The Relationships of Deteriorating Depression and Anxiety with Longitudinal Behavioral Changes in Google and YouTube Usages among College Students in the United States during COVID-19: Observational Study

Boyu Zhang, Anis Zaman, Ehsan Hoque, Vincent Silenzio, Henry Kautz

Submitted to: JMIR Mental Health
on: August 31, 2020

Table of Contents

Original Manuscript 5
Supplementary Files 31
Figures 32
Figure 1 33
Figure 6 34
Figure 7 35
Figure 2 36
Figure 3 37
Figure 4 38
Figure 5 39
Figure 8 40

The Relationships of Deteriorating Depression and Anxiety with Longitudinal Behavioral Changes in Google and YouTube Usages among College Students in the United States during COVID-19: Observational Study

Boyu Zhang[1*] BSc; Anis Zaman[1*] DPhil; Ehsan Hoque[1] DPhil; Vincent Silenzio[2] MD, MPH; Henry Kautz[1] DPhil

[1]Department of Computer Science University of Rochester Rochester US
[2]Department of Urban-Global Public Health Rutgers University Piscataway and Newark US
[*]these authors contributed equally

Corresponding Author:
Boyu Zhang BSc
Department of Computer Science
University of Rochester
2513 Wegmans Hall
Rochester
US

Abstract

Background: Mental health problems among the global population are worsened during the coronavirus disease (COVID-19). Yet, current methods for screening mental health issues rely on in-person interviews, which can be expensive, time-consuming, blocked by social stigmas and quarantines. Meanwhile, how individuals engage with online platforms such as Google Search and YouTube undergoes drastic shifts due to COVID-19 and subsequent lockdowns. Such ubiquitous daily behaviors on online platforms have the potential to capture and correlate with clinically alarming deteriorations in mental health profiles of users in a non-invasive manner.

Objective: The goal of this study is to examine, among college students in the United States, the relationship between deteriorating mental health conditions and changes in user behaviors when engaging with Google Search and YouTube during COVID-19.

Methods: This study recruited a cohort of undergraduate students (N=49) from a U.S. college campus during January 2020 (prior to the pandemic) and measured the anxiety and depression levels of each participant. The anxiety level was assessed via the General Anxiety Disorder-7 (GAD-7). The depression level was assessed via the Patient Health Questionnaire-9 (PHQ-9). This study followed up with the same cohort during May 2020 (during the pandemic), and the anxiety and depression levels were assessed again. The longitudinal Google Search and YouTube history data of all participants were anonymized and collected. From individual-level Google Search and YouTube histories, we developed 5 signals that can quantify shifts in online behaviors during the pandemic. We then assessed the differences between groups with and without deteriorating mental health profiles in terms of these features.

Results: Of the 49 participants, 41% (n=20) of them reported a significant increase (increase in the PHQ-9 score \geq 5) in depression, denoted as DEP; 45% (n=22) of them reported a significant increase (increase in the GAD-7 score \geq 5) in anxiety, denoted as ANX. Of the 5 features proposed to quantify online behavior changes, statistical significances were found between the DEP and non-DEP groups for all of them (P\leq.01, effect sizes \eta_{partial}^2 ranging between 0.130 to 0.320); statistical significances were found between the ANX and non-ANX groups for 4 of them (P\leq.02, effect sizes \eta_{partial}^2 ranging between 0.115 to 0.231). Significant features included late-night online activities, continuous usages and time away from the internet, porn consumptions, and keywords associated with negative emotions, social activities, and personal affairs.

Conclusions: The results suggested strong discrepancies between college student groups with and without deteriorating mental health conditions in terms of behavioral changes in Google Search and YouTube usages during the COVID-19. Though further studies are required, our results demonstrated the feasibility of utilizing pervasive online data to establish non-invasive surveillance systems for mental health conditions that bypasses many disadvantages of existing screening methods.

(JMIR Preprints 31/08/2020:24012)
DOI: https://doi.org/10.2196/preprints.24012

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

The Relationships of Deteriorating Depression and Anxiety with Longitudinal Behavioral Changes in Google and YouTube Usages among College Students in the United States during COVID-19: Observational Study

Abstract

Background: Depression and anxiety disorders among the global population are worsened during the coronavirus disease (COVID-19). Yet, current methods for screening these two issues rely on in-person interviews, which can be expensive, time-consuming, blocked by social stigma and quarantines. Meanwhile, how individuals engage with online platforms such as Google Search and YouTube undergoes drastic shifts due to COVID-19 and subsequent lockdowns. Such ubiquitous daily behaviors on online platforms have the potential to capture and correlate with clinically alarming deteriorations in depression and anxiety profiles of users in a non-invasive manner.

Objective: The goal of this study is to examine, among college students in the United States, the relationships of deteriorating depression and anxiety conditions with the changes in user behaviors when engaging with Google Search and YouTube during COVID-19.

Methods: This study recruited a cohort of undergraduate students (N=49) from a U.S. college campus during January 2020 (prior to the pandemic) and measured the anxiety and depression levels of each participant. The anxiety level was assessed via the General Anxiety Disorder-7 (GAD-7). The depression level was assessed via the Patient Health Questionnaire-9 (PHQ-9). This study followed up with the same cohort during May 2020 (during the pandemic), and the anxiety and depression levels were assessed again. The longitudinal Google Search and YouTube history data of all participants were anonymized and collected. From individual-level Google Search and YouTube histories, we developed 5 features that can quantify shifts in online behaviors during the pandemic. We then assessed the correlations of deteriorating depression and anxiety profiles with each of these features. We finally demonstrated the feasibility of utilizing the proposed features to build predictive machine learning models.

Results: Of the 49 participants, 49% (n=24) of them reported an increase in the PHQ-9 depression scores; 53% (n=26) of them reported an increase in the GAD-7 anxiety scores. The results showed that a number of online behavior features were significantly correlated with deteriorations in the PHQ-9 scores (r ranging between -0.37 and 0.75, $P \leq .03$) and the GAD-7 scores (r ranging between -0.47 and 0.74, $P \leq .03$). Simple machine learning models are shown to be useful in predicting the change in anxiety and depression scores (MSE ranging between 2.37 and 4.22, R^2 ranging between 0.68 and 0.84) with the proposed features.

Conclusions: The results suggested that deteriorating depression and anxiety conditions have strong correlations with behavioral changes in Google Search and YouTube usages during the COVID-19. Though further studies are required, our results demonstrated the feasibility of utilizing pervasive online data to establish non-invasive surveillance systems for mental health conditions that bypasses many disadvantages of existing screening methods.

Keywords: mental health; anxiety; depression; Google Search; YouTube; pandemic; COVID-19

Introduction

Background

Globally, mental health problems such as depression, anxiety, and suicidal ideation are severely worsened during the coronavirus disease (COVID-19) [1–3], specifically for college students [4,5–7]. Yet, current methods for screening mental health issues and identifying vulnerable individuals rely on in-person interviews. Such assessments can be expensive, time-consuming, and blocked by social stigma, not to mention the reluctancy induced by travel restrictions and exposure risks. It has been reported that very few patients in need were correctly identified and received proper mental health treatments on time under the current healthcare system [8,9]. Even with emerging Telehealth technologies and online surveys, the screening requires patients to actively reach out to care providers.

At the same time, because of the lockdown enforced by the global pandemic outbreak, people's engagements with online platforms underwent notable changes, particularly in search engine trends [10–12], exposures to media reports [13,14], and through quotidian smartphone usages for COVID-19 information [5]. Reliance on the internet has significantly increased due to the overnight change in lifestyles, for example, working and remote learning, imposed by the pandemic on society. The sorts of content consumed, the time and duration spent online, and the purpose of online engagements may be influenced by COVID-19. Furthermore, the digital footprints left by online interactions may reveal information about these changes in user behaviors.

Most importantly, such ubiquitous online footprints may provide useful signals of deteriorating mental health profiles, e.g., depression and anxiety, of users during COVID-19. They may capture insights into what was going on in the mind of the user through a non-invasive manner, especially since Google and YouTube Searches are short and succinct and can be quite rich in providing the in the moment cognitive state of a person. On one hand, online engagements can cause fluctuations in mental health. On the other hand, having certain mental health conditions can cause certain types of online behaviors. This opens up possibilities for potential healthcare frameworks that leverage pervasive computing approaches to monitor mental health conditions and deliver interventions on-time. However, the findings of this study do not imply any causal relationship between specific types of online activities and one’s level of anxiety or depression at a given point in time.

Prior Work

Extensive research has been conducted on a population level, correlating mental health problems with user behaviors on social platforms [15,16], especially among young adolescents. Researchers monitored Twitter to understand mental health profiles of the general population such as suicidal ideation [17] and depressions [18]. Similar research has been done with Reddit, where anxiety [19], suicidal ideation [17], and other general disorders were studied [20,21]. Another popular public platform is Facebook, and experiments have been done studying anxiety, depression, body shaming, and stress online [22,23]. In addition, it has been shown that college student communities rely heavily on YouTube for both academic and entertainment purposes [24,25]. Yet, abundant usages may lead to compulsive YouTube engagements [26], and researchers have found that social anxiety is associated with YouTube consumption in a complex way [27].

During COVID-19, multiple studies have reported deteriorating mental health conditions in various communities [1–3,28], such as nation-wise [29,30], across the healthcare industry [31,32], and among existing mental health patients [33]. Recently, it has been shown that greater usages of social media during COVID-19 may induce increasing level of anxiety and depression at both population and individual levels [14,34]. In addition, online behaviors during COVID-19 have been explored, especially for web searches related to the pandemic [10–12] and abnormal TV consumption during the lockdown [13]. Many of the behavioral studies also discussed the effects of online interactions on the spread, misinformation, knowledge, and protective measures of COVID-19, including the roles of YouTube [35–37] and other platforms [38]. [39] investigated hate speech targeting the Chinese and Asian communities on Twitter during COVID-19. Some study in 2009 showed the opposite effect in mental health risk factors: a community-wide crisis may reduce self-harm ideation behaviors [40].

Ubiquitous data has been proved to be useful in detecting mental health conditions. Mobile sensor data, such as GPS logs [41,42], electrodermal activity, sleep behavior, motion, and phone usage patterns [43,44] has been applied in investigating depressive symptoms. [45] found that individual private Google Search histories can be used to detect low self-esteem conditions among college students. [5] examined the longitudinal changes in mental health and smartphone usages through ecological momentary assessments (EMAs) during COVID-19 among college populations. While studies exploring anxiety and depression have been conducted in the past, none of them has leveraged individual-level Google Search and YouTube activities logs to examine the effect of COVID-19 on college students.

Goal of This Study

It has been shown that online platforms preserve useful information about the mental health conditions of users, and COVID-19 is jeopardizing the mental well-being of the global community. Thus, we demonstrate the richness of online engagement logs and how it can be leveraged to uncover alarming mental health conditions during COVID-19. In this study, we aim to examine whether the changes in user behaviors during COVID-19 have a relationship with deteriorating depression and anxiety profiles. We focus on Google Search and YouTube usages, and we investigate if the behavior shifts when engaging with these two platforms signify worsened mental health conditions.

The scope of the study covers undergraduate students in the U.S. We envision this project as a pilot study: it may lay a foundation for mental health surveillance and help delivery frameworks based on pervasive computing and ubiquitous online data. Compared to traditional interviews and surveys, such a non-invasive system may be cheaper, efficient, and avoid being blocked by social stigma while notifying caregivers on-time about individuals at risk.

Methods

Recruitment and Study Design

We recruited a cohort of undergraduate students, all of whom were at least 18 years old and have an active Google account for at least 2 years, from the University of Rochester River Campus, Rochester, NY, U.S.A. Participation was voluntary, and individuals had the option to opt-out of the study at any time, although we did not encounter any such cases. We collected individual-level longitudinal online data (Google Search and YouTube) in the form of private history logs from the participants. For every participant, we measured the depression and anxiety levels via the clinically validated Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7), respectively. Basic demographic information was also recorded. There were in total two rounds of data collection: the first round during January 2020 (prior to the pandemic) and the second round during May 2020 (during the pandemic). During each round, for each participant, the anxiety and depression scores were assessed, and the change in mental health conditions was calculated in the end. The entire individual online history data up until the date of participation was also collected in both rounds from the participants. Figure 3 gives an illustration of the recruitment timeline and two rounds of data collections. All individuals participated in both rounds and were compensated with 10-dollar Amazon gift cards during each round of participation.

Given the sensitivity and proprietary nature of private Google Search and YouTube histories, we leveraged the Google Takeout web interface [46] to share the data with the research team. Prior to any data cleaning and analysis, all sensitive information such as the name, email, phone number, social security number, and credit card information was automatically removed via the Data Loss Prevention (DLP) API [47] of Google Cloud. For online data and survey response storage, we utilized a HIPAA-compliant cloud-based secure storing pipeline. The whole study design, pipelines, and survey measurements involved were similar to our previous setup in [45] and have been approved by the Institutional Review Board (IRB) of the University of Rochester.

To address participation bias, the study had been advertised among the college population via campus wide digital announcements. The text in the study advertisements and consent materials were generic with text such as “help uncover mental health understanding via your online activities.” There was no explicit mention of anxiety or depression in the advertisement. Participation was voluntary with the option to opt out of the study anytime, and their data would not be part of the research study. The intent of the study was clearly explained at the beginning of the recruitment process via one-on-one interviews with the recruiter. We did not have anyone declining to participate or withdrawing in the middle of the study.

