

A Natural Language Processing Framework for Detecting Adverse Drug Reactions in Clinical Structured Drug Reviews

Senobar Naderian¹, Roya Rahmani¹, Taha Samad-Soltani^{1*}

¹Department of Health Information Technology, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

Article History:

Received: December 5, 2023

Accepted: December 18, 2023

ePublished: June 13, 2024

*Corresponding Author:

Taha Samad-Soltani,

Email: samadsoltani@tbzmed.

ac.ir

Abstract

Background: Adverse drug reactions (ADRs) pose significant concerns in healthcare, yet their underreporting remains a challenge. Extracting spontaneous and non-automatic reports from free-text narratives contributes to this low rate of reporting. An automatic ADR detection system can mitigate these issues by identifying, summarizing, and reporting ADRs in a document. This study presents an adverse drug reaction detector (ADRD), a natural language processing (NLP) framework applied to the psychiatric treatment adverse reactions (PsyTAR) dataset. Aiming to automate ADR analysis, the framework explores the relationship between ADRs and patient satisfaction.

Methods: A comprehensive eight-phase approach was employed in the ADRD framework, utilizing Python programming language libraries and NLP tools. The dataset underwent meticulous preprocessing, and the subsequent phases involved data summarization, pattern identification, data cleaning, sentiment calculation, assessment of drug effectiveness and usefulness, analysis of medical conditions, and identification of the most effective and ineffective drugs for each condition.

Results: Analyzing 891 comments related to four unique drugs (i.e., Zoloft, Lexapro, Cymbalta, and Effexor XR) from patients with 285 distinct conditions, the framework offered insights into the dataset structure, statistical indicators, distribution of ratings and ADR counts, the impact of ratings on ADR counts, and length of comments' influence on ratings.

Conclusion: The challenges of extracting ADR reports from free-text narratives have led to their underreporting. ADRD offers an automated and insightful approach for enhancing ADR analysis and reporting processes, making strides toward bridging the gap in ADR reporting.

Keywords: Adverse drug reaction, Informatics, Natural language processing, Drug, Medication review, Clinical informatics

Please cite this article as follows: Naderian S, Rahmani R, Samad-Soltani T. A natural language processing framework for detecting adverse drug reactions in clinical structured drug reviews. *Int J Drug Res Clin.* 2024; 2: e9. doi: 10.34172/ijdr.2024.e9

Introduction

The challenge of improving medication quality is due to the unstructured nature of patient data in electronic health record systems (EHRs).¹ Although some patient information is organized, critical details such as drug reviews are often in free-text narratives.^{2,3} These narratives by healthcare providers offer a comprehensive overview of patients' medical circumstances, enhancing medication treatment quality, and patient outcomes.⁴

Adverse drug reactions (ADRs), defined by the World Health Organization (WHO) as noxious and unintended responses at normal doses, significantly impact patients and healthcare systems, leading to hospital admissions.⁵ Effective prevention is crucial when nearly 50% of ADRs are possibly avoidable.⁶

Despite rigorous preclinical research and clinical trials for novel drugs,⁷ limitations persist, necessitating

pharmacovigilance to identify and predict ADRs in post-marketing studies.⁸ The reporting of ADRs falls below WHO recommendations which is attributed to challenges in extracting non-automatic reports from free-text narratives.^{9,10} The lack of consistent arrangement poses difficulties in comparing medication reviews, affecting patient care.^{11,12} Implementing pharmacovigilance actions becomes challenging due to the absence of standardized information, hindering a comprehensive understanding of a patient's medication routine and ADR discovery.^{13,14}

Manual patient record mining is hindered by a lack of expert personnel, resulting in a high rate of human error.¹⁵ Hence, automated ADR detection systems are needed to overcome these challenges, enabling identification, summarization, and automatic reporting of ADRs in documents.¹⁶ WHO's efforts to address ADRs globally are hampered by the lack of consistency in describing and



evaluating avoidability.¹⁷

Natural language processing (NLP) emerges as a solution capable of extracting concepts from free-text reports.¹⁸ NLP, more robust than keyword searching, allows for contextual extraction.¹⁹ The increasing availability of electronic medical records provides opportunities for mining algorithms to capture ADR information more comprehensively.¹⁷ Recent attention to applying NLP methods in ADR detection across various sources signifies a growing trend.

To conduct this study, a thorough literature review was carried out, and databases such as PubMed and relevant medical journals were searched. The literature review focused on studies related to ADR detection, NLP frameworks in healthcare, and automated analysis of drug reviews. The time period ranged from the inception of relevant databases up to December 2023. This review informed the utilized methodology, ensuring a comprehensive understanding of existing research and methodologies.

Several studies demonstrated NLP efficacy in automatic ADR extraction. Aramaki et al evaluated the automatic extraction accuracy of the standard NLP system on 3012 discharge summaries, achieving high performance.¹⁸ Shang et al employed literature-based discovery (LBD) techniques and NLP tools to identify adverse reactions, showing better accuracy compared to a random baseline.¹⁹ Sarker et al enhanced automatic ADR detection using various NLP techniques and machine learning algorithms, achieving higher classification accuracies.²⁰ Tang et al focused on automating ADR monitoring in pediatric patients, demonstrating promising performance after manual validation.²¹

Tang et al developed the Racial Equity and Policy (REAP) framework, a rule-based NLP system addressing drug adverse events in hospital discharge summaries, achieving 75% precision and 59% recall.²² Nikfarjam et al used a neural network-based named entity recognition (NER) system, DeepHealthMiner, to detect ADR references from social health media posts, achieving high precision.²³ Kim

et al implemented an NLP system to automatically detect medication and ADR information from EHRs, producing superior results.²⁴ Chaichulee et al also concluded a study to develop NLP algorithms for encoding unstructured ADRs in EHRs into institutional symptom terms. Using various NLP techniques, including Bidirectional Encoder Representations from Transformers (BERT) models, their research achieved high performance, demonstrating the potential for automated symptom term suggestion systems.²⁵

To explore NLP methods for ADR detection, a comprehensive overview is provided in Table 1. This table summarizes various studies, detailing the data description and relevance of methods employed in ADR detection.

