The Power Of Natural Language Processing To Transform Real World Data Into Insights

Enhancing drug development with actionable intelligence for better decision-making

In pharma and healthcare, understanding the real world (i.e. outside clinical trials) impact of therapies on patients is critical. Real World Evidence (RWE) can inform all phases of drug development and commercialization.

RWE can shed light on real world clinical effectiveness or safety profiles of products across a broad patient community. It can be used to assess patient-reported outcomes, provide input for product reputation management, help with key opinion leader engagement and more.

However, many Real World Data (RWD) sources contain unstructured text, which hinders easy analysis. **The Linguamatics NLP platform can unlock the value from real world sources such as electronic health records (EHRs), medical information requests, social media posts and customer call transcripts.**

The real world challenge

RWE can inform all phases of drug development, commercialization and drug use in healthcare settings. However, many sources of RWD contain large amounts of valuable information hidden in unstructured text (e.g. in EHRs, patient-reported outcomes such as forums and social media, and insurance claims data). There are many challenges in creating value from RWD. These include:

- data access (which may involve patient privacy issues);
- data structure (e.g. complex grammar in tweets or customer calls);
- data quality (e.g. missing data, coding errors);
- data extraction and integration of structured and unstructured fields; and
- mapping to standards (e.g. medical codes, vocabularies, formats).

The NLP-based text mining solution

The Linguamatics natural language processing (NLP) text mining platform can provide a solution to these challenges—extracting the key facts from these unstructured documents, using relevant ontologies and focused queries, and transforming RWD into actionable intelligence for decision-making.

Queries can be written to extract information on: Treatment patterns, e.g. drug switching or discontinuation; numerics such as lab values or dosage information; or patient data such as history of disease, problem list, demographics, social factors and lifestyle. The agile, iterative nature of query development in the Linguamatics NLP platform means that business rules can be encoded to suit the particular data set, whether sentiments from tweets, or treatment pattern choices and resulting outcomes from EHRs.

BENEFITS TO PHARMA ORGANIZATIONS
For pharma companies, effectively mining RWD provides value from bench to bedside, and feeds enhanced drug discovery, development and post-market delivery.
RWE supports product development and commercial decision-making, based on a better understanding of disease states and treatment patterns across a broad population.
RWE can be used for health economics and outcomes research, comparative effectiveness research and post-market product lifecycle management (disease forum engagement, reputation management, key opinion leader engagement, safety profiles, treatment regime effectiveness).
RWD can be used to understand treatment effectiveness, as well as to provide insight into patterns of care, long-term drug safety, healthcare resource utilization and disease epidemiology.

Robust RWE that weaves together different sources of both structured and unstructured data—such as clinical data, genomic data and socioeconomic data—yields a better picture of individual patient characteristics and responses to therapeutics. This improves our understanding of a drug's ability to treat individual patient needs, and supports precision medicine initiatives.

Data Relevant To RWE Comes In Multiple Types And Forms

For example: Clinical setting data derived from patient medical records, nurses' notes and pathology reports; patient registries; patient-reported outcomes derived directly from the patient experience, such as patient surveys, forum sites and social media; customer-support emails and call centre feeds; and reports in literature and conference abstracts.

Real world use cases

RWD IN PHARMA FOR MEDICAL AFFAIRS: JOHNSON & JOHNSON

Patient- and customer-call transcripts are a rich seam of potential patient-reported outcomes, side effects, drug interactions and more, providing access to insights that can have a huge impact on commercial business decisions, and assist post-launch product marketing and planning.

Johnson & Johnson uses NLP to annotate and categorize "voice of the customer" (VoC) call feeds, to gain insights into the real world use of their drugs. Researchers in the Predictive Analytics group have built an end-to-end workflow to process the call transcripts, using agile text mining to make sense of the unstructured feeds. The calls are categorized and tagged for key metadata such as caller demographics and reason for calling (e.g. complaint, formulation information, side effect, drug–drug interactions). Using the Linguamatics NLP platform in this workflow has more than doubled the efficiency in analysis; the accuracy of the NLP platform mining is at 95%, allowing the Medical Affairs teams to do longitudinal analysis of real world patient outcomes.

BRISTOL-MYERS SQUIBB: MINING ELECTRONIC MEDICAL RECORDS FOR PATIENT STRATIFICATION OF HEART FAILURE RISK

Bristol-Myers Squibb (BMS) wanted to understand more about patient stratification for heart failure risk. Heart failure patients typically exhibit high levels of clinical heterogeneity, which is problematic for treatment and for risk stratification. BMS researchers believed that if they could acquire a deeper understanding of the clinical characteristics of these patients, they could potentially understand how best to treat different patients or populations.

To that end, researchers obtained EHRs and echocardiogram records for about 900 patients, and used Linguamatics NLP to write queries, and extract and normalize approximately 40 different variables around patient demographics, clinical outcomes, clinical phenotypes, and other variables such as ejection fraction and left ventricular mass.

With advanced statistical clustering, BMS researchers identified four classes of patients with discrete clinical and echocardiographic characteristics. The four groups showed significant differences in one- and two-year mortality, and also in one-year hospitalizations. By better understanding how to stratify patient populations for heart failure, BMS has unlocked insights that can potentially improve the design of clinical trials, identify unmet needs and develop better therapeutics.

CONTACT US
+44 (0)1223 651 910 (U.K.) | +1 617 674 3256 (U.S.)
nlp@iqvia.com
linguamatics.com