Anti-Depressant Drugs's Adverse Events by FDA Adverse Event Reporting System (FAERS) Analysis from January 2019 to December 2023 by Using Machine Learning Tools
Research Progress Report

ALEX (THI NGOC BICH) NGUYEN
A1848722

August 10, 2024

Report submitted for **Data Science Research Project A**
at the School of Mathematical Sciences, University of Adelaide

THE UNIVERSITY of ADELAIDE

Project Area: **Data science in medical data**
Project Supervisor: **DR. INDU BALA**

Abstract

Mental health is more prevalent in the modern world with the most issue of depression. With many U.S. Food and Drug Administration (FDA)-approved drugs to treat depressive-related problems, there are unknown adverse effects behind these drugs that we need to identify to have a better treatment for the majority of the population when they are in need. Using FDA's Adverse Event Reporting System (FAERS) data from 2019 to 2023 with diverse text-based values and applying the machine learning model BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding to match adverse keywords with the International Classification of Diseases (ICD) 11 API could help us to understand the side effects of these antidepressants and increase more safety medication.

1 Introduction

Mental health issues have significantly impacted worldwide. Over 40% of Australians have faced mental disorders during their entire life, with 20% of Australians experiencing them during the COVID-19 pandemic [1]. Depression, a common mental disorder, affects some 280 million people globally, which is about 3.8% of the world's population [2]. Those affected experience ongoing symptoms like persistent sadness or loss of interest in life's pleasures. The prevalence and severity of depression highlight the critical need to explore and understand the effectiveness and safety of treatments available. With millions using antidepressants worldwide, understanding the adverse events (AEs) of these drugs is crucial. This study aims to examine the FDA's Adverse Event Reporting System (FAERS) from 2019 to 2023 by using various machine learning-based analyses of AEs related to antidepressant use. Through this analysis, recommendations for medication safety could help to enhance treatment approaches for depression patients.

2 Background

There were 79 FDA-approved medications for neurological disorders in 2015 to treat neuropsychiatric diseases (depression, schizophrenia), neurotraumatic diseases, and neurodegenerative diseases [3]. Depression is diagnosed if the symptoms affect a person for at least two weeks. There are several types of depression, including major depression that interferes with daily life, persistent depressive disorder, perinatal depression during pregnancy or after giving birth, seasonal affective disorder, and depression with symptoms of psychosis of delusions and hallucinations [4] (schizophrenia, bipolar disorder, Parkinson's disease, Alzheimer's disease, and related dementias [5]).

Some antidepressant drugs are also FDA-approved for other medical disorders treatment [6]. There are various medications available for treating depression, categorised mainly as Selective Serotonin Reuptake Inhibitors (SSRIs), Serotonin and Norepinephrine Reuptake Inhibitors (SNRIs), Tricyclic and Tetracyclic Antidepressants, Atypical Antidepressants, Monoamine Oxidase Inhibitors (MAOIs), N-methyl D-aspartate (NMDA) Antagonists, and Neuroactive Steroid Gamma-Aminobutyric Acid (GABA)-A Receptor Positive Modulators [7]. These medications are vital for managing depressive disorders but can also cause various side effects. Common adverse effects include nausea, dizziness, dry mouth, constipation, and more severe issues like seizures, abnormal bleeding, and liver failure [7].

For tracking the AEs of medical devices and drugs, the FDA Adverse Event Reporting System (FAERS) is introduced to record these events across countries with quarterly collections from manufacturers [8]. The data files are available in ASCII or SGML format per quarter with case report's information about demographics of the patients (`DEMO`), drugs (`DRUG`), and indications for reported drugs. (diagnoses) (`INDI`), reactions (`REAC`), patient outcome (`OUTC`), source (`RPSR`), drug therapy start dates and end dates (`THER`) with an instruction (`README`) of utilising these data files [8].

