Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data

Alex B. Fine, Patrick Crutchley, Jenny Blase, Joshua Carroll, & Glen Coppersmith
Qntfy

{alex.fine, patrick, jenny.blase, josh, glen}@qntfy.com

Abstract

Prevailing methods for assessing population-level mental health require costly collection of large samples of data through instruments such as surveys, and are thus slow to reflect current, rapidly changing social conditions. This constrains how easily population-level mental health data can be integrated into health and policy decision-making. Here, we demonstrate that natural language processing applied to publicly-available social media data can provide real-time estimates of psychological distress in the population (specifically, English-speaking Twitter users in the US). We examine population-level changes in linguistic correlates of mental health symptoms in response to the COVID-19 pandemic and to the killing of George Floyd. As a case study, we focus on social media data from healthcare providers, compared to a control sample. Our results provide a concrete demonstration of how the tools of computational social science can be applied to provide real-time or near-real-time insight into the impact of public events on mental health.

1 Introduction

Measurements of the mental health of large populations often become quickly outdated, given traditional techniques for data collection, analysis, and dissemination. For example, estimates of suicide rates in the United States are often delayed by two years (Hedegaard et al., 2018). More up-to-date information about population-level mental health could provide clinicians and other decision-makers with crucial warning signals of shifts in mental health or burgeoning public health crises. Continuous access to sound estimates of population-level mental health variables could also provide a mechanism for evaluating community-level interventions.

The dramatic social upheavals of 2020 provide a visceral illustration of how specific communities are psychologically affected by specific events. For example, the COVID-19 pandemic, which took root in the United States in February and March of 2020, in addition to threatening the health of a broad swath of the population, placed particularly heavy demands on healthcare providers charged with responding to a highly contagious and deadly novel virus, often under resource-constrained circumstances. Anecdotal reports made it clear that the surge in cases–coupled with factors such as underfunded clinics and lack of a coordinated federal response–was leading to acute psychological distress and burnout among healthcare providers such as nurses and physicians. In addition, the killing of George Floyd on May 25, 2020 elicited nationwide responses of grief and anger, and is widely believed to have surfaced latent psychological trauma in large swaths of the American and international population. In both instances, we saw that there was and is no scalable technique for collecting population-scale data to quantify changes in mental health over time, to ask which segments of the population are most severely affected by the situation, or to determine which psychological symptoms are changing in prevalence and therefore what interventions should be prioritized by the community.

Here, we focus on healthcare providers (HCPs) as a case study, and present a framework for monitoring signs of psychological distress in a continuous, scalable, and ethical fashion (Mikal et al.) using public social media data. We use models of anxiety, depression, and suicide risk, trained on a separate data source, to produce longitudinal es-

Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, pages 50–54
Online, November 20, 2020. ©2020 Association for Computational Linguistics
https://doi.org/10.18653/v1/P17

timates of the prevalence of symptoms associated with these conditions among HCPs and a comparison sample.

The model-derived estimates of symptom prevalence show relative changes in mental health aligned with the timing of events related to COVID-19 and the killing of George Floyd among HCPs in the US. For example, we were able to observe the particularly negative impact of the COVID-19 pandemic across the population. Furthermore, we find no evidence that rescinding stay-at-home orders reversed the deleterious effects of the pandemic on mental health, nor do we find evidence that either healthcare workers or the general population had returned to their respective pre-COVID levels of anxiety, depression, and suicide risk at the time of writing.

Moreover, we find evidence that the killing of George Floyd and subsequent civil unrest across the United States had a measurably deleterious effect on all aspects of mental health measured in both the HCP and control populations.

These findings constitute, we believe, a persuasive proof of concept for the use of transparently and ethically collected social media data in providing aggregated, real-time, population-level estimates of emotional and psychological distress, extending the capabilities of what is commonly known as infoveillance (Paul and Dredze, 2011; Eichstaedt et al., 2015; Paparrizos et al., 2016; Eysenbach, 2009). (For a review of different approaches to assessing population-level mental health, see Aoun et al. (2004)) We believe the data collection and modeling techniques reported here can inform and improve public and private efforts to promote population-level mental health.

2 Data

All analyses were performed using public social media data collected from Twitter between January 1 and June 1, 2020. Analyses are based on two groups: healthcare professionals and a community sample group. Healthcare professionals (*HCPs*, $n = 25,040$) are comprised of providers working directly with patients (e.g., nurses, doctors) and those in adjacent roles (e.g., epidemiologists and hospital administrators). Users were geo-located using self-stated location in the user profile, and only US-based users were included in the analysis. In order to determine which individuals in our sample were HCPs, we used techniques modeled on those reported by Beller et al. (2014), who automatically identify profession and other fine-grained social roles on the basis of self-disclosure. Here, we manually constructed a corpus of HCP professional labels (e.g., “physician”, “doctor”, “nurse”, “RN”) and searched for strings containing these labels in contexts demonstrated by Beller et al. to indicate that the author identifies with that role (e.g., “I’m a ___”, “As a ___ I think”). This classification was then manually assessed by human annotators and found to have a 95% true positive rate. The control sample used in these analyses comprise a sample of the general population in the United States (henceforth *Community*, $n = 10,000$) that did not self-identify as HCPs, selected randomly from users for whom geographic data was available (either through a geotagging algorithm or disclosure of their location in their public profile). Users with fewer than 100 posts between the start of the year until the end of May were excluded from the analysis.

