SSM - Population Health 15 (2021) 100851

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: A Twitter analysis

Danyellé Thorpe Huerta [a,*], Jared B. Hawkins [b], John S. Brownstein [a,b], Yulin Hswen [c,d]

[a] Harvard Medical School Department of Biomedical Informatics, Boston, MA, 02115, USA
[b] Boston Children's Hospital Computational Epidemiology Lab, Boston, MA, 02215, USA
[c] University of California, San Francisco, Department of Epidemiology and Biostatistics, San Francisco, CA, 94158, USA
[d] University of California, San Francisco, Bakar Computational Health Sciences Institute, San Francisco, CA, 94158, USA

ARTICLE INFO

Keywords:
Twitter
Social media
Public policy
COVID-19
Public health
Sentiment analysis

ABSTRACT

As policies are adjusted throughout the COVID-19 pandemic according to public health best practices, there is a need to balance the importance of social distancing in preventing viral spread with the strain that these governmental public safety mandates put on public mental health. Thus, there is need for continuous observation of public sentiment and deliberation to inform further adaptation of mandated interventions. In this study, we explore how public response may be reflected in Massachusetts (MA) via social media by specifically exploring temporal patterns in Twitter posts (tweets) regarding sentiment and discussion of topics. We employ interrupted time series centered on (1) Massachusetts State of Emergency declaration (March 10), (2) US State of Emergency declaration (March 13) and (3) Massachusetts public school closure (March 17) to explore changes in tweet sentiment polarity (net negative/positive), expressed anxiety and discussion on risk and health topics on a random subset of all tweets coded within Massachusetts and published from January 1 to May 15, 2020 (n = 2.8 million). We find significant differences between tweets published before and after mandate enforcement for Massachusetts State of Emergency (increased discussion of risk and health, decreased polarity and increased anxiety expression), US State of Emergency (increased discussion of risk and health, and increased anxiety expression) and Massachusetts public school closure (increased discussion of risk and decreased polarity). Our work further validates that Twitter data is a reasonable way to monitor public sentiment and discourse within a crisis, especially in conjunction with other observation data.

1. Introduction

1.1. Background

In an effort to respond to the COVID-19 response, public health policies have been issued with a focus on the physical health of populations during the pandemic as evidenced by governmental mandates of social distancing, quarantining, and school closures to minimize the spread of the virus. However, preliminary research suggests that prolonged mandates of quarantine and social distancing are associated with a negative physical health experiences (including disruptive sleep) and negative mental health outcomes (including a sense of loneliness, and an increased worldwide suicide rate) (Lewis, 2020; Nikolaidis et al., 2020; Zhang, Gao, Gross, Shrum, & Hayne, 2020a, 2020b). As policies are adjusted throughout the pandemic to balance multiple – and at times, conflicting – aspects of health, there is a need for continuous observation of public sentiment and deliberation in order for officials to respond according to public needs so as to encourage adherence and address negative mental health outcomes of the pandemic such as heightened anxiety and perceptions of risk. Twitter data provides a forum for observation that is instantaneous, freely available and does not require logistic planning, unlike public opinion polls. These characteristics are critical during a crisis to provide instantaneous evidence to draw from when officials need to make rapid-response policy. Additionally, Twitter has the capacity to provide an avenue to improve care quality and safety, communication efficiency, cost-effectiveness and convenience of access to interventions (including education and advertising health care services) (Zhou et al., 2018).

* Corresponding author.
E-mail address: dthorpe@post.harvard.edu (D. Thorpe Huerta).

https://doi.org/10.1016/j.ssmph.2021.100851
Received 17 February 2021; Received in revised form 17 June 2021; Accepted 17 June 2021
Available online 19 June 2021

Table 1
Description of variables kept for analysis. Variable definitions are procured from an NLP model's publication, website and/or operator's manual (vaderSentiment 3.3.1 | The Python Package Index (PyPI), n.d.; LIWC | Linguistic Inquiry and Word Count, n.d.; Interpreting LIWC Output, n.d.; Pennebaker, Boyd, et al., 2015).

NLP Model	Variable Range	Variable Name	Variable Description
VADER	[-1,1]	Compound	*Describes the negativity/positivity of the text. -1 means that the text is completely negative, 0 signifies neutral text and 1 means that it is completely positive.*
LIWC	[0,100]	Analytic	*Reflects how analytical the author's thinking is. A high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking. This is a standardized score that has been converted to percentiles from 0-100.*
		Health	*Reflects any mention of health terms in the text such as clinic and medication. The score reflects the percentage of total words within the text that have these words.*
		Risk	*Describes references to dangers, concerns, things to avoid. The score reflects the percentage of total words within the text that have these words.*
		Anxiety	*Reflects any anxiety, worry or fear expressed in the text. The score reflects the percentage of total words within the text that have these words.*

1.2. Twitter use in public health

According to the Pew Research Center, 22% of the American adults use Twitter. These users tend to be representative of the general population except for some considerations. Twitter users tend to identify as Black or Hispanic more than as White, are more likely to be located in an urban or suburban area (Perrin & Anderson, 2019), younger, and are more likely to be Democratic and have higher levels of household income and educational attainment than the general population (Wojcik & Hughes, 2019).

