Journal of Biomedical Informatics 90 (2019) 103091

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

# A systematic approach for developing a corpus of patient reported adverse drug events: A case study for SSRI and SNRI medications

Maryam Zolnoori[a,b,c,*], Kin Wah Fung[b,*], Timothy B. Patrick[a], Paul Fontelo[b,*], Hadi Kharrazi[d], Anthony Faiola[e], Yi Shuan Shirley Wu[f], Christina E. Eldredge[g], Jake Luo[a], Mike Conway[h], Jiaxi Zhu[i], Soo Kyung Park[j], Kelly Xu[f], Hamideh Moayyed[k], Somaieh Goudarzvand[l]

[a] Department of Health Sciences, University of Wisconsin Milwaukee, Milwaukee, WI, United States
[b] Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States
[c] Section of Medical Informatics, Department of Health Science Research, Mayo Clinic, Rochester, MN, United States
[d] Department of Health Policy and Management, Johns Hopkins University, Baltimore, MD, United States
[e] Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, United States
[f] School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States
[g] School of Information, University of South Florida, Tampa, FL, United States
[h] Department of Biomedical Informatics, Utah University, Salt Lake City, United States
[i] Emmes Corporation, Rockville, MD, United States
[j] Department of Epidemiology, Johns Hopkins University, Baltimore, MD, United States
[k] College of Letters and Science, University of Wisconsin Milwaukee, WI, United States
[l] School of Computing and Engineering, University of Missouri-Kansas, Kansas City, MO, United States

## ARTICLE INFO

*Keywords:*
Annotated corpus
Adverse drug events
Drug effectiveness
Online healthcare forums
Patients narratives
Psychiatric medications
SSRIs
SNRIs
Drug safety
Social media
Information extraction
Semantic mapping
SNOMED CT
UMLS
Text mining
Machine learning

## ABSTRACT

"Psychiatric Treatment Adverse Reactions" (PsyTAR) corpus is an annotated corpus that has been developed using patients narrative data for psychiatric medications, particularly SSRIs (Selective Serotonin Reuptake Inhibitor) and SNRIs (Serotonin Norepinephrine Reuptake Inhibitor) medications. This corpus consists of three main components: sentence classification, entity identification, and entity normalization. We split the review posts into sentences and labeled them for presence of adverse drug reactions (ADRs) (2168 sentences), withdrawal symptoms (WDs) (438 sentences), sign/symptoms/illness (SSIs) (789 sentences), drug indications (517), drug effectiveness (EF) (1087 sentences), and drug infectiveness (INF) (337 sentences). In the entity identification phase, we identified and extracted ADRs (4813 mentions), WDs (590 mentions), SSIs (1219 mentions), and DIs (792). In the entity normalization phase, we mapped the identified entities to the corresponding concepts in both UMLS (918 unique concepts) and SNOMED CT (755 unique concepts). Four annotators double coded the sentences and the span of identified entities by strictly following guidelines rules developed for this study. We used the PsyTAR sentence classification component to automatically train a range of supervised machine learning classifiers to identifying text segments with the mentions of ADRs, WDs, DIs, SSIs, EF, and INF. SVMs classifiers had the highest performance with F-Score 0.90. We also measured performance of the cTAKES (clinical Text Analysis and Knowledge Extraction System) in identifying patients' expressions of ADRs and WDs with and without adding PsyTAR dictionary to the core dictionary of cTAKES. Augmenting cTAKES dictionary with PsyTAR improved the F-score cTAKES by 25%. The findings imply that PsyTAR has significant implications for text mining algorithms aimed to identify information about adverse drug events and drug effectiveness from patients' narratives data, by linking the patients' expressions of adverse drug events to medical standard vocabularies. The corpus is publicly available at Zolnoori et al. [30].

* Corresponding authors at: Section of Medical Informatics, Department of Health Science Research, Mayo Clinic, 200 First Street SW, Rochester, MN, United States. Tel.: 1 3175151950 (M. Zolnoori); Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States (K.W. Fung and P. Fontelo).
*E-mail addresses:* Zolnoori.Maryam@Mayo.edu (M. Zolnoori), kfung@mail.nih.gov (K.W. Fung), pfontelo@mail.nih.gov (P. Fontelo).

https://doi.org/10.1016/j.jbi.2018.12.005
Received 25 February 2018; Received in revised form 30 September 2018; Accepted 17 December 2018
Available online 04 January 2019

## 1. Introduction

Controlled conditions on clinical trials and post-pharmacovigilance systems managed by regulatory agencies detect < 10% of Adverse Drug Events (ADEs) due to voluntary nature of the systems and patients unwillingness to share their experiences with the agencies. These limitations resulted in emergency visits, death, and significant burden on individual and healthcare systems [11,28]. However, pharmacovigilance has special importance for psychiatric medications, particularly SSIs and SNRIs, because: (1) most psychiatric medications are frequently associated with ADEs; (2) clinical trials involving psychotropics are conducted under restricted inclusion and exclusion criteria; (3) publication bias in clinical trials; (4) psychiatric medication directly affects the brain that can cause life threatening behaviors, such as suicide attempts [19].

Studies have shown that patient self-reporting of ADEs and drug efficacy to pharmacovigilance surveillance systems have the same quality as reports made by healthcare professionals [26]. Indeed, patients' self-reports are used as a reliable source for risk discovery at the FDA's MedWatch program and the UK MHRA's Yellow Card Scheme. However, patients often do not report to these systems: due to negative attitudes toward clinical providers and drug makers, not recognizing the availability of these systems, or due to the severity of their illness [28]. Instead, patients often report and discuss their experience with prescribed medications in various social media platforms, such as online support groups and message boards. Accordingly, these social media platforms have turned out to be a prime source for discovering various aspects of the risks and benefits of drugs, such as adverse drug effects [4], drug effectiveness, and drug impacts on quality of life.

### 1.1. Challenges in automatic extraction of adverse drug events (ADEs) from patients generated data

Due to the large volume of social media data, methods to automate the extraction and analysis of consumer health information from this type of data has received considerable attention in recent years [20]. However, the performance of these methods is often affected by the inherent complexity of the consumer posts that include colloquial phrases, ambiguous terms, and layperson language instead of professional medical terms used in expressing drug effects.

Colloquial phrases introduce innate bias in conventional text mining methods for named entity recognition and therefore, detecting boundaries of terms and phrases require using an alternative lexicon-based approach to named entity recognition [14,2,16]. The lexicons mainly compiled using the Unified Medical Language System (UMLS) Metathesaurus terminologies and database of adverse drug events, such as Side Effect Resource (SIDER) database [12], Drug Bank database, Consumer Health Vocabularies (CHV) [24]. Nevertheless, ambiguous terms and layperson language often lead to low recall rates and high frequency of undetected desirable terms (e.g., ADRs and withdrawal symptoms). In response to the challenges of dictionary-based approaches, supervised machine learning methods were trained on corpora of adverse drug events (ADEs). For example, Huynh, He, Willis, and Rüger [9] trained Neural Network on a Twitter corpus and Medline corpus of ADEs for automatic classification of ADEs assertive text segments. Comparing the results of both lexicon-based approach and machine learning systems with privately held annotated corpora of patients expressions of ADEs showed that a significant number of ADEs mentioned by patients in healthcare forums still remained undetected [10,21].

Text mining tasks related to pharmacovigilance for psychiatric medications are more challenging due to layperson language expression, which is often subjective. Subjective expression, specifically for psychological ADEs, creates high semantic variations for a specific ADE concept. This variation can significantly reduce the performance of both the lexicon-based approaches and machine-learning methods for ADEs identification and extraction in online healthcare forums.

### 1.2. Contributions

In response to these challenges, we developed a corpus, "Psychiatric Treatment Adverse Reactions" (PsyTAR), which evaluates the pharmacological effects of psychiatric medications, mainly SSRIs and SNRIs. We followed a systematic approach to develop a corpus consisted of three main components: sentence classification, entity identification, and entity normalization. In the sentence classification component, the review posts were split into sentences, and then the sentences were classified for the presence of adverse drug reactions (ADRs), withdrawal symptoms (WDs), sign/symptom/illness (SSIs), drug indications (DIs), drug effectiveness (EF), and drug ineffectiveness (INF). In the entity identification phase, the span (boundary) of four entities including ADRs, WDs, SSIs, and DIs in the sentences were identified. In the entity normalization component, the identified entities were mapped to the corresponding medical concepts in both UMLS and SNOMED CT. In addition, the entities were further classified as physiological, psychological, cognitive, and functional problems (e.g., limitation in daily functioning, social activities, or inter-personal relationships) [31].

Identifying ADRs, WDs, SSIs, DIs, EF, and INF is important for informing patients and clinicians about psychiatric medications' benefits and risks. This corpus can also aid in testing hypotheses concerning the impact of pharmacological factors on patients' attitudes and behaviors toward the psychiatric medications specified in this study [29,31]. Furthermore, this is the first rich, publicly open and annotated corpus focused on linking layperson descriptions of psychiatric medication effects to professional terminologies by identifying semantic links among the expressions of medical terms (e.g. "lack of care for anything" and "nothing moves or excites me" as overlapping in meaning with "apathy"). Overall, this corpus is unique in that it is open access and may be used as a benchmark to train and evaluate the performance of automatic systems aimed to identify ADEs and measure drug effectiveness from online healthcare forums, particularly for psychiatric medications. It may also be used in electronic health records (EHR) to facilitate the seamless exchange of information between patients and clinicians.

## 2. Methodology

The methodology of our work was composed of four major phases: data collection, data pre-processing, annotation, and corpus evaluation. The annotation phase consisted of three main stages including sentence classification, entity identification, and entity normalization. Fig. 1 shows a schematic view of methodology for the development of the PsyTAR corpus.

### 2.1. Dataset information

#### 2.1.1. Data source

The data source of this study is a healthcare forum called "askapatient.com", which collects patients' self-reported experiences for a wide range of medications. The "side-effects" and "comments" fields of the forum collect patients' experiences for various aspects of medications. The duration of usage, reason for prescription, age, gender, and patients 'satisfaction for drugs ranged from 1 (minimum) to 5 (maximum) are collected in other fields. All the data in askapatient.com are anonymous and publicly available.