Online Data Processing and Feature Extractions

The Google Takeout platform enables users to share the entire private history logs associated with their Google accounts, and as long as the account of the user was logged in, all histories would be recorded regardless of which device the individual was using. Each activity in Google Search and YouTube engagement logs were timestamped, signifying when the activity happened to the precision of seconds. Furthermore, for each Google Search, the history log contained the query text input by the user. It also recorded the URL if the user directly input a website address to the search engine. For each YouTube video watched by the user, the history log contained the URL to the video. If the individual directly searched with keyword(s) on the YouTube platform, the history log also recorded the URL to the search results.

In order to capture the change in online behaviors for the participants, we first introduced a set of features that quantified certain aspects of how individuals interact with Google Search and YouTube. The set of features was calculated for each participant separately. Individual-level behavior changes were then obtained by examining the variations of the feature between January to March 1st, 2020 (a week before the state of emergency in the New York state) and March 28th to the May of 2020 (after the outbreak, following the lockdown and mandated social distancing).

We excluded the online data generated between March 1st, 2020 and March 28th, 2020 to account for any acute or temporal behavior changes concentrated around the initial lockdown or due to adapting to remote work from home. We focused on the persistent and stabilized online behaviors throughout the time after the lockdown. Furthermore, the Spring Break at our institution started from March 7th, and the state of emergency in the New York state was issued at the same day. All students were asked to leave campus starting the Spring Break and complete the rest of the semester remotely.

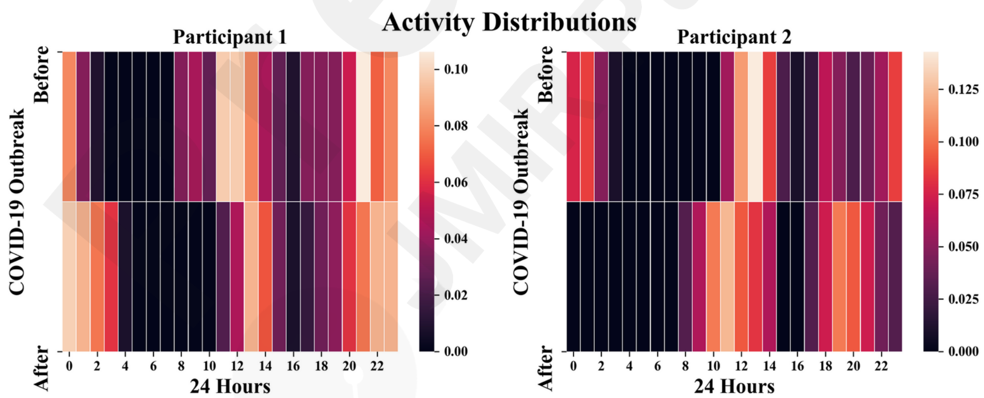
Concretely, we defined 5 features and cut the longitudinal data of each participant into two segments, one from January 1st to February 29th, and the other segment from March 29th to May 31st. Each segment spanned 2 months. We excluded online data from March 1st to March 28th in order to account for the fact that not all individuals were transitioning into work from home on a specific date and practice a strict social distancing lifestyle, although all of our participants are residents of the New York state. The same feature was extracted from both segments of data, and the change was calculated. Such change was referred to as the behavior shifts during the pandemic and lockdown. Figure 3 gives an illustration of data segmentations and feature development pipelines.

Online Activity Distributions

First, we considered, for each participant, how the Google Search and YouTube activities were distributed across the 24 hours of a day before and after the lockdown, given the dates defined above. For each trimmed data segment, we cumulated the total number of activities, regardless of Google Search or YouTube, happened in each of the 24 hours. Thus, we obtained two 24-bin histograms, representing the activity distributions before (D_{before}) and after (D_{after}) the lockdown, respectively.

Figure 1 showcases the normalized distributions before and after the outbreak for two participants, each cumulates 2 months of data. For participant 1 (PHQ-9 increased by 8, and GAD-7 increased by 3), before the outbreak, a few activities started to appear at 8 A.M. After the outbreak, these early morning activities disappeared. In addition, a considerable amount of online activities appeared during late night hours. These patterns most likely indicate a delay in bedtime. For participant 2 (PHQ-9 decreased by 2, and GAD-7 decrease by 6), there were several activities during late night hours before the lockdown. Followed by a long absence from Google Search and YouTube, the next event usually appears around noon. After the lockdown, the first activity of the day started to appear in early morning, and those late-night activities disappeared. Similarly, participant 2 may also have afternoon classes at around 3 to 4 P.M. Notice that these two random cases are simply chosen to represent the fact that study participants reacted non-uniformly to the lockdown.

Figure 1. The normalized activity distributions over 24 hours before and after the outbreak of COVID-19 of two example participants.



After that, for each of the 24 hours *h* of a user, we calculated the percentage, i.e., relative, change of online activities before and after the lockdown:

Equation 1. The percentage change of online activities before and after the COVID-19 outbreak for the hour bin h.

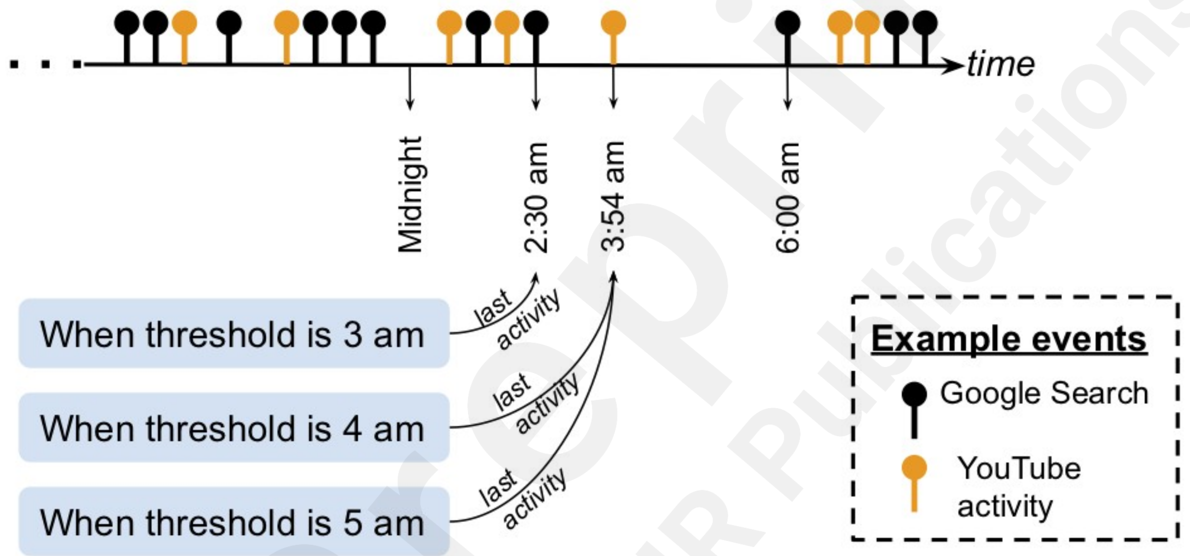
$$\%\text{ change of } D(h) = \frac{D_{after}(h) - D_{before}(h)}{D_{before}(h)} \quad \mathbf{(1)}$$

For the rest of the study, any mentioned percentage or relative changes of features were calculated the same way as above.

Last Seen Activities

We further considered the last seen activity, regardless of Google Search or YouTube, of each user in a day. It is reasonable to assume that, given the nature of our college student population, the last event before they go to bed does not necessarily happen before 12:00 A.M. Strictly speaking, our goal is to capture the last event before they go to bed. Therefore, we set a threshold at late night or early morning and considered the last online activity before it. Since a discrete threshold was used, we tried several cutoff hours to perform sensitivity analyses. For our study population, we observed that the hourly volume of Google Search and YouTube activities started to decrease after 12 A.M., and it reaches the minimum at 5 A.M. This pattern is periodic and persistent across our longitudinal data. Motivated by this observation, we tried a cutoff hour of 0, 1, 2, 3, 4, and 5 A.M. and counted the last events before these thresholds for each participant, respectively. Different from the *Online Activity Distributions* above which measures the volume of activities on Google Search and YouTube hourly, the last seen events focus solely on participant staying up late. An example illustration of the *Last Seen Activities* is provided in Figure 2.

Figure 2. An example to demonstrated how the Last Seen Activities are selected for different threshold hours.



With each threshold, we obtained two distributions of the timestamps of last seen events before and after the lockdown from each participant. On a continuous scale, we then picked the two medians of the last seen event timestamps before and after the lockdown, respectively, and took the difference. For example, a difference of 1.5 hours means that the median time of last seen events shifted 1.5 hours later after the lockdown. A difference of -0.3 hours means the median time of last seen events shifted 0.3 hours earlier after the lockdown. All the time differences are in the unit of hours. There is no need to distinguish between Google Search or YouTube for this feature as we are merely looking for the last event, which could be either.

Short Event Intervals

We defined a short event interval (SEI) as the period of time that is less than a certain threshold, e.g. 5 minutes, between two adjacent events. It usually occurs when one is consuming several related YouTube videos or searching for similar contents in a roll. Taking into consideration that YouTube and Google Search may have different thresholds to define a user session, we adapted the method in [48] to identify proper thresholds for consecutive activities on each of the platforms. After obtaining the session thresholds through mixture models, we counted the total numbers of such short event intervals for each participant before (SEI_{before}) and after (SEI_{after}) the outbreak, respectively. We calculated the relative change of SEI the same way as Equation 1 and used it as a behavioral feature.

LIWC Attributes

The Linguistic Inquiry and Word Count (LIWC) is a toolkit used to analyze various emotions, cognitive processes, social concerns, and psychological dimensions in a given text by counting the numbers of specific words [49]. It has been widely applied in research involving social media and mental health. For the complete list of linguistic and psychological dimensions LIWC measures, see [49(pp3-4)]. We segmented the data log for each participant by the dates mentioned above as two blobs of texts and analyzed the words using LIWC.

Since the contexts and linguistic properties of Google Search and YouTube may be distinct, we extracted the LIWC features from them separately. For Google Search, we input the raw query text; for YouTube, we input the video title and the YouTube query text, if any. There are in total 51 different LIWC attributes. LIWC outputted the count of words falling in each dimension among the whole text. We quantified the shift in behavior by calculating the percentage change of words in each dimension after the outbreak.

Google Search and YouTube Categories

We labeled each Google Search query with a category using the Google NLP API [50]. We utilized the official YouTube API to retrieve the information of videos watched by the participants, including the title, duration, number of likes and dislikes, and default YouTube category tags. For a comprehensive list of Google NLP category labels and default YouTube category tags, please refer to [51,52]. There were several categories overlapping with the LIWC dimensions, such as ‘Health’ and ‘Finance’, and we regarded the LIWC dimensions as a more well-studied standard. Instead, we focused on the number of activities belonging to the ‘Adult’ (specifically originating from Google Search logs) and ‘News’ categories, which were not presented in the LIWC.

Concretely, activities such as visiting a porn site (identified via the URL) and searching explicitly for information related to porn and mature contents are labelled as ‘Adult.’ There is no other ambiguous non-pornographic material being categorized as ‘Adult.’ We used Google Cloud Content Classification API for labeling the search queries and used the Webshrinker [53] API to categorize the domain of every URL an individual visited. We calculated the relative changes of activities in these two categories as the behavior shifts for each participant, the same as Equation 1.

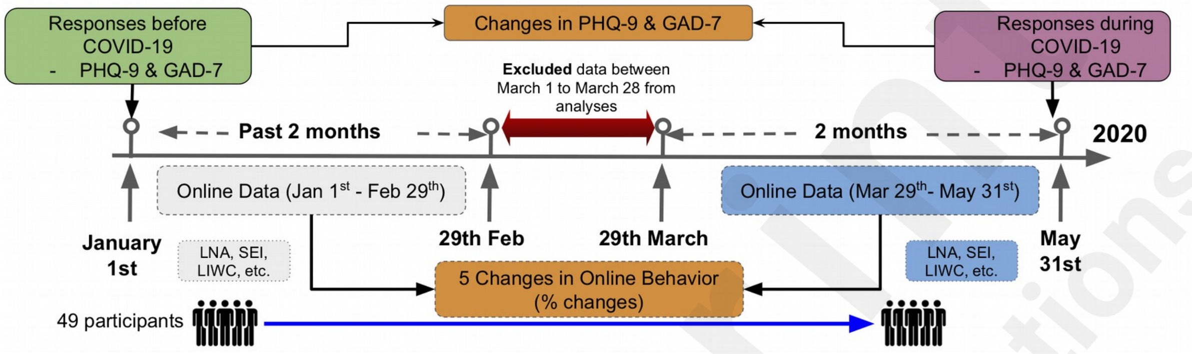
We now present a qualitative example of behavior changes in Google Search and YouTube categories. Table 1 showcases, for a single example participant, the top 5 Google Search and YouTube categories before and after the lockdown, defined by the percentages out of the total activity volume. For Google Search, we observed the disappearances ‘Food & Drinks’ (including searching for restaurants) and ‘Shopping’ from the list. In contrast, the numbers of search related to ‘Beauty & Fitness,’ ‘Home’ (including kitchen and cooking subcategories), and ‘Health’ topics increased during the quarantine. The ‘Reference’ category is largely composed of academic contents such as dictionaries, humanity and history references, and scientific proceedings. For YouTube, videos belong to the ‘Education’ category boosted during the remote learning period after the lockdown, so did ‘Film & Animation.’ ‘Travel & Events’ and ‘Sports’ topics vanished from the list. Notice that this is merely a single example, and the traits reflected here may be personal, uncorrelated to anxiety or depression, or prevalent amongst everyone.

