Fortnightly meeting 13th August 2019

Mark Nelms

13/8/2019

Contents

Questions from last time

After our last meeting we had 4 questions to answer:

- What is the proportion of inhalation TTC studies that are subchronic compared to those studies that are chronic (inc. repro/dev studies)
  - Compare this to the proportion of subchronic studies from our chronic, oral TTC dataset
- Check to see whether we can recreate the Escher et al and Carthew TTC values
- Generate bar plots comparing ToxPrints between our inhalation TTC dataset, the Escher dataset, and the Carthew dataset
- Plot local and systemic NO(A)ELs against one another
  - Do separately for Escher and Carthew studies
  - From this - is there anything we could use to make predictions about which chemicals from our set are local and which are systemic?

1 Inhalation TTC

Filtering criteria for inhalation studies:

- 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, NOEL, NO(A)EC, NOEC

1.1 Comparison of % of subchronic studies in Inhalation and Oral TTC datasets

Within this comparison I'm considering all reproductive/develomental studies to be chronic studies.

After using the filtering criteria above (the same ones from last time). There are **410** chemicals with a subchronic study as their lowest NO(A)EL/NO(A)EC in the ToxValDB inhalation TTC dataset.
Overall, subchronic studies account for **85.8%** of the **478** studies that would be utilised in the generation of an inhalation TTC.

Meanwhile, there were **351** chemicals with a subchronic study as their lowest NO(A)EL in our oral, chronic TTC dataset from ToxValDB.
This equates to **26.9%** of the **1,304** studies within the oral, chronic TTC dataset.

It should also be noted that there is a slight difference in how the study lengths of the two datasets were derived:

- For the oral, chronic TTC dataset we used the original study type information provided in the ToxValDB download
- For the inhalation TTC dataset we wrote code that would enable us to identify additional studies as being subacute, subchronic, or chronic based upon the study length information

However, this shouldn't give such large discrepancies in the percentage of subchronic studies present in the inhalation TTC dataset compared to the oral TTC dataset.

The inhalation dataset from Escher et al contains **74** chemicals with a subchronic study as the lowest NOEC. This corresponds to **36.5%** of the **203** studies in the Escher inhalation dataset.

Meanwhile, there are **69** chemicals with a chronic study driving the NOEC, which corresponds to **34%** of the 203 studies.

Unfortunately, I can't do the same thing for the Carthew et al inhalation dataset because they don't provide that level of detail in their Appendix.

1.2 Recreating the Escher and Carthew TTC values

1.2.1 Preparing Escher and Carthew datasets

Before I could start working on recreating the TTC values I had to get both datasets into a more suitable format. For each dataset I:

- Copied and pasted the table from the Appendix of the paper into Notepad,
- Copied from Notepad into Excel - otherwise everything was spread over only 4 lines in Excel
- Added a Cramer class column and assigned the chemicals to the correct Cramer class based upon the information in the Appendix.

Additionally, for the Escher data I had to:

- Name the General, Systemic, and Local NOEC columns (they were moved onto a separate line)
- Change the numbers they had used to reference each study duration to the corresponding character string (e.g. 1 == subacute)
- Change the way the genotoxic and OP chemicals were identified
  - Instead of having an "x" to identify chemicals I used regex to amend the correct rows to say "Genotox Alert" or "OP"
- Adjust the chemicals were the information was spread over 2 rows (happened when 2 references were present for a single chemical)

1.2.2 Retrieving MW, SMILES, and QSAR-ready SMILES

For each dataset I took either the CAS number (Escher) or the chemical name (Carthew) and retrieved the SMILES, QSAR-ready SMILES, and average mass from the CompTox Chemicals Dashboard.

I used CAS from the Escher paper because it actually returned more information than using the name they provided. It also had the added benefit of side-stepping the use of α- and o- in some chemical names by Escher et al. By doing this I retrieved data for all but 2 chemicals: Diphenylmethane diisocyanate and Dipropylene glycol monomethyl ether. For these 2 chemicals, I searched PubChem using the CAS & compared the resulting name/DTXSID and where there was a match I took the average mass value from PubChem.

After running the Carthew dataset through the dashboard, there were a number of instances where the chemical couldn't be found using the name present in the article. Where this was the case I searched for the chemical using PubChem, ChemSpider, or ECHA to see if there were synonyms that were present in the dashboard. These issues typically arose because a second name or abbreviation was added in parentheses to the end of the chemical name.

