A Social Media Study on the Effects of Psychiatric Medication Use

Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, Munmun De Choudhury

1 Georgia Tech, 2 Harvard Medical School, 3 NYU Shanghai, 4 Microsoft Research
1 koustuv.saha, bsugar, munmund @gatech.edu, 2 jtorous@bidmc.harvard.edu, 3 bd58@nyu.edu, 4 emrek@microsoft.com

Abstract

Understanding the effects of psychiatric medications during mental health treatment constitutes an active area of inquiry. While clinical trials help evaluate the effects of these medications, many trials suffer from a lack of generalizability to broader populations. We leverage social media data to examine psychopathological effects subject to self-reported usage of psychiatric medication. Using a list of common approved and regulated psychiatric drugs and a Twitter dataset of 300M posts from 30K individuals, we develop machine learning models to first assess effects relating to mood, cognition, depression, anxiety, psychosis, and suicidal ideation. Then, based on a stratified propensity score based causal analysis, we observe that use of specific drugs are associated with characteristic changes in an individual’s psychopathology. We situate these observations in the psychiatry literature, with a deeper analysis of pre-treatment cues that predict treatment outcomes. Our work bears potential to inspire novel clinical investigations and to build tools for digital therapeutics.

Introduction

Psychiatric medications are key to treat many mental health conditions, including mood, psychotic, and anxiety disorders. 1 in 6 Americans take psychiatric medications and they account for 5 of the top 50 drugs sold in the U.S (drugs.com). These drugs are designed to correct underlying neuro-pathological disease processes by restoring neural communication by modulating the brain’s chemical messengers and neurotransmitters (Barchas and Altemus 1999). These changes can be accompanied by debilitating neurological impairments and life-threatening effects as severe as suicidal ideation (Coupland et al. 2011) which reduce psychosocial functioning, and make social capital and vocational development less available to these individuals. Given the pervasiveness of their use, psychiatric medications can either alleviate or exacerbate mental illness burden on both personal and societal levels (Rosenblat et al. 2016).

One reason behind the mixed success of psychiatric medications stems from the fact that the mechanisms by which they modify the brain operation are poorly understood. In practice, their effects vary across individuals, and often do not achieve the intended result. Without any biological markers to match patients with the most appropriate medication, the selection of drug treatments is based primarily on trial-and-error (Cipriani et al. 2018; Trivedi et al. 2006). Unsurprisingly, frustration with treatment and side effects often causes treatment discontinuation (Bull et al. 2002).

Consequently, literature in precision psychiatry has emphasized the need to understand the psychiatric effects of these medications (Cipriani et al. 2009). Presently, most knowledge of drug reactions comes from clinical trials and reports of adverse events; e.g., the FDA’s Adverse Event Reporting System (open.fda.gov/data/faers) clinical trial database. However, these trials can be biased, being conducted and funded by pharmaceutical companies, and are rarely replicated in large populations (Lexchin et al. 2003). In addition, these clinical trials suffer from limitations such as non-standardized study design, confounding factors, and restrictive eligibility criteria (Lexchin et al. 2003). For example, an analysis found that existing inclusion criteria for most trials would exclude 75% of individuals with major depressive disorders (Blanco et al. 2008). Even well-designed clinical trials can suffer from low statistical power, or limited observability of effects due to short monitoring and study periods, spanning just weeks or months.

Contributions Our work seeks to address these gaps and complements existing methodologies for understanding the effects of psychiatric medications. We report a large-scale social media study of the effects of 49 FDA approved antidepressants across four major families (SSRIs, SNRIs, TCAs, and TeCAs) (descriptions in (Lopez-Munoz and Alamo 2009)). Our analysis is conducted using two years of Twitter data from two populations: 112M posts from 30K self-reported users of psychiatric medications and 707M posts from 300K users who did not. Adopting a patient-centered approach (Shippee et al. 2012), in this paper, we seek to study the effects of these drugs as reflected and self-reported in the naturalistic social media activities of individuals.

Accomplishing this goal involves meeting several technical challenges, importantly addressing causality, and our work offers robust and validated computational methods for the purpose. We first develop expert-validated machine learning models to assess psychopathological states.
known to be affected by psychiatric medications, including mood, cognition, depression, anxiety, psychosis, and suicidal ideation, as given in the literature (Coupland et al. 2011). Using initial social media mentions of drug intake, we then identify individuals likely beginning treatment. Based on a stratified propensity score analysis (Olteanu et al. 2017), we compare post-treatment symptoms in treated individuals to large untreated control population. With an individual treatment effect analysis, we study the relationship between pre-treatment mental health signals and post-treatment response.

Findings Our results show that most drugs are linked to a post-treatment increase in negative affect and decrease in positive affect and cognition. We find varying effects both within and between the drug families on psychopathological symptoms (depression, anxiety, psychosis, and suicidal ideation). Clinically speaking, SSRIs are associated with worsening symptoms, whereas TCAs lead to improvements. Studying the individual-specific outcomes, our analyses help associate drug effectiveness with individuals’ psycholinguistic attributes on social media.