Table 1. The top 5 Google Search and YouTube categories for an example participant before and after the lockdown.

Top 5 Google Search Categories	Top 5 YouTube Video Categories

Before	After	Before	After
Art & Entertainment	Art & Entertainment	Music	Music
Reference	Reference	Travel & Events	Education
Food & Drinks	Beauty & Fitness	Sports	Film & Animation
Shopping	Home	News & Politics	News & Politics
Beauty & Fitness	Health	Education	Pets & Animals
Finance	Food & Drinks	Film & Animation	Comedy

Figure 3. The study recruitment procedure and feature development process. All of the participants moved to remote learning on March 7th, the same day as declaring a state of emergency in the New York State. To avoid any acute behavior during the transition to remote learning, we excluded the data from March 1st to 28th.



Measurement Outcomes

Measurements for Changes in Online Behaviors

There were in total 5 scalar continuous dependent variables measuring various aspects of the changes in online behavior for each participant, as defined above. These variables were extracted from two segments of the online data logs, namely the data before and after the pandemic outbreak. All of the measurements were all in percentage changes. For the *Online Activity Distributions*, there are 24 measurements for each hour of a day. For the *Last Seen Events*, there are several thresholds for sensitivity analyses. For the *Short Event Intervals*, Google Search and YouTube activities are considered separately with their own fitted session intervals.

Measurements for Mental Health Conditions

For both rounds of the data collection, anxiety levels were assessed using the GAD-7 survey, and depression levels were assessed using the PHQ-9 survey. With two rounds of surveys reported before and after the outbreak, the change in mental health conditions of each participant was obtained. According to [54,55], an increase greater than or equals to 5 in the GAD-7 score may be clinically alarming. Similarly, as stated in [56], an increase greater than or equals to 5 in the PHQ-9 score may indicate the need for medical interventions.

Demographics

In addition, the online data and mental health surveys, we also collected basic demographic information such as school year, gender, and nationality.

Statistical Analysis

Before any analysis of mental health conditions, in order to eliminate the possibility of annual confounding factors interfering with the shifts in online behaviors, two-tailed paired independent *t*-tests were performed. We inspected that, in terms of the five quantitative features, whether the online behavior changes happened every year, such as due to seasonal factors, or only during COVID-19 for the whole study population. As mentioned above, we collected the entire Google history log back to the registration date of the Google accounts of all participants.

We now use the example of *Short Event Intervals* to illustrate the idea. For each participant, we obtained 4 *Short Event Intervals* counts from 4 periods of time: 2019 Jan. 1st to 2019 Feb. 28th, 2019 March 29th to 2019 May 31st, 2020 Jan. 1st to 2020 Feb. 29th, and 2020 March 29th to 2020 May 31st. These counts are represented as 4 points on a Cartesian coordinate plane where the y-axis represents the counts, and the x-axis represents the time. We then calculated the slope S of the line connecting the two points from the same year. With the above process, we achieved two measurements for each participant, namely S_{2019} and S_{2020}. Viewing all the participant as a cohort, we computed the S_{2019} and S_{2020} for all features and performed multiple paired t-tests. This enabled us to estimate the seasonal confounding factors. We cannot perform the paired t-tests on the changes of behavioral features directly because it may only validate a change in the intercept (baseline) amount of activity while ignoring the slope for each feature.

In the main experiments, with each of the above 5 features and various thresholds, we investigated the correlation of online behavioral changes with deteriorations in the GAD-7 and PHQ-9 scores, which does not require arbitrary discretization decisions. The dependent variables were the 5 behavior changes extracted from the longitudinal individual online data. Experiments were carried out in a one-on-one fashion: anxiety or depression condition was the single independent variable, and one of the 5 online behavior changes was the single dependent variable each time. Both of them are continuous variables.

Results

Study Population Statistics

We recruited 49 (N=49) participants in total, and all of them participated in both rounds of the study (response rate=100%). On average, each participant made 2,357 (95% CI 2,106.28-2,433.45) Google Searches and 2,901 (95% CI 2,556.92-3,248.67) YouTube interactions from January to February 29th, and 2497 (95% CI 2,069.45-2,901.34) Google Searches and 3105(95% CI 2,702.48-3,487.56) YouTube interactions from March 29th to the end of May. Of the 49 participants, 49% (n=24) of them reported an increase in the PHQ-9 score; 53% (n=26) of them reported an increase in the GAD-7 score. 41% (n=20) of them reported an increase in the PHQ-9 score ≥ 5; 45% (n=22) of them reported an increase in the GAD-7 score ≥ 5.

Figure 4 shows the baseline (collected from Jan. 1st, 2020) and follow-up post-lockdown (collected from May 31st, 2020) distributions of depression and anxiety scores in our sample student population. The PHQ-9 scores are shown on the left, ranging from 0 to 27. The GAD-7 scores are shown on the right, ranging from 0 to 21. Each dot represents a participant. The x-axis represents the baseline score in January, and the y-axis represents the follow-up score during the lockdown in May. Figure 5 shows the distributions of the change in PHQ-9 depression and GAD-7 anxiety scores before and after the lockdown. Again, the PHQ-9 scores are shown on the left, and the GAD-7 scores are shown on the right. The changes were calculated as the follow-up scores in May subtracted by the baseline scores in January. Putting into the context of the pandemic, the deterioration in anxiety or depression levels may be triggered by the fear of getting infected, loss of jobs, the death of family members or friends, and many other negative impacts from the COVID-19 virus. Particularly, for college students, other major reasons may be the pressure of online learning, loss of financial aids, and living alone. In contrast, students underwent quarantines with their families safely may not show signals of deteriorating anxiety or depression, comparing to the high stress levels during normal school days.

Figure 4. The distributions of PHQ-9 depression and GAD-7 anxiety scores before and after the lockdown. The PHQ-9 scores are shown on the left, and the GAD-7 scores are shown on the right.

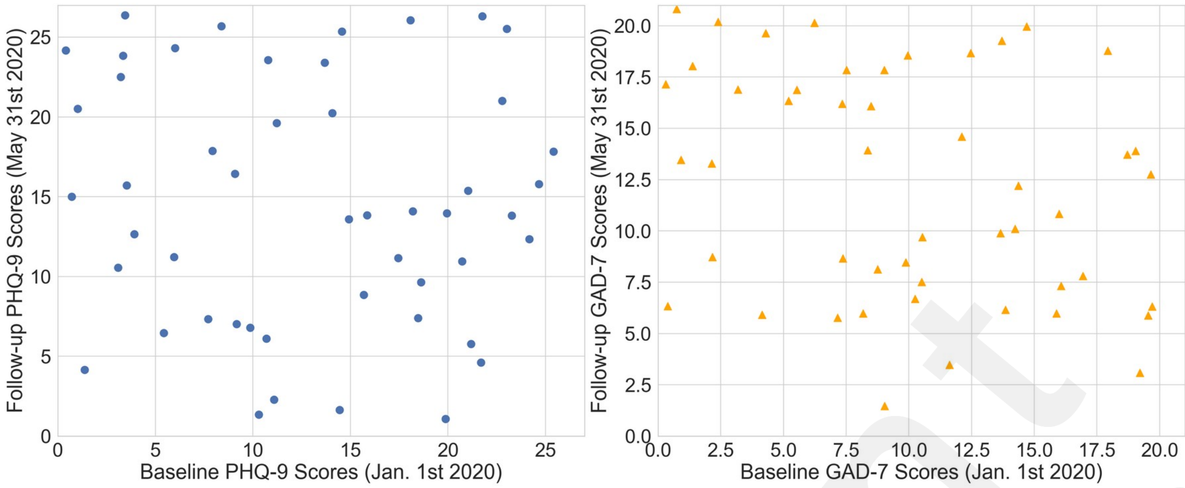
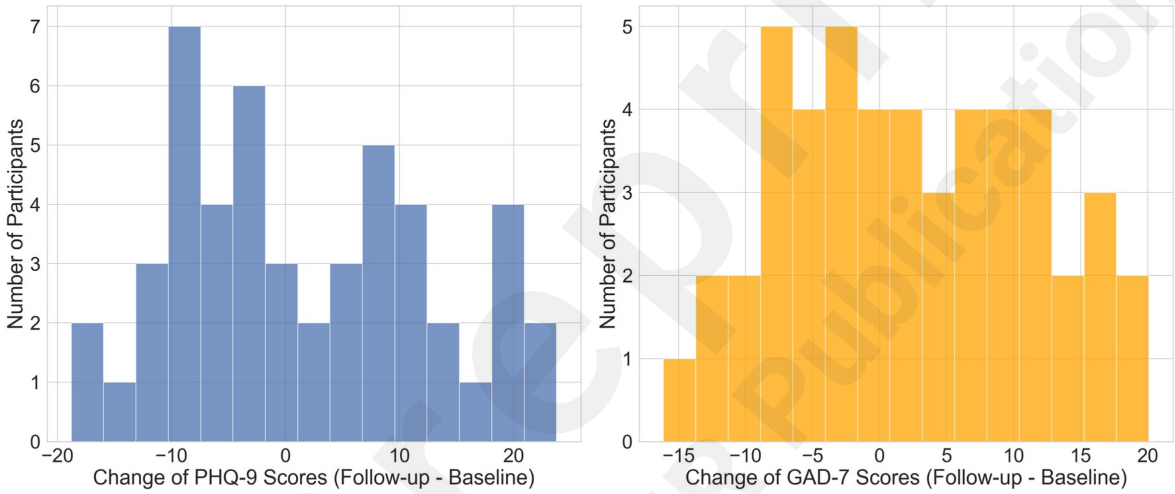


Figure 5. The distributions of the changes in PHQ-9 depression and GAD-7 anxiety scores before and after the lockdown. The PHQ-9 scores are shown on the left, and the GAD-7 scores are shown on the right.



Of the 49 participants (N=49), 61% (n=30) of the them were female; 35% (n=17) of the them were male; the rest 4% (n=2) reported non-binary genders. First, second, third, and forth-year students occupied 22% (n=11), 41% (n=20), 31% (n=15), and 6% (n=3) of the whole cohort, respectively. 80% (n=39) of the participants were U.S. citizens, and the rest (n=10) were international students. A complete breakdown of demographics with respect to the deteriorating anxiety and depressive disorders are given in Table 2.

Table 2. Demographics of the study population.

Demographic	Increased PHQ-9 (n=24)	Increased GAD-7 (n=26)
Female, n (%)	18 (75)	22 (85)
U.S. citizen, n (%)	18 (75)	20 (77)
1st-year students, n (%)	5 (21)	4 (15)
2nd-year students, n (%)	11 (46)	12 (46)
3rd-year students, n (%)	7 (29)	8 (31)
4th-year students, n (%)	1 (4)	2 (8)

Evaluation Outcomes

The two-tailed paired independent *t*-tests mentioned at the beginning of the Statistical Analysis section was designed to rule out seasonal factors in online behavior changes but focus on COVID-19 before any of the main experiments. They reported *P*<.003 for all features. Hence, the presence of annual or seasonal factors accountable for online behavior changes was neglectable, and it was safe to carry out the following main experiment. This is consistent with one of the main conclusions in [5] that, when comparing the longitudinal data between different years, behaviors during COVID-19 shifted drastically.

We calculated the Pearson product-moment correlations *r* of online behavior shifts with deteriorations in anxiety and depression levels, respectively. We reported the correlation coefficients with *P* values and 95% CI obtained for each of the feature proposed above.

Online Activity Distributions

For *Online Activity Distributions*, we calculated the percentage changes in the volume of activities for all 24-hour bins separately. We obtained 24 measurements for each participant. We then evaluated, for each hour of a day, the correlation between the relative hourly activity change and the change in the PHQ-9 depression scores. Significant correlations were found for all hours between 10 P.M. to 6 A.M. of the next day, inclusively. All of them are positive correlations (*r* ranging between 0.32 and 0.75, $P \leq .04$). The correlations started increase after 10 P.M., reached the maximum at 3 A.M., and started to decrease afterwards. It entails that greater late-night online activity volumes after the lockdown may be a signal of deteriorating (increasing) depressive levels.

Next, we observed similar results from the analysis between the hourly activity change and the change in the GAD-7 anxiety scores. Significant correlations were found for all hours between 11 P.M. to 6 A.M. of the next day, inclusively. All of them are positive correlations (*r* ranging between 0.39 and 0.74, $P \leq .006$). The correlations started increase after 11 P.M., reached the maximum at 3 A.M., and started to decrease afterwards. It implies that greater late-night online activity volumes after the lockdown may represent deteriorating (increasing) anxiety levels. For the detailed hourly correlations, 95% CI, and comparisons between the deteriorating PHQ-9 and GAD-7 groups, see Table 3.