Previous work incorporating Twitter data has proven useful in early recognition and characterization in a public emergency (Cassa et al., 2013) and sleep profiles possibly linked with psychosocial issues (McIver et al., 2015), influenza predictions (Nagar et al., 2014) and identification of food poisoning (Harris et al., 2017). Additionally, Twitter data analysis has been consistent with clinical characteristics of autism spectrum disorder (Hswen et al., 2019) and schizophrenia (Hswen, Naslund, Brownstein, & Hawkins, 2018a, 2018b). Literature suggests the potential to leverage Twitter data in order to support public health initiatives such as early illness detection (Hswen et al., 2018a, 2018b) and suicide prevention (Hswen et al., 2019). In this study, we aim to explore if Twitter may be leveraged during the COVID-19 pandemic so as to provide easily available, instantaneous public feedback amid multiple mandates to officials so they may adjust accordingly for the benefit of the public, especially in dense, urban areas.

1.3. Natural language processing (NLP)

Sentiment analysis utilizes contextual information in order to identify subjective details of a text (i.e., intent and feelings). Natural language processing (NLP) models including Linguistic Inquiry and Word Count (LIWC) and Valence Aware Dictionary for sEntiment Reasoning (VADER) have been validated as a quick and accurate sentiment analysis method in many contexts (Pennebaker et al., 2001; Pennebaker et al., 2017; Hutto, Gilbert, 2014) including within social media (Kramer et al., 2014). VADER is geared specifically toward sentiment polarity of social media text and has been employed to demonstrate how social media can be used as powerful cues for health outcomes (Tamesrsoy et al., 2015) and as a way to quantify mental health (Loveys, 2017). VADER uses a lexicon and rule-based sentiment analysis that is *specifically attuned to*

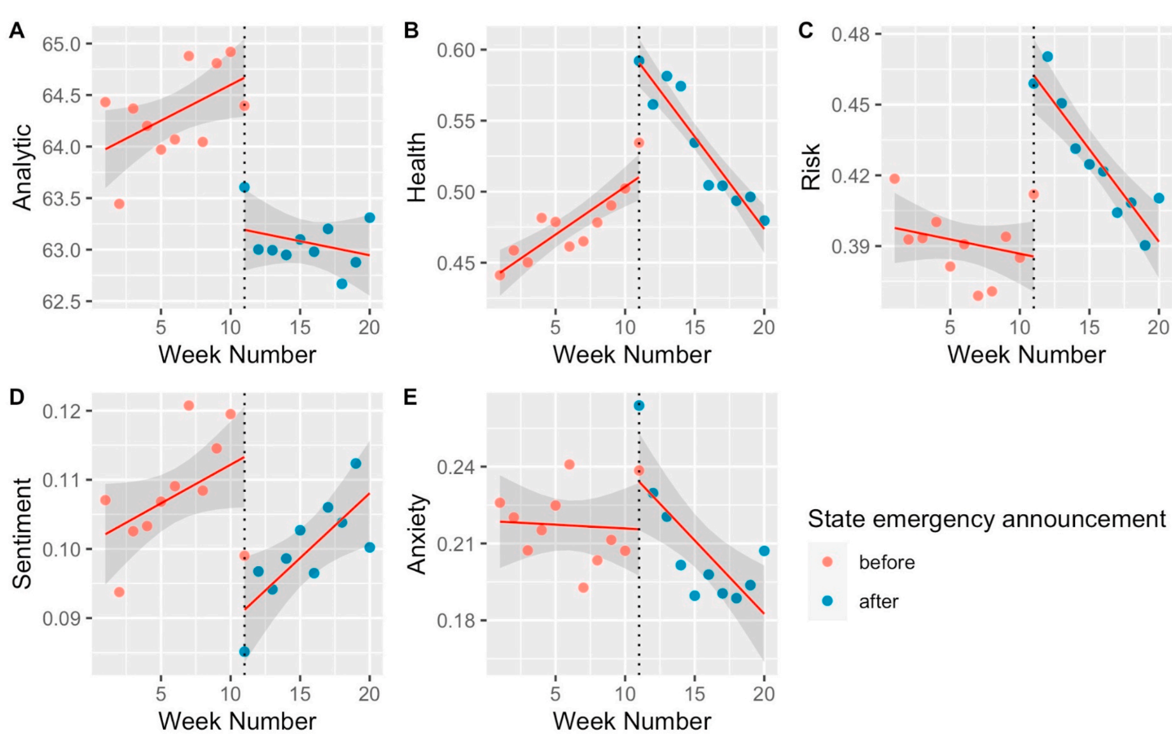

Fig. 1. Plot of average scores by week and Interrupted Time Series linear models for (A) LIWC Analytic score, (B) LIWC Health score, (C) LIWC Risk score, (D) VADER sentiment compound score, and (E) LIWC Anxiety score in relation to Massachusetts State of Emergency (declaration/effective date indicated by dotted line on week 11). 95% confidence intervals indicated by gray area around the linear equation in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Massachusetts state of emergency interrupted time series model summaries.

