

# Pharmacovigilance signal detection from social media

Xiaoyi Chen<sup>1</sup>

<sup>1</sup>Systems Biology Lab, Institut Pasteur, Paris, France

Pharmacovigilance aims to monitor marketed drugs' adverse effect (AE). The principle database through with adverse drug reaction (ADR) signals can be detected is the Spontaneous Reporting System (SRS). Several statistical methods for signal detection have been developed based on measures of drug-AE association describing the reporting disproportionality in SRS, including frequentist methods and Bayesian methods. These methods have been evaluated in both comparative simulation study and in empirical study, and have been used by different regulatory agencies and drug safety monitoring systems. Recently, patients' posts on social media are being considered as a complementary source to SRS. In the framework of ADR-PRISM project, a corpus of 21 million posts extracted from five open French patient forums has been explored for pharmacovigilance. For each post, drug names and medical concepts, as well as the semantic relation between these two entities have been identified with natural language processing tools. Different characteristics between social media data and SRS data can be observed: (i) social media data contains more information about common drugs, while SRS tends to include more reports on newer drugs; (ii) social media data contains large discussions on annoying but less serious AEs while SRS is more interested in serious or less-known AEs. In order to explore ADR signals in social media, we adapted the signal generation modeling for social media data characteristics, compared four common used signal detection methods via an intensive simulation study, and then applied them to ADR-PRISM dataset. Three use cases have been studied.