Abstract no. 1689

Validating proposed improvements to phototoxicity expert alerts against proprietary data

David J Ponting*, Martin P Payne*, Helga H J Gerets†, James S Harvey‡, Jonathan R Howe‡, Alexis D C Parenty§, Angela T White‡, Richard V Williams*, **Donna S Macmillan*.**

** Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, United Kingdom; † UCB BioPharma SRL, Development Science, Chemin du Foriest, 1420 Braine L'Alleud, Belgium*

‡ GlaxoSmithKline, Park Road, Ware, Hertfordshire, SG12 0DP, United Kingdom; § Novartis Institutes for Biological Research, Department of Preclinical Safety, Klybeckstrasse, Basel, Switzerland

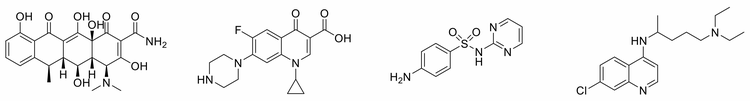
Introduction

Development of expert alerts for toxicological endpoints is a time-consuming process and subsequent improvements are often made after analysis of new data. If refinements are requested based on information from proprietary data, this poses a problem, as this data may be confidential and cannot be shared with the original implementer. We present a method for the acceleration of this process, analogous with the student-teacher model[1] used for statistical methods in similar situations.

The endpoint of phototoxicity was chosen to demonstrate this due to its increasing regulatory concern for pharmaceuticals (ICH S10[2]).

Curation of public data

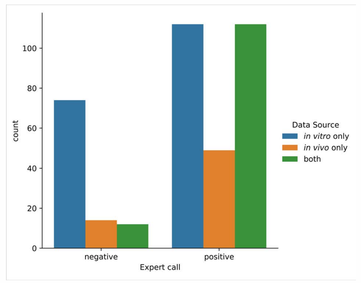
A dataset of 400 publicly available structures with phototoxicity data was collated. This includes a number of drugs on the WHO list of essential medicines with known phototoxic potential, such as Doxycycline (**1**), Ciprofloxacin (**2**), Sulfadiazine (**3**) and Chloroquine (**4**).



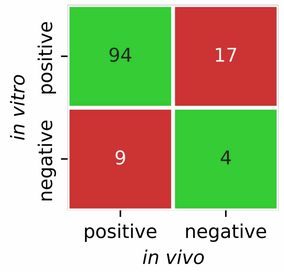
1 **2** **3** **4**

Data from a number of different species and assays were curated together; this includes systematic studies in humans and pre-clinical animals, case reports and *in vitro* assays including the 3T3 NRU (OECD TG432[3]). An expert call was derived from this data, prioritizing human data over pre-clinical and *in vivo* over *in vitro*. The composition of the dataset is described below.

Dataset composition



In vitro/in vivo differences



References

1: Papernot *et al* (2017), arXiv:1610.05755v4 [stat.ML]

2: ICH Guideline S10. Available at https://database.ich.org/sites/default/files/S10_Guideline.pdf

3: OECD (2019), *Test No. 432: In Vitro 3T3 NRU Phototoxicity Test*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris. Available at https://doi.org/10.1787/9789264071162-en.

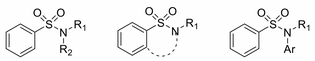
4: Bickers *et al* (2005), *Food. Chem. Toxicol.*, **43**, 799-836

Key toxicophores

Working from the knowledge base of expert alerts contained within Derek Nexus 2018.1 (Lhasa Limited, Leeds, UK), a number of alerts/toxicophores were created or modified on the basis of the public data previously curated. Some key toxicophores are given below.

❖ Aryl sulfonamide

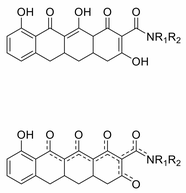
The original alert in Derek Nexus for aryl sulfonamides was broad (**5**), giving a positive predictivity of 77%. Analysis of the new phototoxicity data indicated that certain sub-classes contain features which reduce the rate of the critical S-N bond cleavage. Separating these increased the positive predictivity of the original alert – now "**5** except (**6** or **7**)" – to 95% and creates two additional alerts (**6**,**7**) with 50% positive predictivity and a correspondingly lower level of reasoning.



