Fortnightly meeting 23rd July 2019

Mark Nelms

23/7/2019

Contents

Questions from last time

After our last meeting we had 4 questions to answer:

- What pKa values are associated with corrosive activity?
- Do chemicals with respiratory sensitisation alerts have local effects *in vivo*?
  - Presence of respiratory alert (or extreme pKa) and local *in vivo* effects in ToxVal
- How do our 5th percentile values look converted to TTC values and body dose units (μg/kg-day)?
  - How do these compare to what has been previously been published and vs. the oral values?
- Are there chemicals that overlap between oral and inhalation routes?

1 Inhalation Data

1.1 What pKa values are associated with corrosive activity?

According to pg 170 of ECHA's Guidance on Information Requirements and Chemical Safety Assessment a substance can be considered a skin/eye corrosive/irritant if they are:

- a strong acid (pH < 2.0), or
- a strong base (pH $\geq$ 11.5)

I couldn't find any information that gave the values in terms of pKa, only in terms of pH. Obviously, the pH of a solution can change depending on its concentration, whereas its pKa is a fixed value. Should I just use these values as they are and say if a chemical has a pKa beyond these boundaries then they are likely to be a skin/eye corrosive substance?

1.2 Do chemicals with respiratory sensitisation alerts have local effects *in vivo*?

I started off by searching through the entire Inhalation dataset to see how the critical effects that related to the respiratory system were described. Due to the unique way toxicologists describe the same/similar effects, I found a number of critical effects that I had to search for, quite a few of which were just different ways of saying the same thing. They were:

- Respiratory/respiration
- Olfactory
- Nose/nasal
- Pulmonary
- Lung
- Trachea
- Brochi (covers bronchiolar and bronchiectosis)

- Alveol (covers alveoli and alveolar)
- Local (I've assumed with Inhalation the local effects would affect relevant organs, i.e. pulmonary/eyes)
- Atelectasis
- Altered breathing rate/tachypnea (unusually rapid breathing)/dyspnea (difficult breathing)/labored breathing
- Rales (clicking/rattling sounds in lung)
- Eye/ocular
- Conjunctival congestion
- Lacrimation (excessive secretion of tears)
- Larynx

There were three others I didn't include, "congestion" and "sneezing/coughing", which I don't know if these are really considered local effects. The third effect I didn't include was "red or crusty deposits around eyes, ears, nose, or mouth", do you think this should be included?

I then used the critical effects outlined above to filter the chemicals from the inhalation dataset that fell into Cramer class I, II, or III that had local effects *in vivo*. There were a total of 91 chemicals with local effects. Of these 91 chemicals, only 14 contained a respiratory sensitisation alert from the QSAR Toolbox.

Therefore, very few of the chemicals with local effects actually got picked up as being respiratory sensitisers by the Toolbox. However, I'm not sure how many of the chemicals are out of domain for the alerts: a chemical is either assigned an alert or "No alert found" is returned. I wonder how many of the chemicals assigned "No alert found" are actually out of domain?

I decided to look at how many chemicals in the wider inhalation Cramer set had been assigned a respiratory sensitisation alert from the Toolbox. Of the 478 chemicals with inhalation data that were assigned into one of the three Cramer classes, there were only 41 chemicals in this set that "triggered" a respiratory sensitisation alert in the Toolbox. Subsequently, I had a look at the critical effects associated with these chemicals:

Table 1: Critical effects of chemicals with respiratory sensitisation alert

Chemical name	Critical effect	Respiratory sensitisation alert
4-Hydroxybutyl prop-2-enoate	(local effects)	Michael Addition
Styrene	local	Pro-SN2
2-Hydroxyethyl acrylate	local	Michael Addition
Trimellitic anhydride	local and systemic	Acylation
Dimethylaminoethanol	local effects	Pro-Schiff base formation
Butyl methacrylate	local effects	Michael Addition
Ethyl methacrylate	local effects (Histopathology, olfactory epithelium)	Michael Addition
Furfuryl alcohol	local irritant effects	Pro-Schiff base formation
Methyl acrylate	parental local toxicity	Michael Addition
Isobutyl acrylate	parental local toxicity	Michael Addition
2-(Dimethylamino)ethyl acrylate	parental local toxicity	Michael Addition
Butyl acrylate	parental local toxicity	Michael Addition
tert-Butyl acrylate	parental local toxicity	Michael Addition
2-Ethylhexyl acrylate	parental local toxicity	Michael Addition

As you can see in Table 1 almost half of the chemicals didn't have a critical effect outlined. For the remaining chemicals very few of the rows actually describe where or what the effects are. For example, what does "for increased liver copper concentrations" mean, were the researchers only focussed on an increase in copper concentration in the liver and didn't care if there were effects in other organs at lower concentrations? Or "hypertrophy|necrosis|hyperplasia", in which organ did these effects occur?