For 9 chemicals from the Carthew dataset I couldn't retrieve the SMILES, QSAR-ready SMILES, or average mass. Most of the chemicals that I couldn't retrieve this information for from the CompTox Chemicals dashboard were because they were typically mixtures (e.g. Mineral oils) or the chemical wasn't present in the dashboard (e.g. Benzene alkylate 225).

1.2.3 Re-calculating the 5th percentiles

Once I had that information for all the chemicals I could calculate the 5th percentile and TTC values from each paper (shown below in Table 1).

For the Escher dataset I calculated the 5th percentile and TTC values for general (lowest NOEC for either systemic or local effects), systemic, and local effects based on ppm and μg/person/day, respectively.

For the Carthew dataset I calculated the 5th percentile and TTC values for systemic and local effects. For the 5th percentile values systemic effects were measured in mg/kg-day and local effects were measured in mg/m^3.

Table 1: Comparison of Published and Recalculated 5th percentiles

Type of TTC	Cramer class	5th %ile units	Published 5th %ile	Recreated 5th %ile
Escher				
General	1	ppm	0.21	0.253
	2	ppm	0.028	0.026
	3	ppm	0.003	0.003
Systemic	1	ppm	0.25	0.557
	2	ppm	0.52	0.156
	3	ppm	0.006	0.004
Local	1	ppm	0.021	0.13
	2	ppm	0.028	0.019
	3	ppm	0.003	0.002
Carthew				
Systemic	1	mg/kg-day	0.41	0.44
	3	mg/kg-day	0.07	0.044
Local	1	mg/m3	1.4	1.229
	3	mg/m3	0.47	0.141
ToxVal				
General	1	ppm	NA	0.032
	2	ppm	NA	0.042
	3	ppm	NA	0.014

The NOECs present in the Escher and Carthew papers seem to already be adjusted to account for subacute/subchronic -> chronic extrapolation. Carthew et al, do actually state this in their article, but it's nice that we can verify it for ourselves.

For example, if we adjust the NOEC values reported in the Escher dataset for the Cramer class I chemicals and then calculate the 5th percentile, we end up with a 5th percentile value of 0.0837, which is much lower than the publish 5th percentile of 0.21.

