Typology of drug misuses created from information available in health fora

Élise BIGEARD [a,b,1], Natalia GRABAR [a], Frantz THIESSARD [b]

[a] *Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France*
[b] *Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France*

Abstract. Patients seldom report the misuse of drugs to their physicians. Hence, other sources of information are necessary for studying these issues. We assume that online health fora can provide such information and propose to exploit them for building a typology of drug misuses. The misuses detected are structured according to the goals of patients: we distinguished three types of non-intentional misuses and 14 types of intentional misuses. This work will be used to guide future task of automatic detection of drug misuses.

Keywords. Drug misuse, Patient safety, Pharmacovigilance, France

1. Introduction

If between 3% [1] and 20% [2] of emergency admissions are caused by adverse drug reactions (ADRs), drug misuses are also harmful. Misuses may happen when physicians make prescription errors or when patients do not follow the prescriptions: incorrect dosage (overuse or underuse), drugs used for indications other than those prescribed, etc. Patients are then exposed to risks. Moreover, the discovery of misuses is a difficult task because patients do not report them to physicians or authorities. Hence, the situation is even worse than with the ADRs reporting, which does not exceed 5% [3,4]. In order to study drug misuses, we need to use different sources of information, such as health fora: within the anonymity and without any particular effort, patients willingly talk about their disorders and doings [5], and thus they may give clues about their actions and well-being.

Aside from the detection of ADRs [6], quite a few works are dedicated to the observation of drug use in social media. We can mention the observation of non-medical use of drugs [7], and the creation of drug overuse ontology [8]. The objective of our work is to create a typology of drug misuses. This typology is intended to guide future work on automatic detection of misuses. To design this typology, we study the posts in health fora dedicated to drugs. Our hypothesis is that patients write about their healhcare process and treatments, which may also include drug misuses, be they intentional or non-intentional.

[1] Corresponding Author: Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France. email: elise.bigeard@u-bordeaux.fr

2. Material

Drugs names. We use a set of commercial drug names and DCIs, associated with their ATC codes [9]. We also use the CNHIM database Thériaque[2], base publique du médicament[3] and database Medic'AM from the French healthcare insurance[4]. Thériaque is especially useful because it includes short names of drugs, such as *doliprane*.
Forum corpus. We build the corpus from the French health website Doctissimo. We collect posts written between 2010 and 2015 from the two most active categories, dedicated to drugs[5] and pregnancy[6]. We chose to study those two fora because they contain a large number of messages. We keep only posts that mention at least one drug, which gives a total of 119,562 posts (15,699,467 words). In each post, the drugs are identified and the drug classes are defined by the 3 first characters of the ATC codes. As expected, some drug classes are very frequent. For instance, up to 60% of posts are concerned with the birth control pills, and 15% with antidepressant and anxiolytic drugs. This set of posts is used to create three corpora to be annotated manually: $C1$ with 200 posts, $C2$ with 1,200 posts, and $C3$ with 500 posts. Posts with more than 2,500 characters are excluded because they contain heterogeneous information and are difficult to analyze and to annotate. $C1$ and $C2$ contain randomly selected posts independently on the drug classes they mention. $C1$ is annotated by two annotators and is used for the computing of inter-annotator agreement. Because some drug classes are more frequent than others, $C3$ is built so that it contains a larger variety of drugs: for each of the 50 most frequent drug classes, we randomly select 10 posts. We assume indeed that some misuses can be typical to some drug classes. This motivates the diversification of the analyzed corpus. The annotation rationale of these 1,900 posts and results are presented in the following sections. These corpora provide the reference data for our study.

3. Methods

The annotator task is to assign each post to one of the following categories:

- \+ contains normal drug use: *Anyway the question I'm asking is whether it is normal that loxapac I'm taking needs hours to do someting???*
- \- does not contain drug use: *ouch boo, above all take a break, he didn't prescribe aspegic for the baby??*
- ! contains drug misuse. When this category is selected, the annotator is asked to shortly explain in free text the nature of the misuse (*i.e.* , overuse, dosage, brutal quitting): *well me miss blunder and with head in the clouds I had to start the "utrogestran 200" at d16 and I forgot of course! well I took it this evening!!!!*
- ? unable to decide.

[2] http://www.theriaque.org
[3] http://base-donnees-publique.medicaments.gouv.fr
[4] https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/donnees-statistiques/medicament/medic-am/medic-am-mensuel-2017.php
[5] http://forum.doctissimo.fr/medicaments/liste_categorie.htm
[6] http://forum.doctissimo.fr/grossesse-bebe/liste_categorie.htm

Two annotators are involved in the annotation: one is a medical expert, another is a computer scientist familiar with medical texts and annotation tasks. Each post of $C1$ is annotated by the two annotators independently. This allows to compute the inter-annotator agreement [10]. Afterwards, thanks to discussions between annotators, the consensus has been reached on posts with different annotations or with no decision. $C2$ is devided in two halves, each being annotated by one of the annotators. As for the $C3$ corpus, it is only annotated by the medical expert. The typology of misuses is created from these annotations. We particularly stress on identification of the reasons which lead to the misuses.