Maybe the fact that fewer chemicals with an respiratory sensitisation alert had local effects is partially down to the fact that only a few of the chemicals even had relevant information as to what the critical effect was and in which organ the effect was observed?

1.3 How do our 5th percentile values look converted to TTC values (mg/m^3) and body dose units (μg/person/day) and how do these compare to other published inhalation TTC values?

1.3.1 Converting to common units

My first task was to get the inhalation TTC data that I know of into the same units so that we can compare our results here to those that have previously been published.

I started off by converting our 5th %ile NO(A)EC values for Cramer class I and III to TTC values in both mg/m^3 and body doses (μg/person/day)

First I calculated the TTC values in terms of mg/m^3, using Equation (2) present in Escher et al. (2010). The equation to do this was:

$$TTC(mg/m^3) = \frac{NOEC\,(ppm) \times MW\left(\frac{g}{mol}\right)}{24.45\left(\frac{l}{mol}\right)}$$

$$TTC(mg/m^3) = \frac{5th_{percentile} NOEC\left(\frac{mg}{m^3}\right) \times DailyExposure}{10 \times 2.5} \tag{1}$$

Equation (1) could also be used to calculate the TTC in ppm, all that is needed would be to replace the 5th percentile in mg/m^3 with the 5th percentile in ppm.

DailyExposure is used to convert the dose strategy from 6hrs/day and 5days/week to 24hrs/day and 7days/week by doing (6/24 x 5/7). This was done because most inhalation studies typically have an exposure duration of 6hrs/day for 5days/week. As we don't have this *DailyExposure* information in ToxVal, I assumed this dosing strategy was used for all chemicals.

Next, I converted the mg/m^3 values into body doses, using these equations:

$$\begin{aligned}\text{NO(A)EL}\,(\mu g/kg/day) = 5th_{percentile} NOEC\left(\frac{mg}{m^3}\right) \times DailyExposure \\ \times \left(\frac{V_{human.resp}\left(\frac{m^3}{d}\right)}{BW_{human}\,(kg)}\right) \times 1000\end{aligned} \tag{2}$$

$$\text{NO(A)EL}\,(\mu g/person/day) = \frac{5th_{percentile} NOEC\,(\mu g/kg/day) \times BW_{human}\,(kg)}{10 \times 2.5} \tag{3}$$

Whereby, the $V_{human.resp}$ is the standard human respiratory volume of 20m^3 provided by ECHA (2008) and the BW_{human} is the average human body weight of 60kg from Munro et al. (1996). I used the body weight of 60kg because even though the average has increased since the Munro et al. paper that is what both Carthew et al. (2009) and Escher et al (2010) have used for their analysis so it enables us to better compare our results to theirs.

I then took the TTC data in terms of both mg/m^3 and μg/person/day, where possible, from the Escher et al. (2010) and Carthew et al. (2009) papers. Luckily, the Escher et al. paper already had their TTC values in these units.

The Carthew et al paper, meanwhile, only had the 5th percentile NOAEC in mg/m^3 for the local effects and the μg/person/day for the systemic effects.

Therefore, for the Carthew et al. data I had to calculate:

- TTC(mg/m^3) for local effects
- TTC(μg/person/day) for local effects, and
- TTC(mg/m^3) for systemic effects

Calculating the TTC for local effects from Carthew et al. was relatively simple:

- For converting to TTC(mg/m^3) I just had to plug the 5th %ile NO(A)EC into Eq (1). above.
- For converting to TTC(μg/person/day) I just had to plug the 5th % NO(A)EC into Eqs (2) and (3) above

To calculate the TTC in terms of mg/m^3 for the systemic effects was a bit more complicated because it involved rearranging Eq (2) above so that the 5th percentile NOEC (mg/m^3) was the subject of the equation (shown below) and then plugging the results from this into Eq (1) to calculate the TTC(mg/m^3).

$$5th_{percentile}NOEC\left(\frac{mg}{m^3}\right) = NO(A)EL \times \left(\frac{1}{DailyExposure}\right) \times \left(\frac{BW_{human}\,(kg)}{V_{human.resp}\left(\frac{m^3}{d}\right)}\right) \text{ OR}$$
$$= \frac{(NO(A)EL \times 24hrs \times 7days \times BW_{human}\,(kg))}{\left(6hrs \times 5days \times V_{human.resp}\left(\frac{m^3}{d}\right)\right)}$$

1.3.2 Comparison of inhalation results from our study to Escher et al. and Carthew et al.

Below is the table comparing our inhalation TTCs to those published by Escher et al., Carthew et al., and the oral TTC values published by Munro et al.