Table 3. The correlation coefficients between the change in the volume of online activities of each hour in a day and deteriorating PHQ-9/GAD-7 scores. Significant ones are in bold.

Hour	Deteriorating PHQ-9				Deteriorating GAD-7			
	Correlation Coefficient *r*	*P* value	2.5% CI	97.5% CI	Correlation Coefficient *r*	*P* value	2.5% CI	97.5% CI
12 A.M.	**0.54**	**<.001**	0.38	0.72	**0.45**	**.001**	0.20	0.65
1 A.M.	**0.64**	**<.001**	0.45	0.81	**0.52**	**<.001**	0.28	0.70
2 A.M.	**0.66**	**<.001**	0.47	0.85	**0.60**	**<.001**	0.39	0.75
3 A.M.	**0.75**	**<.001**	0.59	0.94	**0.74**	**<.001**	0.58	0.92
4 A.M.	**0.63**	**<.001**	0.43	0.80	**0.63**	**<.001**	0.43	0.77
5 A.M.	**0.45**	**.001**	0.29	0.65	**0.58**	**<.001**	0.35	0.74
6 A.M.	**0.34**	**.02**	0.21	0.54	**0.39**	**.006**	0.12	0.60
7 A.M.	0.27	.06	0.01	0.54	0.24	.10	-0.05	0.48

8 A.M.	-0.26	.07	-0.50	0.02	-0.28	.05	-0.52	0.00
9 A.M.	-0.28	.05	-0.52	0.00	-0.26	.07	-0.50	0.02
10 A.M.	-0.27	.06	-0.51	0.01	-0.21	.05	-0.46	0.07
11 A.M.	-0.22	.13	-0.47	0.07	-0.25	.08	-0.49	0.03
12 P.M.	-0.19	.19	-0.44	0.09	-0.20	.16	-0.45	0.08
1 P.M.	0.23	.10	-0.05	0.48	-0.10	.51	-0.36	0.48
2 P.M.	0.17	.24	-0.12	0.43	0.17	.24	-0.12	0.19
3 P.M.	0.24	.09	-0.04	0.49	-0.22	.12	-0.47	0.06
4 P.M.	0.18	.23	-0.11	0.43	0.16	.26	-0.12	0.42
5 P.M.	0.15	.31	-0.14	0.41	0.15	.31	-0.14	0.41
6 P.M.	-0.16	.27	-0.42	0.13	-0.15	.30	-0.41	0.13
7 P.M.	-0.12	.39	-0.39	0.16	-0.24	.10	-0.49	0.04
8 P.M.	-0.23	.12	-0.47	0.06	-0.23	.11	-0.47	0.06
9 P.M.	0.25	.08	-0.03	0.50	0.24	.09	-0.04	0.49
10 P.M.	**0.32**	**.04**	0.02	0.55	0.27	.06	-0.01	0.51
11 P.M.	**0.41**	**.003**	0.23	0.62	**0.39**	**.005**	0.13	0.61

Last Seen Activities

For last seen activities, we measured the shift of the median time of last seen activities after the lockdown. Last seen events are determined by the last activity performed before the threshold hour. First, we calculated correlation coefficients between the time shifts and changes in the PHQ-9 depression scores with different cutoff threshold hours. There is a positive correlation overall between the shift of the median time and deteriorating PHQ-9 scores. Thus, the more positive the shift, i.e., the median time of the last events moved to later hours, the greater the deterioration. Specifically, for cutoff hours at 2, 3, 4, and 5 A.M., the correlations were significant (*r* ranging between 0.35 and 0.59, $P \leq .01$). The correlation was strongest for last seen events before 5 A.M. These cutoff values are motivated by the periodic hourly online activity volumes we observed from the study population, as described in the Last Seen Activities section.

Next, we observed similar results when exploring the deterioration in GAD-7 anxiety scores. The positive correlation shows that staying up late tends to signal deteriorations in anxiety levels. For cutoff hours at 2, 3, 4, and 5 A.M., the correlations were significant (*r* ranging between 0.30 and 0.57, $P \leq .03$). The last seen events before 4 A.M. showed the most significant correlation. For the detailed hourly correlations, 95% CI, and comparisons between the deteriorating PHQ-9 and GAD-7 groups, see Table 4.

Table 4. The correlation coefficients between the change in the median time of last seen events and deteriorating PHQ-9/GAD-7 scores at different threshold hours. Significant ones are in bold.

Threshold	Deteriorating PHQ-9				Deteriorating GAD-7			
	Correlation Coefficient *r*	*P* value	2.5% CI	97.5% CI	Correlation Coefficient *r*	*P* value	2.5% CI	97.5% CI
12 A.M.	0.21	.15	-0.07	0.46	0.25	.08	-0.03	0.50
1 A.M.	0.26	.07	-0.02	0.50	0.25	.09	-0.03	0.49
2 A.M.	**0.35**	**.01**	0.08	0.58	**0.30**	**.03**	0.03	0.54

3 A.M.	**0.42**	**.002**	0.15	0.65	**0.45**	**.001**	0.18	0.67
4 A.M.	**0.58**	**<.001**	0.31	0.79	**0.57**	**<.001**	0.30	0.78
5 A.M.	**0.59**	**<.001**	0.33	0.80	**0.56**	**<.001**	0.29	0.78

Short Event Intervals

By considering Google Search and YouTube activities separately, we found different short inter-event session thresholds using the method established in [48]. For Google Search, the boundary between the in-session and between-session mixtures was 1 hour, which is consistent with the finding in the original paper. Thus, we set the threshold of 1 hour and considered all consecutive searches on Google within 1 hour as in the same session. By counting the numbers of adjacent searches within 1 hour before and after the lockdown and calculating the percentage change, we did not find significant correlations between the number of short Google Search intervals and deteriorating PHQ-9 depression scores (*r*=-0.22, *P*=.06, 95% CI -0.43-0.01). Nor did we find significant correlations between the number of short Google Search intervals and deteriorating GAD-7 anxiety scores (*r*=-0.21, *P*=.08, 95% CI -0.40-0.02).

In contrast, we found that the threshold for the in-session and between-session mixtures for YouTube watching histories was 3.2 minutes, much shorter than that of Google Search. The average YouTube video interval time was 21 minutes. This threshold indicated that two adjacent videos consumed within 3.2 minutes of idle time shall be considered as in the same session, i.e., consecutive consumption. We then calculated the relative change of the number of such short YouTube intervals after the lockdown. We found a significant positive correlation between the increase of YouTube short intervals and deteriorating PHQ-9 scores (*r*=0.57, $P \leq .001$, 95% CI 0.38-0.76). Similarly, a significant positive correlation was found between the increase of YouTube short intervals and deteriorating GAD-7 scores (*r*=0.41, *P*=.001, 95% CI 0.20-0.62).

In addition, we include visualizations of the behavioral measurements for *Short Event Intervals* over time. As mentioned in the Measurements for Mental Health Conditions section, an increase ≥ 5 in PHQ-9 or GAD-7 score is clinically alarming. Thus, we first separated the samples by groups with and without an increase ≥ 5 in the PHQ-9 depression scores. We then plot the 7-day moving average total number of short YouTube event intervals, i.e., consecutive video consumption, of the two groups as a function of dates. We overlay the series with the activity data from the same time period (January 1st to May 31st) in 2019 in dashed lines for contrast. As shown in Figure 6, after the lockdown in mid-March, both groups are having increasing amounts of short YouTube intervals. The participants with severe deteriorating depression outrun others significantly. A similar trend was found when we separated the groups by an increase ≥ 5 in the GAD-7 anxiety scores, shown in Figure 7. The group with significantly deteriorated anxiety disorders tends to have higher numbers of consecutive YouTube consumption. The shaded area represents 1 standard deviation. Most importantly, these patterns are stabilized throughout the time after the lockdown, reflecting a meaningful behavioral shift instead of mere acute or temporal observations. No such phenomenon was observed in 2019 for any group, and we further argue the fact that seasonal factors are accountable for the behavioral difference.

Figure 6. The 7-day moving average time series of the total amount of short YouTube activity intervals between groups with and without significant increases in the PHQ-9 depression scores.

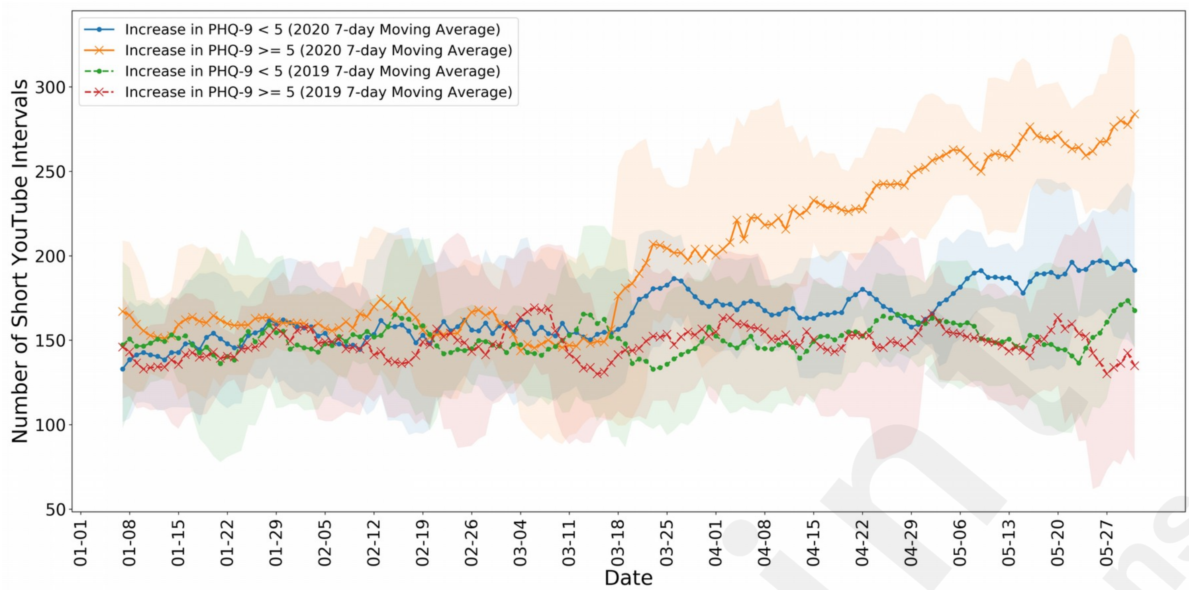
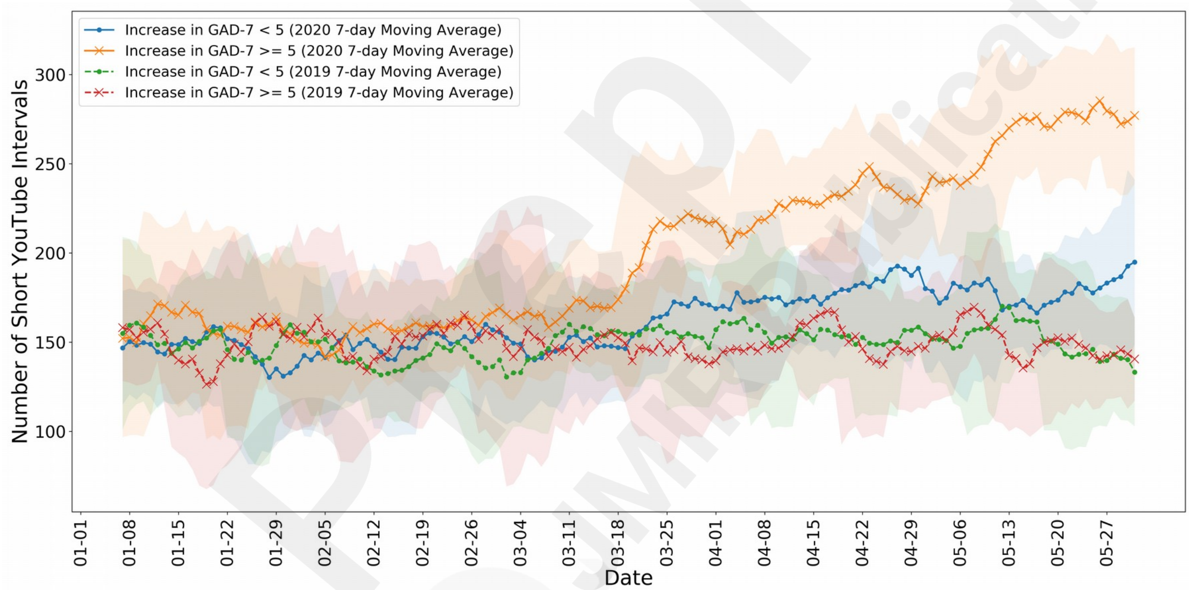


Figure 7. The 7-day moving average time series of the total amount of short YouTube activity intervals between groups with and without significant increases in the GAD-7 anxiety scores.



LIWC Attributes

As the contexts and linguistic properties of Google Search and YouTube may be different, we extracted the LIWC features from them separately. For Google Search, we found that search queries in the 'Work,' 'Money,' and 'Death' categories under the 'Personal Concerns' dimension showed significant positive correlation with deteriorating PHQ-9 depression scores (*r* ranging between 0.33 and 0.59, *P*≤.02). Similar results were found for deteriorating GAD-7 anxiety scores (*r* ranging between 0.41 and 0.51, *P*≤.003).