Clinically, our findings reveal signals of the most common effects of the psychiatric medications over a large population, with the potential for improved characterization of their occurrence. Technologically, we show the potential of novel technologies in digital therapeutics, powered by large-scale social media analyses, to support digital therapeutics (Vieta 2015). These tools can improve the identification of adverse outcomes, as well as the behavioral and lifestyle changes in the heterogeneous outcomes of psychiatric drugs.

Privacy, Ethics, and Disclosure Given the sensitive nature of our work, despite working with public social media data, we are committed to securing the privacy of the individuals in our dataset. We use paraphrased examples of content and avoid personally identifiable information. Our findings were corroborated with our co-author who is a board-certified psychiatrist. However, our work is not intended to replace clinical evaluation by a medical professional, and should not be used to compare or recommend medications.

Background and Related Work

Psychiatric Drug Research and Prescriptions The mechanisms of action of many psychiatric drugs and the basis for specific therapeutic interventions, are not fully understood. Among other hypotheses, the monoamine hypothesis postulates that these drugs target the neurotransmitters serotonin, norepinephrine and dopamine, associated with feelings of well-being, alertness, and pleasure (Barchas and Altmanus 1999). From the monoamine standpoint, medications are classified into families, based on their brain receptor affinities, which distinguish their mechanism of action.

Antidepressant research has grown tremendously, ever since Imipramine, and other Tricyclic Antidepressants (TCAs) were discovered and found to be effective (Gillman 2007). However, TCAs have a broad spectrum of neurotransmitter affinities, which may often lead to undesirable side effects, such as liver toxicity, excessive sleepiness, and sexual dysfunction (Frommer et al. 1987). Several other compounds have since been introduced whose development was guided by the idea that increasing the selectivity of the target of action to individual neurotransmitters would, in theory, limit the incidence of side effects while maintaining the effectiveness of TCAs (Lopez-Munoz and Alamo 2009). These include Tetracyclic Antidepressants (TeCA), Serotonin Norepinephrine Reuptake Inhibitors (SNRI), and Selective Serotonin Reuptake Inhibitors (SSRI).

Given these biochemical underpinnings, historically psychiatric care has adopted a “Disease-Centered Model” (Moncrieff and Cohen 2009), one that justifies prescribing medications on the assumption that they help correct the biological abnormalities related to psychiatric symptoms. However, this model neglects the psychoactive effects of the drugs. Consequently, a “Drug-Centered Model” has been advocated (Moncrieff and Cohen 2009), enabling patients to exercise more control over their pharmacotherapy, and moving treatment in a collaborative direction between clinicians and patients. Our work builds on this notion towards a “Patient-Centered Model” (Shippee et al. 2012), where psychiatrists could leverage complementary techniques (such as stratifying users on their naturalistic digital footprints) to prescribe medications.

Understanding Effects of Psychiatric Drugs The efficacy, safety, and approval of psychiatric drugs are typically established through clinical trials. In one such trial, the randomized controlled trial (RCT), participants are randomly assigned to a treatment or a control group, where the former receives a particular drug, and the latter receives a placebo (eg. a sugar pill with no drug content). Then, the effects of the treatment are measured as a difference in the two groups following the drug intake. A major weakness of these trials is that they are often conducted on individuals who may significantly differ from actual patients, and often, they are not externally validated to a larger and a more representative population (Hannan 2008). As an alternative, a study design that has gained interest is observational study (Hannan 2008). The advantage here is that they enable the researchers to conduct subset analyses that can help to precisely identify which patients benefit from each treatment. Similarly, we use large-scale longitudinal data and a causal approach to not only examine the effects of psychiatric drugs, but also to provide a framework that finds insights about their effectiveness across strata of populations.

Pharmacovigilance, Web, and Social Media

Pharmacovigilance is “the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem” (WHO 2002). Over the years, pharmacovigilance has become centered around data mining of clinical trial databases and patient-reported data. Recently, patient-generated activity online has also been used to understand pharmacological effects in large populations (Harpaz et al. 2017). White et al. (2016) found that web search logs improve detection of adverse effects by 19%, compared to an offline approach.

Social media studies of drug and substance use, including behavioral changes, adverse reactions, and recovery have
garnered significant attention in HCI (Chancellor et al. 2019, Kicman et al. 2018, Liu et al. 2017). Recent research has studied the abuse of prescription drugs, by leveraging drug forums (MacLean et al. 2015), Twitter (Sarker et al. 2015), and Reddit (Gaur et al. 2018). Social media has also facilitated the identification of adverse drug reactions at the population level using self-reports (Lardon et al. 2015) as well as the mentions of side effects of adverse drug reaction on Twitter (Nikfarjam et al. 2015).

Social media enables individuals to candidly share their personal and social experiences (Kicman et al. 2018, Olteanu et al. 2017, Saha et al. 2019b), thereby providing low-cost, large-scale, non-intrusive data to understand naturalistic patterns of mood, behavior, cognition, social milieu, and even mental and psychological states, both in real-time and longitudinally (Chancellor et al. 2016, Coppersmith et al. 2014, De Choudhury et al. 2013, Dos Reis and Cu-lootta 2015, Saha et al. 2019a, Yoo and De Choudhury 2019). In characterizing drug use, being able to quantify these psychopathological attributes is extremely powerful.