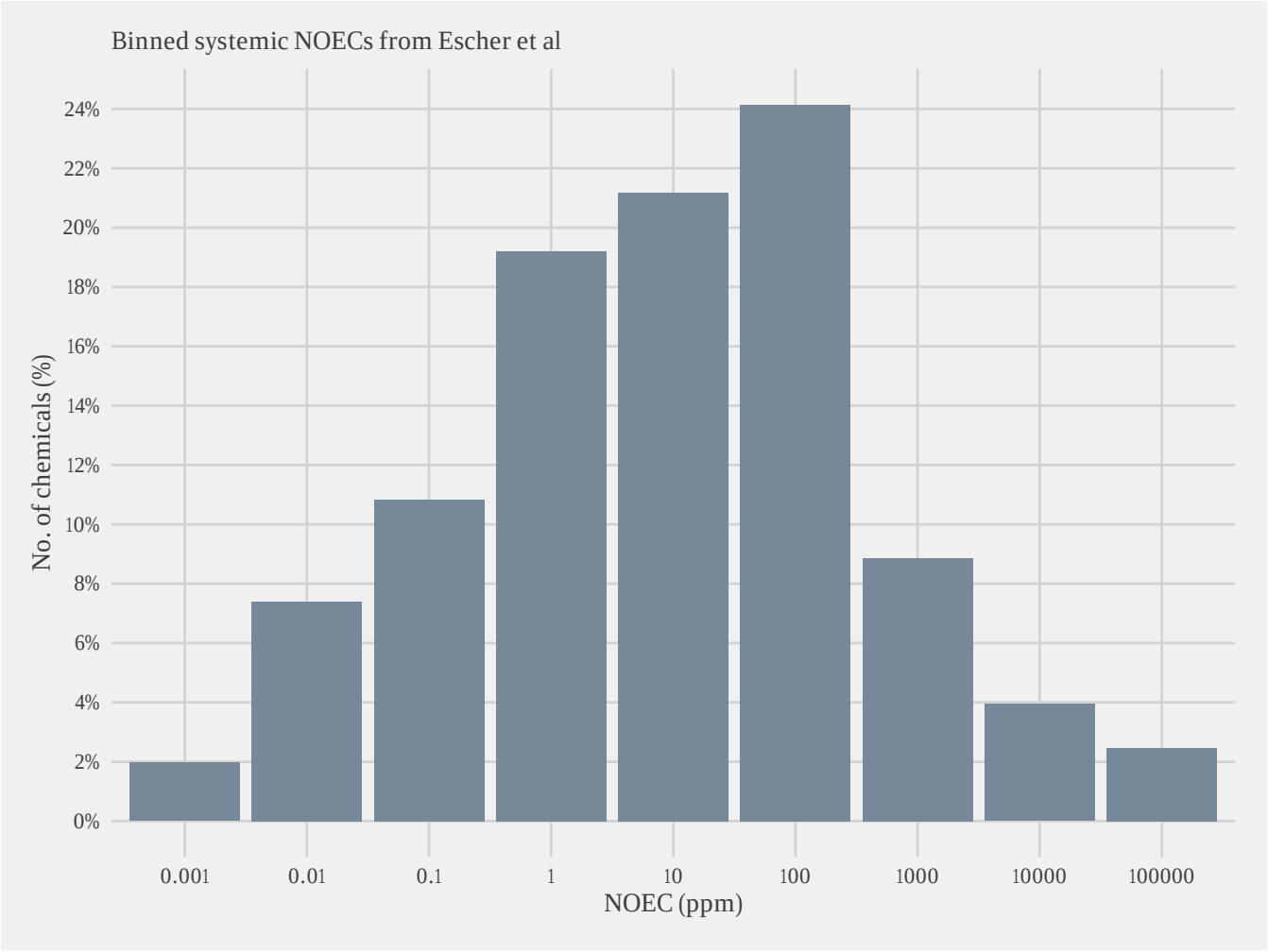

Figure 1: Distribution of 203 Systemic NOECs from Escher et al

Looking at the calculated distribution of the general NOECs in the Escher dataset (Figure 1) compared to the published version, they look practically the same. There is a slight difference in the % of NOECs that fall in the 1, 0.1, and 0.001 bins, but it is very minor.

Unfortunately, I don't have the distribution for the systemic NOECs, because this is where the biggest differences in 5th percentile values are. Therefore, I'm wondering whether the slight variation in terms of 5th percentile value is likely to be due to differences in implementation of the `fitdist()` function and whatever Escher and Carthew used to calculate their distributions. This is because (for the Escher dataset) the calculated geometric mean and median across the 3 effect types are the same as those published (e.g. Cramer class I, general effect type GM = 13.7 and Median = 23.6).

Additionally, for the Escher dataset we had to remove 2 chemicals from consideration for Cramer class III because they had a NOEC of 0.000 (I don't know what the actual NOEC was that Escher used in their calculation - maybe they also removed these chemicals?)

1.2.4 Re-calculating the TTC values

Now that we have the 5th percentile values (or at least our best approximation of them) we can use this information to calculate and compare the TTC values to the published values.

For the Escher dataset I could compare the TTC values in terms of both μg/person/day and mg/m^3, Carthew et al only provided their TTC values in μg/person/day.

Table 2: Comparison of Published and Recalculated TTC values

Type of TTC	Cramer class	Published TTC (mg/m^3)	Recreated TTC (mg/m^3)	Published TTC (μg/person/day)	Recreated TTC (μg/person/day)
Escher					
General	1	3.6e-03	8.02e-03	71	160
	2	4.8e-04	3.81e-04	10	8
	3	1.8e-04	1.66e-04	4	3
Systemic	1	4.8e-03	1.60e-02	95	320
	2	1.1e-02	2.75e-03	214	55
	3	3.2e-04	2.62e-04	6	5
Local	1	6.1e-04	4.42e-03	12	88
	2	4.8e-04	2.90e-04	10	6
	3	1.9e-04	1.04e-04	4	2
Carthew					
Systemic	1	NA	NA	980	1056
	3	NA	NA	170	106
Local	1	NA	NA	200	176
	3	NA	NA	67	20
ToxVal					
General	1	NA	1.24e-03	NA	25
	2	NA	1.52e-03	NA	30
	3	NA	6.28e-04	NA	13

Obviously, the discrepancies in 5th percentile values then get carried over into the TTC values themselves (Table 2).

Additional variation to the TTC in terms of μg/person/day could also be coming from converting the values from ppm to mg/m^3. To convert from ppm to mg/m^3 I've used the average mass from the CompTox Chemicals Dashboard, but these values may differ from those used by Escher et al. (especially in terms of the number of significant figures).