With the FAERS database and FDA drugs information [9], there are not many confirmed associations between antidepressant drugs and severe effects. We believe more FDA-approved antidepressant drugs after 2015 [3] might have other side effects besides the previous studies focusing on the FAERS database with specific medications, including escitalopram oxalate [10] and quetiapine [11]. Moreover, with the trigger of the COVID-19 pandemic, depression increased by 25% in 2020 - one year after the virus appeared [12]. This could be from many reasons, from quarantine and social isolation (starting from February in China [13]), stress to healthcare workers or patients with pre-existing conditions that have no proper treatment due to the lack of healthcare services during the pandemic [12]. To examine the risks behind antidepressant drugs and the impact of the COVID-19 pandemic, we would consider the following:

1. What are the actual symptoms of diagnosed depression?
2. Which medications are used for diagnosed depression patients?
3. Are there any related side effects of these medications on depression?
4. Are there any changes in side effects before (from 2019 to March 2020) and after the outbreak of COVID-19 (from April 2020 to December 2023)?

3 Methods

With the extensive database from FAERS, the author uses the illustrated process in Figure 1 from data collection, preprocessing, and identification of AEs. The details are explained as follows.

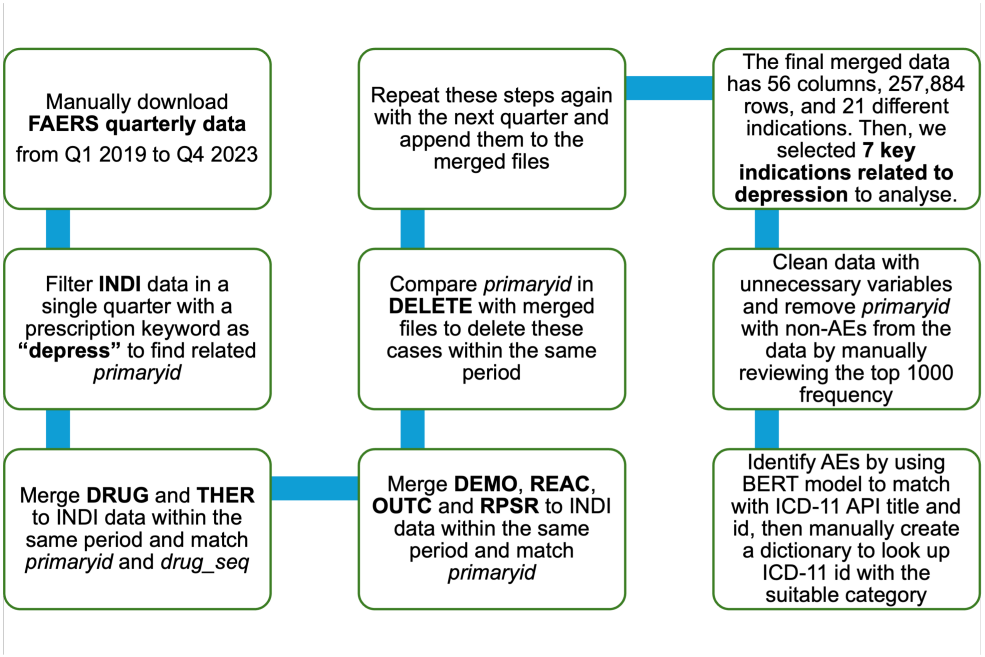

Figure 1: Project's process from data collection to identification AEs

3.1 Data collection and preprocessing

As mentioned above, to examine the AEs of FDA-approved antidepressants, the FAERS quarterly data was manually downloaded from the database [8] from January 2019 (Q1 2019) to December 2023 (Q4 2023) to analyse the impact of the COVID-19 pandemic.

Given the extensive data for many medical devices and drugs from the FAERS and the limitations of resources and computing, the author used the `INDI` data of each quarter with a prescription keyword in `indi_pt` of `"depress"` to find related case reports with `primaryid`. After filtering the data based on indication, the `DRUG` and `THER` data in the same period were merged with matching `primaryid` and `drug_seq`. Following, the `DEMO`, `REAC`, `OUTC`, and `RPSR` data in the same period were merged with matching `primaryid`. After combining all the data files, we compared the `DELETE` cases to remove the reports requiring confidentiality or wrong inputs. We repeated these steps for all the quarterly data files. The merged data had 257,884 case reports, 56 variables, and 21 different indications, as shown in Table 1.