					Cramer class	
Source	Exposure route	Effect type	Units	Number of chems	I	III
ToxVal	Inhalation	General	mg/m3	478	0.00124	0.00063
			ug/person/day		25	13
Carthew		Local	mg/m3	92	0.01	0.0034
			ug/person/day		200	67
		Systemic	mg/m3		0.0492	0.0492
			ug/person/day		980	170
RepDose		General	mg/m3	136	0.0089	0.00018
			ug/person/day		180	4
Munro	Oral		ug/person/day	611	1800	90

As you can see our TTC values via the inhalation route are at least 3-fold lower than any of the previously published TTC values - especially the Munro oral values. The fact that much lower TTC values have been derived for each of the inhalation studies compared to Munro's oral study points to inhalation exposures being likely to lead to toxic effects. One hypothesis is that this may be because of differences in ADME properties. For example a lack of first pass metabolism or an increase in absorption upon inhaling a substance.

As both Carthew et al and Escher et al use subacute data as well as subchronic data, I think we should consider including it in our anaylsis (after adding a 6-fold uncertainty to the NO(A)EC). I'm not sure how many extra chemicals it will give us though.

1.3.3 Are there chemicals that overlap between oral and inhalation routes?

There are 311 chemicals that have both chronic, oral and inhalation data.

To be consistent for the chronic, oral data I have used the ToxVal and TTC categories from the chronic, oral TTC work and for the inhalation data I have used the most up-to-date version of ToxVal (and the latest TTC categories). From the scatter plot below there doesn't seem to be much consistency between the two data sets.