#### 2.1.2. Drug source

We used this forum to collect patients' self-reported experiences of four psychiatric medications from two classes of drugs: Zoloft[1] and Lexapro[2] from SSRI[3] class, and Effexor[4] and Cymbalta[5] from SNRI[6] class. According to IMS[7] Health, these four drugs were among the top 10 antidepressants prescribed in the United States between July 2013 and June 2014 [8].

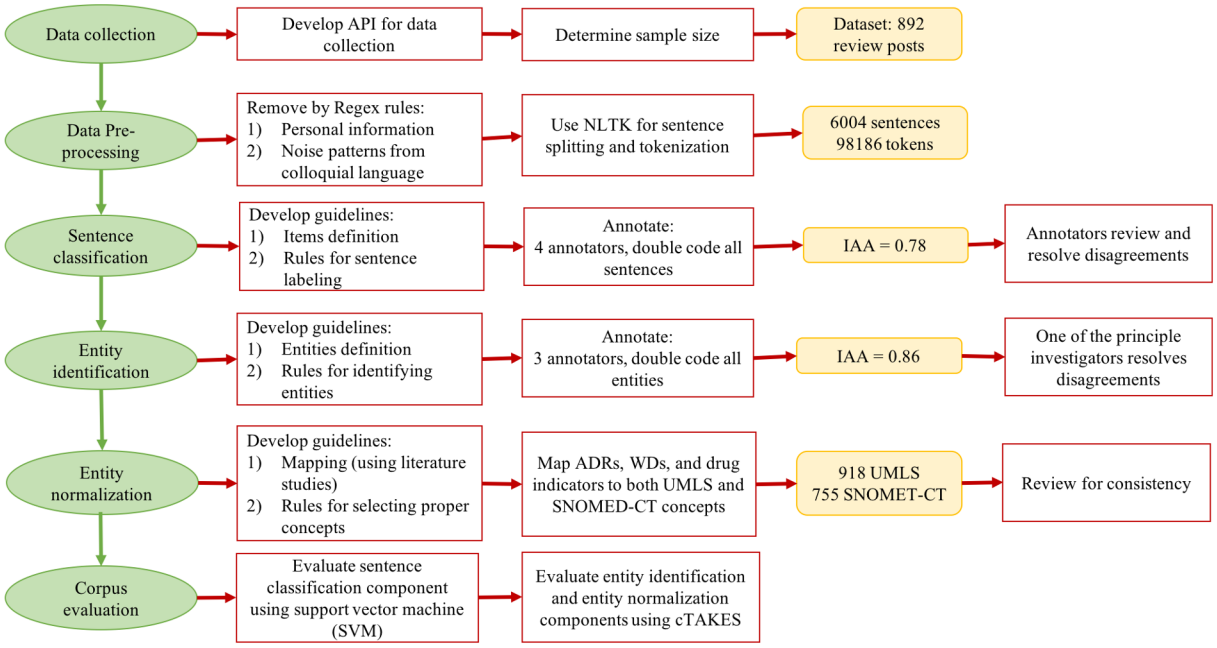



**Fig. 1.** Methodology for developing the corpus. API: Application Programming Interface; ADR: Adverse Drug Reaction; WD: Withdrawal Symptoms; IAA: Inter-Annotator-Agreement; cTAKES: Clinical Text Analysis and Knowledge Extraction System.

### 2.2. Data collection

To collect information automatically from this healthcare forum, we developed a web-crawler. This forum has a built-in search function that facilitates filtering drug reviews for a specific drug.

#### 2.2.1. Calculating the sample size

In order to select a subset of data that sufficiently represents the whole dataset in the healthcare forum, we used the following sample size formula [3]:

$$Sample\ \ Size = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + (\frac{z^2 \times p(1-p)}{e^2 N})} \quad (1)$$

where $Z$ is the confidence interval, $e$ is the margin of error, $P$ is the standard deviation, and $N$ is the population size. For this study, $Z$ is 90%, $e$ is 5%, $P$ is 0.5 to ensure the sample is large enough, and $N$ is the number of post reviews for each drug in the healthcare forum. The calculated sample for Zoloft, Lexapro, Cymbalta, and Effexor XR were 213, 219, 231, 228 respectively.

1. Generic name: Sertraline
2. Generic name: Escitalopram
3. Selective Serotonin Reuptake Inhibitor (SSRI)
4. Generic name: venlafaxine
5. Generic name: duloxetine
6. Serotonin Norepinephrine Reuptake Inhibitor (SNRI)
7. IMS Health is an American company that provides information, services and technology for the healthcare industry.

### 2.3. Data preprocessing

The majority of drug review posts in our sample were composed of multiple sentences. Each sentence may cover multiple aspects of patient experiences with drugs (e.g., drug effectiveness and ADRs). Hence, we set the unit of analysis at the sentence level. However, the sentences were written in colloquial English language, in which patients often did not follow proper grammatical and punctuation rules. To prepare the data for annotation, we formulated regular expression rules to remove: (a) any personal information including emails, phone numbers, and URLs from the reviews; (b) noisy patterns, such as punctuation errors in sentence structure. We further used the open-source Natural Language Toolkit (NLTK) to split posts into sentences. Statistics on posts, sentences, and tokens are presented in the results section.

### 2.4. Annotation

#### 2.4.1. Guidelines

To maintain consistency and uniformity of annotation across the corpus, we developed guidelines for sentence classification, entity identification, and entity normalization.

*2.4.1.1. Component 1 – sentence classification.* We annotated sentences for the presence of ADRs, WDs, SSIs, DIs, EF, and INF. If a sentence did not contain any information about these items, we labeled it as Others. In addition, in a drug review, if the reviewer referred to the experience of another patient with a drug, we labeled it as others.

Table 1 presents definitions of the items with examples from review posts. Distinguishing ADRs from WDs is important from the perspective of clinical trials and interventions designed specifically to help patients manage the process of drug discontinuation. Identifying drug indications from patient posts allows us to better recognize the purpose of drug prescriptions, which may be useful for drug repurposing. Moreover, it will help us to handle the challenge of ambiguity in addressing semantic types of terms (ADR vs. WD vs. SSI Vs. DI) presented in a drug review posts.

*2.4.1.2. Component 2 – entity identification.* In the guidelines for entity identification, we focused on definitions and rules for identifying and extracting mentions of ADRs, WDs, SSIs, DIs, and qualifiers representing severity (QS) and persistency (QP) of the entities. Identifying the qualifiers can help healthcare providers estimate the debilitating effects of ADRs/WDs on patients' quality of life and designing interventions to reduce the effects.

For ADRs and WDs extraction, we used the guidelines developed in our preliminary study [31]. The guidelines include definitions and rules for ADRs identification and extractions. According to the rules, patients' subjective complaints (e.g., "*body move in coordination with other people's*

**Table 1**
Definition of items in the sentence classification guidelines with an example for each item.

| Items | Definition | Example |
|---|---|---|
| ADR | A sentence should be labeled as ADR (Adverse Drug Reaction) if there is an explicit report of any sign/symptom that the patient explicitly associated them with the drug consumption | "I couldn't take Effexor XR. It gave me horrible nightmares and I kept waking up" |
| WD | A sentence should be labeled as WD (Withdrawal Symptom) if there is an explicit report of any sign/symptoms that the patient explicitly associated them with the process of dosage reduction or the drug discontinuation | "But then came the "Weaning Off" period!!! OMG. Cold sweats, vomiting, dizziness, brain zaps (with noise)" |
| EF | A sentence should be labeled as EF (Effectiveness) if it contains an explicit report that the patient's health condition has been improved or his/her symptoms were resolved after the drug consumption. | -"For the first few weeks, it helped me feel better".-"My stress disappeared after I used Zoloft for a few days." |
| INF | A sentence should be labeled as INF (Ineffectiveness) if it contains an explicit report that the patient's health status did not improve or became worse, or remained the same | "Did not work at all"<br>"I am still very anxious" |
| SSI | A sentence should be labeled as SSI (Sign/Symptom/Illness) if it contains an explicit report of any sign/symptoms/illness that the patient experienced before/during/after the drug consumption and they are not the drug's ADRs or WDs | "This drug has worsened my anxiety"<br>"I visited my doctor for my depression" |
| DI | A sentence should be labeled as DI (Drug Indication) if it contains any sign/symptoms/illness that the patient explicitly mentioned as the reason for a drug prescription or the sign/symptoms/illness were resolved because of the drug consumption | "My stress disappeared after I used Zoloft for a few days" |
| Others | Not applicable to any of the defined items | "I started Effexor three weeks ago" |

*bodies*" (Echopraxia)), and functional problems (e.g., limitations on daily functioning, work performance, and social activities) that patients directly attributed to ADRs and WDs need to be identified. Any of these symptoms have pharmacologically related affective components that may contribute to patients negative perceptions towards drugs and eventually drug non-adherence. In addition, collecting this information enables clinicians to recognize the impacts of the medications on their patients' quality of life. Understanding these impacts may enable clinicians to design more effective therapeutic interventions. In this study, we extended the guidelines to define the rules for identifying SSIs (sign, symptoms, illness) and DIs (see Table 2).

*2.4.1.3. Component 3: Entity normalization*. We mapped all the identified entities to the corresponding concepts in Unified Medical Language System (UMLS) and SNOMED CT terminology. If no proper concept was available, we assigned "no-codes" to the entity. The detail of developing the guidelines with examples is explained in Section 2.4.3 entity normalization.

### 2.4.2. Annotation process

*2.4.2.1. Sentence classification*. Four annotators participated in the annotation process. Two annotators were pharmacy students and two annotators had a background in health sciences. Each document was annotated by two annotators. The sentences in each document were labeled using the annotation guidelines.

To measure the observational error, we calculated the Inter-annotator agreement (IAA) using Cohen's Kappa:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (2)$$

where *Pr(a)* is the relative observed agreement among annotators and *Pr(e)* is the hypothetical probability of chance agreement. The overall IAA for the sentence classification component was 0.78, indicating substantial agreement between annotators [13]. Table 3 shows the IAA for each item accompanied with examples of disagreement between annotators.