Coefficient	LIWC Analytic			VADER Compound			LIWC Health			LIWC Risk			LIWC Anxiety		
	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value
Intercept	63.90	63.44–64.37	<0.001	0.10	0.09–0.11	<0.001	0.44	0.42–0.46	<0.001	0.40	0.38–0.42	<0.001	0.22	0.20–0.24	<0.001
Massachusetts State of Emergency Mandate Interventions	−0.41	−1.74 – 0.92	0.524	−0.03	−0.06–−0.00	**0.023**	0.30	0.24–0.35	**<0.001**	0.15	0.10–0.20	**<0.001**	0.08	0.01–0.14	**0.019**
Week of the Year	0.07	0.00–0.14	**0.048**	0.00	−0.00 – 0.00	0.093	0.01	0.00–0.01	**<0.001**	−0.00	−0.00 – 0.00	0.359	−0.00	−0.00 – 0.00	0.852
Interaction term: Intervention & Week	−0.10	−0.20 – 0.01	0.068	0.00	−0.00 – 0.00	0.440	−0.02	−0.02–−0.02	**<0.001**	−0.01	−0.01–−0.00	**0.004**	−0.01	−0.01–−0.00	**0.036**
Observations	21			21			21			21			21		
R^2/R^2adjusted	0.817/0.784			0.511/0.424			0.906/0.890			0.793/0.756			0.406/0.301		

sentiments expressed in social media (Elbagir & Yang, 2019). LIWC is a model developed to study a text's "various emotional, cognitive, and structural components" (Pennebaker, 2015) and has successfully been employed to encode topical discussion (e.g., health) and sentiment (e.g., anger, sadness and anxiety) in short, limited and open-ended text (Tov et al., 2013) and blog posts (Gill et al., 2008), suggesting it can also be successfully applied to social media. The LIWC2007 Dictionary is composed of 2290 words and word stems. Each word or word-stem defines one or more word categories or subdictionaries and if it is found in the target text, each of these subdictionary scale scores will be incremented (Kovach Computing Services, n.d.; *LIWC | Linguistic Inquiry and Word Count, n.d.*).

1.4. Aim & significance

In this study, we aim to explore public response in Massachusetts via social media so as to provide feedback to officials so they may adjust mandates accordingly for the benefit of the public. More specifically, we ask: *Is there a shift in discussion and a difference of Massachusetts public sentiment (as studied through the sentiment of Twitter posts) during the initial response to the COVID-19 pandemic?* Namely, we explore responses following the Massachusetts and national State of Emergency declarations and Massachusetts school closures.

The significance of this project is two-pronged: this study (1) addresses validation of social media data as a way to monitor public sentiment and emotion during times of distress, and (2) investigates public health communications and mandates feedback during a public health crisis.

We hypothesize our work will add to the literature that demonstrates that the analysis of the Twitter data aligns with that of the more traditional data used as a current standard. This time-based analysis of aspects of public sentiment as provided by VADER and LIWC can inform policies and practices on how public discussion and sentiment is affected by public communications. With knowledge of the effects observed after public communications, suggestions can be formed for how and when information is communicated in order to balance public communication and downstream effects on the public sentiment that reflects public mental health. Additionally, officials may be able to better respond based on public needs such as deploying mental health resources to address concerns of higher social media expression of anxiety or depression.

2. Methods

2.1. Dates of interest identification

We aimed to identify early response efforts that would theoretically impact Massachusetts residents the most. We searched the Massachusetts Governor's press releases on the official website of the Commonwealth of Massachusetts (Office of Governor Charlie Baker and Lt. Governor Karyn Polito, n.d.) and identified the date of enforcement of two state-wide mandates of interest.

(1) On **March 10, 2020**, Massachusetts Governor Charlie Baker declared a State of Emergency that was effective immediately (*Declaration of a State of Emergency to Respond to COVID-19*, n.d.). The same day, the Baker-Polito Administration issued a press release that was among the first steps toward social distancing, encouraging at-risk citizens to avoid large crowds and events, and employers to adapt telework where possible, cancel any work-related travel, and cancel or hold virtually any event that would have large attendance (*Governor Baker Declares State of Emergency to Support Commonwealth's Response to Coronavirus*, n.d.).
(2) On **March 17, 2020**, further Baker-Polito Administration emergency actions in the response to the COVID-19 outbreak went into

