

Differences Between Social Media and Regulatory Databases in Adverse Drug Reaction Discovery

Chen Wang Sarvnaz Karimi
CSIRO, Australia
{chen.wang,sarvnaz.karimi}@csiro.au

ABSTRACT

Information extraction from social media for a variety of applications, such as collecting people opinion about a product or a political party, has been widely studied and justified. Extracting information for health related applications however is less justified especially because of sensitivity of health issues, difficulty in establishing the value and trust in lay people to judge their health problems. Using social media to discover adverse drug reactions is one of the most controversial topics. It is difficult to establish the causality between an adverse drug reaction and a drug when the context information such as patient condition is missing. We compare official reports of adverse drug reactions with reports on medical forums related to two different drugs to discuss the potential and challenges in this research area.

1. INTRODUCTION

Monitoring Adverse Drug Reactions (ADRs) has direct relationship with the public health and healthcare costs around the world. An incident of an adverse drug reaction can lead to serious disability or death [2, 4]. In the past 50 years, post-marketing adverse drug reaction surveillance has been mainly done through Spontaneous Reporting Systems (SRS) [1], maintained by national regulators such as the Food and Drug Administration (FDA) or the Therapeutic Goods Administration (TGA). However, it is difficult to achieve “real-time” ADR detection through SRS databases. Recently, as data regarding ADRs becomes increasingly available online, monitoring ADR related data from multiple sources shows promise to speed the ADR detection process. Social media data, such as drug discussion forums, is one potential candidate. It provides first-hand consumer reviews and experiences that may not otherwise reach to the authorities. However, how useful this type of data is for ADR detection is yet unclear as there are a number of challenges associated with data from online resources such as medical forums: informal language, lack of medical expertise of most consumers, noisy data that can include spam or incorrect information, and lack of information on patients’ history or possible interacting medications. Regulators on the other hand require actionable ADR reports that enable them to

establish whether a drug actually caused a reaction.

To get an insight on the potential of social media, especially patient forums, in discovery of adverse drug reactions, we compare posts in drug discussion forums with SRS reports.

2. DATASETS

We use the following four different datasets for two drugs, *Diclofenac* and *Lipitor*, in our comparison.

1. FDA dataset: Reports collected from 2010 quarter 1 to 2012 quarter 3 in FDA Adverse Event Reporting System (FAERS). It contains 2,880 reports about Diclofenac and 46,473 reports about Lipitor;
2. TGA dataset: Reports collected from Adverse Drug Reaction System Database (ADRS)¹ from 1997 to May 2013. It contains 2,175 reports about Diclofenac and 2,702 reports about Lipitor;
3. AskaPatient² dataset: Consumer posts entered from 2001 to September 2013. It contains 264 posts about Diclofenac and 1,057 posts about Lipitor;
4. Treato statistics: Treato³ is an Internet company that collects patient-written health experiences from blogs and forums. We use its statistics about the “top concerns” for Diclofenac and Lipitor as a reference in our comparison.

3. ADR COVERAGE COMPARISON

We list frequently reported ADRs that we extracted from each dataset in Table 1. We also list ADRs in the product labels of these two drugs as a reference. Even though not all the reported ADRs are listed in product labels, the most common ones are. Therefore, comparing the coverage of frequently reported ADRs in different data sources against those in product labels could be an indicator of the data quality. Table 2 shows the coverage. The FDA and TGA reports have different coverage of known ADRs among their top 15 ADRs. We note that the terminology used in the two regulators for the same ADR are often different too. This is mainly due to subjective differences among officers that code the reports using controlled vocabulary known as MedDRA⁴. The top 15 ADRs listed by Treato has similar coverage ratio of Diclofenac ADRs to the FDA reports. For Lipitor, its coverage ratio of the known ADRs is between that of the FDA and the TGA. Given Treato has access to a

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

SoMeRA’14, July 06-11 2014, Gold Coast, QLD, Australia.
Copyright 2011 ACM ACM 978-1-4503-3022-0/14/07 ...\$15.00.