To resolve the disagreement, instances of disagreement were reviewed and discussed by the same annotators who annotated the respective document earlier. For a specific item, annotation was added or removed if they were marked by any of the annotators, given that they both agreed on the decision. Otherwise, the sentences were labeled as *Others*.

*2.4.2.2. Entity identification*. The process of identification and extraction of the span (boundary) of ADRs, WDs, DIs, and SSIs from sentences were conducted by participating four annotators having health-related backgrounds. The disagreement between annotators for the identified entities was resolved by one of the principal investigators. Prepositions and possessive adjectives were excluded to improve consistency in spans of identified entities. For example, in "*my anxiety became worse*," '*my*' was not extracted as part of anxiety. Pair-wise agreement equation [23] was used for calculating the IAA is:

$$Agreement(A_i, A_j) = \frac{match(A_i, A_j, \alpha, \beta)}{max(n_{A_i}, n_{A_j})} \quad (3)$$

where $A_i$, and $A_j$, represents the dataset annotated by annotator *i and j* respectively, and $n_{A_i}$ and $n_{A_j}$ are the size of identified entities in $A_i$, and $A_j$,. $max(n_{A_i}, n_{A_j})$ indicates the maximum number of identified entities. Parameters $\alpha$and $\beta$ present span strictness and tag strictness of identified entities. Since entities were identified from the sentences with predefined labels (tags), they have the label of the sentences. Therefore, we excluded parameter $\beta$ from the calculation. For parameter $\alpha$, we set the span on strictness. For example, for identifying ADRs from this sentence: "*the headache was terrible, I could not continue the drug.*", if annotator *A* identified "*headache*" and annotator *B* identified "*headache was terrible*" as ADR, the matching between annotators *A* and *B* is "0". The computed pairwise agreement for the strict match was 0.86 for ADRs, 0.81 for WDs and 0.91 for SSIs and DIs, with the average 0.86 for the whole dataset. The reason for using PA for computing IAA rather than conventional measures, such as Kappa, is that PA is calculated at the level of entities. Since the identification task requires identifying the entities and determining their correct boundaries, the chance agreement is effectively zero.

### 2.4.3. Entity normalization

Annotation of health-related entities in drug review posts facilitates the process of developing and evaluating text-mining and machine-learning systems aimed at automatic extraction of the entities. However, generating and testing medical hypotheses on the dataset needs normalization step, in which the identified entities translate to equivalent concepts in medical standard vocabularies. This translation fills the gap between identified entities and provides unambiguous standard information for data collection and analysis. To normalize the entities in our corpus, we mapped the identified entities including ADRs, WDs, SSIs, and DIs to their corresponding concepts in both UMLS and SNOMED CT.

*2.4.3.1. Guidelines for mapping*. The semantic match between layperson expression of ADRs/WDs and standard vocabularies is a subjective process. To conduct the mapping systematically, we expanded the guidelines of mapping in our previous study [31]. To develop the guidelines, we used clinical trials report addressing pharmacological effects associated with psychiatric medications and qualitative studies

**Table 2**
Definitions and rules for identifying SSIs and DIs from patients' comments.

| Entity | Definitions | Example | Rules for identification | Example |
|---|---|---|---|---|
| SSIs | If patients explicitly mentioned experience of any sign/ symptoms/illness that were not ADRs or WDs of the drugs | "I suffered from **body pain (SSIs)** and **depression (SSIs)**" | **Worsened Symptoms:** If the patient's sign/symptom became worse by drug consumption, the sign/symptom should be labeled as both SSI and ADR of the drug | "Zoloft increase my anxiety **(SSIs and ADR)**" |
| DIs | Any sign/symptom/illness that the patient mentioned as the reason for the drug consumption/prescription | "My doctor prescribed Zoloft for my **depression (DIs)**" | **Treated symptoms:** If patient explicitly reported that the signs/symptoms/illness were resolved by drug consumption, the signs/symptoms/illness is the drug indication | This drug reduced my **sense of guilt** (DI) |

investigating themes of patients experiences with the drugs. The guidelines were iteratively updated to include the new expressions of pharmacological effects of the drugs.

In some cases, the equivalent UMLS/SNOMED CT concept for a lay person's expression is more general and broader in scope. One example is the UMLS concept "apathy". Behavioral indifference, as a clinical feature of "apathy", is manifested by patients' lack of desire to engage in activities, lack of desire to make any changes, not caring about anything or similar symptoms. Therefore, we mapped any patient complaints reflecting behavioral indifference, such as "just don't care," and "just lived day by day" to apathy as a more general concept.

To conduct systematic mapping of lay person's expression of ADRs/ WDs/DIs/SSIs to both UMLS and SNOMED CT, the guidelines of mapping also include procedure and requirements for mapping (Table 4 and Appendix A).

Appendix A shows the procedure for identifying proper UMLS and SNOMED CT concepts for the identified entities. Table 4 presents the requirement for selecting proper UMLS and SNOMED CT concepts. In the case of availability of multiple UMLS concepts for the original term, the proper concept needs to include the expression of the most recent versions of SNOMED CT. If multiple UMLS candidates met the mentioned requirements, the proper UMLS concept has the SNOMED CT expression that syntactically matches with the patient's expression (original term). Using the flowchart (Appendix A) and requirements for finding proper concepts, we mapped identified entities to both UMLS and SNOMED CT concepts.

If the identified UMLS concept (UMLS (1)) for the original term lacks the SNOMED CT concept, we searched for another UMLS concept that semantically matches with UMLS (1) concept and includes an SNOMED CT expression. For example, for the original term "Zombie-like", the closest syntactically and semantically UMLS concept (UMLS (1)) is "felt like a zombie" [C0857486] that does not include a SNOMED CT concept. Therefore, we used equivalent UMLS concept (UMLS (2)) with a SNOMED CT expression that is "Emotionally detached" [C0233754]. Table 5 provides examples of mapping patients expressions of ADRs to both UMLS and SNOMED-CT concepts.

We used UMLS Terminology Services (2017) and a lexicon that was created based on our guidelines to map patients' expressions to the UMLS and SNOMED CT concepts.

### *2.5. Evaluation of the PsyTAR corpus (PsyTAR in use)*

We demonstrate the potential usability of the PsyTAR corpus by developing a pipeline, which includes three components of text processing, sentence classification, and entity identification and normalization. We used the PsyTAR sentence classification component to automatically train supervised machine learning classifiers including support vector machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR) to identifying text segments with the mentions of ADRs, WDs, DIs, SSIs, EF, and INF. We also showed how augmenting the clinical Text Analysis and Knowledge Extraction System (cTAKES) core dictionary with the PsyTAR dictionary can improve cTAKES performance in the identification of ADRs and WDs from patients' drug reviews for psychiatric medications. In addition, we provided a summary of our prior studies with focus on testing the association between patients' experiences of different types of ADRs and patients attitudes towards SSRI and SNRI medications.

## 3. Results and corpus statistics

This study includes 891 drug review posts. The majority of the reviews were posted by females. The average age for the whole sample was 37, and the duration of usage was between 1 day to 20 years. The overall number of sentences in the corpus was 6004. In the sentence classification component, 36% of the sentences were labeled as "ADRs", followed by sentences labeled as others (not applicable) (35%), and

**Table 3**
Computed IAA for each item in sentence classification component with examples of disagreement between annotators.

| Items | IAA | Example of disagreement | Annotator 1 | Annotator 2 |
|---|---|---|---|---|
| ADR | 0.81 | "I stopped taking the lexapro, the anxiety has lightened a little bit, but the crying hasn't" | ADR | WD |
| WD | 0.69 | "However, I have tried to taper and quit several times and CAN'T" | WD | Others |
| EF | 0.82 | "Helped a great deal then put on the generic and had a totally negative reaction" | EF | EF & INF |
| INF | 0.65 | "At the end of 6 weeks I still felt no difference but my best friend said they noticed a difference" | EF | INF |
| SSI | 0.91 | "I prefer to be depressed than having no feelings" | SSI | Others |
| DI | 0.76 | "I really didn't think that I was depressed, but I think since taking the lexapro, my mod has lifted" | EF | EF & DI |

**Table 4**
Requirements for selecting proper/preferred UMLS and SNOMED-CT concepts.

| Requirements | Description | Example |
|---|---|---|
| (1) Definition | (1) Definition of a proper concept should cover the patient's specific physiological, behavioral, emotional, cognitive, or functional problem.<br>(2) If the patient did not specify severity or type of experienced adverse effect, we used the most general concept (code) that represents the patient's reported problem. | For the patient complaint "takes a long time to get to sleep," the proper UMLS concept is "initial Insomnia." "Sleeplessness" is not a correct concept, because it includes all phases of insomnia. |
| (2) Semantic type | The semantic type of proper concept includes "finding," "sign or symptom" or "mental or behavioral dysfunction." However, in some cases, other semantic types, such as "individual behavior" for the concept "aggressive behavior" is a proper map. | |
| (3) Hierarchical structure (ancestors and children) | The UMLS concept should not convey any additional meaning inherited from ancestors (parents) that are not related to patient's complaint. | For the patient complaint "inability to eat," "Aphagia" [C0221470] is not a proper map because this concept is linked to the ancestors of swallowing finding in SNOMED CT. While the patient with depression does not have any problem with swallowing, but they have a problem with loss of appetite. So "Loss of appetite" [C1971624] is a proper map. |
| (4) Inclusion of SNOMED-CT concept | In a case of existing multiple UMLS candidates, the preferred concept is a concept that includes the most recent version of SNOMED CT concept. | For the patient complaint "memory loss," concept of "Amnesia" [C0002622] compared with "Memory loss" [C0751295] is the proper concept because memory loss includes the obsolete version of SNOMED CT concept. |
| (5) Syntax match with SNOMED-CT concept | In the case of existing multiple UMLS candidates that meet the requirement (1), (2), and (3), the preferred match has SNOMED CT Concept that syntactically match with the patient's expression (original term). | For patient complaint "restlessness", the UMLS concept candidates are "Agitation" [C0085631], "Restlessness" [C3887611], "Psychomotor Agitation" [C3887612], "Akathisia" [C0392156]. The proper match would be "Restlessness" [C3887611] due to including SNOMED Concept preferred term (PT) with the same syntactic match. |

sentences labeled as drug effectiveness (EF) (18%). In the entity identification component, 7414 entities were identified, of which 40% were duplicate. Physiological entities had the highest frequency (53%) followed by psychological entities (40%), cognitive entities (5%), and functional problems (2%). The majority of the identified entities were ADRs (65%), followed by SSIs (16%), DIs (11%), and WDs (8%). In the component of entity normalization, all the identified entities were mapped to 918 unique UMLS concepts and 755 unique SNOMED CT concepts. The number of UMLS concepts in the corpus exceeds that of SNOMED CT concepts because some of the entities were originally mapped to the closest syntactically match UMLS concept that did not contain a SNOMED CT concept. In the next step, the entities mapped to the equivalent UMLS concepts including SNOMED CT concept. We used the sentence classification component of this corpus to automatically train a range of supervised machine learning classifiers to identifying text segments with the mentions of ADRs, WDs, DIs, SSIs, EF, and INF. SVMs classifiers had the highest performance with F-Score 0.90. We also measured performance of the cTAKES (clinical Text Analysis and Knowledge Extraction System) in identifying patients' expressions of ADRs and WDs with and without adding the PsyTAR dictionary to the core dictionary of cTAKES. Augmenting cTAKES dictionary with this corpus improved the F-score of cTAKES by 25%.

