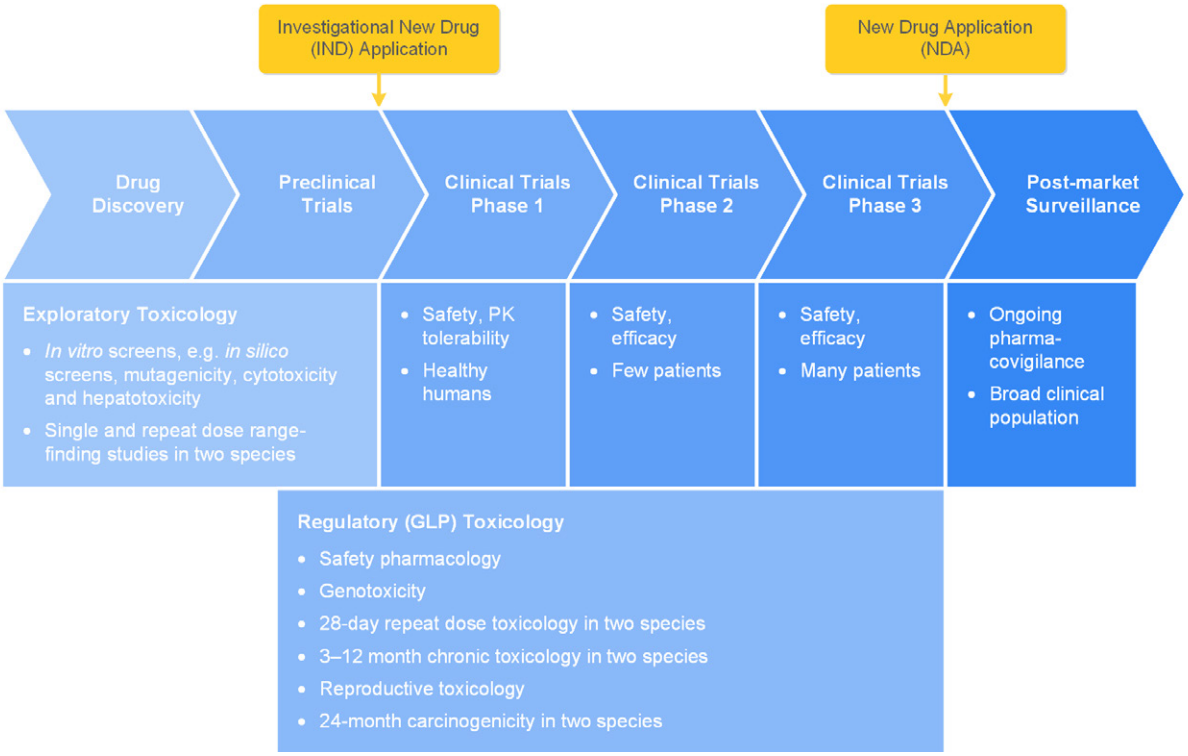
Using Linguamatics Natural Language Processing To Extract Actionable Drug Toxicity And Safety Information

Generating structured data from multiple sources throughout the safety lifecycle of a drug

Identification of safety issues with a drug is crucially important to pharma/biotech companies throughout a drug's development, from preclinical safety/toxicity studies, through clinical safety, to post-marketing surveillance in wider patient populations. Natural language processing (NLP) can efficiently and accurately surface such safety/toxicity signals that are often buried in a wide variety of internal and published sources, to generate structured, actionable information and insights.

Schematic Of The Timing Of The Main Safety Assessment Studies During Drug Discovery And Development, Through Discovery, Clinical Development, And Into Post-Market Surveillance And Pharmacovigilance



At all stages, project teams need the most comprehensive view of relevant data available; text mining plays a key role in access to actionable insights for safety.

Challenge

Lack of clinical efficacy is a major cause of attrition in drug development, but a poor safety profile at any point in the development pipeline is also a critical concern. Rigorous safety and toxicity testing occurs throughout the lifecycle of a drug, from initial lead candidate through preclinical and clinical development, to post-marketing surveillance.

There are essentially three main stages of this assessment: preclinical safety/toxicology in animal models, clinical safety in human subjects and post-market pharmacovigilance scanning for safety signals across wide patient populations, as shown in the schematic above.

At every point in this pipeline, safety and toxicity data is being generated and being searched for in unstructured text—internal safety reports, scientific literature, public data sources (such as CT.gov, MAUDE and FAERS), case safety reports, clinical investigator brochures, patient forums, social media and conference reports. Attempting to locate and extract meaningful, actionable information from this mass of unstructured text using traditional searching techniques is rendered almost impossible by multiple challenges:

- Data volume is increasing at such a rate that manual scanning is ineffective.
- Data structure may include complex grammar and linguistics, and data may be buried inside tables.
- Data standards may be lax, with unorthodox or inconsistent terminology and a failure to use accepted vocabularies, ontologies and formats.