R1,R2 = C,H **5** R1 = C,H **6** R1 = C,H **7**

❖ Tetracycline

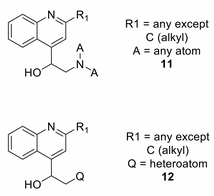
The phototoxicity of tetracyclines is well understood, and no changes are needed to the intended scope (**9**) of alerts for them; however, the number of potential tautomers challenges any structure representation or fingerprinting method. This alert and related ones have therefore been modified to allow single/double bonds (as **10**).



R1, R2 = C, H **9**

R1, R2 = C, H
5 O atoms are -OH or =O
10

❖ 4-Quinolinyl methanol



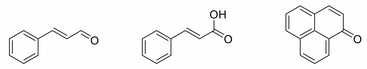
R1 = any except C (alkyl)
A = any atom
11

R1 = any except C (alkyl)
Q = heteroatom
12

For these compounds, a very different concern needed to be raised – the majority of public data for this class comes from a non-standard assay with limited concentration of test compound and range of wavelengths of light investigated; negative data points in this assay may be unreliable and the scope has been expanded (from **11** to **12**).

❖ α,β-Unsaturated β-aryl carbonyl

This new alert captures substituted derivatives of cinnamaldehyde (**13**) and cinnamic acid (**14**) with further conjugation and thus enlarged chromophores. These compounds, without additional substitution, have been reviewed as non-phototoxic[4], however derivatives such as phenalenone (**15**) are observed to be.



13 **14** **15**

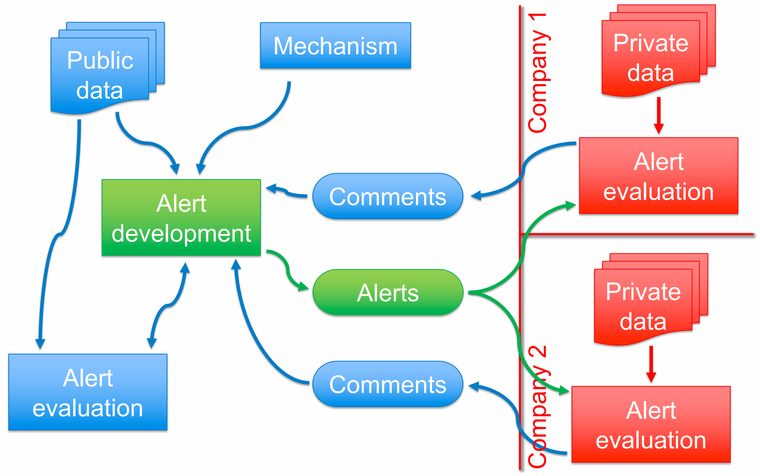
Performance against public data

Alerts	BA	MCC	F-score	Sens	Spec	PPV
Old	76%	45%	75%	63%	87%	92%
New	78%	53%	83%	79%	79%	90%

BA = Balanced Accuracy, MCC = Matthews Correlation Coefficient, Sens = Sensitivity, Spec = Specificity, PPV = Positive Predictivity

Validation against proprietary data

The proposed alert revisions were sent to a selection of pharmaceutical companies and processed against their internal proprietary datasets for phototoxicity, with the following process:



Key evaluation comments from the members concerned are:

- The changes to the aryl sulfonamide alert reduced the prediction of some false positives that could not be entirely excluded from plausible to equivocal, while also flagging a number of as-yet-untested compounds as concerning.
- A new alert for α,β-unsaturated β-aryl carbonyls caught a number of compounds that one company had tested to be positive in the 3T3 NRU assay at the cost of creating new false positives with another company's test results.

The implementation of these results – both analyzing the results from the different datasets and incorporation into Derek Nexus – is ongoing.

Conclusions

- Expert alerts benefit from periodic review against both newly available public and proprietary data
- Significant performance improvements against public data may be achievable without degrading performance against proprietary data, and *vice versa*
- Utilization of multiple proprietary datasets yields additional scope refinement