals: those for which the measured LLNA EC3 is < than the predicted EC3; over-predicted chemicals: Those with measured LLNA EC3 > than the predicted value.

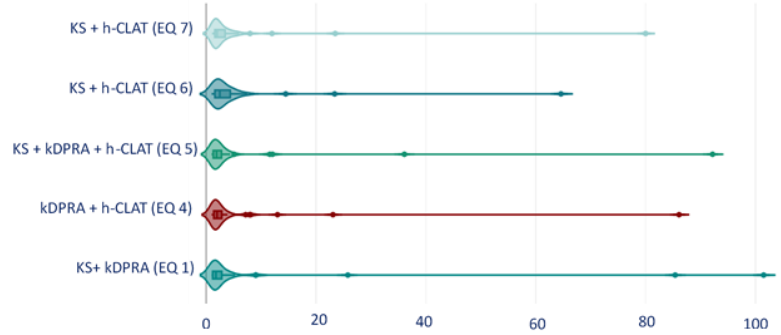
Uncertainty of predictions by regression models: Fragrance chemicals

- The overall uncertainty assessment is including the very strong and extreme sensitizer – at the lower end of the scale, the fold-misprediction tend to be larger
  - Thus for our fragrance-based assessment, focusing on Fragrance materials may be more interesting

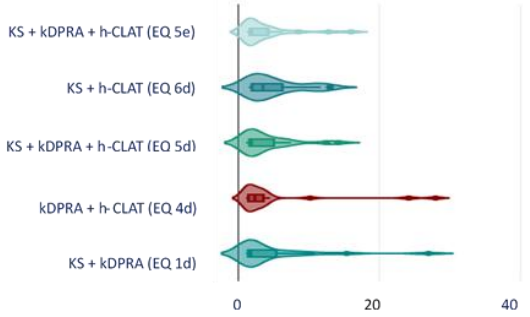
RIFM conducted a study which is in press:

- **Predicting points of departure and potency categories for fragrance ingredients integrating OECD in vitro models.** Isabelle Lee, Mihwa Na, Maura Lavelle, Isabella Schember, Cindy Ryan, G Frank Gerberick, Andreas Natsch, Anne Marie Api; Food and Chemial Toxicology, in press.

	Median fold-misprediction vs **LLNA** based on LLNA models (RIFM study)	Median fold-misprediction vs **NOEL** based on PV models (RIFM study)	Median fold-misprediction vs **PV** based on PV / LLNA models	Median fold-misprediction vs **PV** based on GARD DRP data
EQ 1 (kDPRA and KS)	1.8	1.7	To be done withing IDEA project with extended database	To be done withing IDEA project with extended database
EQ 4 (kDPRA and hClat)	2.0	2.2		
EQ 5 (all 3 tests)	1.9	1.8		



Fold misprediction for predicted EC3 vs experimental EC3 (RIFM study)



Fold misprediction for predicted PV vs human NOEL (RIFM study)

Uncertainty for the chemicals in the RCPL list

- Shown are below predictions for the fragrance molecules with a PV in the RCPL list (Excluding negatives), n = 21 regression models, n= 20 GARD, n = 19 overlap GARD / regression models

Comparison to PV values

	Geometric Mean fold misprediction	Median fold misprediction
Equation 5 trained on LLNA	3.03	3.2
Equation 5e trained on extended PV list	2.66	2.24
GARD dose response	2.91	2.62

Comparison to LLNA EC3 values

	Geometric Mean fold misprediction	Median fold misprediction
Equation 5 trained on LLNA	2.48	2.11
Equation 5e trained on extended PV list	3.32	2.33
GARD dose response	3.04	2.54