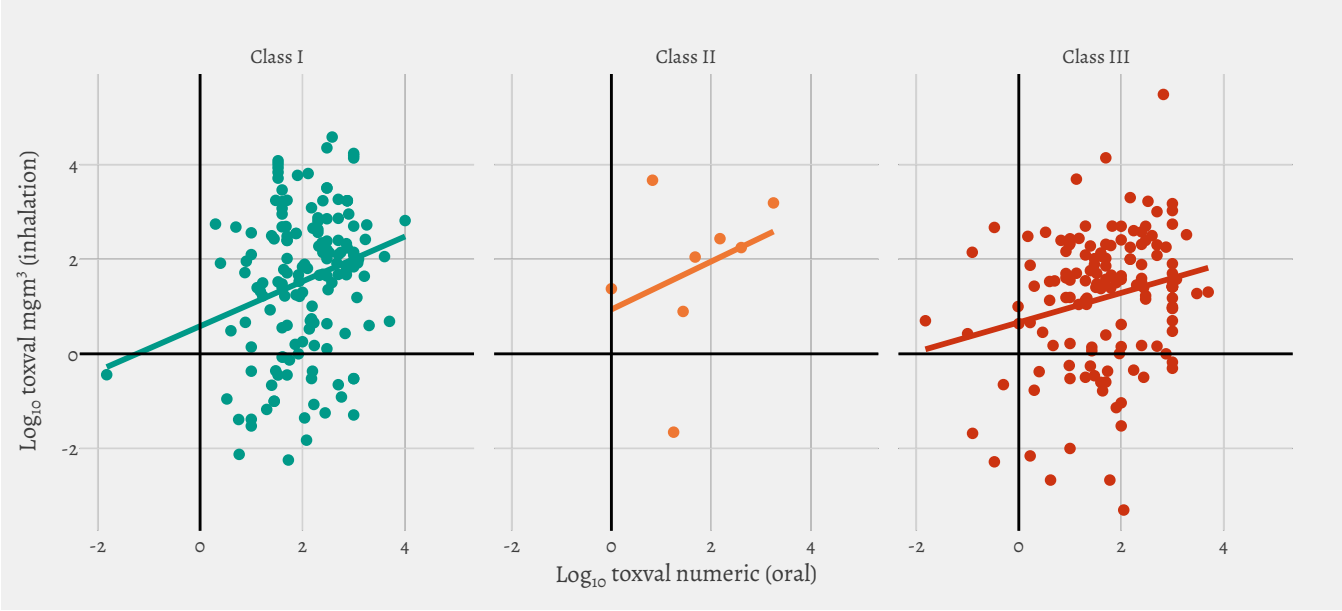


Figure 1: Comparing oral to inhalation toxicity values

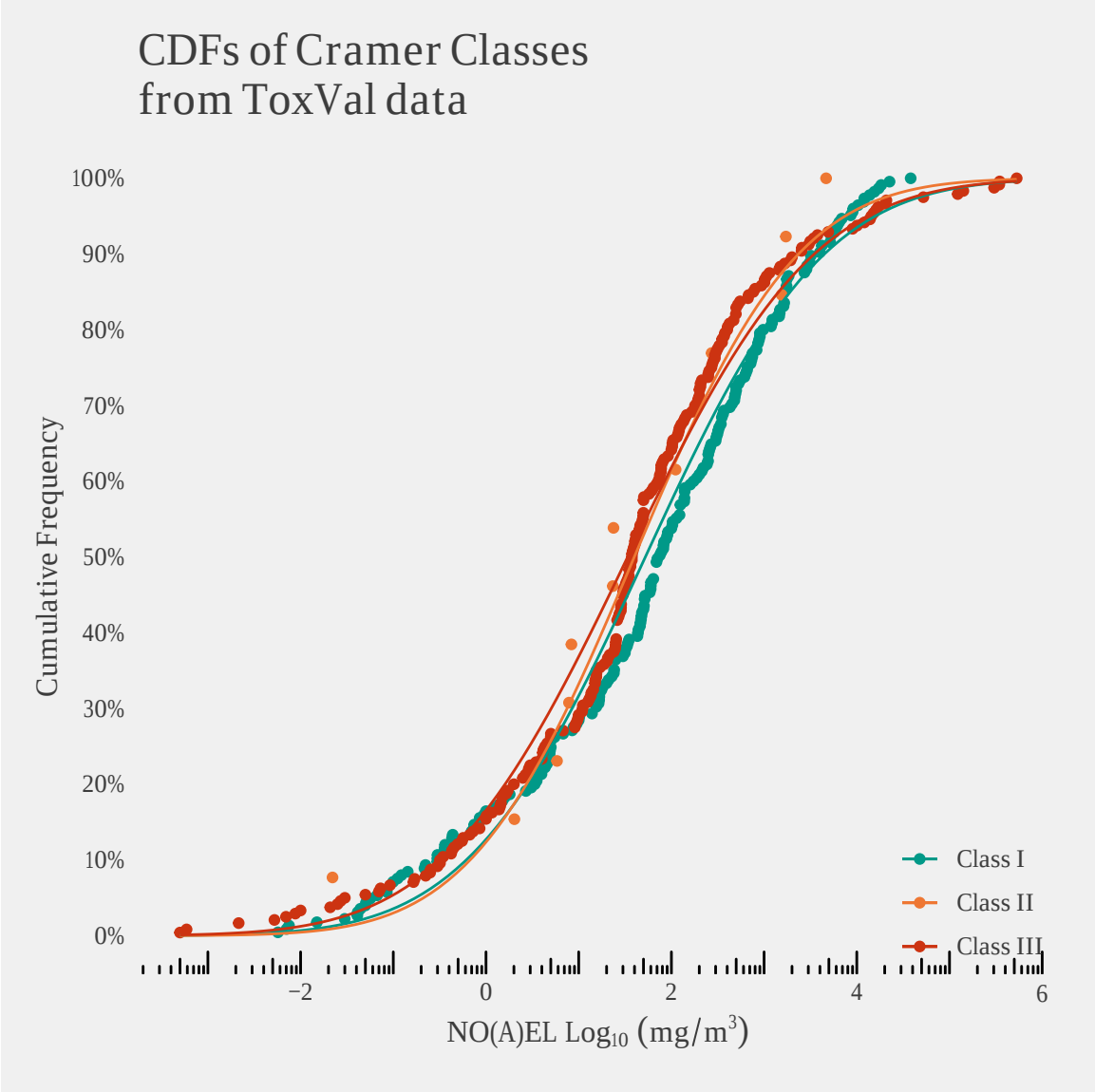
I then ran a linear regression (using `lm()`) for each Cramer class to see if there was an underlying relationship between the inhalation ana oral TTC data.