These studies highlighted the effectiveness of diverse NLP methods in extracting ADR information. Despite sophisticated approaches, the study emphasizes the role of simpler and user-friendly NLP techniques in enhancing patient medication reviews and pharmacovigilance efforts. The study introduces an NLP framework, namely ADRD, applied to the psychiatric treatment adverse reactions (PsyTAR) dataset.²⁶ The preprocessing codes and framework are provided in our GitHub repository (https://github.com/senonaderian/ADRD-an-ADR-Detection_NLP-framework.git).

Methods

Data-Driven Approach and Natural Language Processing Implementation

The current study adopted a data-driven approach, leveraging a pre-collected dataset, that is, the PsyTAR dataset. Unlike traditional experimental designs, this study centered around the development and application of an NLP framework, ADRD, to analyze patient comments on psychiatric treatments. As such, the study did not involve specific experimental conditions in the conventional sense. Instead, it focused on data preprocessing, NLP implementation, and subsequent data analysis to extract valuable insights from the PsyTAR dataset. The following sections detail our methodology, emphasizing the steps

Table 1. Studies on NLP Methods for ADR Detection: Data Description and Method Relevance

Study	Brief Description of Reported Data
Aramaki et al ¹⁸	Evaluated automatic extraction accuracy on 3,012 discharge summaries and demonstrated high performance in automatic ADR extraction.
Shang et al ¹⁹	Employed LBD techniques and NLP tools for identifying adverse reactions and showed better accuracy compared to a random baseline.
Sarker et al ²⁰	Enhanced automatic ADR detection using various NLP techniques and machine learning algorithms and achieved higher classification accuracies.
Tang et al ²¹	Focused on automating ADR monitoring in pediatric patients and demonstrated promising performance after manual validation.
Tang et al ²²	Developed the REAP framework, a rule-based NLP system addressing drug adverse events, and achieved 75% precision and 59% recall.
Nikfarjam et al ²³	Used a neural network-based NER system, DeepHealthMiner, to detect ADR references from social health media posts and achieved high precision.
Kim et al ²⁴	Implemented an NLP system to automatically detect medication and ADR information from EHRs, and produced superior results.
Chaichulee et al ²⁵	Developed NLP algorithms for encoding unstructured ADRs into institutional symptom terms, used various NLP techniques, including BERT models, achieved high performance, and demonstrated potential for automated symptom term suggestion systems.

Note. ADR: Adverse drug reaction; NLP: Natural language processing; LBD: Literature-based discovery; EHR: Electronic health record; REAP: Racial equity and policy; NER: Named entity recognition; BERT: Bidirectional encoder representations from transformers.

taken to preprocess the dataset and implement the NLP framework.

Dataset

The selection of the four drugs (i.e., Zoloft, Lexapro, Cymbalta, and Effexor XR) was based on the availability of a pre-collected dataset, namely, the PsyTAR dataset, compiled by Zolnoori et al. These medications are commonly prescribed for psychiatric conditions. Zolnoori et al compiled this excel-formatted dataset in four steps. In the initial stage, an API was used to collect an example of 891 drug comments for four psychiatric drugs: Zoloft, Lexapro, Cymbalta, and Effexor XR.²⁶

Each comment included information about the patient's personal data, length of treatment, and satisfaction with the drugs. The reviews were then divided into 6009 separate sentences and labeled for the frequency of ADR, withdrawal symptoms (WDs), signs/symptoms/illness (SSIs), drug indications (DIs), drug effectiveness (EF), drug ineffectiveness (INF), and others (not applicable).

Objects with each indicator were identified and mapped from the charted text to the corresponding UMLS Meta thesaurus and SNOMED CT concepts in the final stages. The study's inclusion criteria involved the selection of some rows from two PsyTAR sheets (sample and ADR_ Identified), including index, drug ID, condition, rating, comment, date, and ADRs.

Preprocess

Considering that the ADR representative columns were scattered and written in several rows in the original dataset, it was necessary to specify the number of ADRs related to each patient and their comments in a separate row. Consequently, with the help of two major Python libraries (Pandas and NumPy), the ADRs of each patient with a specific drug ID were placed in a single row; as a result, the number of scattered rows was reduced from 2166 to 819 single rows for each patient. Then, the number of unique ADRs such as weight gain and WEIGHT GAIN was counted and placed in a new column called ADR_count. Table 2 details the conclusive presentation of the dataset, illustrating its final appearance. The final appearance of the dataset was as follows:

Main Process

Following data preprocessing, an NLP framework (i.e., ADRD) was developed to analyze the data in eight phases using various Python libraries.

1. *Reading the dataset:* After converting the dataset's format, the framework was employed to read the

data, inspect the row headers, and explore several significant columns using the Pandas and NumPy libraries.

2. *Summarizing the dataset and categorical data:* To extract significant information, the framework summarized the dataset using statistical analysis on the "rating" and "ADR count" columns. The algorithm examined the number and name of drugs with a -0- ADR count, the number of drugs with no ADR count and a rating greater than or equal to 4 (according to the purposes of the primary dataset, PsyTAR), and the average rating of drugs with no ADR count. Subsequently, we summarized the categorical dataset, removed all records where the condition was missing, and used the Pandas and NumPy libraries to verify all missing values.
3. *Unveiling hidden patterns from the data:* The rating and ADR distribution were examined using the Matplotlib library. This segment of the algorithm displayed the figure size of the primary dataset and generated two subplots displaying the distribution of each rating and the ADR count, respectively. Then, utilizing a subplot, the effect of ratings on ADR counts was analyzed. Next, the framework examined whether the length of a review influences the ratings of the drugs. To calculate the length of the reviews, a new column had to be created; then, using the Matplotlib, Seaborn, NumPy, and Pandas libraries, the longest review was identified.
4. *Cleaning the reviews:* As the comments contained numerous extraneous elements such as stop words, punctuation, numbers, and other elements, it was necessary to remove them using RE and Natural Language Toolkit (NLTK) library tools such as stopwords and word_tokenize.
5. *Calculating the sentiment from reviews:* The NLTK library's Vader Lexicon was utilized for the Sentiment Analyzer. Vader is an NLP open-sourced package within the NLTK that mixes a sentiment lexicon method as well as syntactic instructions and agreements for stating sentiment divergence and strength. This study calculated the sentiment from comments, examined the effect of sentiment on comments, and removed the unique ID, date, comment, length, and sentiment columns.
6. *Calculating the effectiveness and usefulness of drugs:* First, the values of the rating column within the interval [0,1] were normalized by the min-max normalizer (equation 1), and a new column titled "eff_score" was created. The usefulness score was