Table 1: Frequency of Drugs Indication

Indication	Frequency
Major depression	128891
Antidepressant therapy	71373
Mixed anxiety and depressive disorder	19345
Schizoaffective disorder depressive type	18527
Persistent depressive disorder	8545
Adjustment disorder with depressed mood	4720
Perinatal depression	3430
Respiratory depression	785
Post stroke depression	445
Antidepressant drug level	434
Adjustment disorder with mixed anxiety and depressed mood	405
Electrocardiogram ST segment depression	222
Agitated depression	220
Antidepressant drug clearance	206
Menopausal depression	135
Myocardial depression	88
Antidepressant discontinuation syndrome	85
Congenital hypoplasia of depressor angularis oris muscle	9
Childhood depression	9
Neonatal respiratory depression	6
Antidepressant drug level therapeutic	3

After examining the complexity and many factors of psychosis disorders, perinatal depression symptoms and not relevant common depression, the author decided to focus on standard depressive disorders to analyse further. The focused symptoms are 'major depression', 'antidepressant therapy', 'mixed anxiety and depressive disorder', 'persistent depressive disorder', 'adjustment disorder with depressed mood', 'antidepressant drug level', and 'adjustment disorder with mixed anxiety and depressed mood', as shown in Table 2. The data with focused indications had 233,713 case reports and 56 variables.

Table 2: Frequency of Focused Drugs Indication

Indication	Frequency
major depression	128891
antidepressant therapy	71373
mixed anxiety and depressive disorder	19345
persistent depressive disorder	8545
adjustment disorder with depressed mood	4720
antidepressant drug level	434
adjustment disorder with mixed anxiety and depressed mood	405

Due to resource and computing limitations, we removed some unactionable and duplicated values, resulting in 46 variables in total, as shown in the author's project code. Then, we manually reviewed the top 1,000 AEs that appeared for the first time in each patient (`pt`) to remove reactions that are not AEs in the interim data file. Non-AEs were removed, as shown in Figure 2.

After filtering indication and actual AEs, the final data has 154,328 case reports and 46 variables, as shown in Figure 3.

```
# filter none adverse events
values_to_remove = ['drug ineffective', 'drug interaction', 'off label use', 'toxicity to various agents',
                    'product use in unapproved indication', 'condition aggravated', 'intentional product misuse',
                    'product packaging difficult to open', 'loss of personal independence in daily activities',
                    'therapeutic product effect incomplete', 'product use issue', 'economic problem', 'treatment failure',
                    'therapy nonresponder', 'treatment noncompliance', 'inappropriate schedule of product administration',
                    'drug ineffective for unapproved indication', 'product dose omission issue', 'incorrect dose administered',
                    'covid19', 'foetal exposure during pregnancy', 'product substitution issue', 'prescribed overdose',
                    'loss of employment', 'wrong technique in product usage process', 'maternal exposure during pregnancy',
                    'disease recurrence', 'homeless', 'therapy cessation', 'underdose', 'bankruptcy', 'prescribed underdose',
                    'divorced', 'product dose omission', 'drug withdrawal syndrome neonatal', 'therapy interrupted', 'adverse event',
                    'disease progression', 'therapeutic product effect decreased', 'device issue', 'product prescribing error',
                    'adverse drug reaction', 'small for dates baby', 'hospitalisation', 'theft', 'shoplifting',
                    'prescription drug used without a prescription', 'insurance issue', 'therapeutic response unexpected',
                    'unevaluable event', 'wrong product administered', 'road traffic accident', 'product dispensing error',
                    'accidental overdose', 'device malfunction', 'symptom recurrence', 'no adverse event', 'product quality issue',
                    'accident at work', 'drug effective for unapproved indication', 'quality of life decreased',
                    'personal relationship issue', 'multiple drug therapy', 'international normalised ratio increased',
                    'inadequate analgesia', 'incorrect route of product administration', 'unintended pregnancy',
                    'product prescribing issue', 'drug effect less than expected', 'inappropriate affect', 'therapeutic response shortened',
                    'covid19 pneumonia', 'product dose omission in error', 'on and off phenomenon', 'substance use',
                    'labelled drugdrug interaction medication error', 'therapeutic response changed', 'antipsychotic drug level above therapeutic',
                    'drug effect incomplete', 'false positive investigation result', 'adverse reaction', 'expired product administered',
                    'nonspecific reaction', 'selfmedication', 'drugdisease interaction', 'wrong technique in device usage process',
                    'labelled drugdrug interaction issue', 'device leakage', 'post procedural haemorrhage', 'product complaint',
                    'contraindicated product prescribed', 'coronavirus infection', 'product administration error', 'application site reaction']
```