3 Methods

We estimate the impact of various national events in 2020 on population mental health. To do so, we compare measures of average anxiety, depression and suicide risk before and after each event. We will refer to the “Pre-Lockdown Baseline” as the time period from January 1-February 29. The national emergency declaration from the White House came on March 13, 2020, and many stay-at-home orders were put in place around that time. We define “Early Lockdown” as March 15 to March 31, as it signifies a time when people were adjusting to the changes induced by the lockdown including job loss, homeschooling, and working from home. We refer to the period of April 15 to April 30 as “Mid Lockdown”.[1] States took a varied approach to lifting lockdown restrictions, and each followed their own timeline. We suspect the lifting of stay-at-home guidance may have impacted people’s mental health, and obtained the state-specific dates on which those orders were lifted. On May 25, George Floyd was killed in police custody, setting off protests and unrest across the United States. We examine one week prior to and after his death (May 18-25; May 26-June 2).

We use classification models, trained on sepa-

[1] These time periods were specified before the analyses reported below. We did not experiment with multiple time windows.

rate data sets from the one described above, to score each Tweet in the sample with an estimate of the probability that Tweet was authored by a person experiencing anxiety or depression or who had attempted suicide. The labels used in the training were derived via self-stated diagnosis: a user was considered to be living with anxiety, depression, or suicidality if they explicitly reported that they had received a diagnosis of an anxiety disorder or depression or had previously attempted suicide, respectively. Examples of self-statements include disclosures such as, "As a person who has been diagnosed with general anxiety disorder, I can tell you...", "today marks one year since I tried to take my own life". Self-statements were found using manually constructed search terms and regular expressions; we then confirmed their plausibility and validity using human annotators with clinical training. Logistic regression with character n-gram features were trained on three separate samples (anxiety, depression, suicide) to distinguish users with a self-stated mental health diagnosis from control users reporting no such diagnoses. We employed the same models reported in our previous work, using the anxiety and depression models from Coppersmith et al. (2015) and the suicide model from Coppersmith et al. (2018). AUC scores for the anxiety, depression, and suicide models were .84, .72, and .73, respectively.

For each measure (anxiety, depression, suicide), we computed the mean of all messages per user per day. Each user is thus represented as the mean of their per-day estimates. This allows for matched-sample t-tests between time periods, and independent t-tests between groups within the same time period. User data was de-identified prior to being submitted to these models, and all statistical analysis was conducted over aggregated user data.

4 Results

Baseline scores for each mental health variable were higher (i.e., more severe) for HCPs than the Community population. This suggests that, prior to COVID-19 lockdowns, HCPs were experiencing anxiety, depression, and suicide risk at higher rates than the general population ($p < 0.001$; note that the figure below, for the sake of comparison, shows by-group z-scores so that this Pre-Lockdown difference is not apparent).

To get a sense of how each event affected each group relative to their Pre-Lockdown baseline, we calculated by-group Z-scores from this baseline. This is illustrated in Figure 1, along with the time periods under consideration.

First, note that every time period after lockdown exhibits higher scores for all mental health conditions we examined. Furthermore, the killing of George Floyd appears to have had a significant effect on mental health across all groups.

Longitudinal changes in depression for HCPs and Community do not differ reliably ($p > 0.1$). HCPs exhibit less change in their anxiety over time compared to Community (though HCPs are still at a higher base-rate of anxiety). Interestingly, HCPs show a larger change in suicide-related risk during Early Lockdown. This disappears in Mid Lockdown and gets closer to returning to baseline rates towards the end of May (note, again, that baseline rates for HCPs remain higher than for Community).

5 Discussion

Real-time information about the population's mental health is critically important, especially in times of crisis. Our work is relevant to government agencies or other organizations with the resources to craft population-scale public health interventions or policy recommendations. The current study provides a proof of concept of how publicly available social media data might be used to assess population-level mental health in a way that could support these organizations.

We hasten to emphasize that this work represents a proof of concept, and raises several questions for future research. First, the population of social media users does not perfectly mirror the general population, and it is plausible that those who do not engage in social media were affected differently by COVID and the killing of George Floyd. We can only speculate about how such a bias might influence our results. Second, we did not correct for population demographic rates in the creation of the community group, but did take care to capture a geographically diverse population.