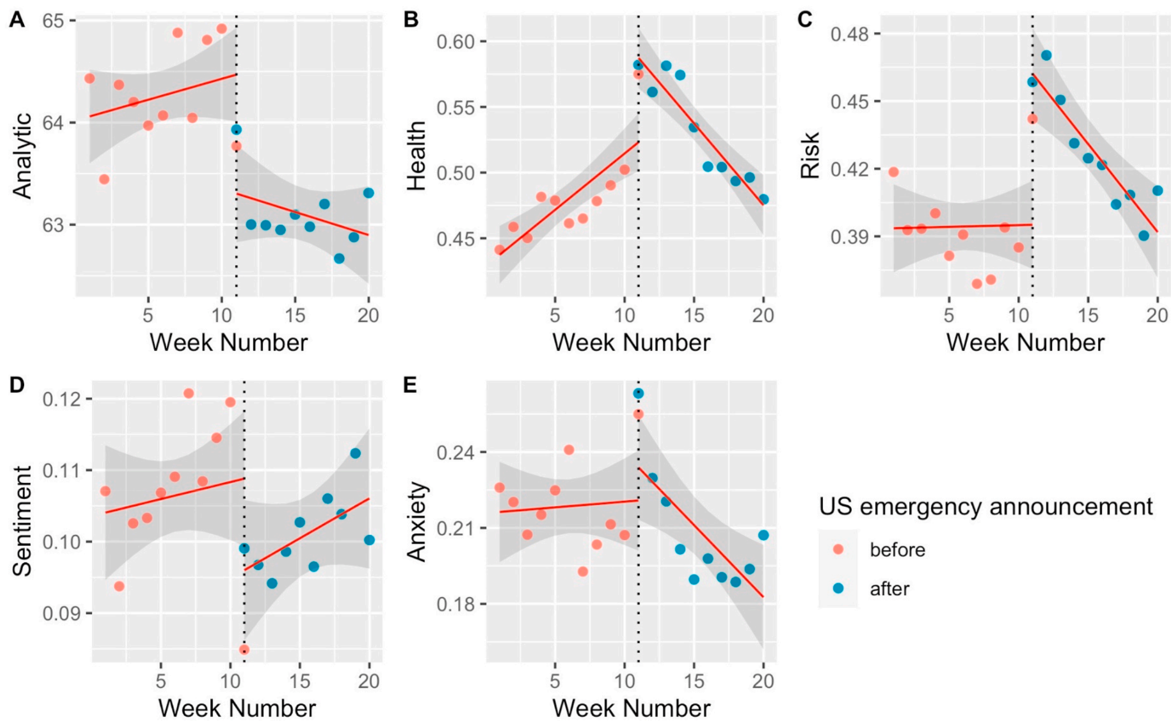

Fig. 2. Plot of average scores by week and Interrupted Time Series linear models for (A) LIWC Analytic score, (B) LIWC Health score, (C) LIWC Risk score, (D) VADER sentiment compound score, and (E) LIWC Anxiety score in relation to US State of Emergency (declaration/effective date indicated by dotted line on week 11).

effect, specifically "including mandatory school closures and prohibiting gatherings of 25 people or more" (*Baker-Polito Administration Announces Emergency Actions to Address COVID-19,* n.d.). These restrictive mandates were announced in a press release on March 15, 2020.

We searched the national press releases on the official United States White House site for any national dates of interest. On **March 13, 2020** in an official proclamation, COVID-19 was retroactively declared a national emergency that began March 1, 2020 and emergency authority was given to the Secretary of Health and Human Services (*Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak,* n.d.).

2.2. Data collection

The Boston Children's Hospital Computational Epidemiology Lab regularly collects a random subset of all Twitter posts (tweets) that include geo-information via Twitter's application programming interface (API) and stores it in a geocoded database, a historical Twitter database from 2012 to present. A similar approach analyzing a geocoded database to inform policy has been used by the St. Louis Department of Health to improve foodborne illness reporting and provide time-sensitive education and mobilizing information (Harris et al., 2017).

On May 18, 2020, we collected the tweets that were publicly available and published in Massachusetts from the time period of January 01, 2020–05/15/2020 from the pre-collected geocoded tweets within the aforementioned database. We chose this date range since it included the dates of mandate implementations and roughly two months before and after as a method to further contextualize any response that may be in association to these early shelter-in-place response efforts within Massachusetts. Tweets were included in analysis based on date of publication and were not filtered farther by relation to COVID-19. Tweets included both text and emojis.

We used Python (version 3.7.6) for data formatting (Python 3.7.6, n.d.). Using Dask for scalable functions, we formatted the data so as to remove any tweets predicted as written in a non-English language, and we labeled features of when a tweet was published including the number of the week (where the first full week of January 2020 was labeled as 1, and so on).

2.3. VADER

We ran the formatted dataset through the VADER model (vaderSentiment 3.3.1 | The Python Package Index (PyPI), n.d.) in order to assign a sentiment compound score to each tweet. VADER compound scores range from −1 to +1 and capture sentiment polarity by encoding not only sentiment expression (i.e., whether a tweet is positive or negative in nature as expressed by the score's positive or negative sign, respectively) but also sentiment intensity (i.e., if a tweet is very or somewhat positive/negative as expressed by the magnitude of the score). The more negative a tweet is, the closer it will be to −1 and the more positive, to positive 1. A tweet with neutral sentiment will receive a score close to 0. Each tweet received a compound score that encoded both expression (via a positive or negative sign) and intensity (via magnitude).

2.4. Linguistic Inquiry and Word Count (LIWC)

The formatted dataset was downloaded and ran through the LIWC desktop application (version LIWC 2015) for linguistic analysis (LIWC | Linguistic Inquiry and Word Count, n.d.). Since we were looking for general trends, we chose not to further refine the dataset for only tweets that reference COVID-19. We looked at the LIWC scores for health (based on words related to clinical terms such as cough, symptom and hospital), risk (based on words related to risk such as danger and doubt) and anxiety (based on words related to anxiety including worry and fear) to see the changes in these emotions before and after the mandates. We used analytical thinking score as a LIWC control as it was assumed the effects of the shelter in place would not change a user's analytical thinking. (A description of all of the variables examined can be found in Table 1.) Tweets were aggregated by week to allow for exploration of the

Table 3
US state of emergency interrupted time series model summaries.