Figure 2: Identification of non-AEs to remove

```
<class 'pandas.core.frame.DataFrame'>
Index: 154328 entries, 0 to 257882
Data columns (total 46 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   primaryid      154328 non-null  int64
 1   caseid         154328 non-null  int64
 2   drug_seq       154328 non-null  int64
 3   role_cod       154328 non-null  object
 4   drugname       154328 non-null  object
 5   prod_ai        153929 non-null  object
 6   val_vbm        154328 non-null  object
 7   route          117690 non-null  object
 8   dose_vbm       111721 non-null  object
 9   cum_dose_chr   3920 non-null    float64
 10  cum_dose_unit  3626 non-null    object
 11  dechal         103414 non-null  object
 12  rechal         27798 non-null   object
 13  lot_num        19364 non-null   object
 14  exp_dt         739 non-null     float64
 15  nda_num        45895 non-null   float64
 16  dose_amt       95276 non-null   float64
 17  dose_unit      95276 non-null   object
 18  dose_form      50401 non-null   object
 19  dose_freq      57427 non-null   object
...
 44  outc_cod       146744 non-null  object
 45  faers_date     154328 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(12), int64(7), object(26)
memory usage: 55.3+ MB
```

Figure 3: Information of final data

3.2 Identification of adverse events

To identify AEs, we had a short brief of these events within the report with more than 3,566 adverse events with different keywords, as shown in Figure 4.

```
pt
suicidal ideation                                    3549
suicide attempt                                      3163
serotonin syndrome                                   2765
anxiety                                              2396
depression                                           2280
                                                     ...
oral surgery                                            1
hepatic lesion                                          1
tongue injury                                           1
posterior reversible encephalopathy syndrome            1
alcoholic                                               1
Name: count, Length: 3566, dtype: int64
```

Figure 4: Initial Adverse Events

With the diversity of AEs, we used the International Statistical Classification of Diseases and Related Health Problems (ICD) to systematically classify all these symptoms into common diseases with the ICD-11 API method of programmatic access to the ICD database. [14]. The ICD-11 API compare the AE keywords in our database and the titles of the diseases in ICD, resulting in the ICD ID in our database.

Then, we use "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding to pre-train deep bidirectional representations from the unlabeled text by joint conditioning on both left and right context in all layers" [15]. It "uses the Transformer encoder architecture to process each token of input text in the full context of all tokens before and after" [16]. With limited resources and computing, we used the BERT-Based and Uncased model from the original BERT authors [16] by following the steps from Keras [17] that is proven to effectively normalisation of biomedical and clinical entity [18]. After modelling, the adverse events were from 3,566 different AEs, which decreased to 86 related ICDs, as shown in Figure 5.

```
icd_code
http://id.who.int/icd/entity/67207871      38006
http://id.who.int/icd/entity/2070699808    16375
http://id.who.int/icd/entity/1791890273    12798
http://id.who.int/icd/entity/1866592137    11518
http://id.who.int/icd/entity/455330172     11268
                                           ...
http://id.who.int/icd/entity/359051131         2
http://id.who.int/icd/entity/229280356         1
http://id.who.int/icd/entity/342704940         1
http://id.who.int/icd/entity/1614234080        1
http://id.who.int/icd/entity/198024823         1
Name: count, Length: 86, dtype: int64
```

Figure 5: Unique ICD ID

Following this, we manually checked the ICD ID within the category. We created a dictionary to map symptom categories for further analysis, resulting in 75 categories, as shown in Figure 6.

```
cate_ae
amnesia                                                    38006
post traumatic stress disorder                             16375
upper respiratory tract disorders                          12798
disorders of nerve root, plexus or peripheral nerves       11518
dermatoses provoked by physical or environmental factors   11268
                                                           ...
injuries to the thorax                                         2
injury of other or unspecified intrathoracic organs            2
inflammatory disorders of the female genital tract             1
injury of urinary or pelvic organs                             1
certain skin disorders attributable to bacterial infection     1
Name: count, Length: 75, dtype: int64
```

Figure 6: ICD Category of Adverse Events

4 Results

As results in Figure 7, we identified the most common AEs of antidepressant drugs are amnesia, post-traumatic stress disorder, upper respiratory tract disorders, disorders of the nerve root, plexus or peripheral nerves, and skin disorders from 2019 to 2023.

