Ranking chemicals based on heterogeneous data

Jan 2017

Ranking chemicals based on heterogeneous data

- Different herbicide screens: A, B, C
  - Differ in scale, amount of test chemical required, duration
  - Species tested, eg tropical grasses vs temperate grasses, or broadleaved vs grasses
  - Application timing: seed *i.e*. pre-emergence ***vs*** seedling *i.e*. post-emergence
- Each screen incudes several plants of several species
- Each chemical is tested at a few application rates
- Chemicals are run through screens in batches
- There is run to run variation in results of a screen
- The percent control of each species is scored, ie 0% = the chemical did nothing, 100% = the chemical completely killed all plants. These data are used to fit a logistic regression model and an ED50 (effective dose, 50%) number is calculated

How can we rank all the tested chemicals across a series of screens?

How can we take uncertainty into account?

Ranking chemicals based on heterogeneous data

- Results for 1 chemical in 1 screen

	Species	A	B	C	D	E	F	G	H
Score at each application rate	100g/ha	10	0	0	0	20	0	0	0
	500g/ha	20	40	30	0	80	10	20	0
	1000g/ha	30	90	75	10	100	50	40	10
	ED50	1500	600	750	>1000	280	1000	1200	>1000

Ranking chemicals based on heterogeneous data

Chemical	Screen P1	Screen Q1	Screen R1	Screen P2
Standard				
Best				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				

syngenta

Ranking chemicals based on heterogeneous data

Chemical	Screen P1	Screen Q1	Screen R1	Screen P2
Standard				
Best				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				

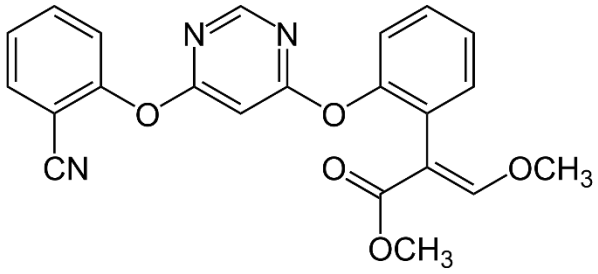
- Traditional approach
  - Pairwise comparisons with “best”
  - If better than the “best” then you have a new winner
- But now we want to rank all chemicals tested
  - To spot trends
  - To better understand chemical space
  - To direct chemistry towards better areas

Ideas for literature starting points

- One of FIFA's systems for ranking national football teams: https://en.wikipedia.org/wiki/FIFA_World_Ranking_system_(1999%E2%80%932006)
- The ELO system used for chess ratings: https://en.wikipedia.org/wiki/Elo_rating_system this article has lots of references including some statistical literature
- "The predictive power of ranking systems in association football" Lasek et al DOI: 10.1504/IJAPR.2013.052339
- Major differences for our case are that
  - we have multiway comparisons, not just two way
  - results have uncertainty (eg chemical A score 80% and B score 70%, but these figures should really have error bars, whereas if Team A beat Team B this is unambiguous)
  - we have a lot of missing data in the overall matrix

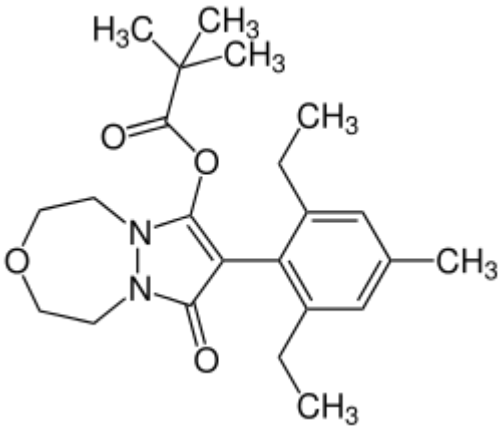
Formulation toxicity

- Syngenta designs, makes and tests organic chemicals, to seek to invent new active ingredients (AIs)



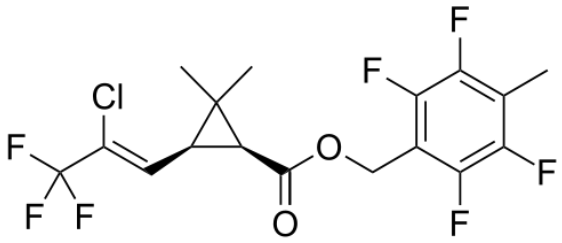
azoxystrobin

fungicide



pinoxaden

herbicide



tefluthrin

insecticide

syngenta

Formulation toxicity

- But we not sell active ingredients – our products are formulations

Quadris
azoxystrobin
A3245B

Axial 45 EC
pinoxaden
A6897D

Force 0.5 GR
tefluthrin
A3214A

Formulation toxicity

- Formulations (formulated products) may contain many ingredients with different roles:
  - 1 or more AIs
  - Everything else, which are collectively called co-formulants
    - solvents, surfactants, preservatives, colour, stabilisers, anti-oxidants, anti-foams, sunscreens
- There is toxicity data for many ingredients, but it is often only available to as a range eg MLD of 300-2000 mg/kg, > 5000 mg/kg
- For every formulation we either test its acute toxicity (6-pack), or we bridge to data for a similar formulation
  - Acute oral toxicity
  - Acute dermal toxicity
  - Acute inhalation toxicity
  - Eye irritation
  - Skin irritation
  - Skin sensitisation (an allergic response)