Coefficient	LIWC Analytic			VADER Compound			LIWC Health			LIWC Risk			LIWC Anxiety		
	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value
Intercept	64.02	63.45–64.58	<0.001	0.10	0.09–0.12	<0.001	0.43	0.40–0.46	<0.001	0.39	0.37–0.42	<0.001	0.22	0.19–0.24	<0.001
US State of Emergency Mandate Interventions	−0.22	−1.84 – 1.40	0.779	−0.02	−0.05 – 0.01	0.226	0.30	0.22–0.37	**<0.001**	0.15	0.09–0.22	**<0.001**	0.08	0.01–0.15	**0.026**
Week of the Year	0.04	−0.04 – 0.12	0.315	0.00	−0.00 – 0.00	0.566	0.01	0.00–0.01	**<0.001**	0.00	−0.00 – 0.00	0.928	0.00	−0.00 – 0.00	0.795
Interaction term: Intervention & Week	−0.09	−0.21 – 0.04	0.172	0.00	−0.00 – 0.00	0.613	−0.02	−0.03–−0.01	**<0.001**	−0.01	−0.01–−0.00	**0.007**	−0.01	−0.01–−0.00	**0.031**
Observations	21			21			21			21			21		
R^2/R^2adjusted	0.718/0.668			0.185/0.041			0.836/0.807			0.667/0.609			0.373/0.263		

relationship between the amount of time since initial enforcement of public health mandates and score magnitude. Scores were aggregated and averaged across the week in which the Tweet was published.

2.5. Interrupted time series

An interrupted time series (ITS) analysis has routinely been used in literature to evaluate public health interventions by employing a linear model to relate an outcome of interest observed across a sequence of equally-spaced population observations (in our case the VADER and LIWC scores) to the intervention at a fixed point in time (the 'interruption' within the series; in our case, mandates going into effect) (Bilgan et al., 2000; Bernal et al., 2017). We prepared the data for an ITS in Python. Weeks were given a binary variable encoding if the week was before or after the mandate. The week of the mandate was split into two groups: tweets published before day of the intervention, and those published after the day of the intervention.

To explore statistical significance of the dates of interest in relation to the scores, we performed the ITS analysis using R (version 4.0.2). We created three linear models for each scoring variable, one for each mandate intervention. For these models, *MA_SOE*, *US_SOE*, *MA_Schools* are the binary variables encoding whether the week of observation is before or after the announcement/enforcement of the state's State of Emergency, nation's State of Emergency and state's school closure intervention mandates, respectively. This coefficient illustrated the change in scores after the mandate intervention. *Week_number* is the continuous variable encoding the week number of the year during which the tweets were published, and its coefficient illustrated the change in scores associated with the change in week as the year progresses. *Slope_change* is the interaction term between the intervention and time passing in the form of week number. This variable reflects the change in slope or the rate of change in the score as time passes either before or after the intervention (Bernal et al., 2017).

Linear models can thus be represented in the following formula: $b_0 - b_1{*}MA\_SOE + b_2{*}Week\_number - b_3{*}Slope\_change$ where, in this example, the model is for the Massachusetts's State of Emergency. The coefficients can be interpreted as follows: b_0 represents the baseline or average score prior to the intervention while b_1 represents the change to baseline after the intervention, b_2 represents the change with each week progression and b_3 represents the rate of change in the score as time passes either before or after the intervention.

3. Results

3.1. Data collection

In total, we collected 2,881,040 tweets with a geocoded label within Massachusetts from the Geocoded database. Of those, 2,847,400 (98.83%) had a proper timestamp and were predicted as being written in English and thus were included for analysis. The earliest tweet collected was published on January 1, 2020 at 12:00:04am EST. The latest tweet collected was published on May 14, 2020 at 11:59:59pm. There was a total of 20 weeks of tweets collected and an average number of 142,370 tweets published per week. Massachusetts and US State of Emergency declarations took place in calendar week 11 and Massachusetts school closure took place in calendar week 12.

3.2. Interrupted time series analysis

3.2.1. Massachusetts State of Emergency mandate intervention model (Table 2)

Coefficients of the LIWC analytic score linear model (Fig. 1a) indicated that there was no significant difference between scores before and after the state's State of Emergency. The only statistically significant coefficient was for the weekly progression.

