Datasets

Case study: Research & Development

Helping a top-10 pharma develop an organoid-based *predictive toxicology model*

ELSEVIER

Advancing human progress together

Introduction

Historically, companies have performed animal testing to identify safety issues pre-clinically. However, depending on the nature of the adverse event, animal safety data may not translate very well to humans. In addition, animal research is incredibly costly, and there are ethical considerations concerning the use of experimental animals.

The growing demand to identify safety issues at an earlier stage of drug development is limited by the number of animal studies that can be conducted simultaneously. Consequently, researchers have been developing more advanced in silico and in vitro systems with the goal of replacing animal research with more accurate, inexpensive, scalable, and reliable alternatives.

Learn how for the R&D Toxicology group at a top 10 pharma, Elsevier combined its data, tools, and data science skills to provide a machinereadable custom data set to inform and organoid-based predictive toxicology model. The data set consisted of a ranked list of drugs with corresponding information on the reported incidence and severity of the adverse event in question, as well as other details essential to establishing an evidence-based calibration curve for more accurate predictions.

Challenge

An R&D toxicology group within a large pharmaceutical company was interested in developing a predictive toxicology model to assess if drug candidates were likely to cause a specific adverse event. The model would involve exposing organoids to drug compounds and measuring their reactions.

However, to identify the most suitable positive and negative controls for the model, the group required an extensive list of drugs and their reported potential to cause the specific adverse event. Although there are publicly available data on adverse drug events, they tend to be unstructured and codified, making them difficult to access and appraise in bulk. Consequently, manually searching for this information would be very time consuming and inefficient.

Solution

The group asked Elsevier whether it could help with data harvesting and structuring. Elsevier has the data, tools, and expertise to undertake such a project, and began to automatically extract relevant data for a wide variety of FDA-approved drugs.

This data included clinical trial, drug-labeling, and post-marketing surveillance data on the incidence and severity of the specified adverse event. Elsevier then cleaned, normalized, and annotated the data to account for factors such as drug name abbreviations, formulation, dose, route of administration, and use as a monotherapy/combination therapy.

Impacts for the company

Elsevier delivered a custom, annotated data set listing approximately 900 FDA-approved drugs, analyzed and ranked for their risk of causing a specific adverse effect. This saved the client significant amounts of time and money, and the data helped to inform a model to screen out compounds that could fail at the clinical stage.

These data provided a launch pad to establish a predictive toxicology model, allowing the team to establish a ‘calibration curve’ for more accurate predictions.

The technique can be easily extended to other adverse drug events – Elsevier can readily provide a list of ‘calibration drugs’ for predictive assays investigating other adverse events.

Insight

Adverse drug events and clinical trial failure

One of the major bottlenecks in drug development is the emergence of adverse events during clinical trials. Such adverse reactions can be very serious for patient health and can also affect drug efficacy. For instance, adverse events that involve gastrointestinal toxicity can negatively affect the absorption of drugs from the gut into the bloodstream, reducing their effectiveness. Adverse events can sometimes be so unpleasant for patients or trial participants that they stop taking the drug or reduce their dose, further affecting treatment efficacy.

Such safety issues are one of the major reasons why drug trials fail in later development stages. This represents a significant waste of resources and time for pharmaceutical companies. Predicting whether a drug candidate is likely to cause safety issues at the pre-clinical level has the potential to save time and money, and to protect trial participants from harmful side effects.

In vitro testing platforms

Historically, companies have performed animal testing to identify safety issues pre-clinically. However, depending on the nature of the adverse event, animal safety data may not translate very well to humans. In addition, animal research is incredibly costly, and there are ethical considerations concerning the use of experimental animals. The growing demand to identify safety issues at an earlier stage of drug development is limited by the number of animal studies that can be conducted simultaneously. Consequently, researchers have been developing more advanced in silico and in vitro systems with the goal of replacing animal research with more accurate, inexpensive, scalable, and reliable alternatives.

Examples of advanced in vitro systems include bioreactors, microfluidic systems containing human cells, and three-dimensional cell culture systems called organoids that are intended to mimic human tissues and organs more precisely than simple cell monolayers. Exposing organoids to drug compounds can elicit a measurable response in terms of protein biomarkers or changes in gene expression that can reveal aspects of their expected activity in humans. This can be useful in terms of learning about their likely efficacy. Moreover, organoids have a role to play in predictive toxicology models that seek to establish safety parameters for drug compounds.