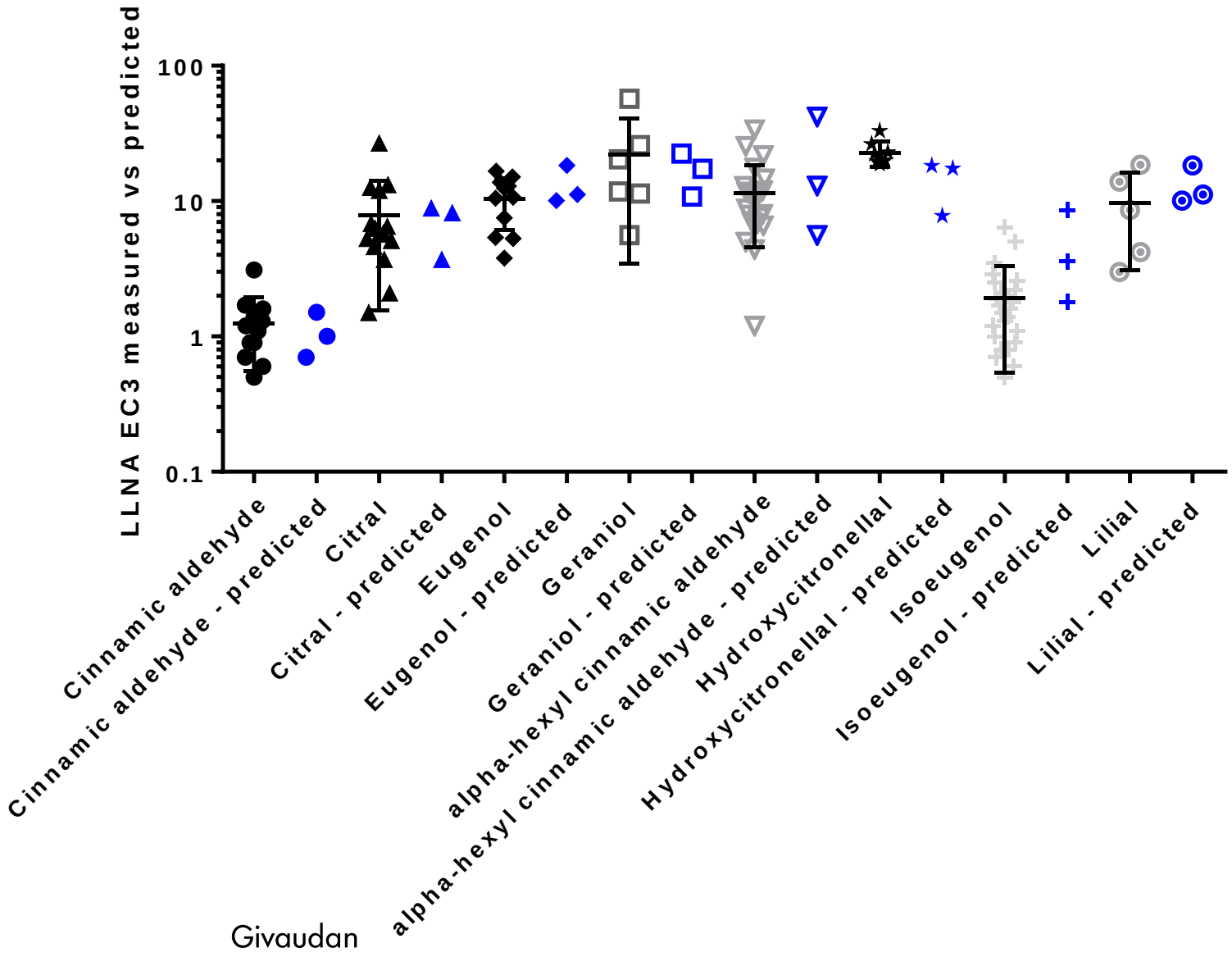
The median fold misprediction for this limited dataset is similar for the two models

For the regression models the median fold-misprediction vs. The PV is similar (just slightly higher) to the assessment of RIFM vs. the NOEL on a larger set of fragrance chemicals

This assesment can now be re-enforced with the larger database with both GARD and regression data and a larger set of PV values

Comparing Uncertainty of LLNA data and PoD predictions

- For an assessment of variability vs. LLNA variability, we can compare the three regression models (EQ1, KS+DPRA); EQ4, hClat and kDPRA and EQ 5 with all three tests) vs. repeat LLNA measurements
- Overall the predictions for these data-rich fragrance ingredient case studies are within the range of the LLNA variability



Options to address uncertainty in NAM-based QRA

- A) **statistical approach** – with the enlarged database, we can assess how much our uncertainty increases with a NAM-based assessment vs. a classical approach using one LLNA measurement and using the EC3 as PoD.
  - Based on this, a fixed assessment factor may be added to QRA2 to account for additional variability introduced by using NAM
  - It is important that this **accounts for *additional* uncertainty** and not overall uncertainty – overall uncertainty is also covered by existing assessment factors.
- B) **Refining uncertainty by exploring related substance**
- Y - Target substance with in vitro data only
- Z - Read-across substance **with in vitro data and in vivo PoD**
- Assess accuracy of assessment of PoD of Z by the NAM approach
- If related read-across substance is well-predicted, a lower assessment factor may be justified
- -> This principle has been explored in below paper, for example see next slides. This can be further explored with the refined data-set of the extended RCPL database