There was a significant decrease in average VADER compound score,

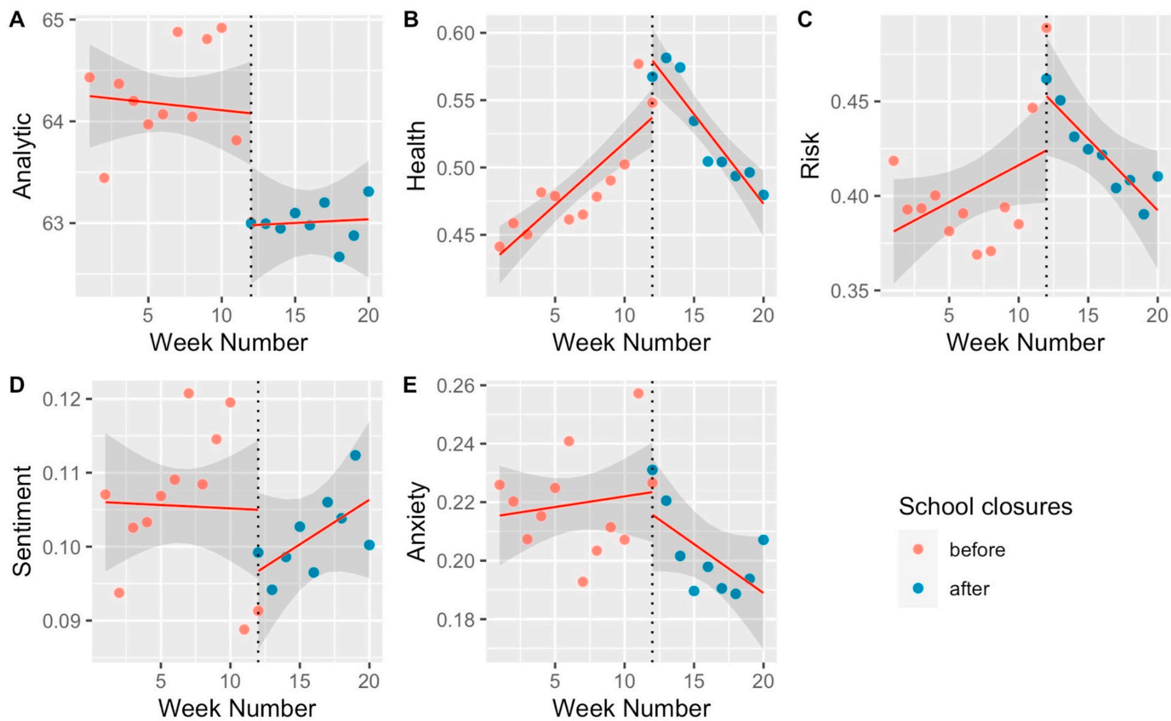

Fig. 3. Plot of average scores by week and Interrupted Time Series linear models for (A) LIWC Analytic score, (B) LIWC Health score, (C) LIWC Risk score, (D) VADER sentiment compound score, and (E) LIWC Anxiety score in relation to Massachusetts School Closure (effective date indicated by dotted line on week 12).

representing tweet sentiment, after the Massachusetts State of Emergency and there was no statistically significant difference in the weekly change nor the rate of change before and after the mandate (Fig. 1b).

All of the LIWC health score linear model (Fig. 1c) coefficients were statistically significant, indicating a rise in the average percentage of words in a tweet pertaining to health after the Massachusetts State of Emergency and with each week progression and a slower rate of change after the intervention.

The average percentage of words in a tweet discussing risk as measured by the LIWC risk score's linear model (Fig. 1d) rose significantly and had a significantly slower rate of change after the state's State of Emergency. The difference with each week progression was not statistically significant.

The LIWC anxiety score linear model (Fig. 1e) indicated that the average percentage of words expressing anxiety showed a statistically significant increase and slower rate of change after the Massachusetts State of Emergency. The difference with each week progression was not statistically significant.

3.2.2. US State of Emergency mandate intervention model (Table 3)

None of the coefficients of the LIWC analytic score and VADER compound score linear models (Fig. 2a and b respectively) for the US State of Emergency were significant, thus, there is no significant difference between these scores before and after the intervention.

The average percentage of words in a tweet pertaining to health as measured by the LIWC health risk score's linear model (Fig. 2c) rose significantly and had a significantly slower rate of change after the US State of Emergency. Scores did also rise significantly with each week progression.

The LIWC risk score's linear model (Fig. 2d) indicates that the average percentage of words in a tweet discussing risk changed significantly after the US State of Emergency and that, after the intervention, there was a statistically slower rate of change in score. There was no difference in scores due to week progression.

Both the change in the average percentage of words expressing anxiety and the slower rate of change after the US State of Emergency were significant as shown by the LIWC anxiety score's linear model (Fig. 2e). There was no difference in scores due to week progression.

3.2.3. Massachusetts School Closures mandate intervention model (Table 4)

None of the coefficients of the LIWC analytic and anxiety scores and VADER compound score linear models (Fig. 3a and b respectively) for the Massachusetts school closures were significant, thus, there is no significant difference between these scores before and after the intervention.

All of the LIWC health score linear model (Fig. 3c) coefficients were statistically significant, indicating a rise in the average percentage of words in a tweet pertaining to health after the Massachusetts school closures and with each week progression and a slower rate of change after the intervention.

The average percentage of words in a tweet discussing risk as measured by the LIWC risk score's linear model (Fig. 3d) rose significantly and had a significantly slower rate of change after the state's school closures. There was no difference in average risk scores with each week progression.

The LIWC anxiety score linear model (Fig. 3e) indicated that the average percentage of words expressing anxiety showed a statistically significant increase and slower rate of change after the state's school closures. The difference with each week progression was not statistically significant.

4. Discussion

Our findings suggest that we may be able to monitor population reactions via geocoded tweets due to the significant differences between tweets, especially with regard to topic discussion around risk and health, as well as in terms of anxiety expressed in the tweet. Due to the similarity in change after each mandate and the many public health mandates in a short amount of time, it is difficult to attribute the differences to be associated with a specific mandate.

Since the control variable (LIWC Analytic score) showed no change

Table 4
Massachusetts school closure interrupted time series model summaries.