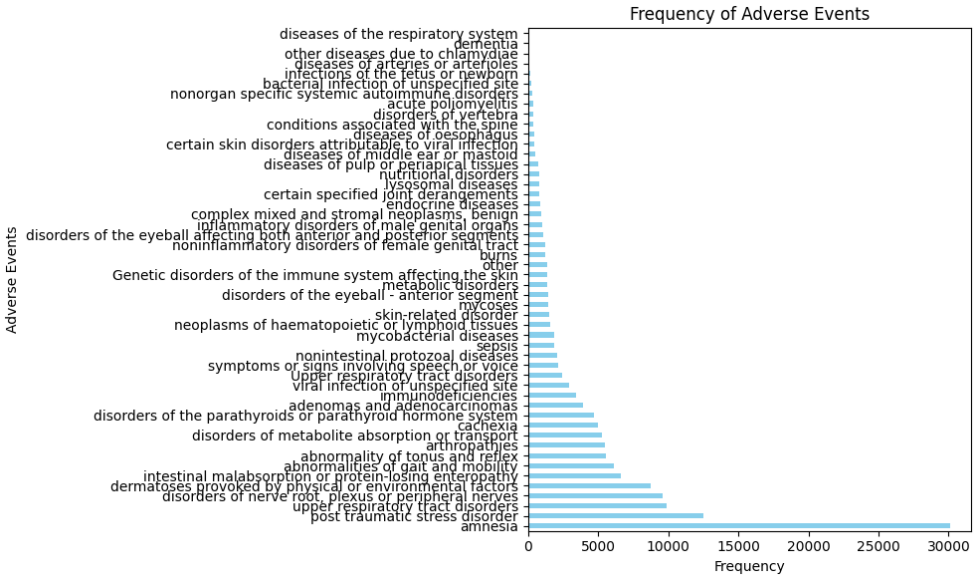

Figure 7: Top 50 adverse events of antidepressants based on ICD category

To illustrate the changes from 2019 to 2023, amnesia and post-traumatic stress disorder were the top 2 AEs that increased significantly in 2022 with doubting of effect from the pandemic with more than 1000 cases in Figure 8. Upper respiratory tract and skin disorders fluctuated around 400 cases before the pandemic and increased to more than 500 cases after 2022. The disorders of the nerve root, plexus or peripheral nerves decreased in 2022, opposite to other symptoms.

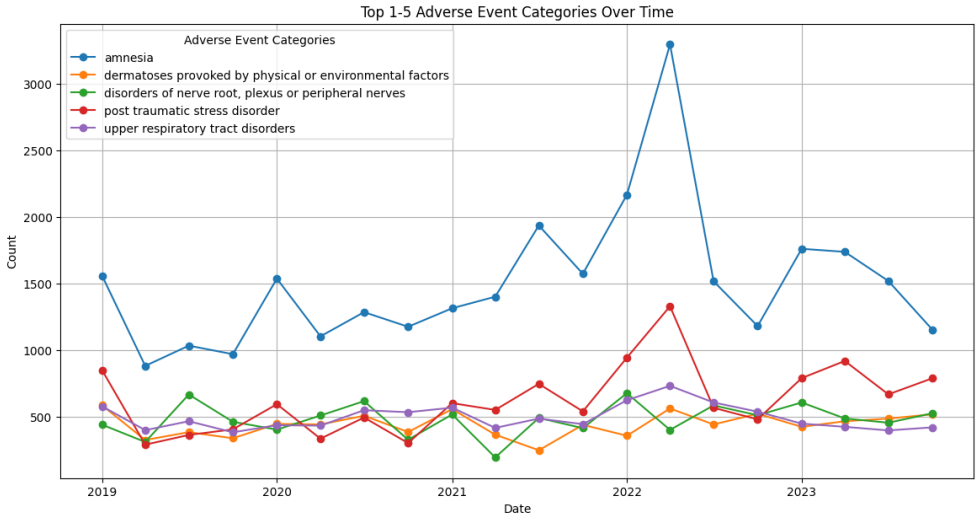

Figure 8: Top 5 adverse events of antidepressants based on ICD category changes over time

We could observe similar trends for the top 6 to 20 AEs in Figures 9, 10, and 11.