Nevertheless, we observe a gap that digital pharmacovigilance studies, particularly those using social media, have largely targeted the named adverse effects of drugs (e.g., “headache”, “palpitations”, “nausea”), and have not measured broader forms of symptomatic changes longitudinally. To fill this gap, our work draws on theoretically grounded methodologies, including lexicon-based and machine learning approaches, to measure the symptomatic outcomes of psychiatric drug use longitudinally, including mood, cognition, depression, anxiety, psychosis, and suicidal ideation.

Data
This work leverages Twitter timeline data of individuals who self-report their use of psychiatric medications. The data collection involves: 1) curating a list of psychiatric medications; 2) using this list to collect Twitter posts that mentioned these medications; 3) identifying and filtering for only those posts where users self-reported about personal medication intake (using a personal medication intake classifier, and 4) collecting the timeline datasets of these individuals who self-reported psychiatric medication intake, and additionally doing that for another set of users who did not self-report psychiatric medication intake. We explain these steps here:

Psychiatric Medication List We scope our work to a list of FDA approved antidepressants and antidepressant augmentation drugs. We crawl a hand-curated set of Wikipedia pages of these drugs, to collect brand names, generic names, and drug family information to obtain a list of 297 brand names mapped to 49 generic names, grouped into four major families: SNRI, SSRI, TCA, TeCA. Our clinician co-author established the validity and relevance of this final list.

Twitter Data of Psychiatric Medication Usage We query the Twitter API for public English posts mentioning these drugs (brand or generic name) between January 01, 2015 and December 31, 2016 to obtain 601,134 posts by 230,573 unique users. A two year period balances concerns about being long enough to avoid confounds by idiosyncratic events and seasonal changes, but short enough to avoid major changes in social media use and drug prescription policies. This also enables us to collect sufficient pre- and post- medication usage timeline data for our ensuing analyses.

Personal Medication Intake Classifier Since mentioning a medication in a tweet does not necessarily indicate its usage, we filter out those posts that were first-person reports of using these medications. For this purpose, we employ a machine learning classifier built in a recent work (Klein et al. 2017). This classifier distinguishes Twitter posts into the binary classes (yes or no) if there is a self-report about personal medication intake. We replicate this model and train it on an expert-annotated dataset of 7,154 Twitter posts (dataset published in Klein et al. 2017). The classifier uses an SVM model with linear kernel and shows a mean k-fold (k=5) cross-validation accuracy and F1-score of 0.82 each.

We use this classifier to label the 601,134 medication mention posts to find that 93,275 of these posts indicate medication self-intake (example posts in Table 1). Figure 1a shows the monthly and overall distribution of the top 20 drugs in our dataset. We find that SSRIs (eg. Sertraline, Escitalopram, Fluoxetine) rank highest in the distribution. This aligns with external surveys on the most prescribed psychiatric drugs in that time which found that the top 5 antidepressants captured over 70% of the prescription volumes (Scripts 2018; Gohol 2018).

Compiling Treatment and Control Datasets The above 93,275 medication usage posts were posted by 52,567 unique users from whom we then collect Twitter metadata such as the number of tweets, followers, followees, and account creation date. To limit our analyses to typical Twitter users, we remove users (e.g., celebrities or typically inactive
users) with more than 5000 followers or followees or posted outside the range of 30 to 300,000 tweets—a choice motivated from prior work (Pavalanathan and Eisenstein 2016). For the remaining 34,518, we collect the timeline data between January 01, 2014 and February 15, 2018, to obtain a total of 112,025,496 posts. Finally, we limit our dataset to those users who posted both before and after their first self-reported use of medication and did not self-report the use before 2015. The resultant timeline dataset of 23,191 users is referred to hereon as the Treatment dataset.

Additionally, we build a Control dataset of users who did not self-report using psychiatric medication. We obtain 495,419 usernames via the Twitter streaming API and prune this list (as above) and remove accounts that did not exist pre-2015. We collect the timelines of the remaining 283,374 users, for a total 707,475,862 posts. Figure 1b shows the mean distribution of Twitter attributes in our two datasets.

**Methods**

**Study Design and Rationale** Recall that our research objective is to examine the effects of psychiatric medications in terms of the changes in mental health symptoms. Effectively answering this question necessitates the use of causal methods to reduce biases associated with the observed effects following the reported medication usage. The effects of drugs are most often measured through Randomized Controlled Trials (RCTs) in clinical settings (Cipriani et al. 2018; Szegedi et al. 2005). Due to the limitations of this approach, noted in the “Background and Related Work” section, and because of the potential advantages of a “Patient Centered Model” that focuses on using the naturalistic self-reports of individuals regarding their psychiatric medication use, this work adopts an observational study design. We do acknowledge that observational studies are weaker than RCTs in making conclusive causal claims like ones needed to accomplish the goals of this paper, but they provide complementary advantages over RCTs in many aspects (Hannan 2008).

Literature in statistics also provides support for these methods and similar frameworks have been leveraged in previous quantitative social media studies (De Choudhury et al. 2016, Kıcıman et al. 2018, Olteanu et al. 2017, Saha et al. 2018).

