Use of Medical Language Processing in Real-World Data to Understand Antidepressant Side Effects and Improve Research in Mental Health

Pedro Alves, BS, Alison Spencer, PhD, Michelle Leavy, MPH, Gary Curhan, MD, ScD, Carl Marci, MD, Costas Boussios, PhD | OM1, Inc, Boston, MA, USA

Background

Data from real-world care settings offer great potential for research aimed at improving diagnosis, treatment, and outcomes for patients with mental health conditions. Unlike data from controlled clinical trials, real-world data (RWD) reflect the realities of mental health care in a fragmented U.S. health system. RWD that includes electronic health records (EHRs) offer a more nuanced view of patient care and often includes broader, more representative patient populations. Appropriate use of RWD requires a thorough understanding of the strengths and limitations of data sources, as well as innovative technologies to link datasets and extract relevant information from structured fields and unstructured clinical notes (1, 2).

Depression is an area where RWD may be particularly useful. Depression is a significant public health issue, and new research is needed to improve diagnosis, treatment selection and sequencing, and patient outcomes. RWD offer opportunities to create large, heterogenous patient populations to support research on depression treatments and calculate standardized depression outcomes (3).

Research on depression treatment and outcomes that uses RWD sources such as EHRs must assess the availability and format of key variables. While some variables such as diagnoses may be available in structured fields, other critical variables such common side effects related to treatment are often recorded in EHR systems as narrative text in unstructured clinical notes. Extraction of this information with a validated computational approach is critical to making EHR data useful for depression research.

Objective

This study examined the feasibility of identifying and extracting common side effects related to selective serotonin reuptake inhibitors (SSRIs) from unstructured clinical notes in a real-world dataset.

Methods

Data for this study were drawn from the OM1 Real World Data Cloud (OM1, Inc, Boston, MA, USA). 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) who had at least one mention of a SSRI in a clinical note (Figure 1). An automated medical language processing-based approach was used to extract mentions of side effects. The approach identifies collections of linguistic patterns and phrases commonly used to record side effects in clinical notes. Language models were constructed and reviewed by clinical experts for validity and reliability.

Results

From a medical notes dataset consisting of 113,072,106 medical notes from 1,038,840 depression patients, a subset of 2,626,866 notes from 276,174 patients explicitly mentioned an SSRI. Out of these, 36,844 explicitly mentioned adverse effects were extracted from 33,332 notes and 7,594 patients (Table 1). Common adverse effects mentioned in the clinical notes were weight gain, nausea, fatigue, sexual side effects, diarrhea, insomnia, headache, agitation, and suicidal thoughts (Figures 2 and 3).

Conclusions

- Common side effects related to SSRIs can be extracted at scale from clinical notes using a medical language processing-based approach.
- The approach used in this study is a scalable and reproducible model for efficiently extracting concepts related to mental health conditions from unstructured clinical notes.
- Application of this efficient approach to other very large data sources and to other concepts should improve the utility of RWD for mental health research

Figure 1. Terms Used to Identify SSRIs in Notes

Celexa	Paxil CR
Citalopram	Pexeva
Escitalopram	Prozac
Fluoxetine	Sarafem
Fluvoxamine	Sertraline
Lexapro	Symbyax
Luvox	Viibryd
Luvox CR	Vilazodone
Paroxetine	Zoloft
Paxil	

SSRI = selective serotonin reuptake inhibitors

Table 1. Inclusion of Patients and Notes in the Extraction Cohort

Cohort	# of Patients	# of Notes
Patients with diagnosis of depression	36,455,543	-
Patients with at least one note and depression diagnosis	1,038,840	113,072,106
At least one SSRI mention in the notes	276,174	2,626,866
At least one side effect mention in the same sentence OR within 7 words of SSRI mention. Includes: • side effect as indication for the SSRI • side effect related to another medication • side effect as an effect of the SSRI	30,222	107,503
Side effect as an effect of the SSRI	7,594	33,332
Side effect confirmed as NOT related to SSRI	2,451	7,371
No relation between side effect and SSRI found	-	67,468

SSRI = selective serotonin reuptake inhibitors

Figure 2. Distribution of Extracted Side Effects

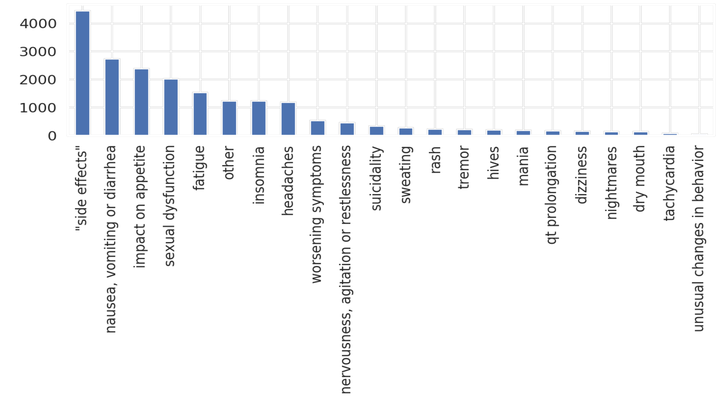

Figure 3. Examples of Side Effect Mentions

*She had to **stop Lexapro because of diarrhea**. The patient still feels kind of depressed.*

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

*She has had poor response to Lexapro in the past and believes that **sertraline caused stomach cramps**.*

*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.*

Conference

Presented at the APA 20223 Annual Meeting. May 20-24, 2023. San Francisco, USA.

References

1. Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA, 320(9), 867–868.
2. Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., LaVange, L., Marinac-Dabic, D., Marks, P. W., Robb, M. A., Shuren, J., Temple, R., Woodcock, J., Yue, L. Q., & Califf, R. M. (2016). Real-World Evidence - What Is It and What Can It Tell Us?. The New England journal of medicine, 375(23), 2293–2297.
3. Gliklich, R. E., Leavy, M. B., Cosgrove, L., Simon, G. E., Gaynes, B. N., Peterson, L. E., Olin, B., Cole, C., DePaulo, J. R., Jr, Wang, P., Crowe, C. M., Cusin, C., Nix, M., Berliner, E., & Trivedi, M. H. (2020). Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice. Annals of internal medicine, 172(12), 803–809.