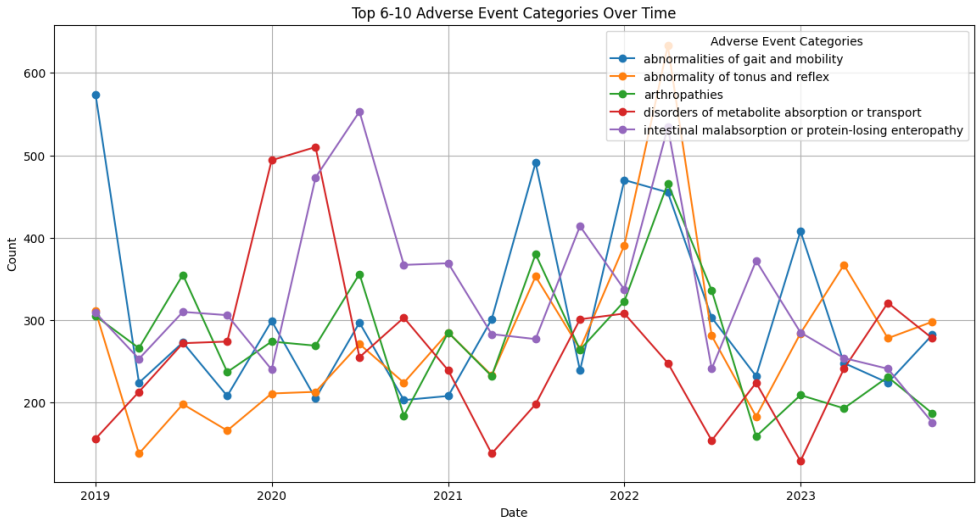

Figure 9: Top 6-10 adverse events of antidepressants based on ICD category changes over time

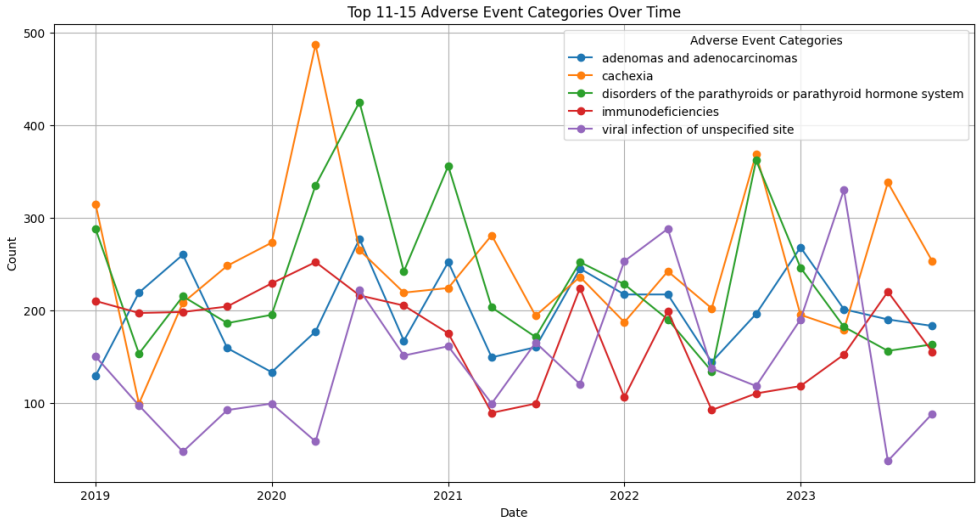

Figure 10: Top 11-15 adverse events of antidepressants based on ICD category changes over time