Specifically, we adopt a causal inference framework based on matching, which simulates an RCT setting by control-
in unlabeled data by transferring a classifier trained on a dif-
ferent labeled dataset (Saha and De Choudhury 2017). To
train such classifiers for use in our work, we identify sev-
eral Reddit communities that are most closely associated
with these measures. That is, the positive examples in our
training data comprise posts shared on r/depression for
depression, r/anxiety for anxiety, r/stress for stress, r/psychosis
for psychosis, and r/SuicideWatch for suicidal ideation. On
the other hand, negative examples are extracted from the
collated sample of 20M Reddit posts gathered from 20
subreddits that appear on the landing page of Reddit dur-
ing the same period of our Twitter data sample, such as
r/AskReddit, r/ww, r/movies, and others.

These classifiers are SVM models with linear kernels
and use 5000 n-grams (n=1,2,3) as features. We use a bal-
anced number of examples for the two classes in training,
and we tune the parameters of the classifiers using k-fold
(k=5) cross-validation (Chandrasekharan et al. 2018). Ta-
ble 2 summarizes the size of the datasets and the accuracy
metrics. Figure 3a shows the ROC curves of these classi-
fiers. These classifiers show a mean cross-validation accu-


racy ranging between 0.79 and 0.88 and mean test accuracy
ranging between 0.81 and 0.91. Table 3 reports the top 10
features in each of the classifiers. Several top n-gram fea-
tures such as depression, stress, hope, help, and feel, are con-
textually related to mental health.

Establishing Model Validity. Since our next goal is to em-
ploy these classifiers, trained on Reddit data, to automatic-
ically infer the symptomatic outcomes in the Twitter user
samples—a platform with distinct norms and posting style,
we present a series of evaluation tests to demonstrate the
validity of the transfer approach and the transferred classi-
fiers. 1) First, motivated from prior work (Saha et al. 2017a),
we conduct a linguistic equivalence test between the Red-


Table 2: Mental health classifiers (training:test data size),
cross-validation and test accuracies.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Dataset Size</th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
<th>CV Test</th>
<th>Test CV</th>
<th>Test</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>40,000</td>
<td>0.88</td>
<td>0.82</td>
<td>0.88</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Anxiety</td>
<td>40,000</td>
<td>0.89</td>
<td>0.92</td>
<td>0.88</td>
<td>0.88</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Stress</td>
<td>5,000</td>
<td>0.79</td>
<td>0.79</td>
<td>0.87</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Psychosis</td>
<td>5,000</td>
<td>0.78</td>
<td>0.87</td>
<td>0.81</td>
<td>0.81</td>
<td>0.87</td>
<td>0.81</td>
<td>0.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>concerns</td>
<td>6 depression</td>
</tr>
<tr>
<td>it looks</td>
<td>5 anxiety</td>
</tr>
<tr>
<td>help</td>
<td>4 stress</td>
</tr>
<tr>
<td>song</td>
<td>4 psychosis</td>
</tr>
<tr>
<td>friends</td>
<td>6 help</td>
</tr>
<tr>
<td>well</td>
<td>4 suicide</td>
</tr>
<tr>
<td>you can</td>
<td>3 depression</td>
</tr>
<tr>
<td>life</td>
<td>2 anxiety</td>
</tr>
<tr>
<td>symptoms</td>
<td>6 help</td>
</tr>
<tr>
<td>end</td>
<td>4 people</td>
</tr>
<tr>
<td>feel</td>
<td>3 suicidal</td>
</tr>
<tr>
<td>attack</td>
<td>3 feeling</td>
</tr>
<tr>
<td>need to</td>
<td>2 medication</td>
</tr>
<tr>
<td>weed</td>
<td>2 feel</td>
</tr>
<tr>
<td>around</td>
<td>3 suicidal</td>
</tr>
</tbody>
</table>

Table 3: Top 10 features in the mental health outcome clas-
sifiers and their corresponding weights (w).

greements have been reached between classifier outcomes
and annotations of mental health experts (a Fleiss’ κ=0.84
was reported in Bagroy et al. (2017)).

Matching For Causal Inference

Matching Covariates When conditioned on high-
dimensional covariate data, matching is known to sig-
ificantly minimize bias compared to naive correlational
analyses (Imbens and Rubin 2015). Our approach controls
for a variety of covariates so that the compared Control
and Treatment groups show similar pre-treatment online
behavior. The 1st set of covariates includes users’ social
attributes (count of tweets, followers, followees, duration
on the platform and frequency of posting). The 2nd set

Finally, to minimize the confounding effects of an indi-


444
Propensity Score Analysis  We use matching to find pairs (generalizable to groups) of Treatment and Control users whose covariates are statistically very similar to one another, but where one was treated, and the other was not. The propensity score model matches users based on their likelihood of receiving the treatment, or the propensity scores. Our stratified matching approach groups individuals with similar propensity scores into strata (Kıçıman et al. 2018). Every stratum, therefore, consist of individuals with similar covariates. This helps us to isolate and estimate the effects of the treatment within each stratum.

To compute the propensity scores, we build a logistic regression model that predicts a user’s treatment status based on their covariates. Next, we discard the outliers in the propensity scores (outside the range of 2 standard deviations from the mean), and segregate the remaining distribution into 100 strata of equal width. To further ensure that our causal analysis per stratum remains restricted to a sufficient number of similar users, we remove those strata with very few Treatment or Control users, as is common practice in causal inference research (De Choudhury et al. 2016). With a threshold of at least 50 users per group in a stratum, this approach gave 63 strata that consisted of 23,163 Treatment and 122,941 Control individuals (Figure 3c).