¹www.tga.gov.au/safety/daen.htm#.U0tZPKZKYTw

²www.askapatient.com/

³www.treato.com

⁴www.meddra.org/

| | Rank | FDA | TGA | Treato | Labelled ADRs |
|----------------|------|-----------------------|--|--------------------|----------------------------|
| (a) Diclofenac | 1 | pain | pruritus | nausea | abdominal pain |
| | 2 | nausea | diarrhea | dizzy | abnormal renal function |
| | 3 | vomiting | rash | stomach irritation | anemia |
| | 4 | fatigue | nausea | bleeding | constipation |
| | 5 | dyspnoea | abdominal pain | drowsy | diarrhea |
| | 6 | headache | urticaria | depression | dizziness |
| | 7 | diarrhea | melaena | stomach problems | dyspepsia |
| | 8 | arthralgia | vomiting | numbness | edema |
| | 9 | renal failure acute | dyspnoea | sciatica | elevated liver enzymes |
| | 10 | dizziness | dizziness | stomach pain | flatulence |
| | 11 | pneumonia | renal failure acute | addiction | GI ulcers and vomiting |
| | 12 | fall | haematemesis | burning | gross bleeding/perforation |
| | 13 | anxiety | hepatic function abnormal | vomiting | headaches |
| | 14 | drug interaction | gastric ulcer | heart attack | heartburn |
| | 15 | malaise | rash erythematous | itching | increased bleeding time |
| (b) Lipitor | 1 | pain | myalgia | muscle pain | arthralgia |
| | 2 | myalgia | blood creatine phosphokinase increased | weight loss | diarrhea |
| | 3 | nausea | hepatic function abnormal | tiredness | dyspepsia |
| | 4 | anxiety | nausea | weakness | nasopharyngitis |
| | 5 | pain in extremity | liver function test abnormal | weight gain | nausea |
| | 6 | fatigue | rhabdomyolysis | cramping | musculoskeletal pain |
| | 7 | death | headache | depression | muscle spasms |
| | 8 | dyspnoea | fatigue | joint pain | myalgia |
| | 9 | arthralgia | amnesia | leg pain | insomnia |
| | 10 | myocardial infarction | paraesthesia | memory loss | pain in extremity |
| | 11 | dizziness | rash | headache | pharyngolaryngeal pain |
| | 12 | headache | pruritus | arthritis | urinary tract infection |
| | 13 | asthenia | abdominal pain | muscle weakness | |
| | 14 | diarrhea | myopathy | muscle problems | |
| | 15 | drug ineffective | arthralgia | insomnia | |

Table 1: Top 15 most frequently reported ADRs for (a) Diclofenac and (b) Lipitor in regulatory databases (FDA and TGA) and Social media (based on Treato). The right hand side box shows the ADRs in the product labels (not ranked).

| Drug name | FDA | TGA | Treato |
|------------|------------|-------------|------------|
| Diclofenac | 7/18 (39%) | 12/18 (67%) | 7/18 (39%) |
| Lipitor | 6/12 (50%) | 4/12 (33%) | 5/12 (42%) |

Table 2: Comparison against drug labels.

large number of medical forums, this close coverage indicates the usefulness of social media data to find common ADRs.

4. ADR DESCRIPTION COMPARISON

We use the posts in AskaPatient for further analysis of the characteristics of ADR information in the social media. We manually annotate the ADR phrases in AskaPatient data regarding Diclofenac, with details explain in [3]. After stemming annotated phrases, we calculate the most frequent phrases used for describing ADRs (Table 3). These posts show the same coverage ratio against the labelled ADRs. However, the overlap between treato terms and AskaPatient phrases for top 15 ADRs is low and only 3 phrases are the same. In comparison, FDA and TGA have 6 ADRs in common even though they use different categories of MedDRA terms to index their reports. Social media tends to use informal terms to describe ADRs. It is not straightforward to translate phrases such as “sick to my stomach” and “extreme gas in stomach and intestines” to the standard ADR terms. The diversity of description is likely to affect the ADR detection. Aggregating more data also may not improve the consistency. As an example, in Table 1(b), Treato contains both “weight loss” and “weight gain” in its list which contradict each other. It is challenging to obtain consistent ADR results from social media.

Conclusions. Our comparison shows that social media contain useful data that reflect known ADRs to a certain degree.

| ADR description | Frequency | Continue.. | |
|------------------|-----------|--------------|----|
| diarrhea | 30 | abdomin pain | 12 |
| nausea | 29 | pain | 10 |
| cramp | 22 | bloat | 8 |
| vaginal bleeding | 20 | sick | 8 |
| dizziness | 19 | gas | 8 |
| stomach pain | 18 | constipation | 8 |
| stomach cramp | 15 | drowsiness | 8 |
| headache | 13 | | |

Table 3: Top ADR in AskaPatient for Diclofenac.

There are however some limitations: firstly, lack of controlled terminology usage makes the ADR statistics from social media data less consistent to those in regulator databases; secondly, social media data often lack of sufficient information to establish the causal effect between a drug and an adverse reaction. However, we believe that social media potentially provides information that can be integrated with other data sources for better ADR detection.

Acknowledgments. We thank our collaborators at the TGA for providing us with their reports dataset. Askapatient provided us with their forum posts.

5. REFERENCES

- [1] P. Coloma, G. Trifiró, V. Patadia, and M. Sturkenboom. Postmarketing safety surveillance. *Drug Safety*, 36(3):183–197, 2013.
- [2] B. Hug, C. Keohane, D. Seger, C. Yoon, and D. Bates. The costs of adverse drug events in community hospitals. *Jt Comm J Qual Patient Saf*, 38(3):120–126, 2012.
- [3] A. Metke-Jimenez, S. Karimi, and C. Paris. Evaluation of text processing algorithms for adverse drug event extraction from social media. In *SoMeRA*, Gold Coast, Australia, 2014.
- [4] E. Roughead and S. Semple. Medication safety in acute care in Australia: Where are we now? part 1: A review of the extent and causes of medication problems 2002-2008. *Australia and New Zealand Health Policy*, 6(1):18, 2009.