

# Drug Side-Effects: What Do Patient Forums Reveal?

Sarvnaz Karimi  
NICTA and Dept. of CSSE  
The University of Melbourne  
skarimi@csse.unimelb.edu.au

Su Nam Kim  
Dept. of CSSE  
The University of Melbourne  
snkim@csse.unimelb.edu.au

Lawrence Cavedon  
NICTA, RMIT University, and  
The University of Melbourne  
lawrence.cavedon@nicta.com.au

## ABSTRACT

Adverse drug reactions are a threat to health, sometimes to the extent of causing death, as well as imposing a substantial time and financial cost from patients and health-care providers. The most common of these negative side-effects are reported by their manufacturers; however, other side-effects may be discovered through different surveillance methods, including volunteer reporting by doctors and pharmacists, and recently, patient self-reports in online forums. Extraction of side-effect information from these forums is challenging but useful for finding indications of adverse reactions that may otherwise go unreported. We outline our vision for this problem, our approach to data collection, and the methodologies we intend to use for side-effect extraction from patient forums.

## Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Health; I.2.7 [Natural Language Processing]: Text analysis

## Keywords

Drug adverse effect extraction, Patient forums

## 1. MOTIVATION

A drug side-effect is an effect that is secondary to the one intended. Despite its predominant use to only describe harmful effects — known as adverse effects/reactions<sup>1</sup> — it can also apply to beneficial, but unintended, consequences of the use of a medicine [2]. Most common and severe side-effects are officially published in the leaflets of the drugs as well as websites such as *Drug Side Effects*<sup>2</sup> or *Drugs@FDA*.<sup>3</sup> The information about the clinical trials through which the listed side-effects were discovered should also be reported to help professionals in their prescriptions. However, not all side-effects are officially identified and listed in these sources due to the limitations of controlled trials, such as the limited number and diversity of participants. The World Health Organization (WHO) has therefore urged patient monitoring and further scientific data collection — referred to as

<sup>1</sup>The term *adverse effects* usually refers to the drug side, and *adverse reactions* to the patient side [2].

<sup>2</sup><http://www.drugs.com/sfx/> (accessed 28 Sep 2011)

<sup>3</sup><http://www.fda.gov/Drugs/default.htm> (accessed 28 Sep 2011)

post-marketing surveillance — after the release of each drug. There is a wide range of organizations that provide facilities for professionals and drug users to report observed side-effects. For example, the US Food and Drug Administration (FDA) established the MedWatch program in 1993 to encourage both patients and professionals to report *severe* adverse effects.

Patient social networks and forums — such as DailyStrength, Ask a Patient, PatientsLikeMe, the Yahoo Health and Wellness group, and iMedx — also collect such feedback directly from drug consumers (either implicitly or explicitly). Automatically extracting information from these forums could still expedite the process of finding side-effects experienced by patients. While data in such forums may be of questionable reliability, it may nevertheless provide indications of real side-effects.<sup>4</sup> Mining health-related information from medical text has been studied for some time, particularly over biomedical literature or databases; using newly emerged social networking data has only been studied recently. Differences between text mining from biomedical literature and social media is the language (formal vs. informal and specialized vs. general), and length of the text (long vs. short). The differences between these two types of texts are the main reasons for current natural language processing tools being less suitable for automatically analyzing the forums.

Leaman *et al.* [3] was one of the first to explore extraction of drug adverse reactions from a health social network, DailyStrength. They created a lexicon of known adverse effects and calculated the similarity of the nominated tokens of patient posts on five different drugs with the lexicon. They indicated the colloquial language used in the posts as one of the main problems in finding the right side-effect candidates.

Chee *et al.* [1] focused on predicting drugs that are likely to be withdrawn from the market by the FDA based on adverse effects posted at Yahoo Health and Wellness Group. They applied a range of classifiers that worked based on the similarity of the drugs under their consideration to those on the watchlist and those withdrawn.

Li [4] proposed creating a hierarchical ontology of adverse effects by mining adverse drug reactions from online forums. She specifically studied statin-class of cholesterol-lowering drugs without using any lexicon, relying solely on patient texts to avoid the low coverage and language gap between them. Such an ontology could potentially assist patients associate their observed side effects to drugs.

<sup>4</sup>We are currently conducting a survey on consumer use and trust of forum data, and will report on this at the workshop.

We will demonstrate our proposed framework for extracting side-effects and related information from patients posts in forums. We will also discuss the challenges and limitations of this project.

## 2. PATIENT FORUMS

Patient forums provide a platform where people can discuss their medical conditions, medicines they take, outcomes, and side-effects they experience. Although not all the information that patients provide is accurate or sometimes even correct, such forums, especially those that are supervised, can keep patients informed about their health conditions, and help them obtain emotional support from their peers. Some of these forums are specifically on drug side-effects and medicine ratings, e.g., Ask a Patient, and some are social networks, e.g., DailyStrength or PatientsLikeMe, and provide more options to the patients. From the point of view of researchers, such forums are rich sources of data from which to extract a variety of information, such as patterns of drug adverse reactions and drug-drug interactions. Accessibility of such data, however, can be restricted to the website owners, making it difficult for researchers to obtain such data. For example, Ask a Patient keeps the posts in a database and only makes some portions of the data available on request. This data generally requires de-identification of patients.