### 3.1. Sample characteristics

The original sample size was 891 reviews, with 432 reviews for the SSRI class (Zoloft (213) and Lexapro (219)) and 460 reviews for the SNRI class (Cymbalta (231) and Effexor XR (228)). Five posts were excluded from the sample due to lack of review content. The final sample is almost equally divided between SSRI and SNRI classes. All the drug reviews in the sample were posted by patients who were prescribed one of the psychiatric medications (Zoloft, Lexapro, Cymbalta or Effexor). If a part of a review was related to the experience of the third-person, we excluded it from the data analysis.

The majority of participants are female for both classes of drugs, which is as expected as the CDC (Centers for Disease Control and Prevention) statistics shows that the prevalence of antidepressants use is almost twice in females compared to males [18]. The mean age 37 may imply that younger patients are more likely to report their experiences through online healthcare forums. Duration of usage ranged from 1 day to 20 years with median five months, suggesting a spectrum of patient tolerability with the drugs of this study. Table 6 shows the statistics for the whole sample, as well as for each class of drugs.

### 3.2. Sentence classification component

On average, patients on SNRI drugs had longer posts, 7.1 versus 6.4 for the number of sentences, and 117.5 versus 103.3 for the number of tokens. Table 7 presents the statistics on posts, sentences, and tokens.

Table 8 shows the counts of annotated items after resolving disagreement for the complete corpus. The frequency of sentence classification for two classes of drugs is almost equally distributed between the two classes of drugs. However, patients on SNRIs reported more WDs (215 vs. 64) than patients on SSRIs. If a sentence did not contain any information about adverse drug reaction (ADR), withdrawal symptoms (WD), drug effectiveness (EF), or drug ineffectiveness (INF),

**Table 5**
Examples of mapping entities to both UMLS and SNOMED-CT concepts.

| Original term | UMLS (1) | SNOMED-CT (1) | UMLS (2) | SNOMED-CT (2) | Qualifier |
|---|---|---|---|---|---|
| ongoing jittery feeling **no feelings at all** | C0549209/Feeling jittery/Sign or Symptom<br>C0456820/Loss of capacity to feel emotions/Mental or Behavioral Dysfunction | No code<br>Loss of capacity to feel emotions (finding) | C0849963/Feeling nervous/Sign or Symptom | Feeling nervous (finding) | Ongoing **(QP)** |
| Extreme fuzzy head | C0541974/Fuzzy head/Finding | No-code | C0423994/Unable to think clearly/Mental or Behavioral Dysfunction | Unable to think clearly (finding) | Extreme **(QS)** |
| left leg felt like it was on fire | C2219739/burning sensation in left leg or foot/Sign or Symptom | No code | C0085624/Burning sensation/Sign or Symptom | Burning sensation (finding) | |
| couldn't even make dinner **Brain Shiver** | C0564332/Difficulty preparing meal/Finding<br>No concept | Difficulty preparing meal (finding)<br>No concept | | | |

**Table 6**
Sample statistics.

| Dataset statistics | Dataset | SSRI | SNRI |
|---|---|---|---|
| Sample Size | 891 | 432 | 459 |
| No. of reviews with text | 887 | 428 | 459 |
| Time span | Feb 2001<br>Sep 2016 | Feb 2001<br>Aug 2016 | July 2004<br>Sep 2016 |
| Rating | 3.16 | 3.33 | 3 |
| Gender | F 669 (76%)<br>M 212 (24%)<br>Missing value (11) | F 310 (72%)<br>M 118 (28%)<br>Missing value (4) | F 359 (79%)<br>M 94 (21%)<br>Missing value (7) |
| Age | Avg. 37<br>Med. 35<br>Missing values (12) | Avg. 35<br>Med. 34 | Avg. 38<br>Med. 37 |
| Age range | 14–83<br>Missing values (3) | 14–73 | 14–83 |
| Duration of usage | Avg. 18 months<br>Med. 5 month | Avg. 19 months<br>Med. 5 months | Avg. 17 months<br>Med. 5 month |
| Duration of usage (range) | 1 day − 20 years | 1 day- 16 years | 1 day − 20 years |

**Table 7**
Statistics on frequency of posts, sentences, and tokens.

| | Corpus | SSRI | SNRI |
|---|---|---|---|
| Number of the Posts | 887 | 428 | 459 |
| Overall number of the sentences | 6004 | 2749 | 3255 |
| Average number of sentences in each post | 6.77 | 6.42 | 7.1 |
| Number of sentences in each post (range) | 1–35 | 1–35 | 1–25 |
| Overall number of tokens | 98,186 | 44,237 | 53,949 |
| Average number of tokens (words) in each post | 110.7 | 103.36 | 117.53 |

**Table 8**
Frequency of items after resolving the disagreements for complete corpus.

| Entities | Corpus | SSRI | SNRI |
|---|---|---|---|
| ADRs | 2168 | 1059 | 1109 |
| WD | 438 | 92 | 346 |
| EF | 1087 | 579 | 508 |
| INF | 337 | 157 | 180 |
| SSIs | 789 | 419 | 370 |
| DI | 517 | 271 | 246 |
| Others | 2107 | 912 | 1195 |

drug indications (DI), signs/symptoms/illness (SSI), then it was labeled it as others (not applicable). Others (35%) after ADR (36%) has the highest frequency in the sentence classification component; this reflects the challenge of machine learning algorithms for automatic classification of sentences (text segments) containing information for items specified in Table 8.

### 3.3. Entity identification component

From 7414 identified entities, 65% are ADRs, 8% are WDs, 16% are SSIs, and 11% are DIs. In total, 40% of the entities were duplicates, with the highest frequency of duplicates for SSIs (62%) and DIs (60%), and the lowest frequency for withdrawal symptoms (21%). This indicates that patients mostly use the same medical terms to describe their illnesses and sign/symptoms. Physiological entities constitute 53% of the total entities, followed by psychological entities (40%), cognitive entities (5%) and functional problems entities (2%). Among ADRs and WDs, psychological and cognitive expressions have higher variability than physiological expressions, possibly is due to the level of subjectivity of these types of entities. For the two classes of psychiatric medications,

**Table 9**
Statistics on entity identification for each class of drugs and each type of entity.

| | Total | | Physiological | | Psychological | | Cognitive | | Functional | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Entities (All) | Unique (% All) | All | Unique (% All) | All | Unique (% All) | All | Unique (% All) | All | Unique (% All) |
| Corpus-entities | 7414 | 60% (4419) | 3975 | 63% (2523) | 2959 | 50% (1483) | 328 | 79% (264) | 152 | 97% (147) |
| ADR Corpus | 4813 | 67% (3180) | 3402 | 62% (2093) | 1072 | 75% (808) | 255 | 78% (198) | 84 | 96% (81) |
| ADR-SSRI | 2247 | 1370 | 1562 | 875 | 523 | 372 | 121 | 84 | 41 | 39 |
| ADR-SNRI | 2566 | 1810 | 1840 | 1218 | 549 | 436 | 134 | 114 | 43 | 42 |
| WD-Corpus | 590 | 79% (468) | 361 | 74% (268) | 171 | 84% (144) | 26 | 100% (26) | 32 | 94% (30) |
| WD-SSRI | 102 | 72 | 55 | 34 | 41 | 33 | 4 | 4 | 2 | 1 |
| WD-SNRI | 458 | 376 | 286 | (222) | 121 | 104 | 21 | 21 | 30 | 29 |
| SSI-Corpus | 1219 | 38% (462) | 133 | 76% (101) | 1036 | 30% (315) | 27 | 85%(23) | 23 | 100% (23) |
| SSI-SSRI | 642 | 217 | 42 | 33 | 577 | 164 | 14 | 11 | 9 | 9 |
| SSI-SNRI | 577 | 245 | 91 | 68 | 459 | 151 | 13 | 12 | 14 | 14 |
| DI-Corpus | 792 | 40% (309) | 79 | 77% (61) | 680 | 32% (218) | 20 | 85% (17) | 13 | 100% (13) |
| DI-SSRI | 419 | 145 | 27 | 21 | 374 | 108 | 12 | 10 | 6 | 6 |
| DI-SSRI | 373 | 164 | 52 | 40 | 306 | 110 | 8 | 7 | 7 | 7 |

SSRI and SNRI, the distribution of ADRs is almost similar. However, patients reported more WDs with the SNRI class. Table 9 shows the frequencies of identified entities for the total corpus, for each class of drugs, and for each type of entity separately. Appendix B includes the top five identified entities for each type of entities and class of drugs.

### 3.4. Entity normalization (mapping) component

The final set of the normalized components contains 918 unique UMLS concepts and 755 unique SNOMED CT Concepts for ADRs, WDs, SSIs, and DIs (see Table 10). The 3180 unique identified ADRs were mapped to 673 UMLS concepts. On average, for each standard ADR concept associated with psychiatric medications, there is 4.7 layperson expressions of that ADR, reflecting the challenge of automatic identification of the ADRs using standard medical lexicons and text mining algorithms. The 462 unique identified SSIs were mapped to 171 UMLS concepts, indicating that patients in drug reviews mostly use the diagnosis terms provided by healthcare professionals to report their illness, sign, or symptoms.