Quality of Matching  To ensure that we matched statistically comparable Treatment and Control users, we evaluate the balance of their covariates. We compute the standardized mean difference (SMD) across covariates in the Treatment and Control groups in each of the 63 valid strata. SMD calculates the difference in the mean covariate values between the two groups as a fraction of the pooled standard deviation of the two groups. Two groups are considered to be balanced if all the covariates reveal SMD lower than 0.2 (Kıçıman et al. 2018), a condition which all our covariates satisfied. We also find a significant drop in the mean SMD from 0.029 (max=0.31) in the unmatched datasets to 0.009 (max=0.05) in the matched datasets (Figure 3d).

Characterizing the Propensity Strata of Users  To understand how the subpopulations across the several strata vary, we characterize their psycholinguistic attributes. Figure 4 plots the usage of affective and cognitive words across all the strata. The propensity score model distributed these users in such a way that the users with a greater tendency to use affective and cognitive words mostly occur in the lower and middle strata, whereas those with a lower tendency to use these words predominantly occur in the higher strata.

Measuring Changes in the Outcomes. To quantify the effects of self-reported psychiatric medication use, we compute the change in the symptomatic outcomes, weighted on the number of Treatment users in each stratum. For this, we first determine the Relative Treatment Effect (RTE) of the drugs per outcome measure in every stratum, as a ratio of the likelihood of an outcome measure in the Treatment group to that in the Control group (Kıçıman et al. 2018). Next, using a weighted average across the strata, we obtain the mean RTE of the medications per outcome measure. We compute the mean RTE for all the drugs and aggregate that for the drug families. An outcome RTE greater than 1 suggests that the outcome increased in the Treatment users, whereas an RTE lower than 1 suggests that it decreased in the Treatment users, following the reported use of psychiatric medication.

Exploring Individual-Specific Effects

We finally aim to study how the drugs affect individuals who vary in their pre-existing psychological state. So once we calculated the treatment effect of the drugs, we explore its relationship with the individuals’ psycholinguistic attributes (as obtained by LIWC). For this, in every stratum, we first build separate linear regression models for all the outcomes of Control users with their covariates as predictors. Using these models we predict the counterfactual outcomes of the Treatment users in the strata – that is, the outcome for each treated user if they had not taken the drug. Next, for every user, we obtain the ratio of the predicted and actual value of the outcome. This essentially quantifies how much a Treatment user is individually effected by treatment, and is referred to as the Individual Treatment Effect (ITE) in individualized and precision medicine literature (Lamont et al. 2018). Finally, we measure the association between pre-treatment psycholinguistic attributes and the ITE values per drug, by fitting a linear regression model. This characterizes the directionality and the effect of a drug on an individual based on their pre-existing psycholinguistic attributes.

Results

Observations about Symptomatic Outcomes

Our first set of results investigates if self-reported psychiatric drug use had a statistically significant effect on the Treatment users. For this, we measure the effect size (Co-
We find that the magnitude of Cohen’s $d$ in the outcome changes between the Treatment and Control users, per drug, per outcome, and per valid strata. We find that $91\%$ of our values fall outside this range, suggesting the Treatment significantly differed from the Control group. An independent sample $t$-test further reveals statistical significance in these differences ($t(\text{df}=9.87, 10.96); p<0.001$), confirming that after the self-reported use of medications, the Treatment users showed significant changes in outcomes.

We then compute the Relative Treatment Effect (RTE) of the psychiatric medications. Figure 5b shows the distribution of RTE across the symptomatic outcomes for the matched Treatment and Control users. We find that the RTE across the outcomes averages at $1.28$ (stdev=$0.61$). We dig deeper into the effects per drug. Figure 6 presents the RTE of the 20 most popular generic drugs and the 4 drug families. We observe many interesting patterns here, such as most medications lead to similar directionality of effects on all the outcomes, e.g., all of the outcomes, depression, anxiety, psychosis, and suicidal ideation increase for the Treatment users in the After period of reported medication use. The similarity in effects across outcomes could be attributed to the comorbidity of the symptomatic outcomes and the clinical presentation of many moods and psychotic disorders (Rosenblat et al. 2016). We also observe that those drugs with similar pharmacological composition, such as Escitalopram and Citalopram, and Desvenlafaxine and Venlafaxine show similar trends in the symptomatic outcomes.

Table 4 summarizes the proportion of Treatment users who showed an increased outcome per drug family. For all these outcomes other than positive affect and cognition (in which case it is the opposite), an increase in the outcome measure also translates to worsened observable mental health condition of the individuals, whereas a decrease suggests an improvement in their mental health condition, as gleaned from Twitter. To study the strata-wise variation for each of these outcomes, we present Figure 7, which shows the RTE per stratum for the four most popular medications.