1.3 Comparing ToxPrints between our inhalation TTC dataset, the Escher dataset, and the Carthew dataset

Figure 2 shows a comparison of the relative frequencies of the ToxPrints for the chemicals present in the ToxVal inhalation, Escher, and Carthew datasets. The chemicals that we considered part of the ToxVal inhalation dataset were those that met the filtering criteria at the top of this document.

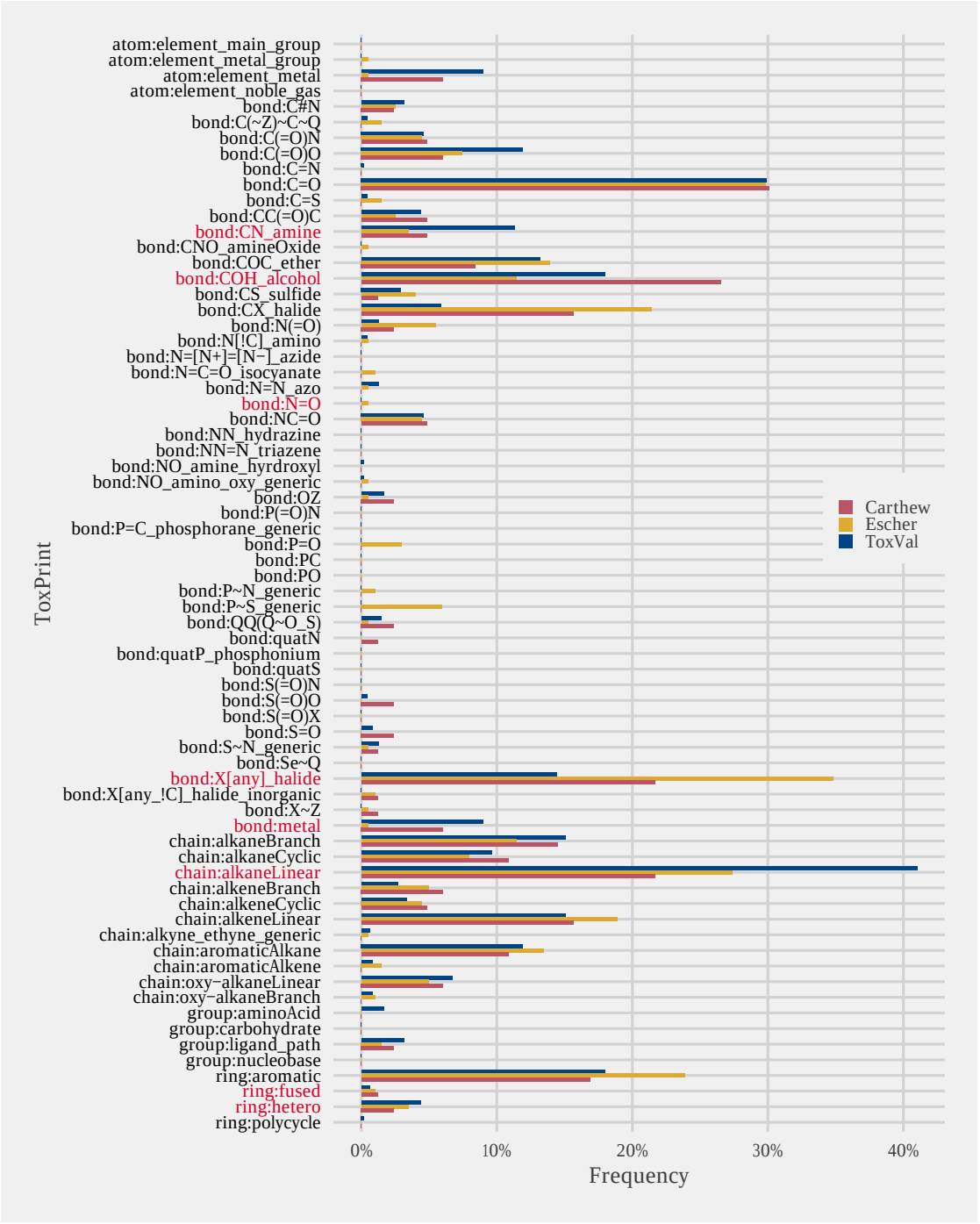

Figure 2: Comparison of frequency of ToxPrints in ToxVal inhalation, Escher, and Carthew datasets

1.3.1 Differences in chemotype proportions

Metal-containing chemicals
As you can see above Carthew and ToxVal both had larger proportions of metal containing compounds (atom:element_metal and bond:metal), but these proportions are still pretty low at 6% and 9%, respectively. This is compared to 0.5% for Escher. However, all five of the Carthew, and most of the ToxVal (50/54), metal-containing chemicals are flagged because they contain silicon: the remaining ToxVal chemicals contain boron.