4. Results of Annotation and Typology

1,900 posts were annotated. Among these, 53% (999) contain normal use, 30% (746) have no use, and 8% (155) contain misuse of drugs. The inter-annotator agreement computed on $C1$ is 0.46, which is a moderate agreement. It also indicates that the task on automatic detection and categorization of misuses may be quite difficult.

The first analysis of the annotated posts indicates that patients can commit misuses of drugs non-intentionally or intentionally. Figure 1 shows the schema of the typology. In case of non-intentional misuse, patients commit mistakes while taking the drug (intake, dosage, contraindication). When patients realize their mistake, they post a message to ask how to mend this situation. However, when the misuse is intentional, patients do not follow the prescriptions and are conscious about it, like in: *J'ai arrêté de moi-même (je sais c'est pas bien) (I stopped by myself (I know this is not good))*; *j'ai descidé de ne pas en reprendre. (I decided to not take it again.)*; *cette fois je rajoute xanax (this time I will take xanax as well)*. For instance, patients can ignore or overlook the prescription without any particular goal: *ma psy m'a bien dit, pas d'alcool, mais j'en bois quand même en week-end quand y'a des occasions (my shrink told me no alcohol, but I drink it anyway during the week-end when I have such opportunities)*. Besides, when committing misuses, patients can: have precise reasons (like the fear of ADRs which leads to underdosage or missed intakes), self-medicate (for which they will try to get the prescriptions and drugs by any means), look for particular effets (like psychotropic effect, weigh loss, or even suicide attemp). Yet another misuse situation occurs when patients become addicted to the drugs they take, which mainly happens with anxiolytic drugs. Yet, it should be noticed that sometimes patients may commit misuse intentionally in a moment of temporary distress, which they regret later and worry about the consequences. From this analysis, we can see that different types of misuses have been detected and that they cover a great variety of situations. Their automatic detection requires specific algorithms and models.

5. Conclusion and Future Work

Because patients do not report the misuses they commit when taking the drugs, we proposed to exploit online health fora to observe such events. Hence, we collect and analyze information written by patients in order to create a typology of drug misuses. This typology reflects the goals of patients, which can be intentional or non-intentional. For instance, patients may refuse the intake of drugs by fear of a probable ADR, they may self-medicate themselves, and even take drugs searching some special effects. In the last case, we can find cases related to suicide attempts, psychotropic effects, weight loss, etc.

