Challenge 28: RespiraTox Surgery Q&As

Q. Is there data available from the Sponsors to train and test the *in silico* model? If so, how much is there and what format is it in, e.g. raw, curated?

A. Human acute inhalation data from a limited number of studies is available from Shell, with subjective and objective measurements of respiratory irritation. The chemicals tested are single chemical entities.

Q. Is there other data available from the web (e.g. repositories, open source) or other companies or organisations which could be used in the project?

A. Published studies contain human and *in vivo* datasets which could be used to develop and test the model:

- Estimating sensory irritation potency of volatile organic chemicals using QSARs based on decision tree methods for regulatory purpose. Gupta et al., Ecotoxicology. 2015 May;24(4):873-86. doi: 10.1007/s10646-015-1431-y. The paper proposes robust and reliable quantitative structure-activity relationships (QSARs) for estimating the sensory irritation potency and screening of the volatile organic chemicals. Approximately 145 chemicals were used in the dataset.

- Development of a database for sensory irritants and its use in establishing occupational exposure limits. Schaper et al., Am. Ind. Hyg. Assoc. J. 1993, 54:488–544. - *295* Airborne materials (single chemicals and mixtures) were obtained from the literature. 154 RD_{50} values were obtained *in vivo* for the 89 chemicals with TLVs (threshold limit values).

- 2016 Society of Toxicology meeting:
  - QSARs for the fiftieth anniversary of the RD50
  - http://www.toxicology.org/education/docs/Alarie_QSAR_RD50%20Lecture_SOT.pdf
  - http://www.toxicology.org/education/docs/Alarie_Table_2_2015_List_of_TLVs_for_chemicals_with_URT_irr_basis_with_no_RD50_values.pdf

- Data is also available from ECHA REACH:
  - https://www.echa.europa.eu/web/guest/information-on-chemicals/registered-substances
  - https://echa.europa.eu/-/data-on-15-000-chemicals-now-available-to-use

- ACGIH TLV (American Conference of Governmental Industrial Hygienist):
  - http://www.acgih.org/tlv-bei-guidelines/policies-procedures-presentations

- US EPA, Japanese OEL database, or Pubmed.

The literature contains anecdotal evidence, and this would need to be mined/curated to be used. Data ownership issues would need to be confirmed, though ECHA data has been used before to underpin commercially available models/software.

The Sponsors are happy to collaborate with the Contractors to help identify these additional sources of information/data, and build upon the current models.

Q. Are the available data sets compatible, or in the same format?

A. No, the datasets would need to be curated, transformed and reformatted to allow for use in the model. Contractors would need to use appropriate tools, or develop new tools to deal with non-uniform datasets.

Q. Does the human acute inhalation toxicity data and the animal data described above align or use the same ontology?

A. Some similar endpoints may be used, but the datasets are arranged on different ontologies. Therefore, comparison of the two will require transformation of the datasets to align.

Q: Do we want to predict the rodent (i.e. model on rodent data) or the human?

A. Having a model to predict respiratory irritation of a chemical in rodent may have 3Rs benefit, as this could remove the need for some animal studies. However, we would want to be able to predict irritation in man, as rodents are fundamentally different to humans, including nasal architecture and lung structure. This may affect how useful the animal, and thus the *in silico* model is to examine human toxicity.

Q: How do the Sponsors expect the new *in silico* models to be developed? For example, using a structured approach based on causality? Or an approach based on interrogating and parsing experimental data to create correlations?

A. This depends on how the team chooses to tackle the problem, but it may be useful to apply structure and physiochemical characteristics (such as acid-base reaction information, solvent properties) and/or -omics data in the model.

Q: How much does any approach have to link into QSARS? Is this a critical requirement?

A. This is a requirement to develop an appropriate QSAR.

Q: How do you define a mixture – how many chemicals would it contain? Would these be chemicals with known or unknown toxicities?

A. Any amount of constituents would give insight into how to model toxicity of mixtures, even with 2-3 chemicals of known irritation potential, but essentially as many as possible. It would be the Holy Grail to be able to test a mixture of chemicals of unknown toxicity and define its irritancy, but lower numbers of chemicals with known respiratory irritation characteristics would be a good starting point.

Q: Is there any data available from the Sponsors or other sources on mixtures?

A. A limited dataset will be available from the Sponsor. The Contractors would need to find additional data sources to develop and test the model. The Sponsors are happy to collaborate with the Contractors to help identify these additional sources of information/data.

Q: Would a mixtures model need to be predictive for individual components or synergistic behaviours?

A. Whatever is feasible. Initially the primary focus should be on individual components.

Q: What do you see as the output of the project? Would it be a finished product, or a series of structural alerts or code which could be incorporated into other systems?

A. In one previous challenge, the QSAR-type *in silico* model has been built and used to underpin a service delivered by a company (KREATiS). In this case, the final product could be web based, an app/standalone piece of software, a service, or whatever best suits the business objectives of the contractor team. The solution would need to deliver the technology to the community, i.e. through commercialisation.

Q: What performance criteria do you use to define 'predictive'?

A. Internal and external validity, and ultimately regulatory confidence in the prediction given by the model.