Amine-containing chemicals
About 11% of chemicals in ToxVal contain an amine group (bond:CN_amine). This is in comparison to 3.5% and 4.8% for the Escher and Carthew datasets.

Alcohol-containing chemicals
Carthew has a much larger proportion of alcohol-containing compounds (27%) than both Escher (11%) and ToxVal (18%).

Halide-containing chemicals
The Escher dataset contains more chemicals with a halide atom (~35%). This is compared to 22% of chemicals in the Carthew dataset and 14% in the ToxVal dataset. Most of these halide atoms seem to be bonded to carbons, rather than being inorganic, as can be seen when looking at the "bond:CX_halide" row.

Nitroso-containing chemicals
Toxtree seems to do a good job at removing all nitroso-containing chemicals (bond:N=O), as there are 0 chemicals in the ToxVal inhalation dataset that contain a nitroso group. Additionally, Carthew and Escher also seem to have done a good job of manually removing nitroso-containing chemicals: only Escher contains a chemical that falls into the bond:N=O Level 2 chemotype.
However, upon closer inspection, this chemical is actually a nitrite- and not a nitroso-containing chemical.

Alkanes/alkenes
All 3 datasets seem to have roughly the same proportion of branched and cyclic alkane and alkene containing chemicals. ToxVal does have a relatively larger frequency of linear alkane-containing chemicals (41%) than the Carthew and Escher datasets (22% and 27%, respectively). Additionally, all 3 datasets seem to have relatively few chemicals with fused/hetero rings.

1.4 Comparing local and systemic NOECs

Now, let's plot each of the chemicals with both a local and systemic NOEC. There are 102 of the 203 chemicals present in the Escher dataset that contain both local and systemic NOEC values. Meanwhile, the entirity of the Carthew dataset contain both local and systemic data.

Be aware that unlike the Escher data which are all in ppm, the Carthew data is mg/m³ for the local effects and μg/kg-day for the systemic effects.