## 3. METHODOLOGY AND DATA

All previously mentioned work is focused on extracting mentions of adverse effects and mostly ignores the contributing factors that are patient-dependent. We are interested in extracting both adverse and beneficial side-effects along with background information on the patients that could contribute to their positive or negative experience, such as age group, gender, or other medications they take, as well as their medical conditions (see Table 1). This is particularly interesting because clinical trials do not cover all possible patient conditions. For example, if mothers of very young boys observe a behavioral disorder when administering a medicine, but mothers of similar age girls do not, it may suggest that gender is a contributing factor and its detection is of value.

We also propose finding patterns of reporting side-effects, both using heuristics and automatically extracted rules, to be able to extract the entities and their relations. The outcome can be used in enriching side-effect ontologies.

One goal of our project is creation of a gold standard collection that allows a platform for fair comparison of different techniques for text mining researchers. We therefore gathered data for ten different drugs from two different forums: Ask a Patient<sup>5</sup> and eHealth Forum<sup>6</sup>. A total of 5,996 posts (40,871 sentences) was collected. We only relied on free-text comments in each post. We decided to extract a variety of relevant information from each post in the form of stand-alone entities and the relationships between these entities, as listed in Table 1. This wide range of entities makes the extraction task complicated; however, this is necessary because finding side-effects in isolation of the patient conditions and the history of their drug consumption could be of limited value. We recruited two annotators to perform the annotations; the annotation effort is ongoing.

<sup>5</sup><http://www.askapatient.com/> (accessed 28 Sep 2011)

<sup>6</sup><http://ehealthforum.com/> (accessed 28 Sep 2011)

Entity	Example/Description
Disease	"After 3 years of having Ativan keep the <i>anxiety</i> in check, ..."
Symptom	"My <i>heart</i> was racing and .."
Drug	"I must be addicted to <i>Xanax</i> "
Duration	"Began taking 5 mg daily(broke the 10mg pill in half) for <i>4 weeks</i> "
Dosage	"Began taking <i>5 mg</i> daily ..."
Frequency	"Began taking 5 mg <i>daily</i> .."
Positive side-effect	"I'm taking this for my back pain but it has been <i>reducing my stress</i> as well."
Negative side-effect	"Sometimes causes <i>drowsiness</i> ."
Lack of negative side-effect	"I feel dizzy and low but no <i>vomiting</i> ."
Lack of positive side-effect	"I was feeling even more <i>energetic</i> initially but it doesnt work like that any more"
Positive outcome	"No apparent side effects thus far and results have been very <i>effective for the pain</i> ."
Negative outcome	"Problem is you build up a tolerance and eventually the <i>drug quits working</i> as has been my case."
Gender of patient	"I was prescribed this for anxiety when my teenage daughter was driving <i>my wife</i> and I into"
Age	"I'm in my <i>forties</i> "
<b>Relation</b>	
Drug-Drug	If a patient explicitly mentions that taking two named drugs together had any effect or no effect, then the two drugs are annotated by a positive, negative, or no effect relation.
Dosage-Frequency	The frequency in which a dosage is taken is annotated by a <i>for</i> relation.
Dosage-Duration	The prolong of intake for a specific dosage is annotated with a <i>for</i> relation.
Drug-Dosage	The dosage which a drug is taken is annotated with a <i>taken</i> relation.

**Table 1: Entities and their relations considered for annotation of forum data.**

One of our contributions will be providing the research community with a rich annotated collection that is large enough for experimentation and diversity in the types of drugs and annotated concepts. The existing literature also does not provide a comparison over previous approaches, mainly due to lack of availability of a standard and publicly accessible dataset. It is therefore difficult to judge which methods work best and why. In the future, we intend to conduct a comprehensive comparison of existing methods as well as our own techniques.

**Acknowledgements:** NICTA is funded by the Australian government as represented by Department of Broadband, Communication and Digital Economy, and the Australian Research Council through the ICT Centre of Excellence programme.

## 4. REFERENCES

- [1] B. W. Chee, R. Berlin, and B. Schatz. Predicting adverse drug events from personal health messages. In *AMIA*, page 10pp, Washington, DC, 2011.
- [2] I. R. Edwards and J. K. Aronson. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet*, 356(9237):1255–1259, 2000.
- [3] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez. Towards Internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *BioNLP*, pages 117–125, Uppsala, Sweden, 2010.
- [4] Y. A. Li. Medical data mining: Improving information accessibility using online patient drug reviews. Master's thesis, MIT, 2011.