Finally, in future work we plan to explore how the outputs of the models reported here can be continuously calibrated and refined using psychometrically validated clinical scales of constructs such as anxiety and depression. We take it as uncontroversial that using methods of the general kind employed here to measure phenomena as complex as

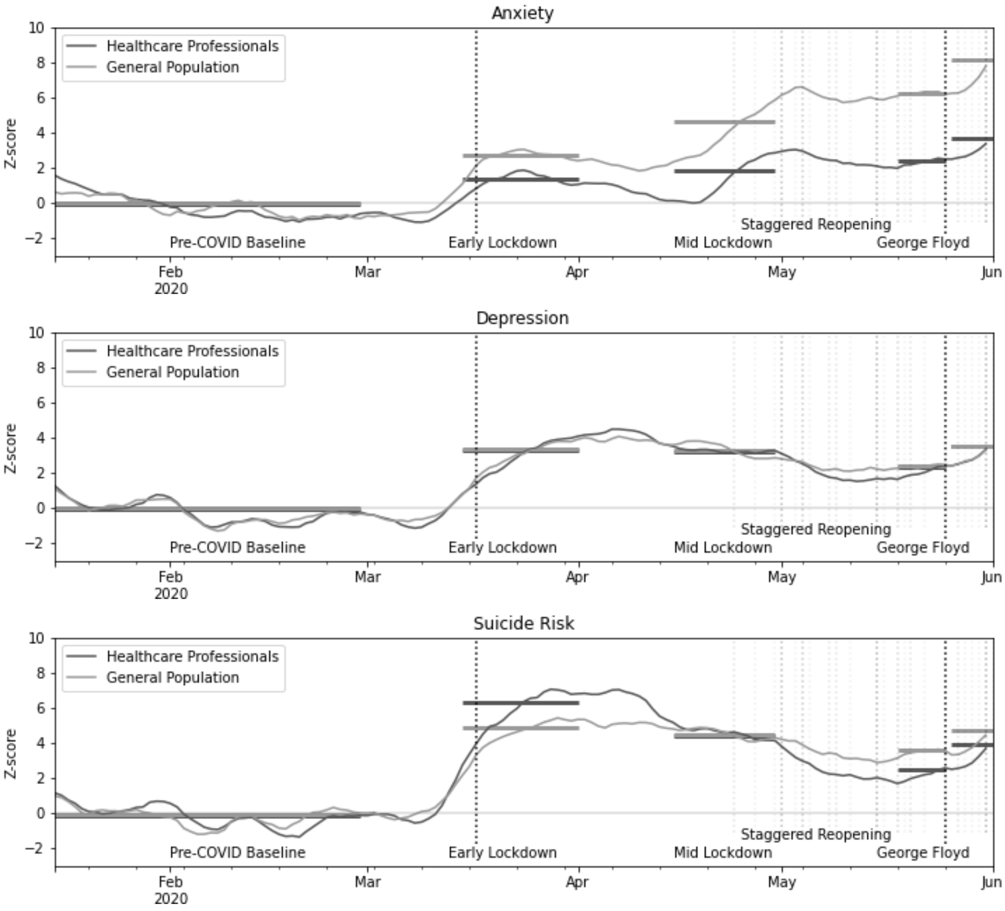

Figure 1: Changes in mental health compared to Pre-Lockdown baseline for HCPs and Community. Y-axis indicates Z-scores compared to each group's Pre-Lockdown baseline; a score of 0 means a return to Pre-Lockdown baseline levels. Time periods for comparison are indicated by thick horizontal bars at the mean for that group across the relevant time period. Significant events are indicated by vertical dotted lines. State reopenings are represented as faded dotted lines.

anxiety, depression, and suicide will demand extensive collaboration and iteration.

6 Conclusion

We have demonstrated the ability to assess population-level mental health constructs in real time, based on publicly available social media data. Quick access to this information could allow lawmakers, mental health practitioners, and others to determine what type of interventions are needed, and where, in the face of rapidly changing conditions. Harnessing this kind of information may be critical to our recovery from COVID-19, and in allowing skillful responses to future crises.

References

S. Aoun, D. Pennebaker, and C. Wood. 2004. Assessing population need for mental health care: A review of approaches and predictors. *Mental Health Serv. Res.*, 6:33–46.

Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell, and Benjamin Van Durme. 2014. I'ma belieber: Social roles via self-identification and conceptual attributes. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 181–186.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, USA. North American Chapter of the Association for Computational Linguistics.

Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.

Johannes C Eichstaedt, Hansen Andrew Schwartz, Margaret L Kern, Gregory Park, Darwin R Labarthe, Raina M Merchant, Sneha Jha, Megha Agrawal, Lukasz A Dziurzynski, Maarten Sap, et al. 2015. Psychological language on twitter predicts county-level heart disease mortality. *Psychological science*, 26(2):159–169.

Gunther Eysenbach. 2009. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of medical Internet research*, 11(1):e11.

Holly Hedegaard, Sally C Curtin, and Margaret Warner. 2018. *Suicide rates in the United States continue to increase*. US Department of Health and Human Services, Centers for Disease Control and . . . .

J. Mikal, S. Hurst, and M. Conway. Ethical issues in using twitter for population-level depression monitoring: a qualitative study. *BMC Medical Ethics*, 17(22).

John Paparrizos, Ryen W. White, and Eric Horvitz. 2016. Screening for pancreatic adenocarcinoma using signals from web search logs: Feasibility study and results. *Journal of Oncology Practice*, 12(8):737–744. PMID: 27271506.

Michael J Paul and Mark Dredze. 2011. You are what you tweet: Analyzing twitter for public health. In *Fifth International AAAI Conference on Weblogs and Social Media*.