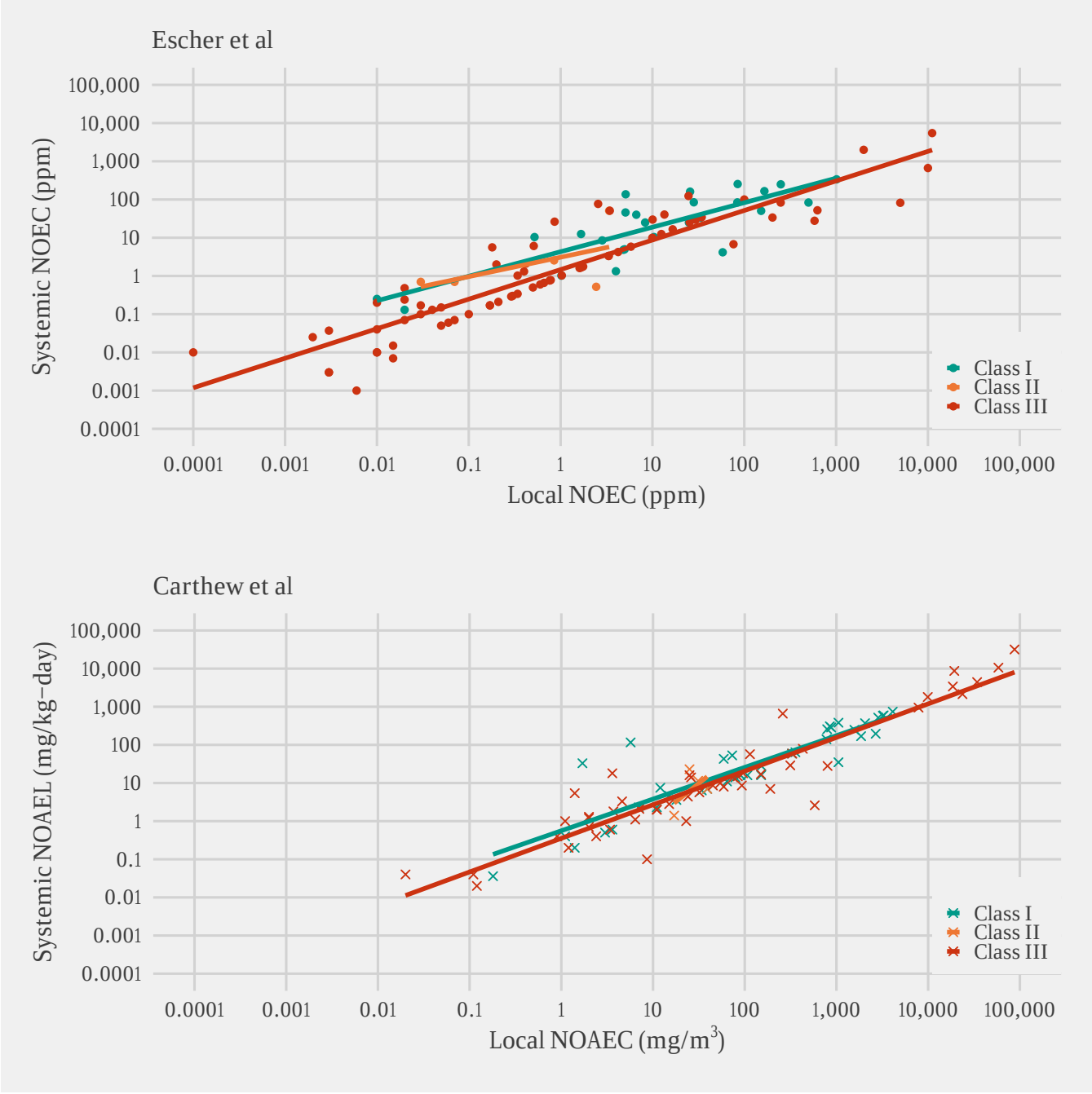

Figure 3: Relationship between Local and Systemic Inhalation NOECs from Escher et al (top) and Carthew et al (bottom)

As you can see from Figure 3. the local and systemic NOECs from both datasets seem to follow a linear trend, irrespective of the Cramer class.

It's all well and good plotting the two datasets onto separate graphs but, even having them grouped together as above, it doesn't enable us to easily compare the data: the fact that they're in different units doesn't help either.

So, I took the NOECs from the Escher dataset and adjusted them to be in the same units as the Carthew data (i.e. local effects in mg/m^3 and systemic effects in $\mu g/kg\text{-}day$) and plotted them together. Just be aware that, as in section 1.3.3, the adjustments to mg/m^3 and $\mu g/kg\text{-}day$ may have changed some of the values from what they "should" be (if they were to have been calculated by Escher et al), but any differences should be minor.

There are a total of 37 chemicals that are present in both datasets; thankfully, they were all categorised as having the same Cramer class.
Of these 37 chemicals, the Escher dataset had both systemic and local NOEC data for 19 chemicals (all the chemicals in the Carthew dataset have both local and systemic data).