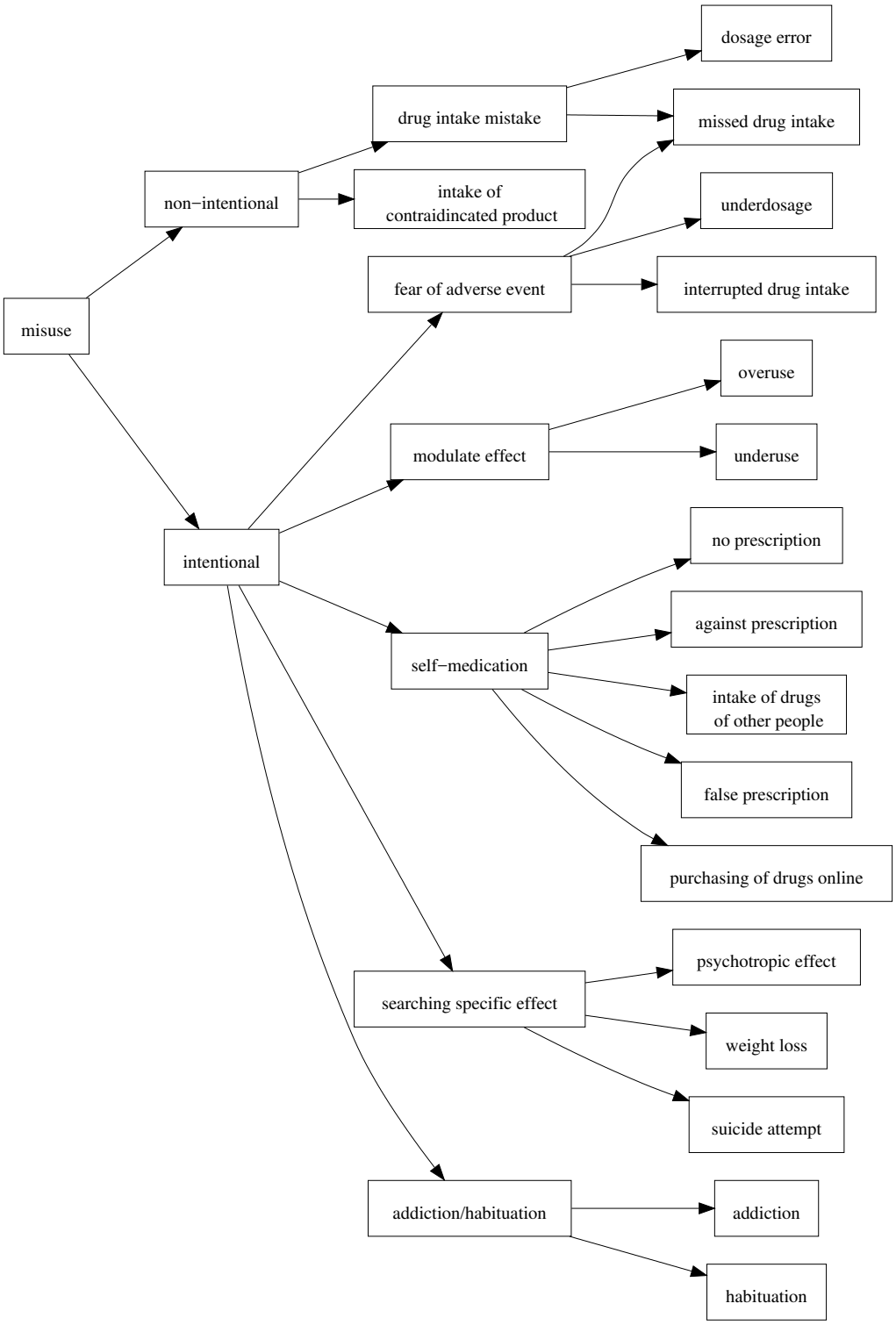

Figure 1. Typology of drug misuses.

This typology is based on posts from the French website *Doctissimo*, and more particularly from fora dedicated to *pregnancy* and *drugs*. Among all the posts, up to 60% of posts are about the birth control pills, and 15% about antidepressants and anxiolytics. This leaves little room for other drug classes. When selecting the posts to process, we tried to adjust the distribution of drug classes, and to include a greater variety of them in the corpus. Yet, it is possible that misuses associated with other types of drugs or pathologies may be missing in the current study. Similarly, in the annotated corpus, we observe only 155 messages containing misuses, and some categories distinguished have only few examples instantiating them. For these reasons, it is necessary to enrich the corpus analyzed in this work with additional posts, which can potentially provide other kinds of misuses, including those that are seldom described.

In future work, we will exploit this categorization and manually annotated posts to propose an automatic method for the detection of misuses in social network texts, and for the extraction of information related to the misuses. Notice that, for some categories of misuses, we have identified typical linguistics patterns, which can also be exploited for the detection of misuses. For example, the cases of overuse are often expressed through the use of quantifiers cooccurring with drug names, such as *3 boites de xanax (3 boxes of xanax)*. Beyond the automatic detection of misuses, the very purpose of this work is to shed light on the various cases of misuses, to alert medical doctors and authorities, and to improve information provided to patients. Through this work, we intend to contribute to the safety of patients and to help prevent misuses they may attempt.

Acknowledgement. This work has been performed as part of the *DRUGSSAFE* project funded by the ANSM.

References

[1] Pouyanne P, Haramburu F, Imbs J, and Bégaud B. Admissions to hospital caused by adverse drug reactions: cross sectional incidence study. French pharmacovigilance centres. *BMJ* 2000;320(7241):1036–.

[2] Queneau P, Bannwarth B, Carpentier F, et al. Emergency department visits caused by adverse drug events: results of a French survey. *Drug Saf* 2007;30(1):81–8.

[3] Lacoste-Roussillon C, Pouyanne P, Haramburu F, Miremont G, and Bégaud B. Incidence of serious adverse drug reactions in general practice: a prospective study. *Clin Pharmacol Ther* 2001;69(6):458–62.

[4] Moride Y, Haramburu F, Requejo Alvarez A, and Bégaud B. Under-reporting of adverse drug reactions in general practice. *Br J Clin Pharmacol* 1997;43(2):177–81.

[5] Gauducheau N. La communication des émotions dans les échanges médiatisés par ordinateur : bilan et perspectives. *Bulletin de psychologie* 2008:389–404.

[6] O'Connor K, Pimpalkhute P, Nikfarjam A, et al. *Pharmacovigilance on twitter? mining tweets for adverse drug reactions*. 2014:924–33.

[7] Kalyanam J, Katsuki T, Lanckriet GRG, and Mackey TK. *Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the twittersphere using unsupervised machine learning*. 2017;65:289–95.

[8] Cameron D, Smith GA, Daniulaityte R, et al. *PREDOSE: a semantic web platform for drug abuse epidemiology using social media*. 2013;46(6):985–97.

[9] Skrbo A, Begović B, and Skrbo S. Classification of drugs using the atc system (anatomic, therapeutic, chemical classification) and the latest changes. *Med Arh* 2004;58(2):138–41.

[10] Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20(1):37–46.