

# Executive Summary

---

## *Data Mining in Drug Development and Translational Medicine*

examines the emerging role of the various flavors of data mining in translational drug development (*i.e.*, the formal stages of preclinical and clinical investigations) and pharmacovigilance (*i.e.*, the surveillance for potential side effects in the postmarketing stage).

Data mining (as opposed to conventional statistical analysis) can uncover patterns and relationships in large data volumes that are completely unexpected. In its strict definition, data mining is strongly correlated with the concept of machine learning; it describes the use of self-directed algorithmical patterns with limited confining parameters and a minimum of human intervention in pools of unstructured data. In accordance with common perception among pharmaceutical scientists, we have, for the purposes of this report, expanded this definition to include advanced modalities of exploratory data analysis, which follows predefined concepts and relies on human input in critical stages.

The pharmaceutical industry has made decisive moves to improve the predictiveness of early-stage drug safety testing. These preclinical efforts generate large amounts of data that can describe anything from toxicity signs in cell-based assays to the behavioral parameters of rodents. Data formats are also widely different, ranging from spreadsheets to video recordings. Clues to safety-related, potential “red flags” can be buried in any of these data, and we discuss the data mining strategies, methods, and tools that have been developed for each of those data types. We give examples of how integrated data mining can make the mandated preclinical evaluations more predictive.

In clinical trials, much more data are captured than are actually analyzed to build the regulatory data file. Clinical databases can thus be mined for information that the respective study was not explicitly designed to provide. This allows developers and regulators to gain insight into the

mechanisms of drug actions, interactions, and side effects in various previously undefined patient subgroups and facilitates the targeted design of subsequent studies. Data mining in clinical trials can also assist in uncovering bias by noncompliant study participants, and even breaches of integrity by clinical investigators.

Importantly, mining of scientific and technical text documents (either from the public domain or from internal corporate sources) has become an important tool in preclinical investigations as well as in the clinical phases. Mining the content of scientific papers and patents can—and should—accompany a molecule throughout its development pipeline, from the preliminary stage of the feasibility study to pre- and post-launch competitor analysis.

Mining postmarketing data for potential drug safety issues that arise during “field use” of approved drugs is a world of its own, driven by international, national, and corporate databases recording adverse events that are (systematically or spontaneously) reported in association with drug or vaccine therapy. This report discusses the reporting systems that feed the national, international, and corporate pharmacovigilance databases, describes the methods and tools for their analyses, and presents case reports.

*Data Mining in Drug Development and Translational Medicine* also presents brief profiles of software and service providers that cater to the pharmaceutical and biotechnology industry, and points to the directions that the use of data mining in transitional drug development and postmarketing pharmacovigilance might take during the next decade. We conclude that the use of data mining in these fields of the life science industry is dynamically emerging yet has huge potential, which will be increasingly realized during the 2010s.