Continuous Post-Market Real World Evidence Generation from Online Drug Reviews using Natural Language Processing

Matthias Hartung[1], Arne Kramer-Sunderbrink[1,2], Soufian Jebbara[1], Yannick Loonus[1], Bassam Mokbel[1], Philipp Cimiano[1,2]

[1]Semalytix GmbH, Bielefeld, Germany
[2]Bielefeld University, Germany

Recent years have seen a marked change in regulatory decision-making in drug approval and health technology assessment: Clinical evidence gathered in randomized controlled trials is no longer considered the single source of truth in order to assess comparative effectiveness of different treatment options and their benefit-risk profiles, but is increasingly complemented by real-world evidence (RWE) derived from a number of data sources that are associated with outcomes in heterogeneous patient populations in real-world settings [1]. Thus, RWE plays a crucial role in shaping pharmaceutical drug development towards patients' real-life needs in the first place [2]. Social media has been identified as a promising source of evidence to reveal insights about patients' subjective experience of outcomes, but also adverse drug reactions, and about how drug interventions improve their quality of life [3].

We present a new methodology that supports the continuous generation of RWE from online drug reviews written by patients. Our approach relies on deep learning methods to extract key information regarding patient-reported outcomes, side effects and adherence problems reported by patients. The method builds on state-of-the-art transformer models that are fine-tuned on the task of extracting these elements based on annotated training data. We evaluate the architecture on four therapy areas (Diabetes Type 2, Obesity, Breast Cancer and Psoriasis) for which 1200 drug reviews were annotated in total, and we test the generalizability on five further diseases (Migraine, Muscle spasm, Depression, Parkinson's Disease, Crohn's Disease).

Our method can be applied continuously to generate periodical reports and supports the direct comparison of all treatments within a drug class wrt. patient-reported effectiveness and safety. We demonstrate the capabilities of our method for RWE generation in three use cases: (i) For comparative effectiveness research, we show how our method allows us to compare the share of patients reporting a certain outcome for various therapies. (ii) For pharmacovigilance research, we show how the method provides a comparative overview of side effects reported by patients on different drugs. (iii) Finally, our method supports the investigation of side effects that are significantly correlated with patient-reported adherence problems, thus providing an explanation for why patients decided to stop their treatments.

References:

[1] Sherman et al. Real-world evidence. What is it and what can it tell us? N Engl J Med 375, 2293-7 (2016).

[2] Patalano et al. Gathering structured patient insight to drive the PRO strategy in COPD. Patient-centric drug development from theory to practice. Adv Ther 37, 17-26 (2020).

[3] McDonald et al. Real-world data and the patient perspective: the PROmise of social media? BMC Med 17, 11 (2019).

Keywords: Real-world evidence, Social media, Patient-reported outcomes, Comparative effectiveness research, Pharmacovigilance

Conflicts of Interest: All authors are employees of Semalytix GmbH which provides Pharos as a real-world evidence generation platform.