An organoid system to predict risk

The R&D toxicology group at the center of this Case Study wished to establish an organoid-based predictive toxicology model to estimate the risk of a particular adverse event associated with lead compounds. This would involve exposing organoids to pre-existing drugs with an established record of causing or not causing the adverse event and then analyzing the resulting biomarkers. For instance, the researchers may expose different organoid samples to one drug with essentially no potential to cause the adverse event, one drug with limited potential to cause the adverse event and one drug with high potential to cause the adverse event; the data can then be used to develop and plot an approximate calibration curve of the resulting biomarkers.

Once the assay has been established in this fashion, it can be used for pre-clinical testing of new drug compounds, and the resulting biomarker data for such drugs can be benchmarked against the pre-existing calibration data. However, to achieve this, the company must first identify drugs with differing established tendencies to cause the adverse event in question, which is not as easy a task as it may sound.

Challenge

Finding needles in many haystacks

The R&D toxicology group did not have data science experience, so obtaining these data themselves would have involved performing manual searches, which would have posed a serious challenge.

While clinical trial data are publicly available, details on adverse events are frequently difficult to find and extract from extensive manuscripts. These data may be unstructured or codified; drugs can be used alone or in combination with others, making it difficult to isolate their individual effects; and the dose and route of administration of a drug can be unclear or presented using diverse abbreviations/units of measurements or other such codification, making it time consuming to find and interpret these data without the use of text mining techniques.

However, determining such details was crucial for the client to apply the drugs as controls in its assay. The situation was further complicated by the desire of the toxicology group to obtain an appraisal of all FDA-approved drugs that met specific criteria.

Rather than just identifying a few drugs that might be useful, the company wished to have an extensive list of candidates to choose from, with a detailed understanding of their potential to cause the specific adverse event. This comprehensive analysis was intended to maximize the chances of success when developing the predictive toxicology model by providing a diverse set of calibration drugs to establish an effective benchmarking system.

Solution

Text mining

Elsevier used text mining techniques to extract information about the reported incidence and severity of the adverse event for drugs with specific molecular properties, as defined by the toxicology group. About 900 drugs were eligible for the evaluation highlighting the enormity – and arguable futility – of manual searches in this context. Elsevier harvested the drug safety data from a variety of sources, including its PharmaPendium database, a comprehensive and curated source of pre-clinical and clinical pharmaceutical data and intelligence from FDA and EMA regulatory documents. Elsevier's biomedical database, Embase, was used to mine literature associated with clinical trials, and ClinicalTrials.gov was used to extract clinical trial results. Elsevier also sourced data from open source drug label information tools and from its PharmaPendium FAERS database, which includes post-market reports recorded within the FDA Adverse Event Reporting System (FAERS).

Data annotation and standardization

Elsevier annotated and cleaned the harvested data including standardizing the name of adverse events, drug names and dosages. Additionally Elsevier further delved into nuances in the safety data, such as whether the drug was administered alone or as part of a combination therapy, and sourcing information concerning elements such as the route of administration, formulation, and dose. The Elsevier team used SciBite's TERMite and other Elsevier search and semantic enrichment tools to achieve this, together with open source natural language processing resources.

The final annotated calibration data set

The final deliverable that Elsevier provided included a machine-readable custom data set, consisting of a ranked list of drugs with corresponding information on the reported incidence and severity of the adverse event of particular interest to the client. Drug-specific pharmacokinetic data were added to the data set to facilitate drug selection for in vitro testing. The report also compared different drug classes in terms of their propensity to cause the identified adverse event. Elsevier prepared the report and data set such that it would be easy for those with limited data science expertise to understand and to use. In the future, Elsevier intends to enhance such deliverables by developing interactive dashboard systems to make it even easier to explore those data sets.

Results

The project created multiple benefits for the toxicology group including the following:

- The machine-readable custom data set contained approximately 900 FDAapproved drugs and details of their tendency to cause the adverse event of specific interest to the customer. The list was ranked by severity.
- These data were an excellent starting point to establish a predictive toxicology model, allowing the team to choose the most appropriate control compounds and establish an evidence-based calibration curve for more accurate predictions. Accurate pre-clinical in vitro models are highly effective in reducing the expense and risk associated with drug development and allow companies to reduce their animal testing.
- The technique is easily extendable to other adverse drug events – if the company wishes to establish additional pre-clinical assays to predict other adverse events, Elsevier can readily provide a similar detailed list of potential calibration drugs.
- The company saved significant time and money by working with Elsevier. Elsevier has the data, tools, and skills to deliver training sets, and often works in adjunct with clients seeking data science expertise and support.

For more information, talk to your Elsevier Sales Representative today,
or visit: elsevier.com/solutions/datasets

ELSEVIER

Copyright © 2023 Elsevier B.V.