Table 2. Final Appearance of the Dataset

Index	Drug Name	Drug ID	Condition	Rating	Comment	Date	ADR Count	Dosage Duration
1	Lexapro	lexapro.1	Depression and anxiety	1	I am detoxing from Lexapro now...	2/21/2011	3	5 years 20 mg 1X D
-	-	-	-	-	-	-	-	-

Note. ADR: Adverse drug reaction.

then computed by multiplying the “rating”, “ADR_count,” and “eff_score” columns. The framework investigated which drugs were beneficial to the greatest number of individuals. This phase utilized the following libraries: interact module from the Ipywidgets library, Seaborn, Matplotlib, Pandas, and NumPy.

$$z = \frac{(x - \min(x))}{(\max(x) - \min(x))} \quad \text{Eq. (1)}$$

7. *Analyzing medical conditions:* The most common conditions and drugs were checked using Pandas and NumPy libraries.
8. *Finding the most effective and ineffective drugs for each condition:* Finally, all duplicates from the dataset were removed, and the interact module from the Ipywidgets, Pandas, and NumPy libraries were used to determine the highest and lowest-rated drugs for each condition.
9. As an additional step, distinct information about each drug set was provided to the framework to achieve greater accuracy and compare the results ([Supplementary file 1](#)).

It should be noted that this framework mainly focuses on data analysis and generating insights rather than implementing classic machine learning or an advanced NLP model for a prediction task. As such, traditional code functioning metrics such as accuracy, F1 score, precision, and recall may not be directly applicable in this case. However, to make sure that the framework is useable for other users, we considered some different methods and techniques as follows:

- **Memory usage:** We added the memory consumption track of the code, especially for the times that it will be used to process large datasets or to perform memory-intensive operations. This will help the users to ensure that the framework is efficient and does not consume large memories.
- **Error handling:** We added try-except blocks to handle errors and exceptions.
- **Modularity and reusability:** We made the framework properly encapsulated into logical modules or classes. In this way, the framework will allow easier use in future projects or codebases.

This version of the framework that includes the mentioned metrics is also available on the repository.

Results

This study analyzed 891 comments gathered for four unique drugs from patients with 285 unique conditions. Each phase of the framework produced a unique result and made the dataset available for analysis. Although an NLP framework was applied to this dataset, the obtained results still required physician approval and further clinical observations.

In the initial phase, the ADRD framework printed dataset columns and primary analyses, including the number of patient comments, the number of unique drugs, the number of unique medical conditions, and the time of data collection, as shown in [Console 1](#).

In the second phase, the framework summarized the dataset by “rating” and “ADR counts” columns and displayed statistical indicators for each column, including their count, mean, standard deviation, and minimum and maximum values.

By summarizing the data based on the “rating” and “ADR count” columns, the maximum number of reported ADRs was determined to be 41, which belonged to the effexor drug set, and the minimum number was 0, belonging to all four unique drug sets ([Table 3](#)). The results for each unique drug set analysis: Lexapro 29, Zoloft 21, Cymbalta 29, Effexor 41 ([Supplementary file 1](#)).

On this console, the framework analyzed ADR and displayed drugs with no ADR, and drugs with the highest ADR counts were illustrated in [Console 2](#).

The following console depicts the analyses of each of the four unique drug sets.

Overall, no ADRs were reported by 85 patients in their comments. Using statistical analysis in the next step, the ratio of patients who did not experience any ADR to the total number of patients related to each drug set was greater for the Lexapro drug set than for the others, with a value of 10.95%. In addition, other ratios for Zoloft, Cymbalta, and Effexor drug sets were 9.90%, 9.09%, and 8.33%, respectively ([Consoles 2 and 3](#)).

According to the PsyTAR dataset definition, the rating column was composed of patient satisfaction ratings ranging from 0 to 5 for the lowest to highest levels. Furthermore, the total number of drugs with high satisfaction (≥ 4) and zero ADR counts in the dataset was 53. Using statistical analysis, the ratio of patients satisfied with each drug set relative to the total number of patients associated with each set was determined. Similar to the results of the previous stage, this ratio was greater for

```
Number of patient comments present in the Dataset: 891
Number of unique drugs in the Dataset: 4
Number of Unique Medical Conditions present in the Dataset: 285

The Time Period of Collecting the Data
Starting Date: date      9-Sep-13
dtype: object
Ending Date: date      1-Apr-04
```

Console 1. The Primary Definition of a Dataset

Table 3. A summary of Dataset

	Rating	ADR Count
Count	891.000000	890.000000
Mean	3.159371	5.283146
Standard deviation	1.488295	4.612870
Minimum	1.000000	0.000000
25%	2.000000	2.000000
50%	3.000000	4.000000
75%	4.000000	7.000000
Maximum	5.000000	41.000000

Note. ADR: Adverse drug reaction.

```
The Number of Drugs with No ADR Count: 85
Number of patients who were satisfied using each drug(rating>=4) and experienced no ADR: 53
Average Rating of Drugs with No ADR Count: 3.74

Analysis on Drugs with most adverse reaction
-----
Maximum ADR count extracted from comment column: 41
Average Rating of Drugs with 20+ ADR Counts: 2.8
```