Moreover, for both depression (*r*=0.31, *P*=.03, 95% CI 0.08-0.49) and anxiety (*r*=0.34, *P*=.02, 95% CI 0.10-0.52) deteriorations, significant correlations were found in the 'Authentic' scores of the search queries, under the 'Summary Variable.' The 'Authentic' scores were developed by [57] to provide a continuous scale for measuring how honest and genuine a given piece of text is. The higher the score, the more authentic the language. It was shown that when individuals address themselves in an authentic manner, they tend to be personal and vulnerable. For instance, a randomly chosen participant searched: "What not to do during at-home exams under proctor surveillance?" For the detailed correlations coefficients and 95% CI, see Table 5.

Table 5. The significant LIWC features from Google Search histories and their correlation coefficients with deteriorating PHQ-9 and GAD-7 scores.

LIWC Attributes	Deteriorating PHQ-9				Deteriorating GAD-7			
	r	*P* value	2.5% CI	97.5% CI	*r*	*P* value	2.5% CI	97.5% CI
Personal Concerns								
Work	**0.33**	**.02**	0.09	0.53	**0.43**	**.002**	0.22	0.60
Money	**0.47**	**.001**	0.28	0.60	**0.51**	**<.001**	0.33	0.63
Death	**0.59**	**<.001**	0.42	0.71	**0.41**	**.003**	0.21	0.56
Summary Variable								
Authentic	**0.31**	**.03**	0.08	0.49	**0.34**	**.02**	0.10	0.52

For YouTube histories, we found that videos containing 'Anxiety' and 'Sadness' keywords under the 'Negative Emotion' dimension showed significant positive correlation with deteriorating PHQ-9 depression scores (*r* ranging between 0.50 and 0.52, $P \leq .001$). Similar results were found for deteriorating GAD-7 anxiety scores (*r* ranging between 0.55 and 0.57, $P \leq .001$). In addition, the 'Friends' keywords from the 'Social Words' dimension showed a significant positive correlation with worsened GAD-7 (*r*=0.41, *P*=.003, 95% CI 0.21-0.56) but not PHQ-9 (*r*=0.27, *P*=.06, 95% CI 0.02-0.47).

Moreover, for both depression (*r*=-0.37, *P*=.009, 95% CI -0.16 - -0.52) and anxiety (*r*=-0.47, *P*=.001, 95% CI -0.28 - -0.60) deteriorations, significant negative correlations were found in the 'Emotional Tone' scores of the videos, under the 'Summary Variable.' The 'Emotional Tone' scores were developed by [58] to provide a continuous scale for measuring the positivity of a given piece of text, and *vice versa*. The higher the score, the more positive the text. For the detailed correlations coefficients and 95% CI, see Table 6.

Table 6. The significant LIWC features from YouTube videos and their correlation coefficients with deteriorating PHQ-9 and GAD-7 scores.

LIWC Attributes	Deteriorating PHQ-9				Deteriorating GAD-7			
	r	*P* value	2.5% CI	97.5% CI	*r*	*P* value	2.5% CI	97.5% CI
Negative Emotion								
Anxiety	**0.52**	**<.001**	0.35	0.64	**0.57**	**<.001**	0.42	0.66
Sadness	**0.50**	**<.001**	0.31	0.63	**0.55**	**<.001**	0.38	0.67
Social Words								

	Friends	**0.42**	**.002**	0.22	0.57	0.27	.06	0.02	0.47
Summary Variable									
	Emotional Tone	**-0.37**	**.009**	-0.16	-0.52	**-0.47**	**.001**	-0.28	-0.60

Google Search and YouTube Categories

The 'Adult' category consists of explicit browser histories of porn-related contents and visiting porn sites (identified via the URL). On one hand, the percentage change of the 'Adult' contents showed a significant positive correlation with deteriorating depression levels (r=0.56, P≤.001, 95% CI 0.35-0.77). On the other hand, the percentage change of the 'Adult' contents did not show a significant correlation with deteriorating anxiety levels (r=0.29, P=.08, 95% CI 0.02-0.56). The 'News' content did not show any significant correlation with deteriorating depression (r=0.25, P=.13, 95% CI 0.04-0.46) nor anxiety (r=0.14, P=.21, 95% CI -0.01-0.29).

Predictive Modeling

In this section, we probe the feasibility of utilizing common machine learning models to predict the change, e.g., deterioration, in PHQ-9 and GAD-7 scores. We frame the task as a supervised regression problem and treat the above behavioral changes as features input to the model. The goal is to predict the change in the PHQ-9 or GAD-7 score. Given our small sample size, we evaluate the model performance by multiple (N=49) leave-one-out train (n=48) and test (n=1) splits. For each of the splits, we tune model hyper-parameters with another complete leave-one-out cross-validation on the 48 training samples.

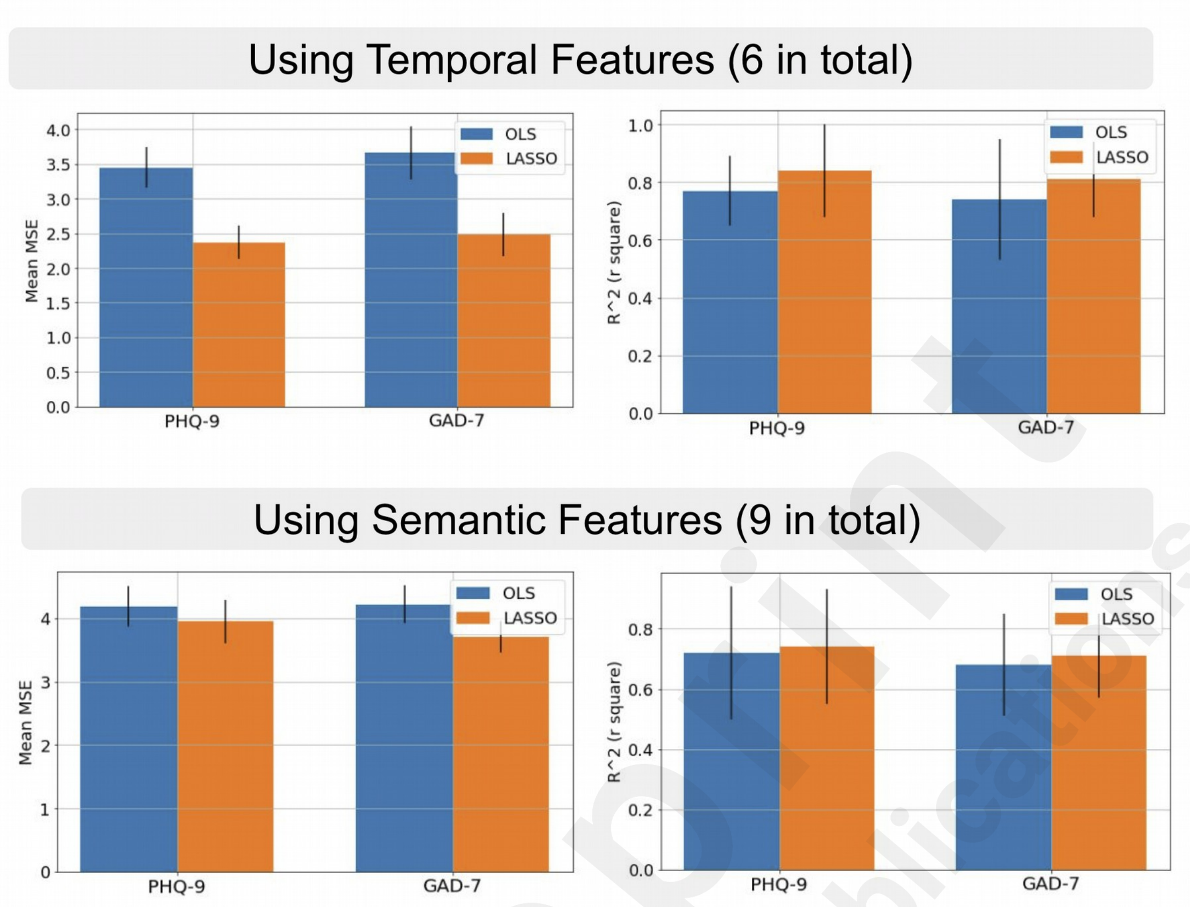
Feature Vectors

We developed the feature vector by concatenating some of the significant behavioral shifts (scalars) above. Nonetheless, given the small sample size, it is reasonable to avoid high-dimensional feature vectors. Thus, we first separated the aforementioned 5 behavioral changes into 2 groups, namely the *temporal* and *semantic* features. Specifically, we picked the *Online Activity Distributions* at 2, 3, and 4 A.M., the *Last Seen Activities* before 4 and 5 A.M., the *Short YouTube Intervals* for the *temporal* feature vector, and hence it has a dimensionality of 6. We then concatenated the 4 significant LIWC categories for Google Search, the 4 significant LIWC categories for YouTube, and the 'Adult' Google Search contents as the *semantic* feature vector. The dimensionality is 9.

Regression Models

We experimented with 2 of the most common linear models: Ordinary Least Square Regression (OLS) and Lasso Regression. In Figure 8, we presented the performances of the regression models. We reported the mean squared error (MSE) and the average coefficient of determination (R^2). The mean and standard deviations were calculated from the 49 leave-one-out splits. Overall, *temporal* features performed better than *semantic* ones, regardless of predicting the change in PHQ-9 or GAD-7. The best average performance in predicting the change of PHQ-9 was achieved by the *temporal* features (MSE=2.37, R^2=0.84). The best average performance in predicting the change of GAD-7 was also achieved by the *temporal* features (MSE=2.48, R^2=0.81).

Figure 8. The performances of the two regression models averaged across 49 leave-one-out splits.



Model Introspection

Below, we introspect the coefficient weights from the linear models for PHQ-9 and GAD-7 predictions. Since Lasso regression performed the best in all cases, we look at the weights from the fitted Lasso models. For the 6 temporal features, as shown in Table 7, *Online Activity Distributions* at 4 A.M. has the most significant importance in both prediction tasks. For the 9 semantic features, as shown in Table 8, keys words related to the 'Anxiety' dimension in the LIWC from YouTube histories have the most significant importance. Moreover, as the above statistical tests have shown that neither the 'Friends' dimension in LIWC nor the 'Adult' *Google Search Categories* was significant for deteriorating GAD-7, the Lasso regression indeed assigned zeros weights to these two features. We used an alpha value of 0.0001 for the regularization term in the Lasso model.

Table 7. The coefficients of 6 temporal features in Lasso regression.

Temporal Features	Coefficient in PHQ-9 Prediction	Coefficient in GAD-7 Prediction
Online Activity Distributions		
2 A.M.	0.08	0.11
3 A.M.	0.34	0.28
4 A.M.	0.68	0.71
Last Seen Activities		
4 A.M.	0.21	0.03

	5 A.M.	0.64	0.54
Short YouTube Intervals		0.00	0.02

Table 8. The coefficients of 9 semantic features in Lasso regression.

Semantic Features		Coefficient in PHQ-9 Prediction	Coefficient in GAD-7 Prediction
LIWC			
	Anxiety	0.63	0.52
	Sadness	0.48	0.47
	Friends	0.02	0.00
	Emotional Tone	-0.23	-0.19
	Work	0.04	0.13
	Money	0.18	0.26
	Death	0.01	0.07
	Authentic	0.02	0.15
Google Search Categories			
	Adult	0.05	0.00

Discussion

In this study, we collected longitudinal individual-level Google Search and YouTube data from college students, and we measured their anxiety (GAD-7) and depression (PHQ-9) levels before and after the outbreak of COVID-19. We then developed explainable features from Google Search and YouTube logs and quantified various online behavior shifts of the participants during the pandemic. We also calculated the change in mental health conditions for all participants. Our experiment examined the correlations of online behavior features with deteriorating anxiety and depression levels, respectively. We finally demonstrated the feasibility of building simple predictive machine learning models with the proposed behavioral signals. To the best of our knowledge, we are the first to conduct observational studies on how anxiety and depression problems and Google Search and YouTube usages of college students are related during COVID-19.

Principal Results

Our results showed that online behavior changes have significant correlations with worsened depression and anxiety profiles during the pandemic. The features we developed based on online activities were all explainable and preserved certain levels of interpretability. For example, the *Short Event Intervals* and *Online Activity Distributions* measured the consecutive usages and hourly activity volumes of Google Search and YouTube, which were inspired by previous studies on excessive YouTube usages [[26]], internet addictions [[59]], and positive associations with social anxiety among college students [[27]]. Our results indicated that individuals with increasing anxiety or depressive disorders during the pandemic tended to have long usage sessions (multiple consecutive activities with short time intervals) when engaging with Google Search and YouTube.

Moreover, the increasing activities during late nigh hours in *Online Activity Distributions* and the positive shifts of medians of *Last Seen Events* corresponded with previous studies in sleep deprivation and subsequent positive correlations with mental health deteriorations [[60,61]]. Our results demonstrated that individuals with worsened anxiety or depressive symptoms during the pandemic were indeed likely to stay up late and engage more online. The above three features captured the temporal aspects of user online behaviors, and they generated best performance in the regression tasks.