For chemicals with both oral and inhalation studies in Cramer classes I and III as the $\log_{10}$(toxval_numeric) increases by 1 log unit the inhalation $\log_{10}$(toxval_mgm3) increases by 0.49 log units for Cramer class I and by 0.33 log units for Cramer class III, both of which are statistically significant at the 0.01 level.

Meanwhile, Cramer class II does not have a statistically significant relationship. This is not surprising given that there are only 6 chemicals that contain both inhalation and oral toxicity data in this Cramer class.

Cramer class	Term	Estimate	Standard Error	Statistic	p-value
Class I	Intercept	0.58438	0.33148	1.76292	0.07987
	toxval numeric	0.47344	0.14881	3.18147	0.00177
Class II	Intercept	0.93746	1.19005	0.78774	0.46082
	toxval numeric	0.50697	0.62320	0.81349	0.44701
Class III	Intercept	0.66470	0.22009	3.02007	0.00299
	toxval numeric	0.31179	0.10967	2.84301	0.00512

1.4 Identifying ToxPrints associated with toxicity levels



As the Cramer classes don't do a good job of separating the cumulative distributions (above), I've started attempting to use the toxval_numeric in ppm to identify ToxPrints that may help to distinguish between chemicals that exhibit high, medium, or low toxicity via inhalation.

I'm not sure whether to use the GHS classification values of toxicity of:

- High toxicity - $\leq$ 15 ppm
- Medium toxicity - > 15 ppm and $\leq$ 80 ppm
- Low toxicity - > 80 ppm

Or the values from Schüürman et al (2016):

- High toxicity - < 0.75 ppm
- Medium toxicity - $\geq$ 0.75 ppm and $\leq$ 12 ppm
- Low toxicity - > 12 ppm

The Schüürman et al (2016) values were derived from analysis of the 296 compounds from the RepDose inhalation set they used to calculate the inhalation TTC values and have been selected to provide bins of roughly the same size.

1.4.1 Association rule mining

So far I've run ARM using `apriori()` on the 584 chemicals that make up our inhalation dataset from ToxVal after filtering for:

- Test species: Rat, mouse, other rodents, or rabbit
- Route of exposure: Inhalation
- Study duration: subchronic, chronic, reproductive, developmental, and multigeneration
- POD type: NO(A)EL/NO(A)EC
- Identifying outliers and keeping only the minimum NO(A)EL/NO(A)EC

After running the ARM algorithm, we currently have 82 rules with an Odds Ratio of ≥ 3 that identify 48 rules for high toxicity, 13 rules for medium toxicity, and 21 rules for low toxicity.

I've done some very preliminary analysis by profiling the same 584 chemicals into High, Medium, and Low toxicity bins using the 82 rules that have been generated. I don't know how well it will work in practice because after profiling the chemicals "only" 383 were binned into one of the three categories:

- 343 High toxicity chemicals
- 5 Medium toxicity chemicals
- 35 Low toxicity chemicals

From these results we still have one category with much fewer chemicals.

Additionally, there were some chemicals that fell into more than one of the 3 categories:

- 13 chemicals were identified as both High and Low toxicity, and
- 2 chemicals were identified as both High and Medium toxicity
- There was no overlap in chemicals between Medium and Low toxicity - not surprising given the low number of chemicals

Maybe using the values from Schüürman et al (2016) to discriminate between high, medium, and low toxicity would be better to use?

Additionally, I'm currently only doing ARM on the inhalation ToxVal data but we could add the Carthew and Escher chemicals to the dataset and run the ARM on them too.

Having those chemicals too would likely enable us to take a random sample of 80%, generate the rules on that training set, and then test the rules on the remaining 20% of chemicals. After some hiccups (names not matching, etc.), I've managed to get what I think are the right QSAR ready SMILES from the CompTox dashboard for each of the datasets.

I've already run the chemicals from both the Escher and Carthew papers through the CompTox Chemicals dashboard to get the DTXSID and QSAR ready SMILES. Whilst doing this I ran into a number of problems with chemicals either having a different name or multiple structures in the dashboard. There were 3 chemicals in Carthew where this was the case. For one chemial I could use the name in parentheses to ID what looked to be the right chemical, for the other two I had to leave them without a structure.

There were eight chemicals in the Escher dataset where we had problems. This was sometimes due to the presence of α symbols in the name that weren't present in the name in the dashboard. Luckily, the Escher data had CAS numbers so I could use these in conjunction with the chemical name to check the structure was correct.