Addressing Gaps in Real-World Data Sources through Automated Extraction of Psychiatric Comorbidities and Medication Side Effects from Clinical Notes

Noa Palmon, BS, Pedro Alves, BS, Michelle Leavy, MPH, Jessica Probst, MPH, Richard Gliklich, MD, and Costas Boussios, PhD | OM1, Inc, Boston, MA, USA

OM1®

Objectives

Depression is a significant public health concern, affecting some 21 million adults and 5 million adolescents in the United States.[1] While multiple treatment options exist, many questions remain about screening and diagnosis, treatment selection and treatment sequencing, treatment discontinuation, and short- and long-term patient outcomes, especially in populations traditionally underrepresented in clinical research. Real-world data (RWD) offer a potential avenue for broad, representative studies of depression diagnosis, treatment, and outcomes to inform patient care and improve outcomes.[2]

Yet, RWD studies are often limited to the data contained in structured fields. Some critical variables, such as psychiatric conditions that commonly co-occur with depression and side effects related to treatment, are often recorded in EHR systems as narrative text in unstructured clinical notes. This study aimed to assess the feasibility of identifying and extracting psychiatric comorbidities and side effects related to treatment from clinical notes drawn from a RWD source.

Methods

Data for this study were drawn from the OM1 Real World Data Cloud (OM1, Inc, Boston, MA, USA), a deterministically linked, de-identified, individual-level dataset containing EMR with medical and pharmacy claims from 2013 to present. All data were de-identified and tokenized, and the study was reviewed and IRB approved. The study cohort was restricted to patients with a diagnosis of major depressive disorder (MDD) for extraction of psychiatric comorbidities and patients with a diagnosis of MDD and at least one mention of a selective serotonin reuptake inhibitors (SSRI) in a clinical note.

A natural language processing-based approach was used to extract mentions of psychiatric comorbidities using a collection of common linguistic patterns to identify comorbidities in clinical notes and categorize them as affirmations (presence) or negations (absence). Language models were then constructed and validated by subject matter experts to ensure reliability of the patterns. This approach was also used to extract mentions of side effects recorded in clinical notes.

Results

Comorbidity Extractions

Using these inclusion criteria, 2.9 million notes were identified for analyses. Of these, 1,703,849 were classified as affirmations. Generalized anxiety disorder was mentioned most frequently, followed by bipolar disorder, anxiety disorder, and substance use disorder (Figure 1).

Negations of comorbidities (e.g., denial of substance use) were also extracted where relevant. Figure 3 presents examples of comorbidity affirmations and negations in clinical notes.

Side Effect Extractions

Using these inclusion criteria, 2.6 million notes were identified for analysis. Of these, 36,844 side effect mentions were extracted from 33,332 notes and 7,594 patients. Common side effects recorded in the clinical notes included weight gain, nausea, fatigue, sexual side effects, diarrhea, insomnia, headache, agitation, and suicidal thoughts (Figure 2). Figure 3 presents examples of side effects mentioned in clinical notes.

Conclusions

- Psychiatric comorbidities that commonly co-occur with depression can be extracted from clinical notes from RWD using a medical language processing-based approach.

- Common side effects related to SSRIs can also be extracted at scale from clinical notes using this medical language processing-based approach.

- The approach used in this study may be applied to other concepts related to depression or other clinical areas.

- Use of this approach may improve characterization of depression patients and understanding of treatment effectiveness in RWD sources used for depression research.