Overall, all four types of entities (ADRs, WDs, DIs, and SSIs) were mapped to 791 UMLS concepts, from which 164 concepts did not include SNOMED expressions. For the UMLS concepts without SNOMED CT expression, we identified equivalent UMLS concepts containing the SNOMED CT concepts. Table 11 lists the most frequent UMLS concepts in the corpus that did not include SNOMED CT expressions. For the UMLS concepts without SNOMED CT expression, we identified equivalent UMLS codes covering the SNOMED CT concept.

### 3.5. Qualifiers indicating intensity and persistency of entities

Identifying the qualifiers can help healthcare providers to estimate the magnitude of debilitating effects of ADRs/WDS on patient quality of life, and whether they need to use any specific interventions to improve patient adherence to medication. Table 12 shows statistics on the qualifiers indicating intensity and persistency of the ADRs, WDs, SSIs and DIs.

### 3.6. PsyTAR in use

The PsyTAR corpus was created as a component-based system to enable training and evaluation of the text mining algorithms, which can be applied to a large dataset of patients' drug reviews on psychiatric medications, particularly SSRIs and SNRIs drugs. A pipeline of the processing of components can be constructed using this corpus to automate text segment classification and identification of effectiveness and a wide-range of ADRs, WDs, and DIs associated with psychiatric medications (Fig. 2). Patients' drug reviews can pass through the pipeline while being analyzed by each component with the results of analysis being available to later components.

The pipeline has three main parts:

**Text processing:** In this component, text (patient's review for a medication) will be processed for removing the personally identifiable information and handling grammatical and punctuation errors. In the next step, text is split into sentences and tokens.

**Sentence classification:** Each sentence is annotated for the presence of ADRs, WDs, DIs, SSIs, EFs, and INFs. We utilized the gold standard annotation of sentence classification to train support vector machine (SVMs) to recognize sentences for the seven items. SVMs are trainable classifiers and they have been shown to be effective in classifying text segments [27]. For our seven sentences labels (ADRs, WDs, DIs, SSIs, EF, INF, and others), we created seven binary classifiers and each applied independently of each other on the annotated sentences. We used SVM implementation system in Scikit-learn (a machine learning library for Python) [17]. A pre-processing step including tokenization, lowering, and stemming were applied on the sentences. As a feature, we computed Tf.Idf (term frequency, inverse document frequency) values for every token in each sentence [15]. The performance of the classifiers, which were computed over 6004 sentences in the corpus using ten-fold cross validation, is shown in Table 13**,** using the standard metrics of Precision, Recall, and F-Score. To show the difficulty of sentence classification task for each item, Table 13 also includes the IAA score for the two independent annotators. Please note that the classifiers were trained on the consensus between annotators annotating sentences. The high F1score indicates the high quality of the annotation sentences of the PsyTAR corpus and also indicates that it can be used as a useful tool for developing text-segment classifiers to address the challenges of ambiguity in semantic types of patients expressions (e.g., ADR vs. WD vs. SSI vs. DI) presented in drug review posts.

We also built other types of binary classifiers including Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR) on the sentence classification component of the PsyTAR corpus. The average F-score for RF, NB, and LR classifiers were 0.87, 0.88, and 0.86 respectively, indicating that SVM classifiers outperform the RF, NB, and LR for this task.

**Entity identification and normalization**: The second and third

**Table 10**
Statistics on entities normalized with UMLS and SNOMED-CT concepts.

| Entities | Frequency of identified UMLS concepts | Physiological | Psychological | Cognitive | Functional |
|---|---|---|---|---|---|
| **ADRs** | **673** | **420** | **188** | **46** | **28** |
| The five most frequent ADRs after normalization | | Sleeplessness (172), nausea (171), weight gain (156), lack of libido (139), headache (102) | Anxiety (92), detailed recall of dream (62), depressed mood (42), apathy (42), feeling suicidal (38) | Foggy feeling in head (46), unable to concentrate (29), amnesia (19), memory impairment (15), forgetful (14) | Difficulty in daily functioning (11), emergency room admission (9), social withdrawal (9), restricted work performance (7), hospitalization (6) |
| **WD** | **218** | **105** | **67** | **18** | **19** |
| The five most frequent WDs after normalization | | Brain shivers (52), Dizziness (44), nausea (33), headache (27), malaise (10) | Irritable mood (16), depressed mood (11), mood swings (11), nightmares (8), severe anxiety (6) | Confusion (4), unable to concentrate (3), mental suffering (3), foggy feeling in head (2), actual low self control (1) | Difficulty in daily functioning (4), bed-ridden (3), restricted work performance (3), difficulty driving a car (2), emergency room admission (2) |
| **SSIs** | **171** | **46** | **103** | **9** | **13** |
| The five most frequent SSIs after normalization | | Pain (15), fibromyalgia (9), sleeplessness (9), lethargy (9), fatigue (8) | Depressed mood (429), anxiety (232), panic attacks (25), feeling suicidal (25), social fear (social anxiety) (21) | Obsessive thoughts (12), unable to think clearly (2), unable to control emotion (1), unable to think clearly (1) | Difficulty in daily functioning (6), housebound (finding) (3), Social withdrawal (2), self-doubt (2), difficulty maintaining relationship (2) |
| **DIs** | **143** | **43** | **84** | **7** | **9** |
| | | Pain (12), Fibromyalgia (6), sleeplessness (5), feeling upset (3), tired (3) | Depressed mood (290), anxiety (142), panic attacks (15), feeling suicidal (18), social fear (social anxiety) (13) | Obsessive thoughts (11), unable to think clearly (2), racing thoughts (2), intrusive thoughts (2), unable to control emotions (1) | Difficulty in daily functioning (4), social withdrawal (2), personal relationship breakdown (2), difficulty maintaining relationships (1), does throw objects (1) |

components of the PsyTAR corpus includes the annotated patients' expressions for the ADRs, WDs, SSIs, and DIs, which were normalized by mapping to UMLS and SNOMED CT concepts. Name entity recognition (NER) systems can be trained on these components to automate the identification of patients' expression for the entities. To demonstrate the usability of the PsyTAR corpus in improving performance of NER systems in identifying patients' expressions of ADRs and WDs from drug reviews, we utilized the clinical Text Analysis and Knowledge Extraction System (cTAKES) [22].

cTAKES has a modular system of pipelined components that was primarily developed for identification and extraction of medical entities from clinical text, by combining rule-based and machine learning methods. cTAKES components include sentence boundary detector, tokenizer, normalizer, part-of-speech (POS) tagger, shallow parser, and named entity recognition (NER) annotator. The cTAKES NER component uses a "terminology-agnostic dictionary look-up algorithm within a noun-phrase look-up window" [22] for name entity recognition. The core dictionary of cTAKES is built on UMLS terminologies.

For the purpose of this evaluation, we randomly split the patients' drug reviews in the PsyTAR corpus into training and test sets, with 80% of reviews used as a training set, and the remainders for measuring cTAKES performance. The identified and normalized ADRs and WDs in the training set were used to augment cTAKES' core dictionary. PsyTAR's dictionary included the patients' expressions of ADRS and WDs and the corresponding UMLS CUIs.

We used the following approach to measure the improvement of cTAKES' performance for identifying patients' expressions of ADRs and WDs using PsyTAR's added dictionary:

(a) Step 1: We customized the NER component in the cTAKES default pipeline to identify signs and symptoms in the text. Drug names, procedure, and other medical entities that do not indicate sign and symptoms in the text were excluded from the NER component.
(b) Step 2: We ran the cTAKES with the customized pipeline on the sentences in the test sets to extract the ADRs and WDs from the drug review sentences. The cTAKES performance was measured using the standard metrics of Precision, Recall, and F-Score, which were 0.19, 0.32, and 0.24 respectively.
(c) Step 3: We evaluated cTAKES with the customized pipeline augmented with PsyTAR's dictionary, using the test set. In this case, the Precision, Recall, and F-Score for extracting ADRs and WDs increased to 0.48, 0.50, and 0.49 respectively, showing the added-value of PsyTAR's dictionary.

Fig. 3 depicts the difference between the performance of cTAKES with and without augmenting with PsyTAR dictionary.

cTAKES was primarily designed to identify medical terms from clinical notes, so it is not surprising that its performance is relatively low for identifying ADRs and WDs from patients' drug reviews due to colloquial expressions and laypersons language. However, after augmenting the core dictionary of cTAKES with PsyTAR's dictionary, the F-score increased by 25%. Note that this improvement occurs while almost 73% of the ADRs and WDs expressions were unique in the PsyTAR's corpus (see Table 9). cTAKES uses dictionary look-up algorithm within a noun-phrase look-up window to recognize NER in the text. For example, cTAKES could recognize the "unable to drive on the freeway" in the test data, while there is no exact syntactic match in the PsyTAR dictionary for this expression. But other terms such as "could not drive" or "could definitely not drive" were in the dictionary. These results indicate that despite high semantic and syntax variation of patients' expressions of ADRs and WDs, PsyTAR's corpus can significantly improve the performance of NER systems for identifying the expressions in online healthcare forums, particularly for psychiatric medications. Improving the performance of such applications, particularly for psychological and cognitive ADRs and WDs in patients' drug reviews required more sophisticated Name Entity Recognition system.