	LIWC Analytic			VADER Compound			LIWC Health			LIWC Risk			LIWC Anxiety		
Coefficient	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value
Intercept	64.26	63.64–64.88	<0.001	0.11	0.09–0.12	<0.001	0.43	0.40–0.45	<0.001	0.38	0.34–0.41	<0.001	0.21	0.19–0.24	<0.001
Massachusetts School Closure Mandate Interventions	−1.37	−3.57 – 0.83	0.206	−0.02	−0.06 – 0.02	0.230	0.31	0.22–0.41	**<0.001**	0.17	0.05–0.29	**0.009**	0.04	−0.03 – 0.11	0.255
Week of the Year	−0.02	−0.10 – 0.07	0.705	−0.00	−0.00 – 0.00	0.901	0.01	0.01–0.01	**<0.001**	0.00	−0.00 – 0.01	0.091	0.00	−0.00 – 0.00	0.594
Interaction term: Intervention & Week	0.02	−0.13 – 0.18	0.761	0.00	−0.00 – 0.00	0.351	−0.02	−0.03– −0.02	**<0.001**	−0.01	−0.02– −0.00	**0.011**	−0.00	−0.01 – 0.00	0.117
Observations	21			21			21			21			21		
R^2/R^2adjusted	0.639/0.576			0.114/−0.042			0.818/0.786			0.400/0.294			0.340/0.224		

associated with any of the interventions, it behaved as expected. While there was a statistical significance for the Massachusetts's State of Emergency's Analytic score linear model's weekly progression coefficient, it is the exception rather than the rule for our three models and we suspect that it may be unrelated to the intervention effect.

The general insignificance among linear models regarding sentiment may indicate the need to separate the sentiment scores into positive and negative scores rather than relying on polarity scores. We discuss this in our limitations section.

The control variable general behavior, along with other expected behaviors such as the consistent increase in discussion around risk and health due to pandemic-related responses and increase in anxiety during uncertain times across the three interventions' linear models, leads us to believe that our work further supports the use of Twitter as a supplemental data source of public wellbeing that may support data-driven community health initiatives, especially with regard to rapid-response emergency mandates when there is less time to gather evidence through traditional polling methods. It is possible that there is no significant difference in anxiety scores after the latest intervention because the public felt more secure as time went on and more interventions went into place while discussion in health and risk remained heightened as conversations about the virus continued.

When aligned with publicly accessible data tracking newly infected counts, our study suggests that the increased health and risk discussion and increased anxiety expression following the mandates may align with slowing the spread of COVID-19. Data reflects that the number of newly infected in Massachusetts peaked in late March before the rate began to fall (COVID-19 Projections | US | Massachusetts, n.d.). The CDC states it may take up to 2 weeks after viral exposure for infected individuals to show symptoms (*Symptoms of Coronavirus*, n.d.) that may prompt testing; thus, it makes sense that the lag time between the intervention and concurrent associated shift in discussion and the decrease in Massachusetts newly infected cases to be about the two weeks between mid-to late-March. This provides an interesting parallel that may provide evidence to how social media can be a signal for real world changes in behavior and environment.

Additionally, a study has found that a predictor of positive uptake of suggested practices, such as social distancing and improved hand hygiene, was fear of COVID-19 (Harper et al., 2020). This could be in alignment with our study since we found that anxiety did increase after mandates were enacted. The increased anxiety may have contributed to adherence to the mandates and, consequently, the fall in new cases after the 2-week lag time.

While we see a shift in public engagement in health and risk topics due to intervention, regardless of the specific mandate, we did not see the same results for the VADER compound score. In these sentiment scores we observed only one significant shift for the Massachusetts State of Emergency. Further work may explore more specific sentiment variables for significance to investigate if the insignificant difference is due to the general nature of the score since it only captures polarity of the text.

4.1. Limitations

It is important to note that, in this study, we are making the assumption that this random subset of tweets is representative of the general public of Twitter users. We are also assuming that a device's GPS that sometimes informs geocode labels are accurate. Further, we acknowledge that, without a network, user and content analysis, we cannot conclude that our data excludes fake news, bot accounts, unreliable sources, echo chambers or other perils of information and informing within our Twitter dataset.

Our study design allows for exploration regarding general trends and is unable to reveal more specific trends in tweets or characteristics about the authors. For example, we cannot conclude that the increased anxiety portrayed in an author's posts is specifically in tweets discussing COVID-

19 topics or if the anxiety is centered around something independent of the pandemic. It is also possible we are losing granularity by looking at average scores on a weekly rather than daily basis.

Additionally, polarity may not be specific enough and we may benefit from further analysis of more targeted sentiment scores such as VADER positive and negative scores and/or other LIWC scores that center on emotion. More specific approaches to sentiment have found that the average positive sentiment scores have been associated with public health uptake behaviors when characterized by quintiles (Wong et al., 2016); thus, it will be important to explore different approaches not only in sentiment scores but also in score characterization.

Lastly, we acknowledge that tweets could be a reflection of a user's thought in that moment in time and not wholly reflective of their overall sentiment throughout the pandemic. However, this feature allows public health officials to get a snapshot of real-time changes in sentiment.