**Effects on Affect and Cognition** Figure 6 and Table 4 together indicate that the top medications and families are associated with an increase in the likelihood of negative affect. However, that the likelihood of positive affect and cognition also decrease for most of these medications, aligns with literature about the inverse relationship observed in the occurrence of these attributes and mental health symptoms (Pennvakker, Mehl, and Niederhoffer 2003). Among the drug families, we find that the TCAs show the greatest improve-
For these second set of outcomes, we observe requires further investigation, beyond the scope of our work. severity of their mental illness. However, to disentangle that served ineffectiveness of the drugs is likely influenced by the processing. Interestingly, these symptoms are also comor-
dividuals with lower affective expressiveness and cognitive is associated with ineffective (or worsening) effects on in-
s suggest that the self-reported use of these medications are not mutually exclusive. That is, an individual can see both increasing positive affect and increasing negative affect if they are using more affective words overall. The higher strata included users who typically showed lower affect and cognition than the rest (see Figure 4). Together, our findings suggest that the self-reported use of these medications is associated with ineffective (or worsening) effects on individuals with lower affective expressiveness and cognitive processing. Interestingly, these symptoms are also comorbid with mood disorders (Rosenblat et al. 2016), and the observed ineffectiveness of the drugs is likely influenced by the severity of their mental illness. However, to disentangle that requires further investigation, beyond the scope of our work.

Effects on Depression, Anxiety, Psychosis, and Suicidal Ideation For these second set of outcomes, we observe varied changes across medications. We observe that reported use of most of the medications are associated with worsening of these outcomes. These also include the most popular medications such as Sertraline, Escitalopram, and Fluoxetine. All of these are classified as SSRIs—the family which shows the most worsening in these outcomes among the drug families. In fact, our dataset reveals that within SSRIs, over 90% of the users were in strata that showed increased anxiety and suicidal ideation. On the other hand, we find improving symptoms in TCAs such as Dosulepin, Imipramine, and Clomipramine. From the perspective of drug families, the TCAs and the TeCAs show the greatest improving effects, with the majority of their users belonging to strata with decreased effects in the outcome measures.

Although most medications show similar effects at an aggregated level, we find differences in their strata-wise effects distributions (Figure 7). For example, in case of Duloxetine, we find minimal effects in the middle region, the one that showed high cognition (Figure 4). In contrast, Fluoxetine showed improving effects in a few lower valued strata. This observation—that the strata-wise effects can be different, inspired our next set of post-hoc analyses, wherein we examine individual-specific effects and drug-specific changes associated with the reported use of the medications.

Understanding Individual-Treatment Effects To understand how pre-treatment psycholinguistic signals correlate with post-treatment response to the drugs, we examine the effects at the individual level. For every Treatment user, we obtained their Individual Treatment Effect (ITE) values for all outcomes. Next, we fit several linear regression models per psychiatric medication to obtain the relationship between the ITEs and the psycholinguistic (LIWC) attributes of the users who reported using the medication. To simplify interpretability, corresponding to every psycholinguistic attribute, we averaged the coefficients of outcomes (preserving their directionality of improvement). For the four most popular drugs, Table 5 reports the coefficients of five psycholinguistic attributes with the greatest magnitudes in improvement or worsening. We summarize a few distinct patterns below, noted by our clinician coauthor to be most salient, based on the clinical literature and experience:

For Sertraline, the use of first person singular and auxiliary verb shows negative coefficients, indicating that this drug might not be effective in those with greater pre-occupation and self-attentional focus—the known characteristics of these two attribute usage, typically prevalent in depressed individuals (De Choudhury et al. 2013). In contrast, Escitalopram and Duloxetine shows better efficacy in those individuals who have greater pre-occupation and lower social integration. Similarly, Fluoxetine and Duloxetine shows better efficacy in those individuals with greater usage of cognitive words—typically those who show lower cognitive impairment, but Sertraline shows the opposite effect in them.

Discussion Our work presents two significant contributions: 1) By detecting the effects of drug use and that these changes are sensitive to drug families, we show a proof of concept that social media is useful as an effective sensor to scalably detect behavioral changes in individuals who initiate treatment via (self-reported) use of psychiatric medication; and 2) our empirical findings include the discovery that people’s online behaviors change in some unexpected ways following drug intake, and these may differ from the named side-effects of these drugs. We discuss the significance and implications of these contributions in the remainder of this section.

Contextualizing the Findings in Psychiatry As highlighted earlier, there are complexities in determining the effects of psychiatric medications in individuals; but at

<table>
<thead>
<tr>
<th>Attribute Coefficient</th>
<th>Attribute Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sertraline</td>
<td>Fluoxetine</td>
</tr>
<tr>
<td>Past Tense</td>
<td>Cognitive Mech.</td>
</tr>
<tr>
<td>Tentative</td>
<td>Present Tense</td>
</tr>
<tr>
<td>1st P. Singular</td>
<td>Relative</td>
</tr>
<tr>
<td>Aux. Verbs</td>
<td>Percept</td>
</tr>
<tr>
<td>Cognitive Mech.</td>
<td>Conjunction</td>
</tr>
<tr>
<td>Escitalopram</td>
<td>Duloxetine</td>
</tr>
<tr>
<td>Article</td>
<td>Cognitive Mech.</td>
</tr>
<tr>
<td>1st P. Singular</td>
<td>Relative</td>
</tr>
<tr>
<td>Social</td>
<td>1st P. Singular</td>
</tr>
<tr>
<td>Bio</td>
<td>Social</td>
</tr>
<tr>
<td>2nd Person</td>
<td>Work</td>
</tr>
</tbody>
</table>
the same time, there are discrepancies in the claims made by clinical studies. For example, Geddes et al. found no major differences in the efficacy of SSRIs and TCAs, whereas other studies found one kind to perform better than others (Cipriani et al. 2018). Other studies found placebos or non-pharmacological care to have outperformed certain antidepressants (Szegedi et al. 2005). These conflicting findings in the literature prevent us from drawing conclusive claims about the validity of our findings.