**Table 11**
The most frequent UMLS concepts in the corpus without SNOMED expression.

| UMLS-Primary Concept | SNOMED-CT Primary Concept | Frequency | UMLS-Equivalent Concept | SNOMED-CT Concept |
|---|---|---|---|---|
| C0859330/Foggy feeling in head/Finding | No Code | 48 | C0683369/Clouded consciousness/Sign or Symptom | Clouded consciousness (finding) |
| C0857507/Spaced out/Finding | No Code | 35 | C0349446/Dissociative trance/Mental or Behavioral Dysfunction | Dissociative trance (disorder) |
| C0392703/Shakes/Finding | No Code | 33 | C0040822/Tremor/Sign or Symptom | Tremor (finding) |
| C0549209/Feeling jittery/Sign or Symptom | No Code | 29 | C0849963/Feeling nervous/Sign or Symptom | Feeling nervous (finding) |
| c0549209/feeling jittery/sign or symptom | No Code | 26 | c0557875/tired/sign or symptom | tired (finding) |
| C0857486/Felt like a zombie/Finding | No Code | 23 | C0233484/Emotionally detached/Finding | Emotionally detached (finding) |

#### 3.6.1. Further applications of the PsyTAR corpus

The PsyTAR dataset can also be used to evaluate the association between different types of ADRs and patient attitude toward antidepressants. A sample of the SSRI and SNRI drugs that was used for developing the PsyTAR corpus includes patients' attitudes (satisfaction) towards the medications with an average of 3.16 (see Table 6). Patients' attitude toward drugs is a strong determinant of non-adherence to medications. In our prior studies [29,31,32], we tested the association between 21 physiological symptoms listed in the Antidepressants Side-Effect Checklist (ASEC) questionnaire [25] and eight most common psychological and cognitive ADRs (of SSRIs and SNRIs drugs) with patient attitudes to medications specified in this study, using Chi-square and Fisher's exact test.

Our findings showed that among 21 physiological ADRs, "dry mouth", 'increased appetite", "weight gain", "problem with sexual functioning", "disorientation", and "palpitation" were associated with patients negative attitudes ($P < 0.05$ for all the ADRs listed). While, there were no significant association between ADRs "headache", "insomnia", "drowsiness", "constipation", "diarrhea", "decreased appetite", "sweating", "increased body temperature", "nausea or vomiting", "vertigo", "light headed", and "problem with urination" and patients attitudes towards the medications ($P > 0.05$ for all the ADRs). All the physiological and cognitive ADRs including "emotional indifference"," apathy", "mood swing", and "anxiety", "difficulty in concentrating", and "memory problem" were strongly associated with negative attitude towards the medications ($P < 0.05$ for all the ADRs listed). In the prior study, we tested the association of subset of ADRs reported by patients in the PsyTAR corpus. As Table 10 shows, patients reported 673 different types of ADRs and 218 types of WDs. Future studies can measure the association between other types of ADRs and WDs along with patient's attitude toward the medications.

## 4. Discussion

We generated a corpus (PsyTAR) of two groups of psychiatric medications, SSRI and SNRI, which is annotated over patients' narrative data from a healthcare forum called "askapatient.com." This corpus, of four of the most common psychiatric medications, consists of three components: sentence classification, entity identification, and terminology association. We followed a systematic approach to ensure reliable and consistent annotated data to support text-mining of consumer health posts for pharmacovigilance purposes.

Analysis of the corpus showed that both classes of SSRIs and SNRIs have almost the same frequency distribution for the components of sentence classification and entities identification. Among identified entities (ADRs, WDs, IDs, and SSIs), the lowest frequency of duplicates occurred for drug indications and signs/symptoms/illness, signifying that patients mostly use the standard diagnosis terms provided by healthcare professionals to report the reasons for their prescription.

The entities were further classified as physiological, psychological, cognitive, and functional problems. Functional problems comprised only 2% of the entire identified set of entities. Regarding the importance of functional problems in understanding ADRs impacts on quality of life, it would be useful if healthcare forums also asked patients to report the impact of drug ADRs on their daily functioning and social activities.

Some applications of the PsyTAR corpus were discussed in Section 3.6.1 "PsyTAR in Use". High performance of SVMs classifiers trained on this corpus for sentence classification and the significant improvement of cTAES performance in identifying ADRs and WDs after augmenting with the PsyTAR dictionary indicate that PsyTAR corpus is a high quality annotated corpus. Furthermore, PsyTAR can have important implications in addressing the challenges of ambiguity and colloquial expressions of pharmacological effects of medications in patients' drug reviews, particularly for psychiatric medications.

### 4.1. PsyTAR corpus innovation

There are also other corpora for ADRs identification, which were mainly built on biomedical literature, such as Medline case reports and abstracts [5–7]. Annotated corpora for mentions of ADEs using biomedical literature and clinical notes in EHR systems have important implications for automatic extractions of ADEs from these resources. However, they may not provide significant performance improvement for ADEs identification in consumer health posts in social media. As Sarker and Gonzalez [21] showed, incorporation of the ADEs corpus (an annotated corpus of ADRs constructed based on biomedical literature) with two corpora constructed based on social media does not provide significant improvement in the accuracy of an ADR assertive sentences-classifier system because the ADE corpus structure is not compatible with the corpora developed using consumer health posts in online healthcare forums. In addition, biomedical literature follows

**Table 12**
Frequency of identified qualifiers associated with ADRs and WDs with example of the top five ones.

| Category | Frequency | Example |
|---|---|---|
| Mild | 128 | Mild (30), Slight (28), a little (15), slightly (12), minor (7) |
| Moderate | 68 | Some (36), Moderate (5), 10 pounds (4), Somewhat (3), almost (2) |
| Severe | 700 | Very (79), Extreme (71), severe (53), So (25), horrible (24) |
| Persistent | 267 | All the time (23), constant (16), constantly (15), always (13), chronic (8) |
| Not-persistent | 376 | Initially (25), at first (24), occasional (12), in the beginning (11), sometimes (11), in the beginning (11) |

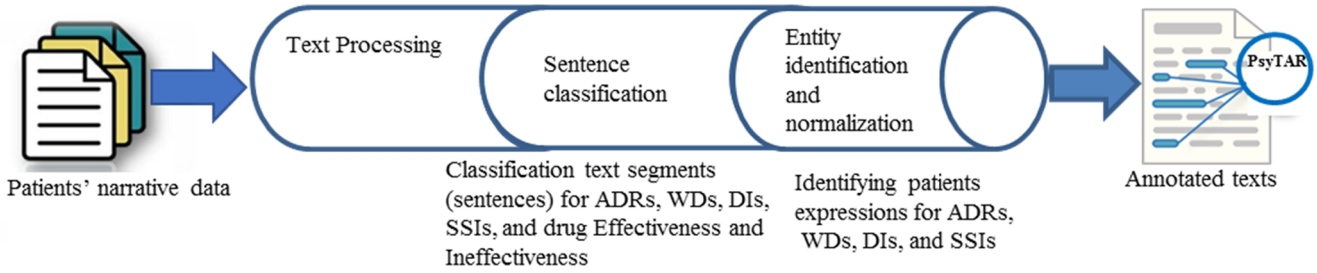



**Fig. 2.** PsyTAR pipeline for automatization of text segment classification and identification of patients' expression of pharmacological effects associated with Psychiatric medications.

**Table 13**
Sentence classification score for SVM classifiers trained on the Ps**yTAR's** sentence classification component.

| Entity | Metric | | | IAA |
|---|---|---|---|---|
| | Precision | Recall | F-Score | |
| Adverse drug reactions (ADRs) | 0.864 | 0.855 | 0.858 | 0.81 |
| Withdrawal symptoms (WDs) | 0.936 | 0.943 | 0.936 | 0.69 |
| Sign/symptom/illness (SSIs) | 0.836 | 0.841 | 0.830 | 0.82 |
| Drug indications (DIs) | 0.973 | 0.975 | 0.964 | 0.65 |
| Drug effectiveness (EF) | 0.924 | 0.928 | 0.922 | 0. 91 |
| Drug ineffectiveness (INF) | 0.927 | 0.933 | 0.929 | 0.76 |
| Overall | 0.91 | 0.912 | 0.906 | 0.75 |

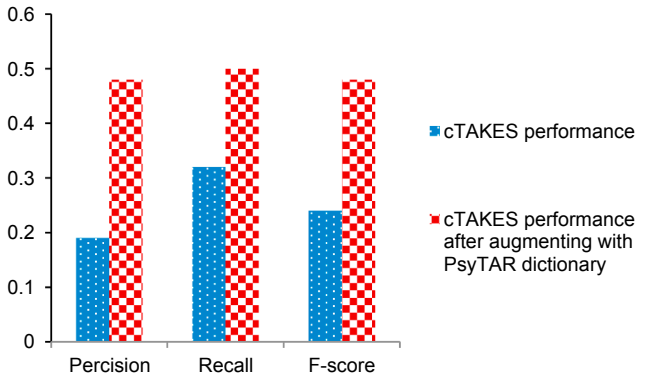



**Fig. 3.** Performance of cTAKES with and without PsyTAR dictionary for identifying ADRs and WDs from patient drug reviews.

grammatical rules, which are not rich in colloquial expression of medical entities. Therefore, they do not address the challenges of health data extractions from consumer health posts.

The CADEC corpus developed by Karimi et al. [11] is the only open source corpus developed using consumer health posts. This corpus consists of 1231 comments for various formulations of Diclofenac and Lipitor. Drug reviews were annotated for entity identification and entity normalization. The identified entities include ADRs, symptoms, disease, and mentions of drugs, which were mapped to SNOMED CT, MEDRA, and AMT. The IAA for entity identification was 60.2 (46.2 for Diclofenac category and 74.2 for Lipitor category) on the strict agreement.

Our corpus is different from the CDADE in several important aspects, including the type of drugs studied, the methodology of development, identified entities, and normalization process. The detail of the differences is as follows:

(1) By following a systematic approach, we set the unit of analysis at the sentence level for annotation which resulted in a higher IAA (86%) for identified entities in the PsyTAR corpus compared to the IAA (60.4%) reported for the CADEC corpus. The higher IAA ensures consistency and reliability of the PsyTAR corpus for machine learning and text mining tasks.
(2) PsyTAR includes SSRI and SNRI drugs that are mainly prescribed for patients with psychiatric disorders (e.g., anxiety disorder, major depressive disorder, posttraumatic stress disorder (PTSD)), while the CADEC focuses on Diclofenac that is mainly prescribed for arthritis pain and migraines, and Lipitor that is prescribed to lower low-density lipoprotein in the blood thereby reducing cardiovascular risks. Therefore, there is no overlap of drug indications between these two corpora.
(3) As we explained in this study, SSRIs and SNRIs drugs are associated with psychological (e.g., "lack of capacity to feel emotion", "apathy", "aggression") and cognitive ADRs ("lack of concentration", and "memory problem"), while Diclofenac is typically associated with physiological ADRs and Lipitor is significantly less associated with psychological or cognitive ADRs.
(4) We also identified functional problems (e.g., "unable to drive","-difficulty in daily functioning") associated with ADRs and WDs in this corpus, while the CADEC corpus does not include this type of entities.
(5) There may be some overlaps between physiological ADRs mentioned by CADEC corpus and PsyTAR corpus. However, comparison of the most frequent physiological ADRs mentioned by CADEC and PsyTAR shows that the overlap is not significant.
(6) Moreover, the PsyTAR includes the drug effectiveness/ineffectiveness and qualifiers representing severity and persistency of the ADRs and WDs, which were not covered by the CADEC. Finally, all the identified entities were mapped to UMLS, which was not reported by the CADEC.