Console 2. ADR Analysis. Note. ADR: Adverse drug reaction

```
Lexapros with No ADR
-----
The Number of lexapros with No ADR Count: 24
Number of patients who were satisfied using
lexapro (rating>=4) and experienced no ADR: 18
Average Rating of lexapro with No ADR Count: 4.33
Analysis on lexapros with most ADR
-----
Maximum ADR count extracted from comment column:
29
Average Rating of lexapros with 20+ ADR Counts:
3.6

Zolofts with No ADR
-----
The Number of zolofts with No ADR Count: 21
Number of patients who were satisfied using
zoloft (rating>=4) and experienced no ADR: 14
Average Rating of zolofts with No ADR Count: 3.90
Analysis on zolofts with most ADR
-----
Maximum ADR count extracted from comment column:
21
Average Rating of zolofts with 20+ ADR Counts:
2.0
```

```
Cymbaltas with No ADR
-----
The Number of cymbaltas with No ADR Count: 21
Number of patients who were satisfied using
cymbalta (rating>=4) and experienced no ADR: 11
Average Rating of cymbaltas with No ADR Count:
3.24
Analysis on cymbaltas with most ADR
-----
Maximum ADR count extracted from comment column:
29
Average Rating of Drugs with 20+ ADR Counts: 1.6

Effexors with No ADR
-----
The Number of Effexors with No ADR Count: 19
Number of patients who were satisfied using
effexor (rating>=4) and experienced no ADR: 10
Average Rating of effexors with No ADR Count:
3.37
Analysis on effexors with most ADR
-----
Maximum ADR count extracted from comment column:
41
Average Rating of effexor with 20+ ADR Counts:
4.0
```

Console 3. Unique Drug ADR Analysis. Note. ADR: Adverse drug reaction

Lexapro with a value of 8.21% compared to 6.6%, 4.76%, and 4.38 % for the Zoloft, Cymbalta, and Effexor drug sets, respectively (Consoles 2 and 3).

Table 4 presents all patients who experienced an ADR count of ≥ 20 , along with the administered drug regimen, the medical condition, and the dosage duration. As can be seen, no significant correlation was found between the columns of this table.

This phase also provided a summary of categorical data, which included the total count of each drug name, condition, comment, and dosage duration, as well as their count, unique condition, frequency, and leading values (Table 5).

According to Table 5, depression was the most prevalent medical condition, with a frequency of 276. Since the patient comments in the dataset were collected narratively and not in separate columns, it was expected that the total

number of comments (N=768) would be close to their unique number, 766, as illustrated in Table 5.

The third phase used graphs to depict the distribution of ratings and ADR count (Figure 1) and the impact of ratings on ADR counting (Figure 2).

Figure 1 depicts the distribution of rating and ADR count. Generally, the number of patients who were the most satisfied with the drug (rating = 4, 5) was greater than 175. This scale also applied to the Lexapro and Zoloft drug sets; however, in the Cymbalta and Effexor drug sets, the number of patients dissatisfied with the drug (rating = 1) was the highest. Moreover, the number of patients who reported between 0 and 10 independent ADRs for specific

drug sets was greater than the number of patients who reported more than 10 ADRs (Figure 1).

Figure 2 demonstrates the effect of patient satisfaction (rating) on the number of ADRs.

As seen in Figure 2, for individual drug sets, the number of ADRs reported was greater for drugs with a satisfaction rating of 3 or less, compared to those with higher ratings. This suggests that lower patient satisfaction is associated with a higher frequency of adverse drug reactions.

To determine whether the length of comments influences the ratings of the drugs or not, a new column titled "Len" was created to calculate the length of the comments. Statistical measures are used in Table 6 to demonstrate the effect of the length of comments on the rating (minimum, mean, and maximum). Table 6 illustrates the effect of comment length on patient satisfaction. As the findings suggest, there is no correlation between these two columns. Moreover, since ratings 5 and 1 exhibit negative sentiments, it can be concluded that there are no significant relationships between ratings and comments, as shown in Table 6.

Then, ADRD examined the longest comment, allowing the framework to identify the longest comment. Console 4 contains the longest comment in the PsyTAR dataset, with 1533 characters. As there was no correlation between comment length and rating, the "len" column was eliminated from the dataset.

In the fourth phase, the framework deleted stop words, punctuation, and numbers, among other elements, using several NLP methods.

In the fifth phase, to calculate sentiment from the dataset, the framework created a new column titled 'sentiment' which received a score within the interval [-1, 1] based on the mean of the NLTK SentimentIntensityAnalyzer tool. Moreover, the influence of each comment's sentiment was determined by its rating and sentiment score, as illustrated in Table 7.

In the sixth phase, after calculating an effective rating,

Table 4. Name, Condition, and Dosage Duration of Drugs with 20+ADR Counts

Drug ID	Condition	Dosage Duration
0 Lexapro.43	Mild depression/grief from loss	3 days
1 Lexapro.140	Depression	1.5 years
2 Lexapro.179	Depression	9 months
3 Zoloft.9	Depression/anxiety/ PTSD	2 years200MG 1X D
4 Zoloft.93	PTSD/anxiety/OCD/ depression	2 months100mg
5 Cymbalta.5	Depression/anxiety	1 day
6 Cymbalta.41	Muscular pain/ depression/ anxiety	1 day 60
7 Cymbalta.112	Depression	3 months
8 EffexorXR.110	Chronic depression	4 months
9 EffexorXR.200	Manic depression/ bipolar/ anxiety	8 years300 MG 1X D

Note. PTSD: Posttraumatic stress disorder; OCD: Obsessive-compulsive disorder.