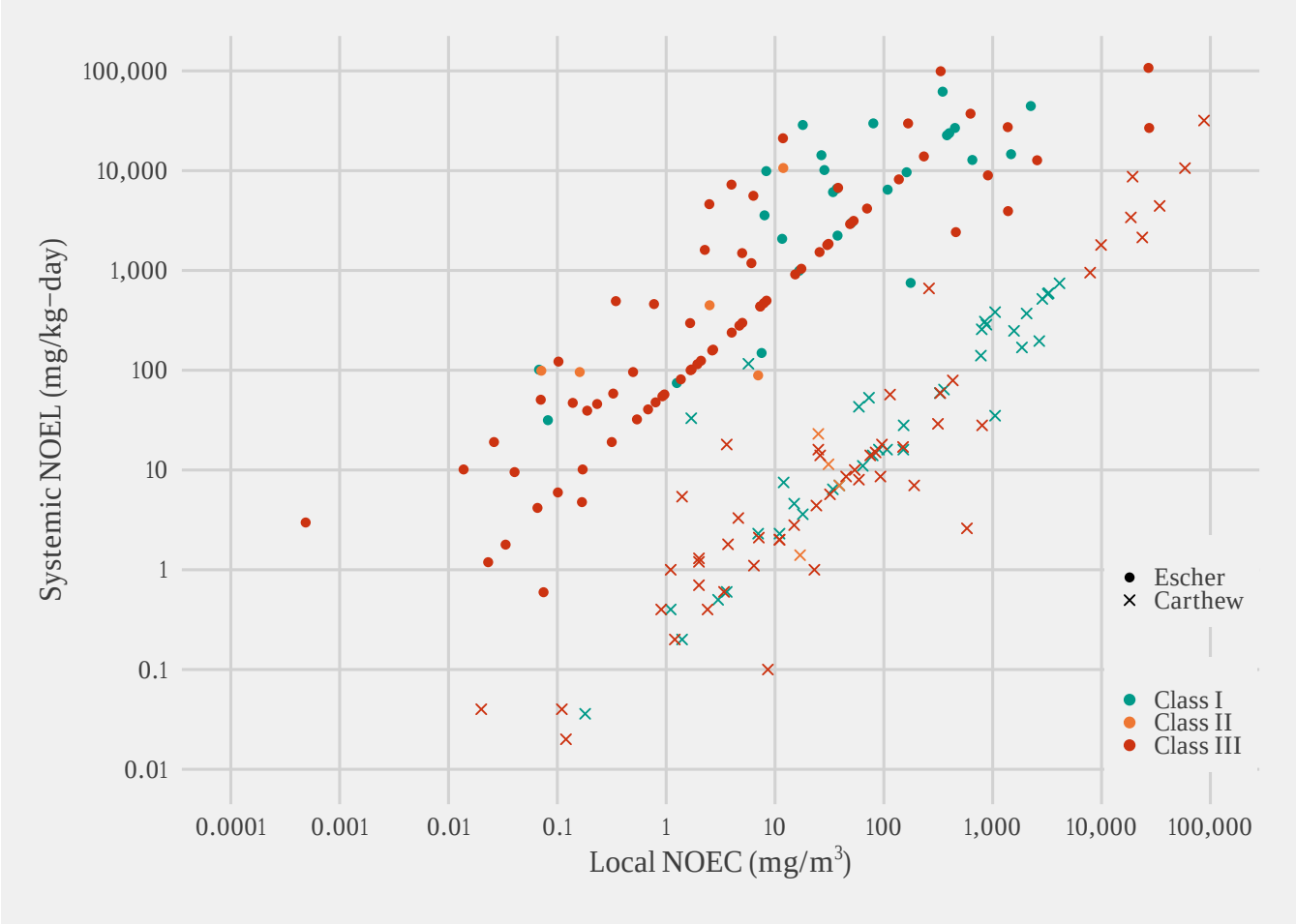

Figure 4: Comparison of Local and Systemic NOECs/NOELs from both datasets

As you can see in Figure 4, both datasets seem to follow a very similar positive linear trend, with the intercept of the Escher dataset being slightly higher (i.e. less toxic) than that of the Carthew dataset.
Maybe that is a bit unexpected given that the Carthew dataset is mainly chemicals that are present in consumer aerosols, whilst the Escher dataset consists of mainly industrial chemicals with some pesticides.

Perhaps more surprisingly, from Figure 4, the Carthew data seems to suggest that in the majority of cases a chemicals local effects are likely to occur **after** the systemic effects (at about a x10 greater "concentration"); whereas, the Escher data suggest that in the majority of their cases the local effects are more likely to occur **before** the systemic effects (at about a x100 lesser "concentration").
Although, this may be driven by the fact that the local and systemic effects are being measured in different units. If you look at the top scatter plot in Figure 3, you'll see that the local effects occur before the systemic effects at low concentrations and this flips as you get to the extremely high concentrations. The majority of chemicals have local and systemic effects occurring at roughly the same

concentration.

To see what effect, if any, having everything in the same units makes I converted the Carthew systemic NOELs from μg/kg-day to mg/m^3 using this equation:

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

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

It looks a lot more complicated that it is, it's basically taking the reciprocal of calculating the NOEL in μg/kg-day.

Next, I plotted the local and systemic NOEC data for both the Carthew and Escher datasets.