Linguamatics' NLP approach

The Linguamatics NLP text mining platform can address all these challenges by efficiently and comprehensively extracting structured data from this mass of unstructured text. This structured data is then amenable to analysis and visualization, and can rapidly generate actionable intelligence on safety and toxicity concerns throughout a drug's safety lifecycle.

Linguamatics NLP can search and extract drug names, dosages, adverse events (AEs), safety indicators, and context such as species and tissue, from large document collections. Queries can be defined using keywords and linguistic expressions. Linguamatics NLP has powerful table processing that enables accurate data to be extracted from tabular preclinical toxicity or drug safety summaries.

By plugging in ontologies, Linguamatics NLP queries will automatically find synonyms or search for entire classes of items. Predefined smart queries can also be used; these are templates that hide complexity from the user by only exposing specific predefined options. Queries can also be combined to answer multiple questions simultaneously, for example by providing systematic profiles of compounds. Text processing workflows can be fully automated to cope with the ever-increasing flood of documents, reports, publications, etc.

Linguamatics NLP presents the structured results in a choice of formats. These include web pages with results classified by drug, dosage and AE. Microsoft Excel spreadsheets, XML files and network graphs are also supported, and allow scientists to visualize direct and indirect relationships between entities. Results can also be presented in formats suitable for export to third-party databases, for further processing and analysis using custom or in-house tools.

In summary, Linguamatics NLP's unique strengths can:

- provide scientists, researchers and analysts with comprehensive, precise and accurate data
- capture precise relationships
- find concepts in their appropriate context
- normalize and extract quantitative data from text and buried in embedded tables
- surface actionable insights and intelligence that might otherwise remain hidden.

Use cases

The following use cases from Merck MSD, Agios Pharmaceuticals, Pfizer and the FDA exemplify the successful application of the Linguamatics NLP platform to extract, detect and monitor safety and toxicity signals throughout the three main phases of drug development: preclinical, clinical and post-market.

ASSESSING NON-CLINICAL SAFETY TO ADVANCE HIGH QUALITY DRUG CANDIDATES AT MERCK MSD

Merck MSD uses an automated Linguamatics NLP process to extract unstructured conclusions and interpretations from final study reports, animal ante- and post-mortem reports, and protocols stored in an official file repository. The NLP queries identify, extract and normalize study annotation metadata and organ pathology findings. The results are combined with structured output and visualized via dashboards by the safety assessment teams. This NLP workflow has helped Merck MSD drive regulatory changes to reduce the safety assessment system workload without compromising human safety, and has provided historical summaries that can help in assessing the significance of new toxicity observations on pipeline compounds.

USING NLP FOR CLINICAL SAFETY AT AGIOS

Agios Pharmaceuticals say that they "use Linguamatics NLP to get decision support as fast and as comprehensively as possible." They apply NLP in clinical safety workflows to mine AE reports and assist with initial coding of reported events and WHO drugs. In a specific case, Agios explored the risk of a rare, yet potentially life-threatening AE, Differentiation Syndrome, in patients on a clinical trial for Agios' IDH-1-inhibitor AG120. Linguamatics NLP was used to extract key information from Serious Adverse Event Report Forms, and the extracted data was visualized as networks in Cytoscape. This enabled clinicians to explore the patterns of symptoms between patients and, critically, identify those at risk of the potentially fatal AE.

TEXT MINING FOR POST-MARKET SURVEILLANCE AT PFIZER

Increasing amounts of real world data for pharmacovigilance and post-market surveillance offer pharma companies and healthcare organizations a rich stream of data to monitor and mine. Many companies use Linguamatics NLP to extract structured information from patient forums, call center feeds, social media tweets, and electronic medical records and patient narratives.

Pfizer used Linguamatics NLP to categorize and tag call center feeds for key metadata such as caller demographic and reason for calling. Using text analytics, medical affairs researchers deepened the relationship for drug–disease associations, by combing the call logs for information on pre-existing conditions and relating these to reported potential side effects. With these associations, analysts related >70% of the reported side effects to underlying pre-existing conditions, rather than to an adverse drug reaction.

ARTIFICIAL INTELLIGENCE/MACHINE LEARNING FOR PREDICTIVE SAFETY AT THE FDA

The U.S. Food and Drug Administration (FDA) is performing ongoing work to develop models that leverage post-market safety data to predict AEs for new drugs coming to market. Researchers use Linguamatics NLP in a pipeline with machine learning ensemble models and classification algorithms to develop predictive AE models. They have shown that pharmacological target AE profiles, based on marketed drugs, can be used to predict unlabeled AEs for a new drug at the time of approval.

Digital transformation for safety

Safety is assessed throughout the lifecycle of a drug, from bench to bedside. The ultimate test is after the drug has been approved and used in clinical settings with many thousands of patients, across broader indications, and combined with other drugs. There are always risks, but the more relevant data that can be extracted, analyzed and transformed into actionable insights and information, the greater the likelihood of lowering that risk.

CONTACT US

+44 (0)1223 651 910 (U.K.) | +1 617 674 3256 (U.S.)

nlp@iqvia.com

linguamatics.com

©2021. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries. 07.2021.RWS