Table 5. A Summary of Categorical Dataset

	Drug Name	Condition	Comment	Dosage Duration
Count	891	891	768	888
Unique	4	285	766	291
Top	Cymbalta	Depression	Bad drug	6 months
Frequency	231	276	2	39

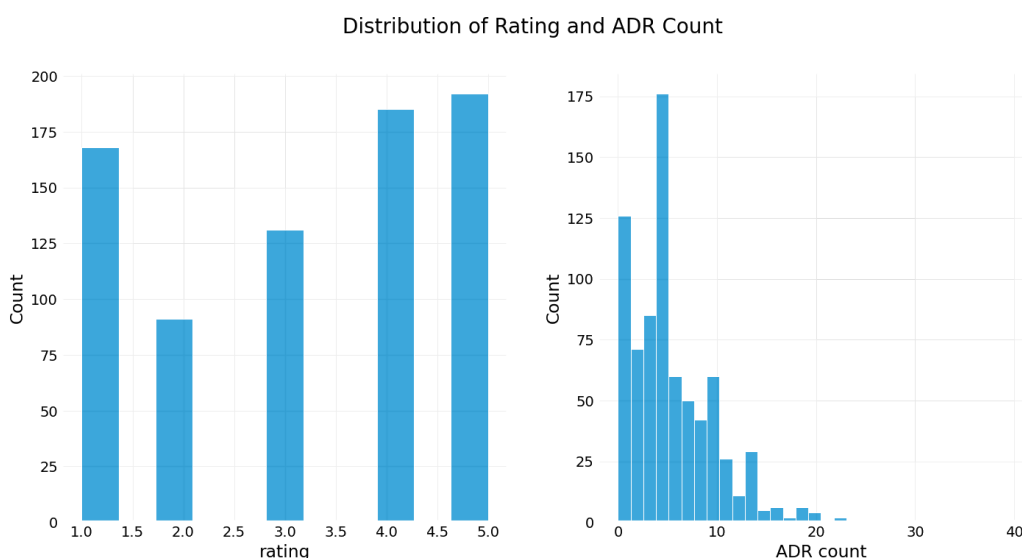


Figure 1. Distribution of Rating and ADR Count. Note. ADR: Adverse drug reaction

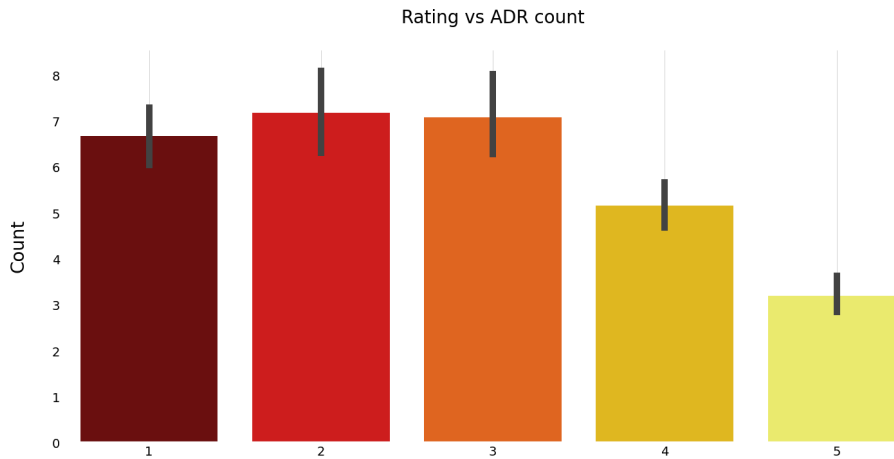


Figure 2. Impact of Ratings on ADR Counting. Note. ADR: Adverse drug reaction

Length of Longest comment: 1533

Longest comment:

I was never one to take any kind of drugs. I would only take aspirin as a last resort. But I have to say that taking the Effexor has made me feel so much better. I read all the side effects and was reluctant to start it, but knew that I needed to take something in order to function in life. I am currently on 75 mgs per day and it has made a huge difference. My advice to anyone considering taking the drug is don't stress out when reading the side effects. All drugs effect people differently. Give it a chance. It took about 14 days for the drug to kick in. The side effects did go away after about 4 days and when I increased my dose from 37.5 mgs to 75 mgs I didn't have any side effects. My Dr. prescribed xanax for me to take along with the effexor for the first week which actually eliminated the shakiness and nervousness immediately. I only had to take the xanax for the first three days on the Effexor. I suggest that if you decide to take the Effexor, having the xanax on hand for the first few days will help. I have not had to increase my dosage as of yet. I'm hoping to be able to stay on the 75 mgs for as long as possible. Reading about the withdrawals is a little scary, but like I said, all people are affected differently by different drugs, so I will cross that bridge if and when I get to it. Effexor has made a huge difference in my life and when and if I have to come off it, if I experience any withdrawals, well, so be it. It would be a small price to pay for being able to enjoy life.

Console 4. The Longest Comment

Table 6. Impact of Length of Comments on Ratings

Rating	Len		
	Minimum	Mean	Maximum
1	9	382.398810	1508
2	6	395.670330	1501
3	4	384.595420	1292
4	16	443.108108	1500
5	3	397.067708	1533

Table 7. Impact of Sentiment on Comments

Rating	Sentiment		
	Minimum	Mean	Maximum
1	-0.9652	-0.085024	0.9938
2	-0.9825	-0.136681	0.9846
3	-0.9766	0.010662	0.9844
4	-0.9929	0.069332	0.9859
5	-0.9893	-0.150158	0.9559

the framework computed the usefulness score and examined the top 10 most useful drugs to the greatest number of individuals possible, along with their respective conditions.

We aimed to identify the top 10 drugs for various conditions based on their usefulness score (Table 8). Patients taking Lexapro for depression and anxiety in some cases found it more effective. Furthermore, the dosage duration ranged from 3 months to 5 years, with no significant correlation with the drug's efficacy.

As a result of the seventh phase of the algorithm's computation, the most prevalent conditions were recorded (Table 9). After calculating the most common conditions (Tables 5 and 9), 24% of all patients were diagnosed with depression. In all drug sets besides Lexapro, depression was the leading identified condition, and the leading condition treated by Lexapro was "depression and anxiety". The specific details regarding the number of

unique conditions are documented in Console 5.