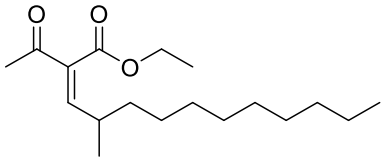
  Natsch, A., et al., Deriving a No Expected Sensitization Induction Level for Fragrance Ingredients Without Animal Testing: An Integrated Approach Applied to Specific Case Studies. Toxicol Sci, 2018. 165(1): p. 170-185.

C) **Other approaches** – to be discussed – ideas welcome, and this is maybe an **important subject for a F2F-workshop!**

Case study on new molecule: Risk assessment without LLNA from Natsch et al, 2018

- Scentaurus clean predicted as sensitizer by TIMES, KeratinoSens and peptide reactivity

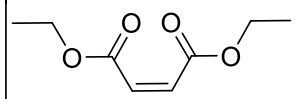
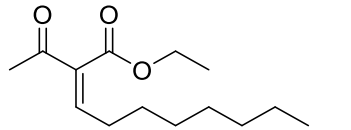
a) Data, assessment with DIP and additional mechanistic tests

Name:	**ethyl (Z)-2-acetyl-4-methyltridec-2-enoate**	**DPRA:**	Cys-depletion: 27.8 % Lys-depletion : 1.3 % **Positive in low category**, ca. **6.6% direct adduct** with Cys-peptide
Structure:		**KeratinoSens:**	EC 1.5: 7.95 µM EC3 not reached due to cytotoxicity IC50: 13.2 µM **Positive**
TIMES parent:	**strong sensitizer, αβ-Carbonyl compounds with polarized double bonds**	**Prediction global model:**	**EC3: 5.1 %**
TIMES metabolite:	strong sensitizer, αβ-Carbonyl compounds with polarized double bonds	**Prediction Local model:**	**EC3: 14 %**
LC-MS:	Cor1C420 depletion: 14 % Adduct: **direct MA adduct** Peptide oxidation predominant	**Additional mechanistic tests:**	Not needed
Domain attribution:	Michael acceptor	**Results mechanistic tests:**	n/a

Case study on New molecule: Uncertainty

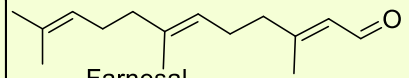
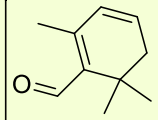
- Uncertainty assessment:
  - Related analogues: Michael acceptors with the double bond activated by two carbonyl groups
  - Well predicted by global and local model, here global model more accurate and on conservative side
  - Use global model for conservative assessment

a) Analysis of close analogues for uncertainty assessment

Close analogue:	 Diethylmaleate	 ethyl (Z)-2-acetyldec-2-enoate
Rationale for selecting close analogue:	Double activated MA-ester	Double activated MA-ester, substructure of target
Prediction close analogue global model:	**EC3 1.4%**	**EC3 3%**
Prediction close analogue local model (MA):	**EC3 3.8 %**	**EC3 5.6 %**
In vivo results close analogue:	**EC3 2.1 %**	**EC3 2.6 %**
Prediction accuracy analogues:	**Good prediction with local and global model, better accuracy for global model for these double activated MA-esters**	

Case study Citral: Uncertainty assessment

- Related β-branched, αβ-unsaturated aldehydes assessed
- Local MA models predicts EC3 within 2-fold error, on conservative side
- **Indicates high certainty of the prediction for Citral**

Close analogue:	 Farnesal	 Safranal
Rationale for selecting close analogue:	β-alkyl-substituted αβ-unsaturated aldehydes	Di-substituted αβ-unsaturated aldehydes
Prediction close analogue global model:	**EC3 2.3%**	**EC3 1.7%**
Prediction close analogue local model (MA):	**EC3 6.9 %**	**EC3 3.4 %**
In vivo results close analogue:	**EC3 11.7 %**	**EC3 7.5 %**
Prediction accuracy analogues:	**Local model predicts within 2-fold error; on conservative side**	

Follow us on social media @givaudan

Givaudan

Human by nature