From the perspective of clinical literature, our results offer varied interpretations. Figure 6 indicates a small impact of antidepressants on cognitive symptoms—an observation consistent with clinical experience and studies (Rosenblat et al., 2016). It is more difficult to explain the variable impact of the drugs on depressive symptoms. For instance, in our post-hoc analysis, Sertraline showed poor effects for individuals exhibiting attributes of depression, despite clinical evidence suggesting the opposite. On the other hand, Duloxetine was associated with positive symptomatic outcomes, as also found in clinical studies (Cipriani et al. 2018). Nevertheless, that these antidepressants have varying effects on individuals across strata finds support in clinical trials which report varying efficacy of antidepressants on different cohorts (Coupland et al. 2011).

Notwithstanding these varied findings, our work highlights the potential of older antidepressants. While TCAs (Imipramine, Clomipramine) are not often prescribed today because of serious toxicity issues that may be fatal in overdose (Kerr, McGuffie, and Wilkie 2001), our results demonstrate their effectiveness with the most favorable responses reported, compared to the other classes of anti-depressants.

Clinical Implications

Patient-Centered Approach to Pharmacological Care

Our findings show that social media can provide valuable complementary insights into the effects of psychiatric drugs. This can complement clinical trials, allowing observations in larger populations and over longer time spans. Further, in psychiatry, medications are still prescribed by trial-and-error, or based on side effect profiles of these medications (Trivedi et al. 2006). Our analysis of individual treatment effects shows that the pre-treatment signals of mental health states appear to be linked to or predictive of individual drug success, raising the possibility of using such signals for precision psychiatry (Vieta 2015). While we use social media to demonstrate that this relationship exists, other sources of mental health signals may be used to complement our analyses, that are reliable and more broadly available.

Drug Repurposing

Our results offer a novel opportunity to advance drug repurposing. Presently the pipeline for new pharmacological agents for mental illnesses is sparse (Dubovsky 2018), apart from ongoing research on ketamine and other potential new antidepressants (Dubovsky 2018). Drug repurposing—finding new clinical applications for currently approved medications, offering the potential of low cost and quicker to market treatments (Corsello et al. 2017). So far drug repurposing efforts in mental illnesses like depression have focused on biological targets (Powell et al. 2017). Although these approaches have been successful in identifying plausible repositioning candidates, a key challenge is providing direct evidence of candidate efficacy in people, rather than relying on surrogate biomarkers or indirect evidence. This is the first research to explore how social media may serve to identify novel targets as well. Our methods highlight how large quantities of real-time data can offer low cost and high volume assessments of people’s own reports and perceptions related to antidepressants’ use.

Technological Implications

Technologies for Regulatory Bodies

Our results offer an important tool in generating “real-world evidence” for incorporation into technologies that can be used by regulatory bodies like the FDA. The FDA seeks to advance its approach to regulate and rely more on real-world evidence in addition to pre-market clinical studies data. As the FDA currently writes its novel digital health software program certification plan, where medical software such as smartphone applications will receive FDA approval without extensive clinical research—a key component is stated to be “monitoring real-world performance”, though it is to be noted that they are still “considering how to best work to collect and interpret information about the product’s safety and effectiveness” (fda.gov). This paper offers a novel technological approach that may meet the evolving needs of the FDA, by being able to identify the uses and effects of various medications as self-reported by people on social media.

Technologies for Drug Safety Surveillance

From a public health perspective, our methods offer the potential to build technologies that surface early warning signs of adverse effects related to psychiatric drug use. The FDA’s current Sentinel Initiative which aims to apply big data methods to medical claims data from over 5.5 billion patient encounters in an effort to flag previously unrecognized drug safety issues and to tackle issues of under-reporting of drug effects, has not still superseded traditional reporting directly from physicians or pharmaceutical companies (Kuehn 2016). The data gathered in this paper—even though it only represents a subpopulation of those who use social media (Saha et al. 2017b), offers a new lens onto specific groups of people who may have less or more extreme reactions to medications. Including this information in technologies for drug safety monitoring can therefore complement traditional sources, and improve awareness regarding emerging safety issues in a spontaneous fashion—serving as sentinels prompting further exploration in pharmacovigilance research.

Technologies to Support Digital Therapeutics

Psychiatrists’ view and knowledge of a patient’s health is often limited to self-reports and information gathered during in-person therapeutic visits (Vieta 2015). This paper provides a new source of collateral information to support digital therapeutics (Fisher and Appelbaum 2017) and enhance evidence-based, personalized pharmacological treatment. In particular, it reveals the potential to build technologies that augment information seeking practices of clinicians, e.g., with patient consent, clinicians can learn about the effects
and symptomatic expressions shared by patients in the natural course of their lives, and beyond the realms of the therapeutic setting. Further, given the risks posed by prescription drug overdose and abuse (McKenzie and McFarland 2007), increased and finer-grained awareness of the effects of psychiatric medications in specific patients can lead to improved toxicovigilance related interventions.