The final phase identified the highest and lowest-rated drugs for each condition. Table 10 presents the top 5 useful results, highlighting the most significant achieved outcomes. Additionally, for a comprehensive analysis, we refer to Table 11, which outlines the bottom 5 useful results, shedding light on areas where improvement may be needed.

Discussion

The present study leveraged NLP techniques to extract valuable insights from patient comments on psychiatric treatments, aligning with previous research that demonstrated the effectiveness of NLP systems in identifying ADRs.¹⁸ The ADRD framework facilitated the analysis of patient narratives, shedding light on ADR counts, ratings, sentiment, and drug effectiveness. In a study by Chaichulee et al,²⁵ NLP algorithms, particularly

Table 8. Top 10 Useful Drugs on Several Conditions

	Drug ID	Condition	Usefulness	Dosage Duration
0	Lexapro.140	Depression	116.0	1.5 years
1	EffexorXR.200	Manic depression/ bipolar/ anxiety	115.0	8 years300 MG 1X D
2	Lexapro.179	Depression	110.0	9 months
3	Lexapro.115	Depression	80.0	11 months
4	Lexapro.121	Depression/ anxiety	64.0	10 years20 MG 1X D
5	Zoloft.143	Anxiety/ possible depression	64.0	5.5 months
6	EffexorXR.166	Depression	60.0	2 years
7	Lexapro.201	Depression/ anxiety	55.0	10 weeks
8	Cymbalta.221	Major depression/ anxiety/ PMDD	55.0	60 days30 mg 1X D
9	Lexapro.114	Depression/ anxiety	52.0	3 months

Note. PMDD: People with premenstrual dysphoric disorder.

Table 9. Top 10 Frequent Conditions Count

Condition	Count
Depression	218
Depression/anxiety	193
Major depression	13
Mild depression	13
Severe depression	11
Fibromyalgia/depression	8
Depression/OCD	5
Mild depression/ anxiety	5
Bipolar/depression	5
Major depression/anxiety	4

Note. OCD: Obsessive-compulsive disorder.

BERT models, were explored for encoding formless ADRs in EHRs. Although our study did not employ advanced NLP models for classification, the outcomes resonate with the potential of NLP algorithms. Moreover, our approach showcased the feasibility of automating the extraction of drug reaction narratives from patient comments, a process that can streamline accessibility and mitigate the need for human coding. Tang et al²¹ delved into utilizing NLP methods to detect ADRs from EHR entries, achieving favorable results for pediatric cases. Similarly, our study harnessed NLP methods to precisely detect ADR-related drug reactions from patient comments. The outcomes reinforce the notion that NLP methods can be instrumental in EHR-based ADR identification. Comparing our approach to studies such as Shang et al,¹⁹ which utilized LBD techniques, our study focused on simpler NLP methods for ADR extraction from drug reviews. The emphasis on practical and interpretable NLP methods contributes to enhancing patient medication reviews and pharmacovigilance efforts. While NLP methods show promise in ADE recognition, as seen in Bayer and colleagues' assessment of ADEs in FDA-approved drug labels,²⁷ our study echoes the importance

Table 10. Top 5 Useful Results

	Drug ID	Usefulness	Dosage Duration
1	Lexapro.140	116.0	1.5 years
2	Lexapro.179	110.0	9 months
3	Lexapro.115	80.0	11 months
4	EffexorXR.166	60.0	2 years
5	Zoloft.135	52.0	3 years 50-150 MG 1X D

Table 11. Bottom 5 Useful Results

	Drug ID	Usefulness	Dosage Duration
1	Lexapro.2	0.0	2 days 10mg 1X D
2	Cymbalta.86	0.0	5 months
3	Cymbalta.95	0.0	5 months
4	Cymbalta.102	0.0	3 months 100 MG 1X D
5	Zoloft.135	0.0	2 weeks

Number of Unique Conditions : 261

Console 5. The Number of Unique Conditions

of human evaluation. Our algorithm demonstrated promising performance; however, the involvement of clinicians remains essential for precise assessment and evaluation. Sarker et al explored advanced NLP methods for ADR detection from diverse sources, including social media data.²⁰ Although the current study did not employ such advanced techniques or use social media data, its ability to detect ADRs from patient reviews showcases the potential of simpler NLP methods in this context.

Dataset summarization and organization, as highlighted by Tang et al,²² underscore the importance of effectively managing data. Consistent with this, our study prioritized meticulous dataset organization, impacting the performance and evaluation of our algorithm. Considering the findings of specific drugs in our dataset, our categorical summaries reinforced the prevalent conditions associated with drugs such as Lexapro and Zoloftin line with Aldrich et al and Chermá and colleagues' studies.^{28,29} In broader contexts, Hughes and colleagues'

study on antidepressant users agrees with the results of our study, the framework of which identified specific conditions related to drug use and variations in treatment satisfaction levels among different antidepressants.³⁰ The integration of our findings with existing literature enhances our understanding of patient experiences and satisfaction with specific drugs. In conclusion, these studies collectively emphasize the potential of NLP techniques for ADR identification and monitoring across various data sources. Aligning with these findings, the current study specifically contributed to addressing drug-related conditions and treatment satisfaction. By relating our results to the existing literature, we showcased the potential of NLP and artificial intelligence in advancing pharmacovigilance practices and enhancing patient safety.

Although this study focused on psychiatric treatments and the ADRD framework, the role of artificial intelligence tools in biomedical data processing, particularly text processing tools, is acknowledged. The study by Bressler et al leveraged NLP on emergency medical service's (EMS's) EHRs to identify variables associated with child maltreatment, showcasing the potential of NLP in extracting valuable information from healthcare datasets. This demonstrates the broader applicability of NLP in enhancing child welfare and suggests future directions for developing screening tools in EMS records.³¹ Additionally, the work by Lewinski and McInnes addressed the exponential growth of nanotechnology literature, emphasizing the need for NLP to catalog engineered nanomaterials. The review identified nine NLP-based tools, underlining the importance of sharing such tools through online repositories to advance engagement in Nano informatics³². These studies collectively reinforce the versatility and significance of NLP in various biomedical applications and highlight the potential for future advancements in the field.