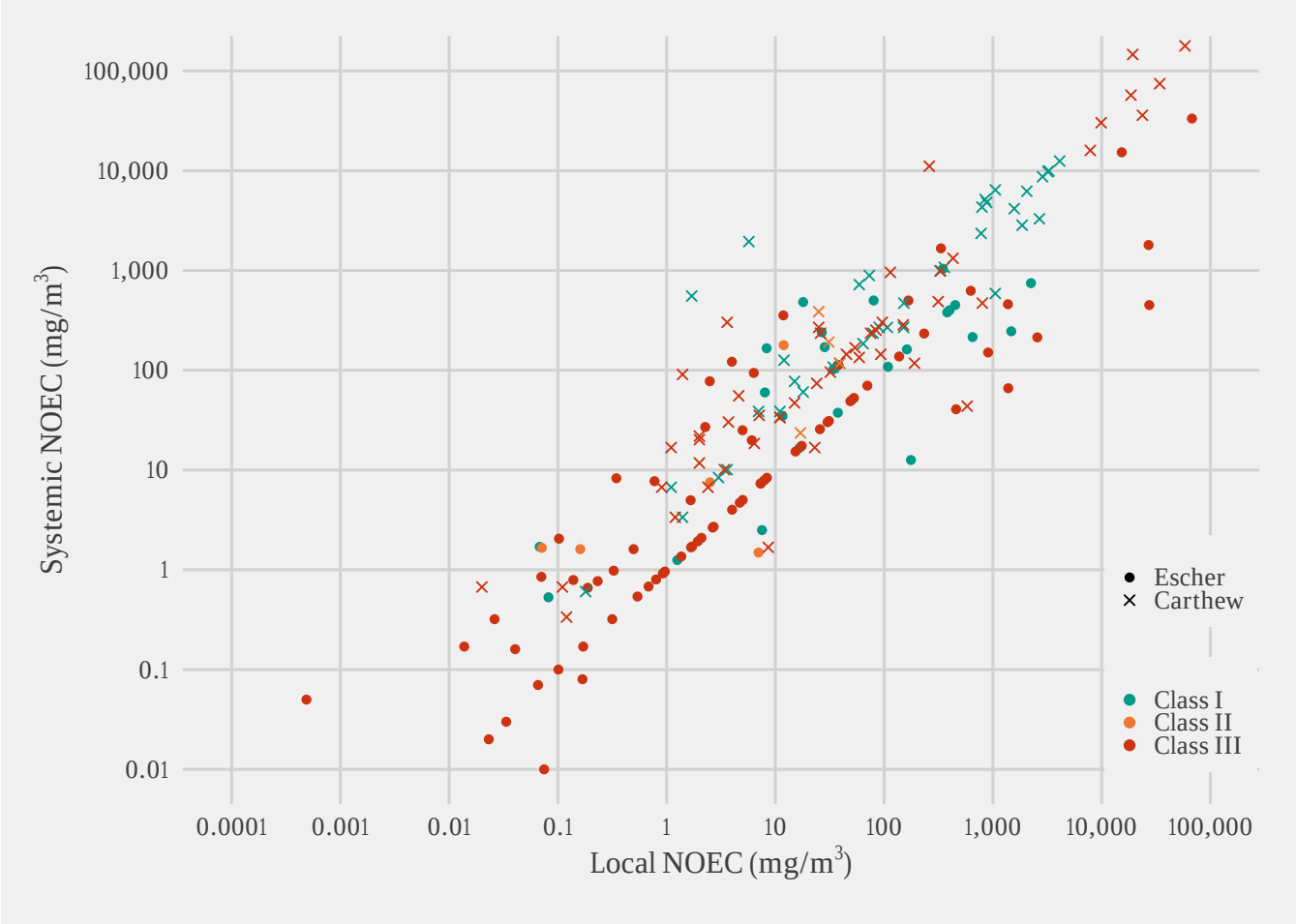

Figure 5: Comparison of Local and Systemic NOECs from both datasets

As you can see from Figure 5, after transforming the data into the same units the Escher data typically have the local and systemic effects occurring at roughly the same concentrations, whilst the Carthew data typically have the local effects occurring at a slightly lower concentration than the systemic effects.

2 Acute TTC

2.1 Questions from last meeting

There were a couple of questions I've not yet answered from a little while ago:

- Is there a constant factor the 5th percentiles are adjusted between the acute and chronic oral data?
- What is the relationship between Cramer class and GHS category?

2.2 Constant factor in 5th percentiles between acute and chronic oral?

Table 3: Comparison of 5th percentiles for Acute and Chronic oral data

Cramer class	Calculated Acute 5th %ile (μg/person/day)	Calculated Chronic 5th %ile (μg/person/day)	Munro 5th %ile (μg/person/day)
Class I	159.405	3.734	3
Class II	131.107	3.459	0.91
Class III	33.879	0.394	0.15