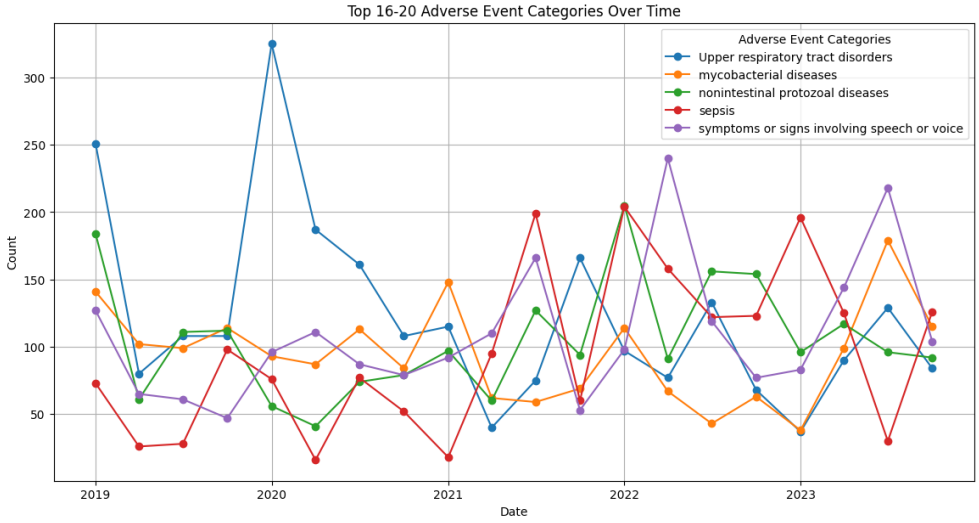

Figure 11: Top 16-20 adverse events of antidepressants based on ICD category changes over time

For the bottom 10 AEs in Figure 12, we could see uncommon AEs, including breast or oral neoplasms and inflammatory disorders in pelvic organs in females. These symptoms could alert medical practitioners to observe closely with their patients.

```
cate_ae
neoplasms of unknown behaviour of breast                          12
disorders of muscles                                              12
neoplasms of unknown behaviour of lip, oral cavity or pharynx      9
neoplasms of unknown behaviour of female genital organs            4
disorders of consciousness                                         3
disruptive behaviour or dissocial disorders                        2
injuries to the thorax                                             2
inflammatory disorders of the female genital tract                 1
injury of urinary or pelvic organs                                 1
certain skin disorders attributable to bacterial infection         1
Name: count, dtype: int64
```

Figure 12: Bottom 10 adverse events of antidepressants based on ICD category changes over time

5 Conclusion and Future scope

The project could identify common categories for AEs of using antidepressants. However, the project is still limited in using only BERT-based models and ICD-11 API, which results in false or too general categories. The authors have yet to analyse the reasons behind these AEs and the lack of evidence supported by other scientists for medical reasons. Further research and analysis are required to make meaningful insights.

For the future scope of the project, these challenges need to be tackled:

1. Research symptoms of depression to classify the symptoms belonging to depression or actual AEs.
2. Establish a list of all the medications used for diagnosed depression patients from 2019 to 2023 based on the Drugs@FDA database [3] and guidelines for treating depression [7] with the keyword 'antidepressant'.
3. Consider having other keywords for indications of antidepressant drugs besides the keyword 'depress' to better extract all the adverse symptoms.
4. Research other NLP models and AEs standardised categories to extract the keywords more accurately and precisely.
5. Analyse the association between medications and categorised adverse symptoms using statistical models.

6. Explore the changes in AEs of anti-depressant drugs before and after the COVID-19 pandemic using machine-learning (ML) models by using the timelines of before COVID-19 (Q1 2019 to Q1 2020) and after COVID-19 (Q2 2020 to Q4 2023). Suggested ML models of Gradient Boosting Machines (XGBoost, LightGBM, CatBoost) and Random Forest could be explored. We also consider using deep learning models if the resources allow, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM).

Acknowledgements

The author acknowledges using OpenAI for recommendations of some computing parts to analyse this dataset to make meaningful insights.

Appendices

This work uses the database from the FDA Adverse Event Reporting System (FAERS) Quarterly Data Extract Files from quarter 1 in 2019 to quarter 4 in 2023.