Moreover, our discussions not only broaden our understanding of the potential clinical applications of NLP in psychiatric treatments and biomedical data processing but also contribute to broader conversations surrounding patient care improvement, exploring therapeutic alternatives, and promoting sustainable biomedical practices.^{32,33} The collective insights from these studies emphasize the multifaceted impact of NLP on enhancing healthcare practices and fostering advancements in biomedical research.

Conclusion

The challenges related to natural and non-automated report extraction from free-text narratives have caused a low rate of ADR reportage. Nevertheless, the framework presented in this study shows the ability to automate the investigation and report the progression of ADRs. Using NLP methods, the framework empowers the mining of valued visions from patient comments, including ADR counts, ratings, sentiment analysis, and drug

effectiveness. Since the outcomes obtained from the framework deliver valued information, it is necessary to highlight that physician approval is still important to verify precise assessment and clarification of the findings. This framework demonstrates the ability of simple NLP algorithms to improve pharmacovigilance practices and enhance patient safety through automatic ADR detection and analysis.

Limitations

- **Selection bias:** The reliance on a pre-collected dataset introduces potential selection bias as it reflects the experiences of individuals who voluntarily shared their views, possibly not representing the broader user population.
- **Data granularity challenges:** The granularity of information in the dataset, especially in categorizing ADRs, may be limited due to variations in reporting styles among healthcare providers and patients.
- **Insights-driven approach:** Our analysis focuses on data analysis for generating insights rather than employing advanced modeling techniques. This approach may limit the depth of predictive capabilities compared to more sophisticated models.
- **Drug selection bias:** The study's selection of Zoloft, Lexapro, Cymbalta, and Effexor XR drugs was based on dataset availability which can potentially restrict the generalizability of findings to other psychiatric medications.

Ethics statement

The research utilized a pre-collected dataset, namely, the psychiatric treatment adverse reactions (PsyTAR) dataset, compiled by Zolnoori et al. This dataset is publicly available, and the study did not involve direct support or intervention from the research committee in terms of data collection.

Disclosure of funding source

This study was supported by the Student Research Committee of Tabriz University of Medical Sciences.

Conflict of interests declaration

There is no conflict of interests.

Data availability statement

The PsyTar repository, <https://github.com/basaldella/psytarpreprocessor>, includes the analyzed datasets for the current study.

Author contributions

Conceptualization: Taha Samad-Soltani, Senobar Naderian.

Data curation: Senobar Naderian.

Formal analysis: Senobar Naderian.

Investigation: Taha Samad-Soltani, Senobar Naderian.

Methodology: Taha Samad-Soltani, Senobar Naderian.

Project administration: Taha Samad-Soltani.

Software: Senobar Naderian, Roya Rahmani.

Supervision: Taha Samad-Soltani.

Visualization: Senobar Naderian.

Writing—original draft: Senobar Naderian, Roya Rahmani.

Writing—review & editing: Taha Samad-Soltani.

Consent for publication

Not applicable.

