

The Robot Toxicologist

The “Toxic Trio” as a case study

Wouldn't it be wonderful if a machine could read all of the world's science literature, decide which substance would trigger new liability exposures, say how much this would cost and who should pay?

After > 10 years of development work, the recent marketing document¹ from Allianz illustrates how far along this path one particular robot has travelled.

UK liability insurers read the report and asked –‘is it better than tossing a coin’? 51% is seen as the minimum requirement for authorising reserves for example.



The task was to compare the fifteen substantial findings in the report (in the context of nail varnish) with the written views of expert toxicology committees produced over several decades.

Is this a fair test?

One of the key features of any advisory tool is the ability to quantify and report the number of true positives, true negatives, false positives and false negatives. It is traditional that such thoroughness is absent from a marketing document. To improve fairness, the comparison work also explored the number of true negatives (things which were justifiably not included) and false negatives (things which should have been in a truly informative report, but weren't).

Fifteen comparisons may not be enough to provide a really clear view of the 50% test. Precision in the probability estimate would be made more meaningful by a larger sample.

The findings

There was some overlap between expert and robot views. A positive predictive value (PPV) was estimated with a precision of $\pm 10\%$ using Monte Carlo methods. With such a precision, one can confidently distinguish between an ideal threshold PPV of 50% and a measured value of 80% and between a measured value of 20% and the ideal 50%. Measured PPV values below 20% and above 80% would provide strong guidance on the usefulness of the Allianz *et al.* robot (AR).

Why were there any differences?

There were five key areas where the ‘quality’ assessment made by the robot disagreed with the quality assessment and insight normally provided by an expert. These are provided in the full report. For example, is it really true that the effects of

¹*Risk Bulletin*. (2018) Vol.3. Allianz Global Corporate & Specialty. The analyses have been brought to the attention of both Allianz and the robot developer. The robot is referred to here as AR.

massively unrepresentative doses will be of the same kind at much lower doses?
Sometimes. It depends.

In the course of the comparison it became clear² that the AR is capable of rooting out and collating some of the evidence which is relevant to an expert liability evaluation. This suggests a promising future for AR.

The AR seems to behave as a semantic popularity poll. However, in expert hands one high quality study trumps the editorial line chosen for various reasons³ to promote tens of poor ones even if they all agree. Popularity poll analysis has no place in court.

Was it a useful question?

The 'Tossing a coin' comparison is often a good test of value-added and is the gateway to formal action in liability risk management business. However, for liability exposure, the potential value of one prediction out of twenty five being correct could very significantly outweigh the time and money wasted on preparing for and investing in another twenty four. Failure to meet the 50% challenge wouldn't necessarily mean all the effort is always wasted. Above 50% and the economic argument looks much more favourable.



For those regimes that adopt a more precautionary standard e.g. environmental regulators and risk averse "failure to warn" regimes, the 50% PPV test is probably too high a hurdle. But it is the right question for common law liabilities.

Summary

Based on fifteen comparisons, an estimate of the positive predictive value was made with an uncertainty of $\pm 10\%$.

Based on the available evidence, would tax authorities, shareholders and insurance regulators give assent to actions justified by the outputs of this AR tool? A positive predictive value below 20% would suggest – No. Above 80% - Yes.

The detailed report provides the PPV value for the 'toxic trio' report and is commercially available in confidence to UK liability insurers and re-insurers. The potential for robots to be of use in toxicology for liability risk assessment is discussed.

andrew@reliabilityoxford.co.uk

²There is the example of a speculative effect of a trace contaminant in toluene (one of the "toxic trio") being used as evidence of synergy. Benzene is a trace contaminant in toluene. The Allianz *et al.* report didn't say that it was but did use benzene as an example where there is tentative evidence of accelerating the possibly harmful biological effects associated with formaldehyde (another one of the trio). It is hard to believe that this example was chosen by good luck, but when asked, the authors made no comment.

³ Science publication is the tradeable currency of research scientists. As with any form of commerce sometimes corners are cut and words are chosen to increase the apparent value. Peer reviewers are usually members of the same club.