Corpora such as PsyTAR with focus on identification of ADEs events and beneficial effects from consumer health posts are expensive to develop. However, when they are developed, they have important implications for subsequent research. As we shown some of the applications in the section "PsyTAR in Use", first, they can improve the recall of dictionary-based systems designed for automatic identification of the pharmacological aspects of drugs. Second, they can aid in developing and evaluating text mining and machine learning systems aimed to identify ADRs, WDs, and drug indications from consumer health posts. Third, they can be used for training of machine learning-based classifiers aimed to distinguish ADRs from other semantic types, such as drugs indications. Fourth, they can be used for developing and testing automatic systems aimed to measure effectiveness and ineffectiveness of psychiatric medications. Finally, they have applications in developing systems targeting automatic mapping between layperson expression of health information to the UMLS and SNOMED concepts.

A structured vocabulary of layperson expressions of adverse effects and drug indications of the PsyTAR may also be used in EHR systems to facilitate the seamless exchange of information between patients and

clinicians. This can be achieved by mapping information in personal health records (PHR) to EHR systems. This can be further used to design patient safety initiatives including decision support tools to reduce the risk of ADRs.

PsyTAR can also be used for measuring the association between different types of ADRs and patients attitudes toward medications, which is a significant predictor of patient adherence behavior. We presented a summary of our findings for testing the associations between 21 physiological ADRs and eight psychological and cognitive ADRs with patients' attitudes towards medications specified in this study. Future studies can focus on analyzing the association of patient attitudes with other types of psychiatric medication ADRs and WDs in the PsyTAR corpus.

### 4.2. Normalization challenge

We managed to improve the accuracy of inference of intended meaning of the colloquial expression of ADRs by expanding guidelines built on clinical trials and qualitative studies, and contextual cues in patients reviews. However, we had some challenges in selecting proper UMLS/SNOMED-CT concepts for ADR expressions. Throughout the corpus, we could not map 128 expressions (out of 7414) to either UMLS or SNOMED concept. For example, the ADR of "tying a bow or a knot very difficult", could not be mapped to any UMLS concepts due to the vagueness of this expression. By this expression, does patient imply muscular weakness or pain, or is it about a feeling of lethargy or fatigue?

Another challenge is that layperson expressions of ADRs are fuzzier than the corresponding UMLS/SNOMED concepts. For synonym concepts, the layperson ADRs expressions are more likely to be "narrower-than" or "broader-than" their closest UMLS concept. For example, "not being able to express sadness" or "could not cry in funeral ceremony" were all mapped to "blunted affects". This happens particularly for psychological systems and functional problems.

Some of the ADRs/WDs expressions, such as "brain shivers" (feeling electric shocks, brain zap) are briefly discussed by literature [1], however no UMLS concept is available for them. This ADR/WD was identified 88 times in the corpus.

### 4.3. Limitations

***Limitation of the sample:*** Sample of this study was collected from the healthcare forum "askapatients.com" for two classes of antidepressants, SSRI and SNRI, that may not be representative of drugs from other classes, such as the TCAs. Also, it is possible that patients' self-reported experiences with these medications may not be a balanced sample of experiences with drugs in other healthcare forums.

***Ambiguity in patients' drug complaints*:** There is a risk that patient complaints about a specified antidepressant could be caused by interaction of the antidepressant with another medication or herbal treatment. In addition, some of the severe ADRs which were reported, such as suicidal attempt or self-harm, could be the result of improper dosing of an antidepressant. Furthermore, there is the risk of patient misinterpretation of mental disorders or signs or symptoms of the ADRs associated with the drugs.

***Lack of specificity in mapping:*** Patient expressions of ADRs/WDs are affected by numerous factors including geographic region, demographic information (such as level of education), general health, environmental conditions, and personal experience with illness and treatment. This problem may influence the accuracy of linking layperson expressions to standard vocabularies, and subsequently, impact the accuracy of performance of automatic systems using the PsyTAR corpus as the gold standard.

***The Risk of non-genuine reportsx in social media:*** although several studies showed the reliability of patient self-reporting in healthcare forums discussing medications, we cannot exclude the risk of fake or inaccurate reporting.

***The Possibility of human errors in data analysis:*** Although the entire data set is double coded, there is still the possibility that annotators did not interpret a sentence correctly and therefore assign a wrong label to it. In addition, the span of the identified entities may include less or more information than necessary. These issues may affect the performance of machine learning systems trained based on this corpus to identify drug effectiveness, ADRs, and drug indications in consumer health posts.

Future research should identify information from review posts that indicate perceived distress from ADRs and WDs. This research could inform healthcare providers of patient perspectives and attitudes towards their medications. These variables are important from the perspective of testing hypotheses concerning attitude and adherence towards medications. This will require identifying drug names from review posts and mapping them to Rx-Norm coding standards. Detecting the span of drugs mentioned in the posts can help researchers to identify the variation of spelling drug names.

## 5. Conclusion

In this study, we developed a corpus of patient posts from an online healthcare forum to address the challenges of automatic extraction of the pharmacological effects of psychiatric medications from patient generated data. We showed that psychological and cognitive adverse events, and functional problems have higher semantic variability and deviation from standard vocabularies compared to physiological ADEs and drug indications. We also showed that there are some gaps in the SNOMED CT coverage of ADEs, which can potentially be considered for addition. Training text-mining and machine learning algorithms on this corpus can significantly improve identification of these ADEs from patient drug reviews, specifically for psychiatric medications. In addition, this corpus can be used to address the ambiguity of the semantic types of identified terms, as same terms may be recognized as an adverse drug reactions, withdrawal symptoms, or drug Indications. The corpus is publicly available at Zolnoori et al. [30].

## Conflict of interest

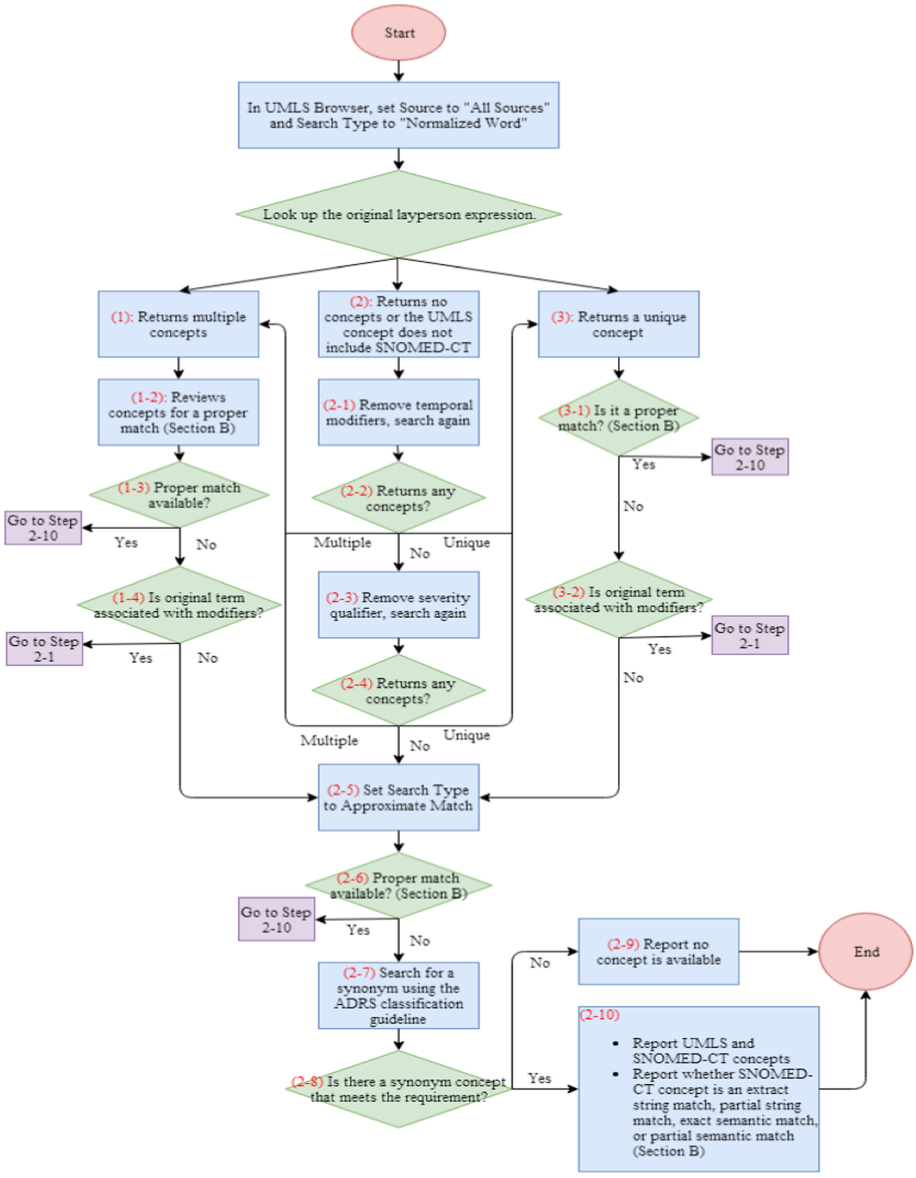
We have no conflict of interest to declare.

## Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM) and Lister Hill National Center for Biomedical Communications (LHNCBC). We thank our colleagues at NLM, University of Wisconsin-Milwaukee, and Mayo Clinic who provided insight and expertise that greatly assisted this research project. We also thank Dr. Greg Downs and Margaret Downs for their insight and support throughout this project.

## Appendix A

Flowchart of Finding Proper Concept for Layperson's Expression of Medical Entities




## Appendix B

The top five identified entities for each class of drug s and type of entities separately

| | Physiological – top (5) | Psychological – top (5) | Cognitive – top (5) | Functional – top (5) |
|---|---|---|---|---|
| SSRI-ADRs | Weight gain (59), nausea (39), dry mouth (30), insomnia (30), fatigue (25) | vivid dreams (20), increased anxiety (10), nightmares (9), suicidal thoughts (9), anxiety (8) | Memory loss (5), brain fog (4), inability to concentrate (3), hard to focus (4), forgetfulness (2) | Called in sick (2), impossible to do my job (1), struggling to learn new things (1), unable to function (1), struggle just to comb my hair (1) |
| SNRI-ADRs | Insomnia (48), constipation (38), weight gain (37), nausea (34), dry mouth (29) | vivid dreams (19), anxiety (14), nightmares (14), irritability (6), crying (6) | Memory loss (6), confusion (4), Foggy brain (3), inability to concentrate (3), couldn't focus on anything (2) | Hospitalized (2), unable to function (1), loss of friends (1) unable to work (1), cannot drive (1) |
| SSRI-Withdrawal symptoms | Dizziness (7), upset stomach (3), brain zap (2), nausea (2), dizzy (2) | Irritability (4), suicidal thoughts (2), aggression (1), crying spells (1), very anxious (1) | Slight confusion (1), Severe mental confusion (1), Couldn't concentrate (1), Agitation (1), mood swings (1) | Ended up in the er (1), 1/2 weeks in phych word (1), cannot function (1) |
| SNRI Withdrawal symptoms | Dizziness (16), nausea (11), brain zaps (8), headaches (6), dizzy (6) | Mood swings (5), nightmares (3), moody (2), feel like a walking zombie (2), aggressive (2) | Dissociative episodes (1), foggy (1), lack of concentration (1), confused (1), cannot think (1), feel spaced out (1) | can't function (1), could not drive (1), Difficulty tidying house (1), Difficulty performing shopping activities (1), Difficulty performing educational activities (1) |
| SSRI-SSIs | Insomnia (3), upset (3), night sweat (2), fibromyalgia (2), not being able to get out of bed (1) | Depression (155), anxiety (120), depressed (36), panic attacks (13), anxious (12) | intrusive thoughts (3), obsessive thoughts (2), confused (1), racing thoughts (1), inner critic (1) | lost my job (2), isolating (1), barely functional (1), dysfunctional (1) |

| | | | | |
|---|---|---|---|---|
| SNRI-SSIs | Pain (11), fatigue (4), seizures (3), fibromyalgia (3), fibro pain (2) | Depression (160), anxiety (84), depressed (22) anxious (8), panic attacks (8) | obsessive thinking (2) Mental clutter (1), memory loss (1), rumination (2), stutter (1) | Loss everyone in my life (1), could not function at work (1), lost everyone in my life (1), harsh edge to my inter-action with others |
| SSRI-Drug Indications | Insomnia (3), upset (2), night sweats (2), fibromyalgia (1) | Depression (109), anxiety (69), depressed (22), panic attacks (7), suicidal thoughts (6) | intrusive thoughts (2), obsessive endless thoughts (1), confused (1), racing thoughts | Unable to function (1), barely functional (1), isolating (1), dysfunctional (1), |
| SNRI-Drug Indications | Pain (9), fibro pain (2), fibromyalgia (2), neuropathy pain (2), joint pain (2), worry (1) | Depression (116), anxiety (52), depressed (11), panic attacks (5), anxious (4) | mental clutter (2), felt in control again (1), obsessive thinking (1), rumination (1), | life filled with doubt (1), threw things (1), lost everyone in my life (1), could not think about holding a job (1), needy |

## References


[1] J. Aronson, Bottled lightning, BMJ 331 (7520) (2005) 824.
[2] A. Benton, L. Ungar, S. Hill, S. Hennessy, J. Mao, A. Chung, J.H. Holmes, Identifying potential adverse effects using the web: a new approach to medical hypothesis generation, J. Biomed. Inform. 44 (6) (2011) 989–996.
[3] J. Charan, T. Biswas, How to calculate sample size for different study designs in medical research? Indian J. Psychol. Med. 35 (2) (2013) 121–126, https://doi.org/10.4103/0253-7176.116232.
[4] S. Golder, G. Norman, Y.K. Loke, Systematic review on the prevalence, frequency and comparative value of adverse events data in social media, Br. J. Clin. Pharmacol. 80 (4) (2015) 878–888.
[5] H. Gurulingappa, R. Klinger, M. Hofmann-Apitius, J. Fluck, An empirical evaluation of resources for the identification of diseases and adverse effects in biomedical literature. Paper presented at the 2nd Workshop on Building and evaluating resources for biomedical text mining (7th edition of the Language Resources and Evaluation Conference), 2010.
[6] H. Gurulingappa, A. Mateen-Rajput, L. Toldo, Extraction of potential adverse drug events from medical case reports, J. Biomed. Semantics 3 (1) (2012) 15.
[7] H. Gurulingappa, A.M. Rajput, A. Roberts, J. Fluck, M. Hofmann-Apitius, L. Toldo, Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports, J. Biomed. Inform. 45 (5) (2012) 885–892.
[8] T. Hrenchir, 10 Most-Prescribed Antidepressant Medications. Retrieved from https://www.newsmax.com/health/health-wire/most-prescribed-antidepressant-medications/2015/09/02/id/673123/, 2017.
[9] T. Huynh, Y. He, A. Willis, S. Rüger, Adverse drug reaction classification with deep neural networks, 2016.
[10] T. Huynh, Y. He, A. Willis, S. Rüger, Adverse drug reaction classification with deep neural networks. Paper presented at the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016.
[11] S. Karimi, A. Metke-Jimenez, M. Kemp, C. Wang, CADEC: a corpus of adverse drug event annotations, J. Biomed. Inform. 55 (2015) 73–81.
[12] M. Kuhn, I. Letunic, L.J. Jensen, P. Bork, The SIDER database of drugs and side effects, Nucleic Acids Res. 44 (D1) (2015) D1075–D1079.
[13] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.
[14] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, G. Gonzalez, Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks, Paper Presented at the Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, (2010).
[15] J. Leskovec, A. Rajaraman, J.D. Ullman, Mining of Massive Datasets, Cambridge University Press, 2014.
[16] X. Liu, H. Chen, AZDrugMiner: an information extraction system for mining patient-reported adverse drug events in online patient forums, Smart Health, Springer, 2013, pp. 134–150.
[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, V. Dubourg, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[18] L.A. Pratt, D.J. Brody, Q. Gu, 2017. Antidepressant Use Among Persons Aged 12 and Over: United States, 2011–2014. NCHS data brief (283), 1–8.
[19] R.P. Rajkumar, G. Melvin, Pharmacovigilance for psychiatrists: an introduction, Indian J. Psychiatry 56 (2) (2014) 176.
[20] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, G. Gonzalez, Utilizing social media data for pharmacovigilance: a review, J. Biomed. Inform. 54 (2015) 202–212.
[21] A. Sarker, G. Gonzalez, Portable automatic text classification for adverse drug reaction detection via multi-corpus training, J. Biomed. Inform. 53 (2015) 196–207.
[22] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, J. Am. Med. Inform. Assoc. 17 (5) (2010) 507–513.
[23] H. Schouten, Measuring pairwise agreement among many observers, Biometrical J. 22 (6) (1980) 497–504.
[24] C.A. Smith, P.Z. Stavri, Consumer health vocabulary, Consumer Health Informatics, Springer, 2005, pp. 122–128.
[25] D. Tallon, N. Wiles, J. Campbell, C. Chew-Graham, C. Dickens, U. Macleod, S. Gilbody, Mirtazapine added to selective serotonin reuptake inhibitors for treatment-resistant depression in primary care (MIR trial): study protocol for a randomised controlled trial, Trials 17 (1) (2016) 66.
[26] R. Uher, A. Farmer, N. Henigsberg, M. Rietschel, O. Mors, W. Maier, A. Placentino, Adverse reactions to antidepressants, Br. J. Psychiatry 195 (3) (2009) 202–210.
[27] S. Winters-Hilt, A. Yelundur, C. McChesney, M. Landry, Support vector machine implementations for classification & clustering. Paper presented at the BMC bioinformatics, 2006.
[28] C.C. Yang, H. Yang, L. Jiang, M. Zhang, Social media mining for drug safety signal detection, Paper Presented at the Proceedings of the 2012 International Workshop on Smart Health and Wellbeing, (2012).
[29] M. Zolnoori, K.W. Fung, P. Fontelo, H. Kharrazi, A. Faiola, Y.S.S. Wu, V. Stoffel, T. Patrick, Identifying the Underlying Factors Associated With Patients' Attitudes Toward Antidepressants: Qualitative and Quantitative Analysis of Patient Drug Reviews, JMIR mental health 5 (4) (2018) p.e10726.
[30] M. Zolnoori, K.W. Fung, T.B. Patrick, P. Fontelo, H. Kharrazi, A. Faiola, et al., (submitted). The PsyTAR dataset: From Social Media Posts to a Corpus of Adverse Drug Events and Effectiveness of Psychiatric Medications. Data in Brief.
[31] M. Zolnoori, T. Patrick, K.W. Fung, P. Fontelo, A. Faiola, Y.S.S. Wu, K. Xu, J. Zhu, C. E. Eldredge, Development of an Adverse Drug Reaction Corpus from Consumer Health Posts for Psychiatric Medications, SMM4H@AMIA, Washington DC, 2017.
[32] M. Zolnoori, Utilizing Consumer Health Posts for Pharmacovigilance: Identifying Underlying Factors Associated with Patients' Attitudes Towards Antidepressants, Doctoral Dissertation, 2017. Retrieved from University of Wisconsin-Milwaukee Digital Library https://dc.uwm.edu/etd/1733/.