Supplementary files

Supplementary file 1. Distinct information about each drug set

References

- Létinier L, Jouganous J, Benkebil M, Bel-Létoile A, Goehrs C, Singier A, et al. Artificial intelligence for unstructured healthcare data: application to coding of patient reporting of adverse drug reactions. *Clin Pharmacol Ther.* 2021;110(2):392-400. doi: [10.1002/cpt.2266](https://doi.org/10.1002/cpt.2266).
- Maletzky A, Böck C, Tschoellitsch T, Roland T, Ludwig H, Thumfart S, et al. Lifting hospital electronic health record data treasures: challenges and opportunities. *JMIR Med Inform.* 2022;10(10):e38557. doi: [10.2196/38557](https://doi.org/10.2196/38557).
- Al Ani M, Garas G, Hollingshead J, Cheetham D, Athanasiou T, Patel V. Which electronic health record system should we use? A systematic review. *Med Princ Pract.* 2022;31(4):342-51. doi: [10.1159/000525135](https://doi.org/10.1159/000525135).
- Hohl CM, Badke K, Zhao A, Wickham ME, Woo SA, Sivilotti MLA, et al. Prospective validation of clinical criteria to identify emergency department patients at high risk for adverse drug events. *Acad Emerg Med.* 2018;25(9):1015-26. doi: [10.1111/acem.13407](https://doi.org/10.1111/acem.13407).
- World Health Organization (WHO). Safety of Medicines: A Guide to Detecting and Reporting Adverse Drug Reactions: Why Health Professionals Need to Take Action. WHO; 2002.
- Woo SA, Cragg A, Wickham ME, Peddie D, Balka E, Scheuermeyer F, et al. Methods for evaluating adverse drug event preventability in emergency department patients. *BMC Med Res Methodol.* 2018;18(1):160. doi: [10.1186/s12874-018-0617-4](https://doi.org/10.1186/s12874-018-0617-4).
- Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc.* 2014;21(2):353-62. doi: [10.1136/amiajnl-2013-001612](https://doi.org/10.1136/amiajnl-2013-001612).
- Mashima Y, Tamura T, Kunikata J, Tada S, Yamada A, Tanigawa M, et al. Using natural language processing techniques to detect adverse events from progress notes due to chemotherapy. *Cancer Inform.* 2022;21:11769351221085064. doi: [10.1177/11769351221085064](https://doi.org/10.1177/11769351221085064).
- Fossouo Tagne J, Yakob RA, McDonald R, Wickramasinghe N. Barriers and facilitators influencing real-time and digital-based reporting of adverse drug reactions by community pharmacists: qualitative study using the task-technology fit framework. *Interact J Med Res.* 2022;11(2):e40597. doi: [10.2196/40597](https://doi.org/10.2196/40597).
- Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: mining adverse drug reactions. *J Biomed Inform.* 2015;56:318-32. doi: [10.1016/j.jbi.2015.06.016](https://doi.org/10.1016/j.jbi.2015.06.016).
- Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to MedDRA: automatically encoding adverse drug reactions. *J Biomed Inform.* 2018;84:184-99. doi: [10.1016/j.jbi.2018.07.001](https://doi.org/10.1016/j.jbi.2018.07.001).
- Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc.* 2015;22(3):671-81. doi: [10.1093/jamia/ocu041](https://doi.org/10.1093/jamia/ocu041).
- Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc.* 2011;18(5):544-51. doi: [10.1136/amiajnl-2011-000464](https://doi.org/10.1136/amiajnl-2011-000464).
- Skentzos S, Shubina M, Plutzky J, Turchin A. Structured vs. unstructured: factors affecting adverse drug reaction documentation in an EMR repository. *AMIA Annu Symp Proc.* 2011;2011:1270-9.
- Sampathkumar H, Chen XW, Luo B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC Med Inform Decis Mak.* 2014;14:91. doi: [10.1186/1472-6947-14-91](https://doi.org/10.1186/1472-6947-14-91).
- Inglis JM, Caughey GE, Smith W, Shakib S. Documentation of adverse drug reactions to opioids in an electronic health record. *Intern Med J.* 2021;51(9):1490-6. doi: [10.1111/imj.15209](https://doi.org/10.1111/imj.15209).
- Chaichulee S, Promchai C, Kaewkomon T, Kongkamol C, Ingviya T, Sangsupawanich P. Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing. *PLoS One.* 2022;17(8):e0270595. doi: [10.1371/journal.pone.0270595](https://doi.org/10.1371/journal.pone.0270595).
- Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. *Stud Health Technol Inform.* 2010;160(Pt 1):739-43.
- Shang N, Xu H, Rindflesch TC, Cohen T. Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J Biomed Inform.* 2014;52:293-310. doi: [10.1016/j.jbi.2014.07.011](https://doi.org/10.1016/j.jbi.2014.07.011).
- Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform.* 2015;53:196-207. doi: [10.1016/j.jbi.2014.11.002](https://doi.org/10.1016/j.jbi.2014.11.002).
- Tang H, Solti I, Kirkendall E, Zhai H, Lingren T, Meller J, et al. Leveraging food and drug administration adverse event reports for the automated monitoring of electronic health records in a pediatric hospital. *Biomed Inform Insights.* 2017;9:1178222617713018. doi: [10.1177/1178222617713018](https://doi.org/10.1177/1178222617713018).
- Tang Y, Yang J, Ang PS, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *Int J Med Inform.* 2019;128:62-70. doi: [10.1016/j.ijmedinf.2019.04.017](https://doi.org/10.1016/j.ijmedinf.2019.04.017).
- Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, et al. Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JMIR Public Health Surveill.* 2019;5(2):e11264. doi: [10.2196/11264](https://doi.org/10.2196/11264).
- Kim Y, Meystre SM. Ensemble method-based extraction of medication and related information from clinical texts. *J Am Med Inform Assoc.* 2020;27(1):31-8. doi: [10.1093/jamia/ocz100](https://doi.org/10.1093/jamia/ocz100).
- Chaichulee S, Promchai C, Kaewkomon T, Kongkamol C, Ingviya T, Sangsupawanich P. Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing. *PLoS One.* 2022;17(8):e0270595. doi: [10.1371/journal.pone.0270595](https://doi.org/10.1371/journal.pone.0270595).
- Zolnoori M, Fung KW, Patrick TB, Fontelo P, Kharrazi H, Faiola A, et al. The PsyTAR dataset: from patients generated narratives to a corpus of adverse drug events and effectiveness of psychiatric medications. *Data Brief.* 2019;24:103838. doi: [10.1016/j.dib.2019.103838](https://doi.org/10.1016/j.dib.2019.103838).
- Bayer S, Clark C, Dang O, Aberdeen J, Brajovic S, Swank K, et al. ADE eval: an evaluation of text processing systems for adverse event extraction from drug labels for pharmacovigilance. *Drug Saf.* 2021;44(1):83-94. doi: [10.1007/s40264-020-00996-3](https://doi.org/10.1007/s40264-020-00996-3).
- Aldrich SL, Poweleit EA, Prows CA, Martin LJ, Strawn JR, Ramsey LB. Influence of CYP2C19 metabolizer status on escitalopram/citalopram tolerability and response in youth with anxiety and depressive disorders. *Front Pharmacol.* 2019;10:99. doi: [10.3389/fphar.2019.00099](https://doi.org/10.3389/fphar.2019.00099).
- Chermá MD, Ahlner J, Bengtsson F, Gustafsson PA. Antidepressant drugs in children and adolescents: analytical and demographic data in a naturalistic, clinical study. *J Clin Psychopharmacol.* 2011;31(1):98-102. doi: [10.1097/JCP.0b013e318205e66d](https://doi.org/10.1097/JCP.0b013e318205e66d).
- Hughes S, Lacasse J, Fuller RR, Spaulding-Givens J. Adverse

-
- effects and treatment satisfaction among online users of four antidepressants. *Psychiatry Res.* 2017;255:78-86. doi: [10.1016/j.psychres.2017.05.021](https://doi.org/10.1016/j.psychres.2017.05.021).
31. Bressler CJ, Malthaner L, Pondel N, Letson MM, Kline D, Leonard JC. Identifying children at risk for maltreatment using emergency medical services' data: an exploratory study. *Child Maltreat.* 2024;29(1):37-46. doi: [10.1177/10775595221127925](https://doi.org/10.1177/10775595221127925).
32. Lewinski NA, McInnes BT. Using natural language processing techniques to inform research on nanotechnology. *Beilstein J Nanotechnol.* 2015;6:1439-49. doi: [10.3762/bjnano.6.149](https://doi.org/10.3762/bjnano.6.149).
33. Fossouo Tagne J, Yakob RA, McDonald R, Wickramasinghe N. Barriers and facilitators influencing real-time and digital-based reporting of adverse drug reactions by community pharmacists: qualitative study using the task-technology fit framework. *Interact J Med Res.* 2022;11(2):e40597. doi: [10.2196/40597](https://doi.org/10.2196/40597).