Additionally, our analysis found that the amount of porn consumption has significant correlations with deteriorating depression, which adheres to previous findings that people suffering from depression and loneliness are likely to consume excessive pornographies [62,63]. For the LIWC features, participants with significant increases in anxiety watched more videos with ‘Anxiety,’ ‘Sadness,’ and negative tone, and previous research showed that negative YouTube videos tended to receive more attention from vulnerable individuals [64]. They also consumed more videos related social activities and ‘Friends’ keywords. On contrary, the ‘Friends’ keywords did show any significance in the depression analysis. This was consistent with studies on patterns of social withdrawal and depression [42,65,66], and social interactions and isolations have been recognized by [67] as one of the priorities in mental illness prevention, especially during COVID-19 [30]. For Google Search, both the participants with significant increases in anxiety/depression searched more contents related to ‘Work,’ ‘Money,’ and ‘Death’, focusing on real-life practices. None of the emotional dimensions was significant in Google Search logs. Instead, LIWC considered the search queries from depressed and anxious individuals more honest and vulnerable, e.g., asking for help, after the lockdown, given the ‘Authentic’ score. While prior research has shown that individuals living with depression tend to use more first-person languages [68], we did not observe any similar pattern. This is probably due to the fact that search queries are more succinct, imperative, and functional, which leaves less necessity for personal references. These attributes captured the semantic aspect of user online behaviors. The prevalence of personal affair, social activity, and negative keywords as well as porn consumption has shown statistically significant correlations.

Many researchers have reported that there has been a significant boost in health and news-related topics, at the population level, in various online platforms during COVID-19. This is partly due to additional measures taken by individuals, various stakeholders, and agencies with regards to preventive measures [11,36,37], daily statistics [10,12,13], and healthcare (mis)information [35,37,38], However, unlike many, our investigation was carried out considering individual-level Google Search and YouTube engagement logs, and our analysis did not reveal any significant spikes in ‘News’ and ‘Health/illness’ categories among individuals with deteriorating anxiety and depression during the pandemic. One possible explanation for such observation can be due to the target population (college students) of our study who may prefer to follow news from other popular platforms such as social media.

Finally, COVID-19 has shaken the foundation of human society and forced us to alter daily lifestyles. The world was not ready for such a viral outbreak. Since there is no cure for COVID-19, it, or an even more deadly viral disease, may resurface at different capacities in the near future. Society may be forced to rely on technologies even more and employ remote learning, working, and socializing for a longer period of time. It is important that we learn from our experience of living through the initial COVID-19 outbreak and take necessary measures to uncover the changes in online behaviors, investigating how that can be leveraged to understand and monitor various mental health conditions of individuals in the least invasive manner. Furthermore, we hope our work paves the path for technology stakeholders to consider incorporating various mental health assessment monitoring systems using user engagements, following users’ consent in a privacy-preserving manner. They can periodically share the mental health monitoring assessment report with respective users based on their online activities, education, and informing users about their current mental health. This can eventually encourage individuals to acknowledge the importance of mental health and take better care of themselves.

Limitations

First, while most of the online behavioral features we developed showed significant correlations, our study cohort only represented a small portion of the whole population suffering from mental health difficulties. Therefore, further studies are required to investigate if the significant behavioral changes still hold among more general communities, not limiting to college students. It is possible that the relationship of worsened anxiety and depression with online activities on Google Search and YouTube of our college population is different from that of other populations whose education and social backgrounds alters a lot. There may also be differences in mental health for sub-student populations, such as those living under harmful environments and those depending on financial aids, all of whom may suffer more from physical and economic crisis during COVID-19. Nonetheless, we argue that the explainable features we constructed, such as late-night activities, continuous usages, inactivity, pornography, and certain keywords, can remain behaviorally representative and be applied universally across experiments exploring the relationship of anxiety and depression with online activities during the pandemic.

Second, in this work, we studied the relationship between user online behaviors and the fluctuations in anxiety and depression conditions during COVID-19. Any causal relationship between online behavior and mental disorders is beyond the scope of this work. As one can readily imagine, online behavioral changes could both contribute to or be caused by deteriorating anxiety or depressive disorders.

We acknowledge that it may be impossible to obtain data without noise because one may seldom, or even never, search on Google or watch YouTube videos. Such concealed information makes it impossible for the proposed model to flag alarming symptoms that are reflected in the PHQ-9 and GAD-7 questionnaires.

Moreover, though we included preliminary demographic information as covariates, there remains the possibility of other confounding factors. In fact, both the shifts in online behaviors and deteriorating mental health profiles may be due to common factors such as living conditions, financial difficulties, and other health problems during the pandemic. Nor there was any causal direction implied between COVID-19 and online behavior changes, which was introduced in the first paragraph of Statistical Analysis as a precaution before the main experiments.

Ethical and Privacy Concerns

Albeit a pilot study, our results indicated that it is possible to build an anxiety and depression surveillance system based on passively collected private Google data histories during COVID-19. Such non-invasive systems shall be subject to rigorous data security and anonymity checks. Necessary measures need to be in place to ensure personal safety and privacy concerns when collecting sensitive and proprietary data such as Google Search logs and YouTube histories. Even in pilot studies, participants shall preserve full rights over their data: they may choose to opt-out of the study at any stage and remove any data shared in the system.

Moreover, anonymity and systematic bias elimination shall be enforced. As an automatic medical screening system based on pervasive data, it has been extensively studied that such frameworks are prone to implicit machine learning bias during data collection or training phases [69–71]. Black-box methods should be avoided as they are known to be vulnerable to adversarial attacks and produce unexplainable distributional representations [72,73]. Anonymizing data and obscuring identity information should be the first step in data debiasing.

In the end, to what extent should caregivers trust a clinical decision made by machines remains an open question. We believe that possible pervasive computing frameworks shall play the role of a smart assistant, at most, to the care providers. Any final intervention or help delivery decision should be made by healthcare professionals who understand both the mental health problems and the limitations of automatic detection systems in clinical settings.

Acknowledgments

This research was supported in part by grant W911NF-15-1-0542 and W911NF-19-1-0029 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO). We acknowledge the contributions by Michael Giardino, Adira Blumenthal, and Ariel Hirschhorn at the beginning phase of the project.

Conflicts of Interest

None declared.

Abbreviations

COVID-19: coronavirus disease
EMA: ecological momentary assessment
LIWC: Linguistic Inquiry and Word Count
SEI: Short event interval

References

1. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. *Int J Soc Psychiatry*. 2020;66(4):317-320. doi:10.1177/0020764020915212
2. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. *New England Journal of Medicine*. 2020;0(0):null. doi:10.1056/NEJMp2008017
3. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. *Asian Journal of Psychiatry*. 2020;52:102066. doi:10.1016/j.ajp.2020.102066
4. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLOS ONE*. 2020;15(7):e0236337. doi:10.1371/journal.pone.0236337
5. Huckins JF, DaSilva AW, Wang W, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. *Journal of Medical Internet Research*. 2020;22(6):e20185. doi:10.2196/20185
6. Liu X, Liu J, Zhong X. *Psychological State of College Students During COVID-19 Epidemic*. Social Science Research Network; 2020. doi:10.2139/ssrn.3552814
7. W C, Z F, G H, et al. The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Res*. 2020;287:112934-112934. doi:10.1016/j.psychres.2020.112934
8. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV Disorders in the National Comorbidity Survey Replication (NCS-R). *Arch Gen Psychiatry*. 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617
9. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):603-613. doi:10.1001/archpsyc.62.6.603
10. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. *JMIR Public*

Health and Surveillance. 2020;6(2):e18941. doi:10.2196/18941
11. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell–related searches during the COVID-19 outbreak. *International Forum of Allergy & Rhinology*. 2020;10(7):839-847. doi:10.1002/alr.22580
12. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. *JMIR Public Health and Surveillance*. 2020;6(2):e19374. doi:10.2196/19374
13. Dixit A, Marthoenis M, Arafat SMY, Sharma P, Kar SK. Binge watching behavior during COVID 19 pandemic: A cross-sectional, cross-national online survey. *Psychiatry Res*. 2020;289:113089. doi:10.1016/j.psychres.2020.113089
14. Gao J, Zheng P, Jia Y, et al. Mental health problems and social media exposure during COVID-19 outbreak. *PLOS ONE*. 2020;15(4):e0231924. doi:10.1371/journal.pone.0231924
15. Seabrook EM, Kern ML, Rickard NS. Social Networking Sites, Depression, and Anxiety: A Systematic Review. *JMIR Mental Health*. 2016;3(4):e50. doi:10.2196/mental.5842
16. Lin L yi, Sidani JE, Shensa A, et al. Association Between Social Media Use and Depression Among U.s. Young Adults. *Depression and Anxiety*. 2016;33(4):323-331. doi:10.1002/da.22466
17. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. *Proc SIGCHI Conf Hum Factor Comput Syst*. 2016;2016:2098-2110. doi:10.1145/2858036.2858207
18. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: *Proceedings of the 5th Annual ACM Web Science Conference*. WebSci '13. Association for Computing Machinery; 2013:47–56. doi:10.1145/2464464.2464480
19. Shen JH, Rudzicz F. Detecting Anxiety through Reddit. In: *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics; 2017:58–65. doi:10.18653/v1/W17-3107
20. Silveira Fraga B, Couto da Silva AP, Murai F. Online Social Networks in Health Care: A Study of Mental Disorders on Reddit. In: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. ; 2018:568-573. doi:10.1109/WI.2018.00-36
21. Gaur M, Kursuncu U, Alambo A, et al. "Let Me Tell You About Your Mental Health!": Contextualized Classification of Reddit Posts to DSM-5 for Web-based Intervention. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. CIKM '18. Association for Computing Machinery; 2018:753–762. doi:10.1145/3269206.3271732
22. Frost RL, Rickwood DJ. A systematic review of the mental health outcomes associated with Facebook use. *Computers in Human Behavior*. 2017;76:576-600. doi:10.1016/j.chb.2017.08.001
23. Zhang R. The stress-buffering effect of self-disclosure on Facebook: An examination of stressful life events, social support, and mental health among college students. *Computers in Human Behavior*. 2017;75:527-537. doi:10.1016/j.chb.2017.05.043
24. Fleck BKB, Beckman LM, Sterns JL, Hussey HD. YouTube in the Classroom: Helpful Tips and Student Perceptions. *Journal of Effective Teaching*. 2014;14(3):21-37.
25. Moghavvemi S, Sulaiman A, Jaafar NI, Kasem N. Social media as a complementary learning tool for teaching and learning: The case of youtube. *The International Journal of Management Education*. 2018;16(1):37-42. doi:10.1016/j.ijme.2017.12.001
26. Klobas JE, McGill TJ, Moghavvemi S, Paramanathan T. Compulsive YouTube usage: A comparison of use motivation and personality effects. *Computers in Human Behavior*. 2018;87:129-139. doi:10.1016/j.chb.2018.05.038
27. de Bérail P, Guillon M, Bungener C. The relations between YouTube addiction, social

anxiety and parasocial relationships with YouTubers: A moderated-mediation model based on a cognitive-behavioral framework. *Computers in Human Behavior*. 2019;99:190-204. doi:10.1016/j.chb.2019.05.007
28. Gunnell D, Appleby L, Arensman E, et al. Suicide risk and prevention during the COVID-19 pandemic. *The Lancet Psychiatry*. 2020;7(6):468-471. doi:10.1016/S2215-0366(20)30171-1
29. Agteren J van, Bartholomaeus J, Fassnacht DB, et al. Using Internet-Based Psychological Measurement to Capture the Deteriorating Community Mental Health Profile During COVID-19: Observational Study. *JMIR Mental Health*. 2020;7(6):e20696. doi:10.2196/20696
30. Holmes EA, O'Connor RC, Perry VH, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry*. 2020;7(6):547-560. doi:10.1016/S2215-0366(20)30168-1
31. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. *BMJ*. 2020;368. doi:10.1136/bmj.m1211
32. Chen Q, Liang M, Li Y, et al. Mental health care for medical staff in China during the COVID-19 outbreak. *The Lancet Psychiatry*. 2020;7(4):e15-e16. doi:10.1016/S2215-0366(20)30078-X
33. Yao H, Chen J-H, Xu Y-F. Patients with mental health disorders in the COVID-19 epidemic. *The Lancet Psychiatry*. 2020;7(4):e21. doi:10.1016/S2215-0366(20)30090-0
34. Ni MY, Yang L, Leung CMC, et al. Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China: Cross-Sectional Survey. *JMIR Ment Health*. 2020;7(5):e19009. doi:10.2196/19009
35. Li HO-Y, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? *BMJ Global Health*. 2020;5(5):e002604. doi:10.1136/bmjgh-2020-002604
36. Basch CE, Basch CH, Hillyer GC, Jaime C. The Role of YouTube and the Entertainment Industry in Saving Lives by Educating and Mobilizing the Public to Adopt Behaviors for Community Mitigation of COVID-19: Successive Sampling Design Study. *JMIR Public Health and Surveillance*. 2020;6(2):e19145. doi:10.2196/19145
37. Basch CH, Hillyer GC, Meleo-Erwin ZC, Jaime C, Mohlman J, Basch CE. Preventive Behaviors Conveyed on YouTube to Mitigate Transmission of COVID-19: Cross-Sectional Study. *JMIR Public Health and Surveillance*. 2020;6(2):e18807. doi:10.2196/18807
38. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol Sci*. 2020;31(7):770-780. doi:10.1177/0956797620939054
39. Lyu H, Chen L, Wang Y, Luo J. Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19. *IEEE Transactions on Big Data*. Published online 2020:1-1. doi:10.1109/TBDATA.2020.2996401
40. Gordon KH, Bresin K, Dombeck J, Routledge C, Wonderlich JA. The impact of the 2009 Red River Flood on interpersonal risk factors for suicide. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*. 2011;32(1):52-55. doi:10.1027/0227-5910/a000051
41. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*. 2016;4:e2537. doi:10.7717/peerj.2537
42. Saeb S, Zhang M, Karr CJ, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J Med Internet Res*. 2015;17(7):e175. doi:10.2196/jmir.4273