Formulation toxicity

- Industry needs to generate information for every formulation, but…
  - Most new formulations are variants of existing ones
  - There is a finite list of ingredients (our cupboards contain much the same set of ingredients), many of which have been tested singly somewhere sometime
  - Bridging arguments, read-across, *in vitro* and *in silico* methods are great, but have had incremental benefits, *i.e.* we still test a lot of animals
- It is time for the regulatory science to come together and end the great majority of acute animal studies for pesticide formulations
- Syngenta and Dow Agroscience are sponsors of an NC3Rs CRACK IT project to seek to achieve this:

  Maximise: maximising confidence whilst minimising data generation for acute hazard classification of mixtures
- However, there is a maths challenge here too, which could help reduce animal testing…

Formulation toxicity

		AI	AI	AI	Solvent	Solvent	Solvent	Surfactant	Surfactant	Surfactant	Anti-foam	Anti-foam	Anti-foam
		1	2	3	4	5	6	7	8	9	10	11	12
Formulation	1	■			■				■		■		
Formulation	2	■				■		■					
Formulation	3	■			■		■						
Formulation	4		■			■							
Formulation	5		■			■						■	
Formulation	6		■				■						
Formulation	7		■		■				■				
Formulation	8			■	■								
Formulation	9			■	■	■							■
Formulation	10			■		■		■					
Formulation	11			■			■					■	
Formulation	12			■			■						
Formulation	13			■						■			
Formulation	14			■									

- We know the toxicity of all the formulations
- We can assume additive toxicity for formulation ingredients
- We know the toxicity of some ingredients tested singly (either as a number or as a range)
- How can be back out toxicity estimates for as many ingredients as possible?
- How can we spot non-additive effects?

More about acute toxicity

- Measures of acute toxicity vary in nature.
  - For acute oral toxicity, the Median Lethal Dose (MLD) is used, which in principle is on a continuous scale.
    - 327 mg/kg – interpolated from results at a range of doses
    - or >3000 mg/kg – no toxicity seen at the top dose (perhaps the only one tested)
    - or <200 mg/kg – lots of toxicity seen at the lowest dose (perhaps only the only one tested)
  - If the data is not ours but someone else's, then often only the resulting acute toxicity class is available, eg Category 4, meaning 300-2000 mg/kg.
  - For other types of acute toxicity, eg skin irritation, usually only class data is available (non irritant vs irritant vs severe irritant), and toxicity on a continuous scale could be thought of as a hidden variable.

More about acute toxicity

- An ATE (acute toxicity estimate) is an estimate of acute toxicity of a formulation based on the toxicity of its components. Toxicity of a mixture is obtained by adding the toxicity of each component, taking into account the proportion of each component. But to make it additive you need to take the inverse of the toxicity value, eg MLD (an MLD of 5 mg/kg is very toxic and 5000 mg/kg is very non-toxic).
- Significant deviation from additivity is rare, and is either synergism (1+1>2) or antagonism (1+1<2). Acute oral tox categories and ATE calculation is shown below.

Exposure routes	Classification category or experimentally obtained acute toxicity range estimate (*see Note 1*)	Converted acute toxicity point estimate (*see Note 2*)
Oral (mg/kg bodyweight)	0 < Category 1 ≤ 5	0.5
	5 < Category 2 ≤ 50	5
	50 < Category 3 ≤ 300	100
	300 < Category 4 ≤ 2000	500
	2000 < Category 5 ≤ 5000	2500
Dermal (mg/kg bodyweight)	0 < Category 1 ≤ 50	5
	50 < Category 2 ≤ 200	50
	200 < Category 3 ≤ 1000	300
	1000 < Category 4 ≤ 2000	1100
	2000 < Category 5 ≤ 5000	2500
Gases (ppmV)	0 < Category 1 ≤ 100	10
	100 < Category 2 ≤ 500	100
	500 < Category 3 ≤ 2500	700
	2500 < Category 4 ≤ 20000	4500
	Category 5 - See footnote to 3.1.2.5.	
Vapours (mg/l)	0 < Category 1 ≤ 0.5	0.05
	0.5 < Category 2 ≤ 2.0	0.5
	2.0 < Category 3 ≤ 10.0	3
	10.0 < Category 4 ≤ 20.0	11
	Category 5 - See footnote to 3.1.2.5.	
Dust/mist (mg/l)	0 < Category 1 ≤ 0.05	0.005
	0.05 < Category 2 ≤ 0.5	0.05
	0.5 < Category 3 ≤ 1.0	0.5
	1.0 < Category 4 ≤ 5.0	1.5
	Category 5 - See footnote to 3.1.2.5.	

Ingredient(s) with unknown toxicity is <= 10 %;

$$\frac{100}{ATEmix} = \sum_{n} \frac{C_i}{ATE_i}$$

Ingredient(s) with unknown toxicity is >10 %;

$$\frac{100 - (\sum C_{unknown} \text{ if } >10\%)}{ATE_{mix}} = \sum_{n} \frac{C_i}{ATE_i}$$

C_i = concentration of ingredient i;

n ingredients and i is running from 1 to n;

ATE_i = Acute toxicity estimate of ingredient i;

syngenta