For analysis, the author writes herself in Jupyter Notebook using Python programming languages and uploads the code to the GitHub repository.

References

[1] Australian Bureau of Statistics, "National study of mental health and wellbeing for 2020-2022." https://www.abs.gov.au/statistics/health/mental-health/national-study-mental-health-and-wellbeing/latest-release, 2023. Accessed on: August 9, 2024. [Online].

[2] World Health Organization (WHO), "Depressive disorder (depression)." https://www.who.int/news-room/fact-sheets/detail/depression, 2023. Accessed on: August 9, 2024. [Online].

[3] P. Bhutani, G. Joshi, N. Raja, N. Bachhav, P. K. Rajanna, H. Bhutani, A. T. Paul, and R. Kumar, "US FDA approved drugs from 2015-June 2020: a perspective," *Journal of Medicinal Chemistry*, vol. 64, pp. 2339–2381, 2021.

[4] National Institute of Mental Health (NIMH), "Understanding psychosis." https://www.nimh.nih.gov/health/topics/depression, 2024. Accessed on: August 9, 2024. [Online].

[5] National Institute of Mental Health (NIMH), "Depression." https://www.nimh.nih.gov/health/publications/understanding-psychosis. Accessed on: August 9, 2024. [Online].

[6] Z. M. Sheffler, P. Patel, and S. Abdijadid, "National Center for Biotechnology Information - Antidepressants." https://www.ncbi.nlm.nih.gov/books/NBK538182, 2023. Accessed on: August 9, 2024. [Online].

[7] U.S. Food and Drug Administration (FDA) Office of Women's Health, "Depression medicines." https://www.fda.gov/consumers/womens-health-topics/depression-medicines, 2019. Accessed on: August 9, 2024. [Online].

[8] U.S. Food and Drug Administration (FDA), "FDA Adverse Event Reporting System (FAERS) Quarterly Data Extract Files." https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html. Accessed on: August 9, 2024. [Online].

[9] U.S. Food and Drug Administration (FDA), "Drug approvals and databases." https://www.fda.gov/drugs/development-approval-process-drugs/drug-approvals-and-databases, 2022. Accessed on: August 9, 2024. [Online].

[10] Y. Jiang, Y. Cheng, Z. Du, Y. Shen, Q. Zhou, Y. Ji, and H. Zhu, "Unveiling potential adverse events associated with escitalopram oxalate: A real-world analysis based FDA adverse event reporting system database," *Journal of Psychopharmacology*, vol. 38, no. 6, pp. 567–578, 2024.

[11] Y. Shu, Y. Ding, L. Liu, and Q. Zhang, "Cardiac adverse events associated with quetiapine: Disproportionality analysis of FDA adverse event reporting system," *CNS Neuroscience Therapeutics*, vol. 29, pp. 2705–2716, 2023.

[12] World Health Organization (WHO), "COVID-19 pandemic triggers 25Accessed on: August 9, 2024. [Online].

[13] World Health Organization (WHO), "Listings of WHO's response to COVID-19." https://www.who.int/news/item/29-06-2020-covidtimeline, 2020. Accessed on: August 9, 2024. [Online].

[14] World Health Organization (WHO), "International Statistical Classification of Diseases and Related Health Problems (ICD)." https://www.who.int/standards/classifications/classification-of-diseases. Accessed on: August 9, 2024. [Online].

[15] J. Devlin, M.-W. Chang, K. Lee, and K. G. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Google AI Language*, 2019. Published in arXiv, maintained and operated by Cornell Tech - Cornell University.

[16] Tensorflow, "Classify text with BERT." https://www.tensorflow.org/text/tutorials/classify_text_with_bert. Accessed on: August 9, 2024. [Online].

[17] Keras, "BERT." https://keras.io/api/keras_nlp/models/bert/. Accessed on: August 9, 2024. [Online].

[18] F. Li, Y. Jin, W. Liu, B. P. S. Rawat, P. Cai, and H. Yu, "Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT)-Based models on large-scale electronic health record notes: An empirical study," *JMIR Medical Informatics*, vol. 7, no. 3, pp. e14830–e14830, 2019.