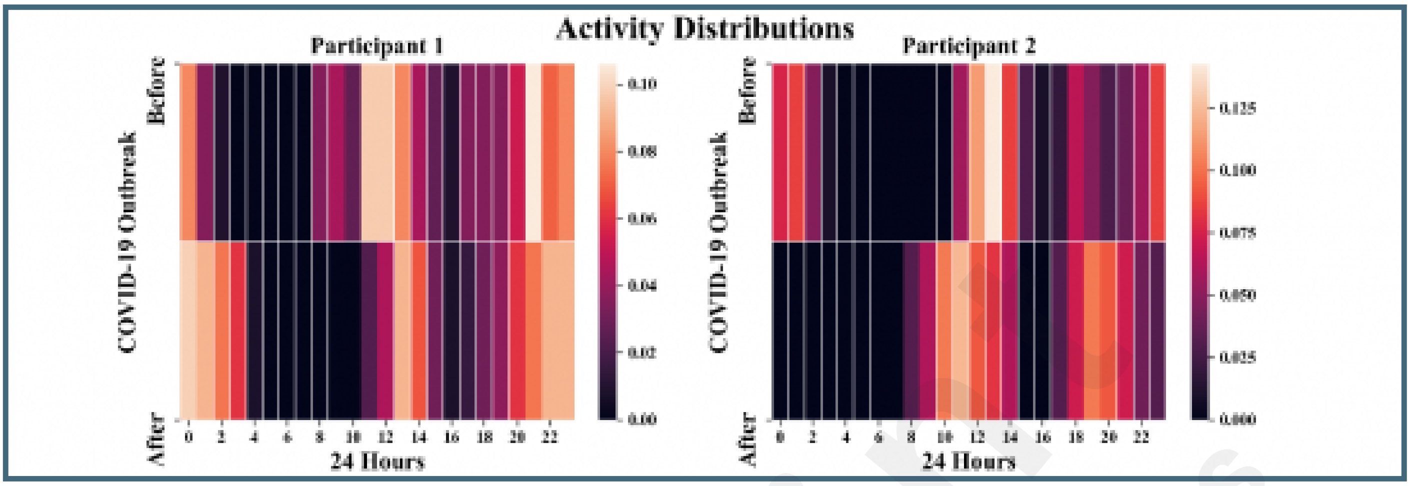
43. Ghandeharioun A, Fedor S, Sangermano L, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. ; 2017:325-332. doi:10.1109/ACII.2017.8273620
44. Wang R, Wang W, daSilva A, et al. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2018;2(1):43:1–43:26. doi:10.1145/3191775
45. Zaman A, Acharyya R, Kautz H, Silenzio V. Detecting Low Self-Esteem in Youths from Web Search Data. In: *The World Wide Web Conference on - WWW '19*. ACM Press; 2019:2270-2280. doi:10.1145/3308558.3313557
46. Google. Google Takeout. Accessed August 7, 2020. https://takeout.google.com/
47. Google. Cloud Data Loss Prevention. Accessed August 7, 2020. https://cloud.google.com/dlp
48. Halfaker A, Keyes O, Kluver D, et al. User Session Identification Based on Strong Regularities in Inter-activity Time. In: *Proceedings of the 24th International Conference on World Wide Web*. WWW '15. International World Wide Web Conferences Steering Committee; 2015:410–418. doi:10.1145/2736277.2741117
49. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Published online September 15, 2015. Accessed August 2, 2020. https://repositories.lib.utexas.edu/handle/2152/31333
50. Google. Classifying Content. Accessed August 7, 2020. https://cloud.google.com/natural-language/docs/classifying-text
51. Google. Content Categories. Accessed August 7, 2020. https://cloud.google.com/natural-language/docs/categories
52. Google. VideoCategories. Accessed August 7, 2020. https://developers.google.com/youtube/v3/docs/videoCategories
53. Webshrinker. Webshrinker API. Webshrinker. Accessed October 16, 2020. https://www.webshrinker.com/apis/
54. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092
55. Rutter LA, Brown TA. Psychometric Properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in Outpatients with Anxiety and Mood Disorders. *J Psychopathol Behav Assess*. 2017;39(1):140-146. doi:10.1007/s10862-016-9571-9
56. Kroenke K. Enhancing the clinical utility of depression screening. *CMAJ*. 2012;184(3):281-282. doi:10.1503/cmaj.112004
57. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying Words: Predicting Deception from Linguistic Styles. *Pers Soc Psychol Bull*. 2003;29(5):665-675. doi:10.1177/0146167203029005010
58. Cohn MA, Mehl MR, Pennebaker JW. Linguistic Markers of Psychological Change Surrounding September 11, 2001: *Psychological Science*. Published online May 6, 2016. Accessed October 18, 2020. https://journals.sagepub.com/doi/10.1111/j.0956-7976.2004.00741.x
59. Kuss DJ, Griffiths MD, Binder JF. Internet addiction in students: Prevalence and risk factors. *Computers in Human Behavior*. 2013;29(3):959-966. doi:10.1016/j.chb.2012.12.024
60. Tarokh L, Saletin JM, Carskadon MA. Sleep in adolescence: physiology, cognition and mental health. *Neurosci Biobehav Rev*. 2016;70:182-188. doi:10.1016/j.neubiorev.2016.08.008
61. Konjarski M, Murray G, Lee VV, Jackson ML. Reciprocal relationships between daily sleep and mood: A systematic review of naturalistic prospective studies. *Sleep Med Rev*. 2018;42:47-

58. doi:10.1016/j.smrv.2018.05.005
62. Willoughby BJ, Busby DM, Young-Petersen B. Understanding Associations between Personal Definitions of Pornography, Using Pornography, and Depression. *Sex Res Soc Policy*. 2019;16(3):342-356. doi:10.1007/s13178-018-0345-x
63. YODER VC, III TBV, AMIN K. Internet Pornography and Loneliness: An Association? *Sexual Addiction & Compulsivity*. 2005;12(1):19-44. doi:10.1080/10720160590933653
64. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a Source of Information on Immunization: A Content Analysis. *JAMA*. 2007;298(21):2482-2484. doi:10.1001/jama.298.21.2482
65. Girard JM, Cohn JF, Mahoor MH, Mavadati SM, Hammal Z, Rosenwald DP. Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image and Vision Computing*. 2014;32(10):641-647. doi:10.1016/j.imavis.2013.12.007
66. Katz SJ, Conway CC, Hammen CL, Brennan PA, Najman JM. Childhood Social Withdrawal, Interpersonal Impairment, and Young Adult Depression: A Mediational Model. *J Abnorm Child Psychol*. 2011;39(8):1227. doi:10.1007/s10802-011-9537-z
67. Leigh-Hunt N, Bagguley D, Bash K, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public Health*. 2017;152:157-171. doi:10.1016/j.puhe.2017.07.035
68. Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting Suicidal Ideation in Chinese Microblogs with Psychological Lexicons. In: *Proceedings of the 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. UIC-ATC-SCALCOM '14. IEEE Computer Society; 2014:844–849. doi:10.1109/UIC-ATC-ScalCom.2014.48
69. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231-237. doi:10.1136/bmjqs-2018-008370
70. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med*. 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763
71. Wang W, Kiik M, Peek N, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLOS ONE*. 2020;15(6):e0234722. doi:10.1371/journal.pone.0234722
72. Kim H, Jung DC, Choi BW. Exploiting the Vulnerability of Deep Learning-Based Artificial Intelligence Models in Medical Imaging: Adversarial Attacks. *J Korean Soc Radiol*. 2019;80(2):259. doi:10.3348/jksr.2019.80.2.259
73. Ma X, Niu Y, Gu L, et al. Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*. Published online May 1, 2020:107332. doi:10.1016/j.patcog.2020.107332

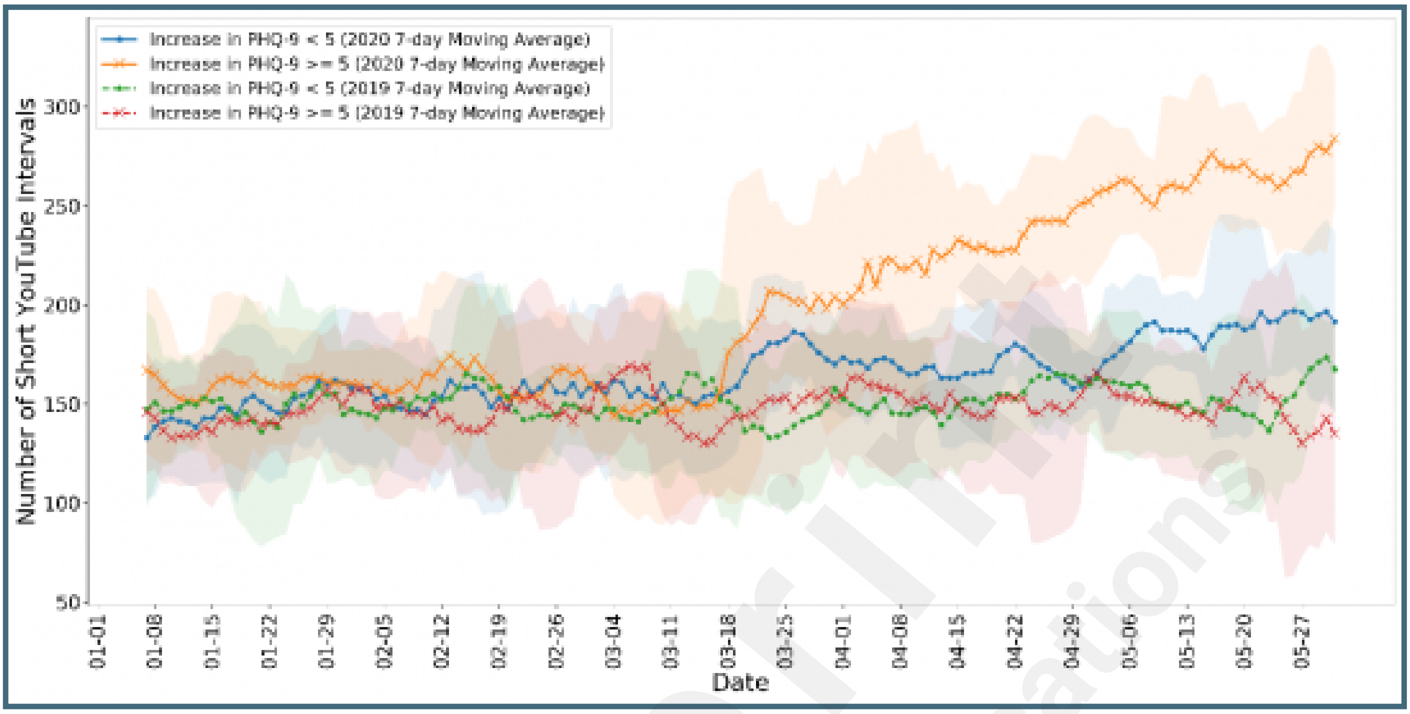
Supplementary Files

Figures

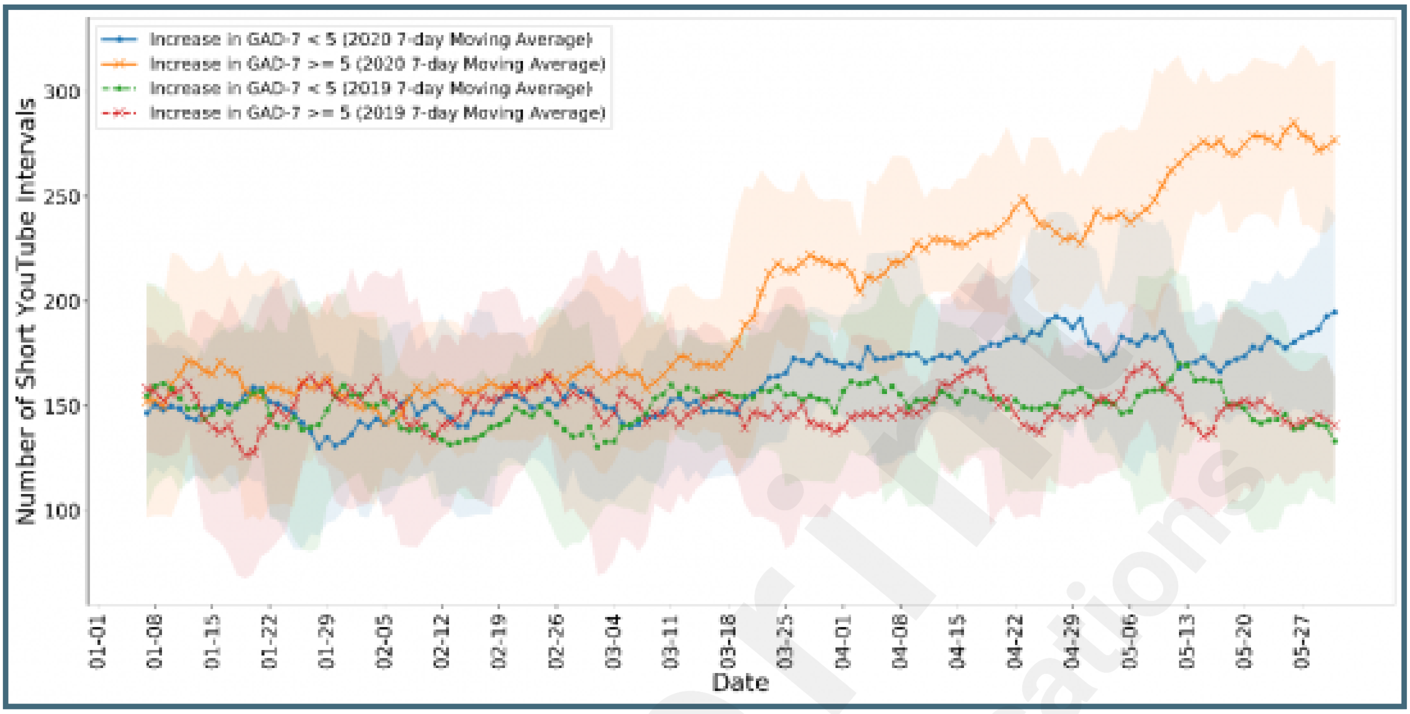
The normalized activity distributions over 24 hours before and after the outbreak of COVID-19 of two example participants.