Policy and Ethics

Despite the potential highlighted above to build novel technologies for regulatory authorities, guidelines on how social media signals should be handled, and their use in the surveillance of the effects of drugs do not yet exist. Although the FDA has released two guidelines on the use of social media for the risk-benefit analysis of prescription drugs (Sarker et al. 2015), they focus on product promotion and “do not establish legally enforceable rights or responsibilities” (FDA 2014). Therefore, the potential (unintended) negative consequences of this work must be considered.

Note that the clinical and technological implications rest upon the names of the medications not being anonymized. We recognize that this surfaces new ethical complexities. For example, while understanding what medications work for which individuals may facilitate “patient-centered” insurance coverage decisions, it can also be (mis)used to decline coverage of specific drugs resulting in “health inequality”. Additionally, patients may blindly adopt these findings creating tension in their therapeutic relationship with their clinicians, causing a decrease in medication adherence. We suggest further research investigating and mitigating such potential unintended consequences of the work.

Limitations and Conclusion

We recognize that this study suffers from limitations, and some of these suggest promising directions for future work. Our results on the varied effects of psychiatric medications are likely to be influenced by selection bias in those who choose to publicly self-report their medication use on social media. This is especially true given the stigma around mental illness (Corrigan 2004), which is a known obstacle to connecting individuals with mental healthcare. We cannot verify if self-reports of medication use corresponded to their actual use (Ernala et al. 2019). Therefore, the users in our data who chose to self-report their medication usage may represent unique populations with lowered inhibitions. Self-report bias further complicates the types of effects that we observed—different individuals respond differently, as shown in our results, however, our observations are limited to only the types of effects that characterize the individuals in our data. For these same reasons of sampling bias, we caution against drawing population-wide generalizations of the effects of psychiatric medication usage.

Despite adopting a causal framework that minimizes confounding effects, we cannot establish true causality, and our results are plausibly influenced by the severity in the clinical condition of the individuals. While we considered many confounders in our propensity score matching approach, there are other latent factors that could impact the effects considered here; e.g., duration, history, dosage, and compliance of using self-reported medications; additional medications or adjuvant treatments one might be using. Further, future work can adopt methods such as location-based filtering to better account for geo-cultural and linguistic confounds. Additionally, self-reporting bias about medications can lead to treatment leakage, where some control individuals may be taking medications, but not mentioning it on Twitter.

Our work is not intended as a replacement for clinical trials. In fact, social media lacks many features that clinical trials possess. First, we do not have the notion of a placebo, used to eliminate the confound that simply the perception of receiving a treatment produces non-specific effects. Second, even though we match users based on several characteristics, we do not pre-qualify individuals as potential beneficiaries of a medication. Last, social media analysis does not allow us to closely monitor the treatment, unlike a clinical trial, which results in high variance in the number of measurements that each individual contributes.

Despite corroboration by a psychiatrist, we are limited by what can be observed from an individual’s social media data. Without complementary offline information (e.g., the people’s physiologies), we cannot ascertain the clinical nature of the mental health outcomes in our data. Further, the symptomatic outcomes themselves, such as measures of depression or suicidal ideation, need additional clinical validation, e.g., based on DSM-5 criteria (APA and others 2013), or the Research Domain Criteria (RDoC) introduced by the National Institutes of Mental Health (Insel et al. 2010). Without dampening the clinical potentials, we caution against making direct clinical inferences. Still, while we acknowledge that the medical community rarely adopts the most innovative approaches for immediate use, this work can inspire replication studies in patient populations.

In conclusion, our work represents a novel dynamic viewpoint onto mental health—limitations notwithstanding, it captures the real-time variation and accounts for dynamic systems theory, network theory, and instability mechanisms (Nelson et al. 2017). Such a new window onto the field clearly contrasts the traditional static viewpoint on the effects of psychiatric medications. It warrants further research in this evolving space and opens up interesting opportunities beyond existing reporting methodologies.

Acknowledgement

We thank the members of the Social Dynamics and Well-being Lab at Georgia Tech for their valuable feedback. Saha and De Choudhury were partly supported by NIH grant #RO1GM112697. Torous was supported by a patient-oriented research career development award (K23) from NIMH #IK23MH116130-01. Abrahao was supported by a National Natural Science Foundation of China (NSFC) grant #61850410536 and developed part of this research while affiliated with Microsoft Research AI, Redmond.

References


De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. In ICWSM.

De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; and Kumar, M. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In CHI.

Dos Reis, V. L., and Culotta, A. 2015. Using matched samples to estimate the effects of exercise on mental health from twitter.


Dutta, S.; Ma, J.; and De Choudhury, M. 2018. Measuring the impact of anxiety on online social interactions. In ICWSM.

Ernala, S. K.; Rizvi, A. F.; Birnbaum, M. L.; Kane, J. M.; and De Choudhury, M. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. CSCW.


Kiciman, E.; Counts, S.; and Gasser, M. 2018. Using longitudinal social media analysis to understand the effects of early college alcohol use. In ICWSM.


Kuehn, B. M. 2016. Fda’s foray into big data still maturing. JAMA.