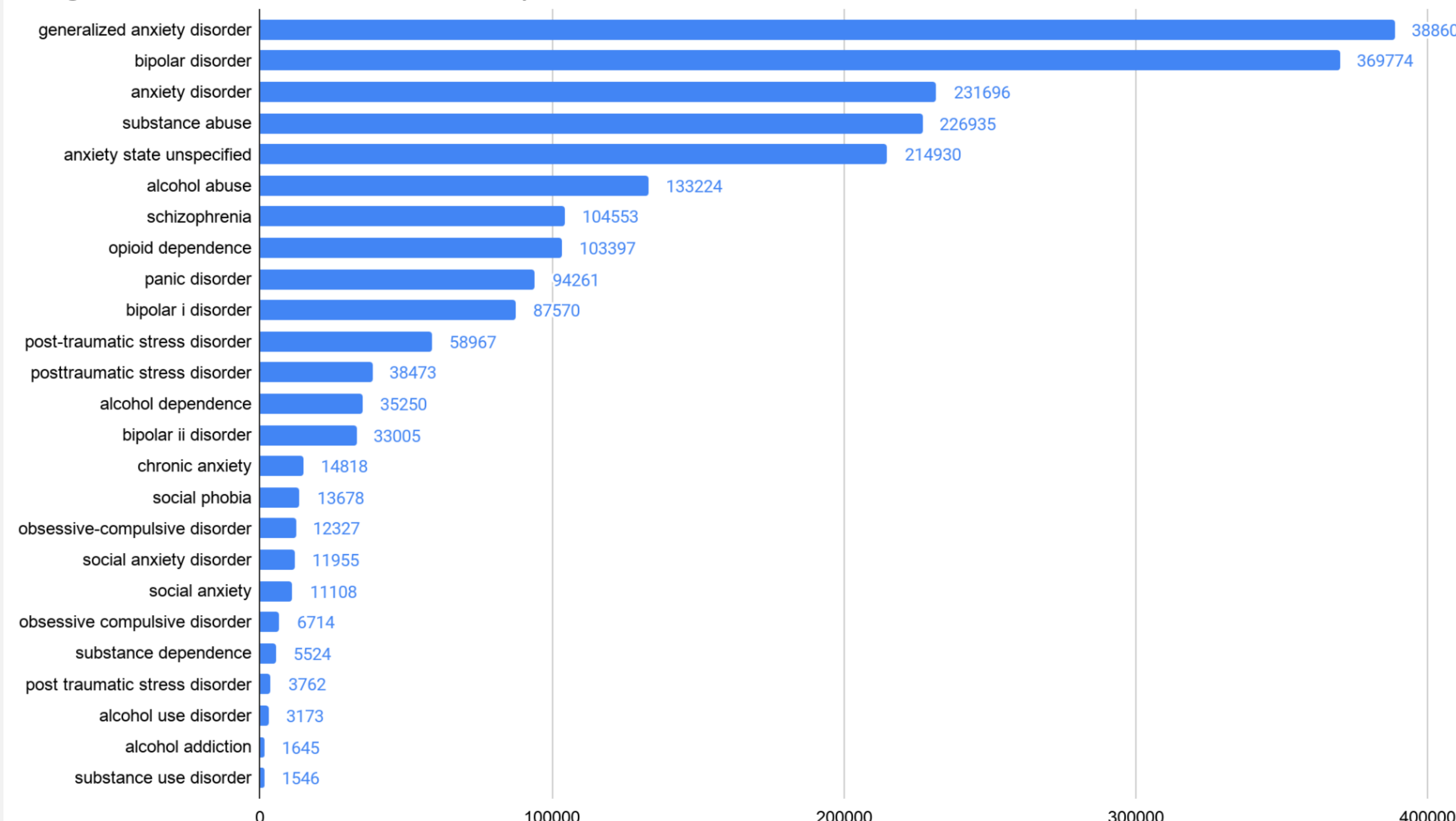**Figure 1.** Distribution of Extracted Psychiatric Comorbidities

Comorbidity	Count
generalized anxiety disorder	388603
bipolar disorder	369774
anxiety disorder	231696
substance abuse	226935
anxiety state unspecified	214930
alcohol abuse	133224
schizophrenia	104553
opioid dependence	103397
panic disorder	94261
bipolar i disorder	87570
post-traumatic stress disorder	58967
posttraumatic stress disorder	38473
alcohol dependence	35250
bipolar ii disorder	33005
chronic anxiety	14818
social phobia	13678
obsessive-compulsive disorder	12327
social anxiety disorder	11955
social anxiety	11108
obsessive compulsive disorder	6714
substance dependence	5524
post traumatic stress disorder	3762
alcohol use disorder	3173
alcohol addiction	1645
substance use disorder	1546