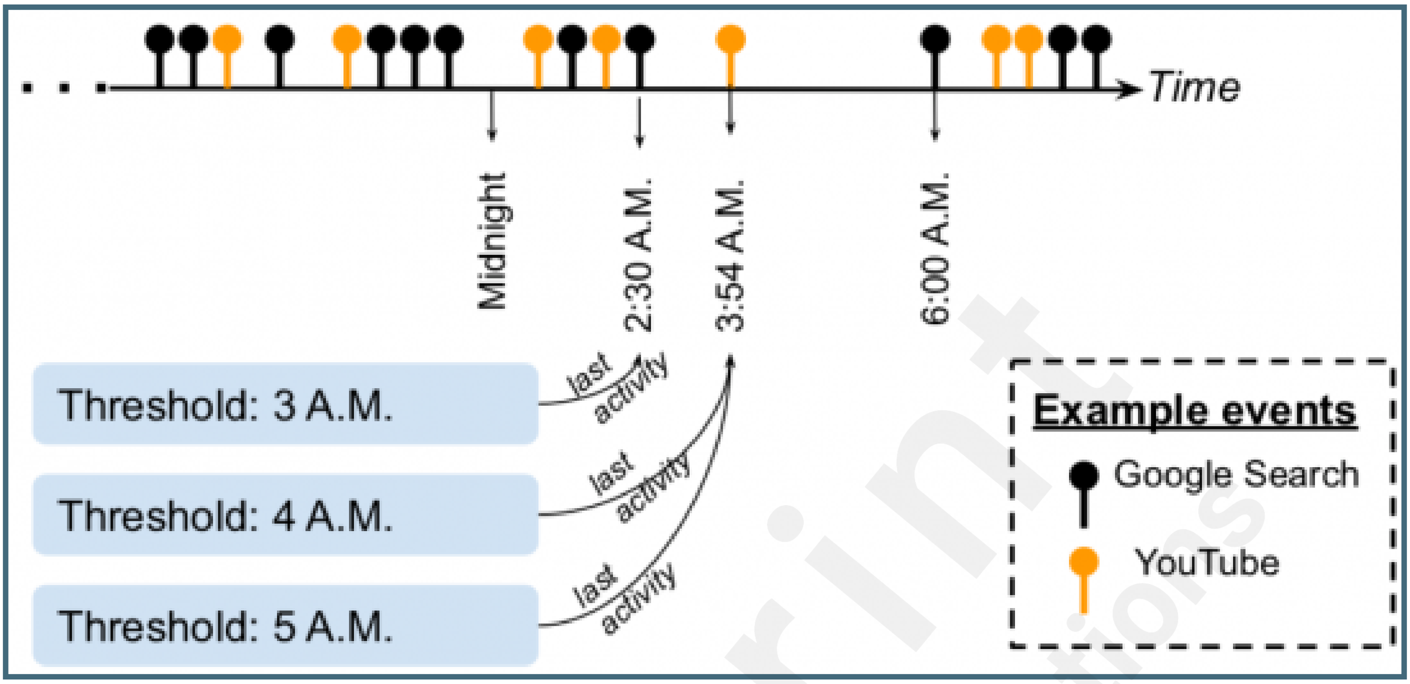
The 7-day moving average time series of the total amount of short YouTube activity intervals between groups with and without significant increases in the PHQ-9 depression scores.



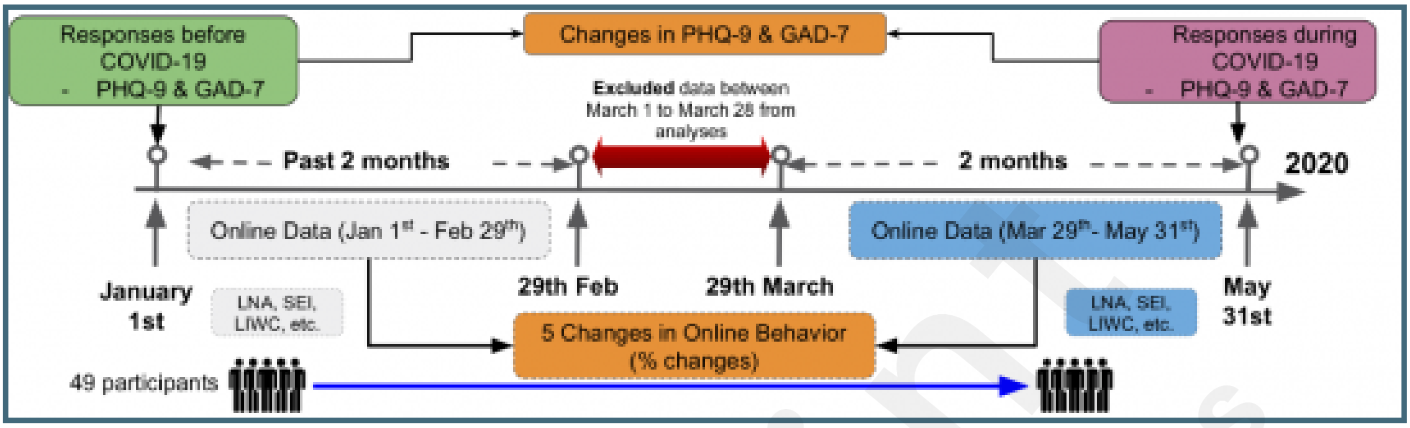
The 7-day moving average time series of the total amount of short YouTube activity intervals between groups with and without significant increases in the GAD-7 anxiety scores.



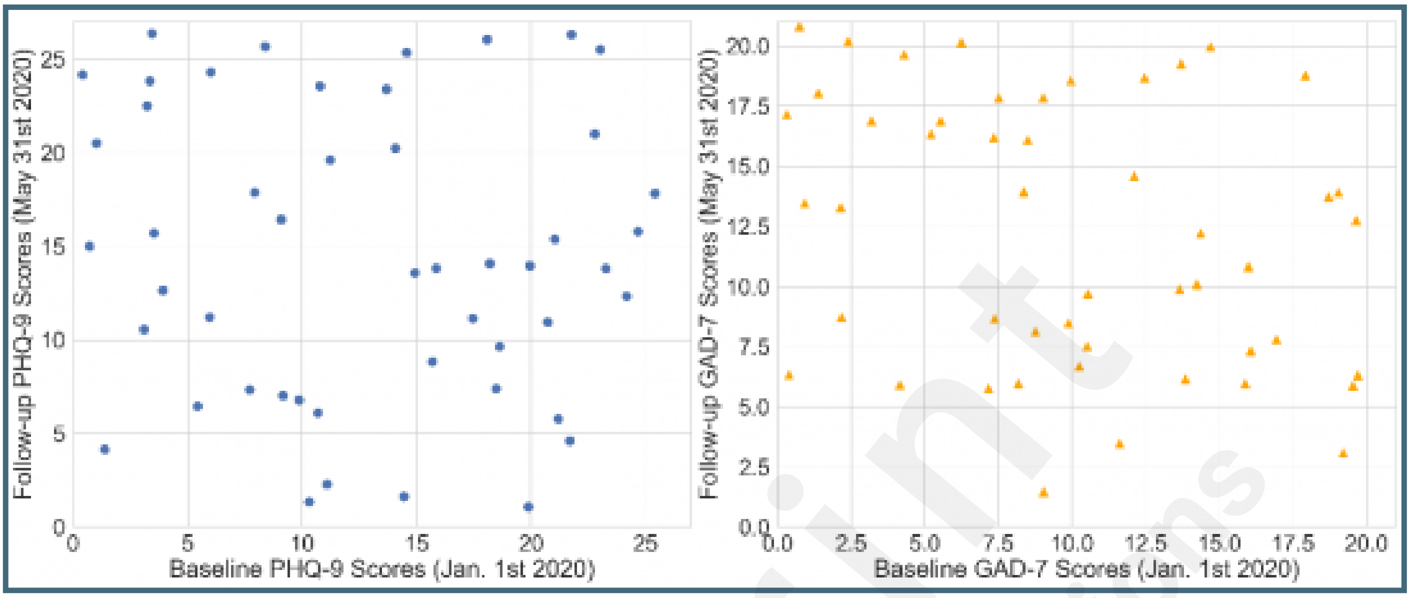
An example to demonstrated how the Last Seen Activities are selected for different threshold hours.



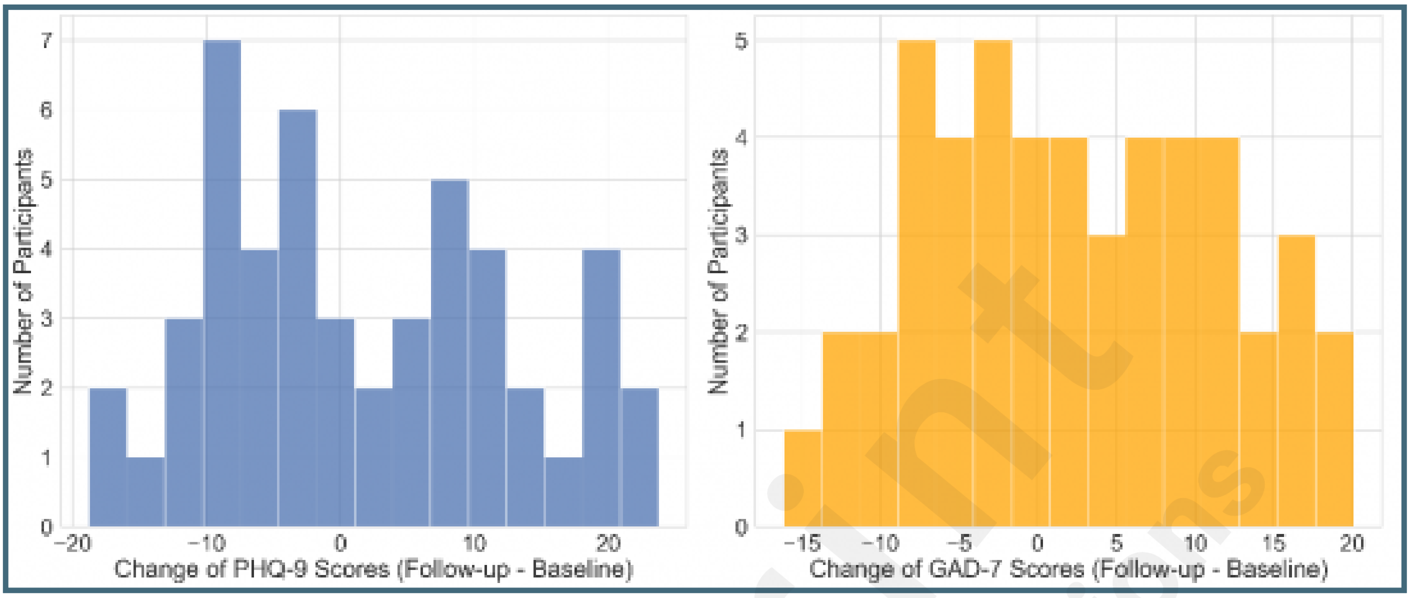
The study recruitment procedure and feature development process. All of the participants moved to remote learning on March 7th, the same day as declaring a state of emergency in the New York State. To avoid any acute behavior during the transition to remote learning, we excluded the data from March 1st to 28th.



The distributions of PHQ-9 depression and GAD-7 anxiety scores before and after the lockdown. The PHQ-9 scores are shown on the left, and the GAD-7 scores are shown on the right.



The distributions of the changes in PHQ-9 depression and GAD-7 anxiety scores before and after the lockdown. The PHQ-9 scores are shown on the left, and the GAD-7 scores are shown on the right.



The performances of the two regression models averaged across 49 leave-one-out splits.