Figure 2. Distribution of Extracted Side Effects

Side effects (listed top to bottom): suicidality, dry mouth, worsening of depression symptoms, tachycardia, nightmares, panic attacks, tremor, unusual changes in behavior, hives/rash, dizziness, nervousness, agitation or restlessness, worsening symptoms, insomnia, headaches, other, sexual dysfunction, impact on appetite, "side effects", nausea, vomiting or diarrhea, fatigue
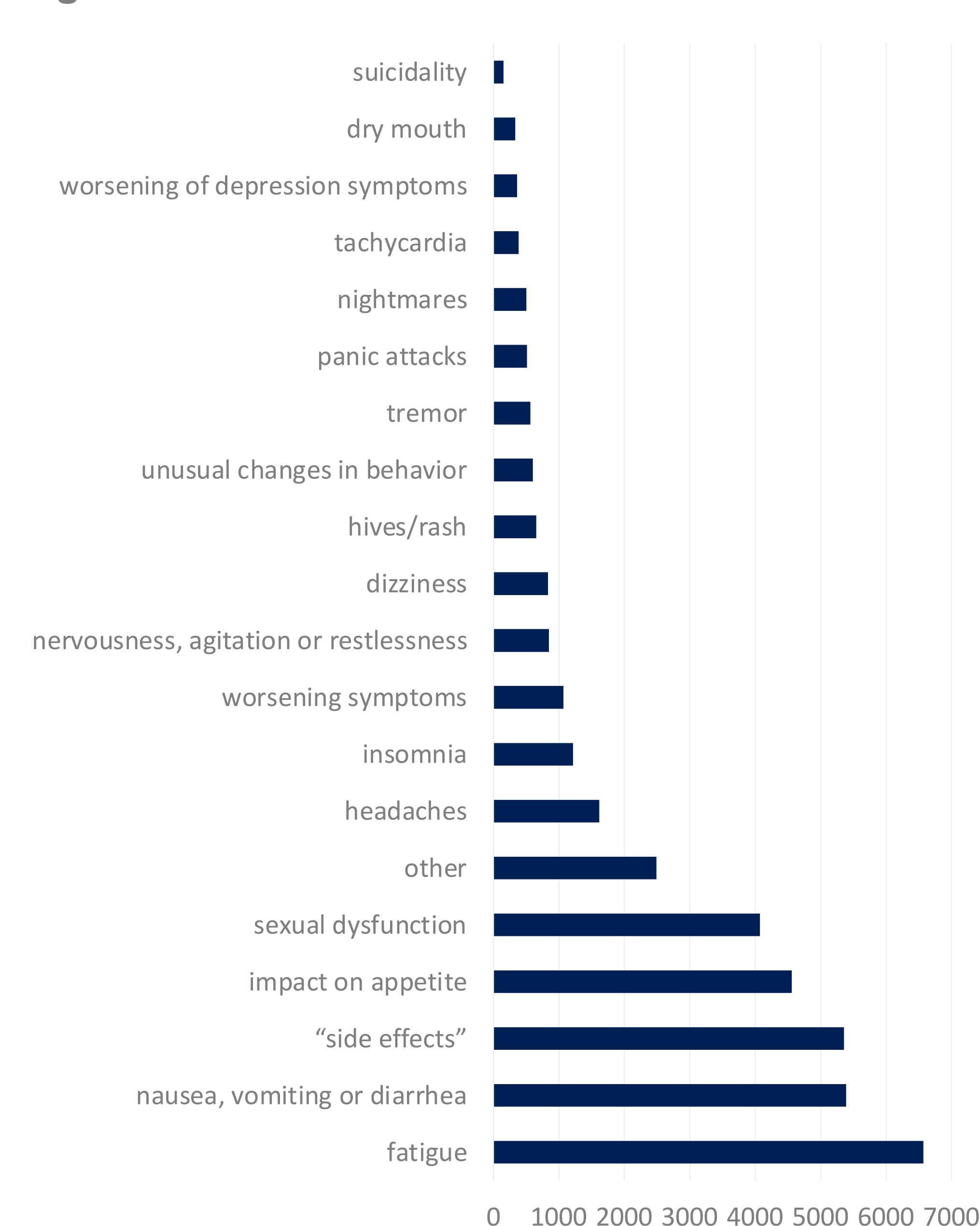
Figure 3. Examples of Comorbidity & Side Effect Mentions in Clinical Notes

Comorbidity Example: We advised the patient to follow a low-cholesterol diet, to exercise at least 3 times a week and to watch the BMI. Generalized anxiety disorder (F41.1): Doing well taking Alprazolam. Will continue to monitor…

Comorbidity Example: Does the patient use any substances, and if so, what are they? tobacco

Side Effect Example: She had to stop Lexapro because of diarrhea. The patient still feels kind of depressed.

Side Effect Example: …then she skipped it for a couple days, and her mood worsened. Fluoxetine 20 mg daily was associated with severe insomnia (up all night)

Side Effect Example: She has had poor response to Lexapro in the past and believes that sertraline caused stomach cramps

Side Effect Example: Pt reports stable mood, reports mild depression but is managing well. Pt reports paroxetine 10 mg caused side effects and since LOV stopped the medication.

References
1. Major Depression. National Institute of Mental Health. Updated July 2023.
2. Gliklich RE, Leavy MB, Cosgrove L, et al. Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice. Ann Intern Med. 2020;172(12):803-809. doi:10.7326/M19-3818.