4.2. Strengths

Our work has the potential to inspire further hypotheses for validating social media for policy and intervention impact. Since this is the first pandemic severe enough to warrant wide-spread school closures and self-isolation mandates in which we have the technology to easily access public sentiment in real time and directly, we have an unprecedented ability to analyze the effect of mandates and communication of information about these mandates in order to preserve public health. Twitter is valuable since feedback is uncensored and provided in real-time, and researchers can have instant access via the Twitter API, unlike the lengthy process involved in collecting surveys. Our work also may encourage officials to look at social media data for evidence to help inform early responses and adjustments to interventions so as to balance public communication and downstream effects in order to encourage adherence.

4.3. Future directions

Future directions may include an analysis that incorporates more granularity. This granularity may be in exploring differences in location characteristics such as rural and urban settings (rural areas tend to have more outdoor space to gather and less risk of spreading due to proximity/population density), or government political affiliation since some research has found that the percentage of Republican residents correlates with changes in state-level mobility restriction (Hsiehchen et al., 2020). Granularity may also be found in a profile analysis to understand if the majority of tweets may come from a small minority of tweeters (Wojcik & Hughes, 2019); topic analysis to better understand the text being measured; investigations into subtopics of conversations within health and risk discussion to illuminate if people are considering adherence to the mandates; and further temporal patterns that may be associated with the loosening and tightening of restrictions as cases decrease and increase respectively.

It would also be important to consider additional mandates and announcements on the city, county, and even global levels. For example, we did not incorporate the World Health Organization reclassification of the COVID-19 outbreak as a pandemic as the rates of infection continued to rise in many locations around the world and across the United States and how that may have affected the score variables as well.

Additional analysis on a daily instead of weekly basis, as well as building a model that takes into account any compounding effects from the mandates occurring so close to one another, would also be beneficial. We see that models appear similar regardless of the source and content of the mandates and we suspect that it is due to their introductions being within an 8-day span. We believe it is appropriate to try to include the interventions in one model to try to control for any compounding effects. It is possible that this will also reduce the variety in R^2 values.

We provide the background for further work on how sentiment may be monitored in this case. We suggest that further work in this area be done to explore if sentiment may be better captured in quintiles instead of raw magnitude values or in ratios of tweets with overall negative and/or positive sentiment.

Lastly, general public awareness of the pandemic began 2020, so the baselines may be skewed. Performing an outer comparison via a difference-in-difference analysis including 2019 data as well would be important to contextualize our current work.

5. Conclusion

Our work further demonstrates that Twitter data provides a venue to observe the association of intervention with public sentiment and discourse, especially in conjunction with other observation data such as case and/or death rates. Tweets provide the potential for a more informal and inexpensive way of monitoring population sentiment, streamlining the collection process and rapid analysis than traditional survey methodology, allowing for policy adjustment in shorter amounts of time depending on public needs in order to encourage adherence.

CRediT authorship contribution statement

Danyellé Thorpe Huerta: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Jared B. Hawkins:** Methodology, Writing – review & editing, grant funding. **John S. Brownstein:** Writing – review & editing, grant funding. **Yulin Hswen:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Acknowledgements

Funding for this study was provided by an NIH/NLM Institute grant (No. 3R01LM011965-03S1). The NIH/NLM had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

References

Cassa, C. A., Chunara, R., Mandl, K., & Brownstein, J. S. (2013). Twitter as a sentinel in emergency situations: Lessons from the Boston marathon explosions. *PLOS Currents Disasters*. https://doi.org/10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b

Baker-Polito Administration Announces Emergency Actions to Address COVID-19 | Mass.gov. (n.d.). Retrieved October 7, 2020, from https://www.mass.gov/news/baker-polito-administration-announces-emergency-actions-to-address-covid-19.

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology, 46*(1), 348–355. https://doi.org/10.1093/ije/dyw098

Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science: The Official Journal of the Society for Prevention Research, 1*(1), 31–49.

Covid-19 Projections | US | Massachusetts. (n.d.). Retrieved November 07, 2020, from https://covid19-projections.com/us-ma.

Declaration of a State of Emergency to Respond to Covid-19 | Mass.gov. (n.d.). Retrieved October 7, 2020, from https://www.mass.gov/news/declaration-of-a-state-of-emergency-to-respond-to-covid-19.

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international MultiConference of engineers and computer scientists 2019, Hong Kong* (Vol. 5). http://www.iaeng.org/publication/IMECS2019/IMECS2019_pp12-16.pdf.

Gill, A. J., French, R. M., Gergle, D., & Oberlander, J. (2008). The language of emotion in short blog texts. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08, 299*. https://doi.org/10.1145/1460563.1460612

Governor Baker Declares State of Emergency to Support Commonwealth's Response to Coronavirus | Mass.gov. (n.d.). Retrieved October 7, 2020, from https://www.mass.gov/news/governor-baker-declares-state-of-emergency-to-support-commonwealths-response-to-coronavirus.

Harper, C. A., Satchell, L. P., Fido, D., & Latzman, R. D. (2020). Functional fear predicts public health compliance in the COVID-19 pandemic. *International Journal of Mental Health and Addiction*, 1–14. https://doi.org/10.1007/s11469-020-00281-5

Harris, J. K., Hawkins, J. B., Nguyen, L., Nsoesie, E. O., Tuli, G., Mansour, R., & Brownstein, J. S. (2017). Using twitter to identify and respond to food poisoning: The food safety STL project. *Journal of Public Health Management and Practice, 23*(6), 577–580. https://doi.org/10.1097/PHH.0000000000000516

Hsiehchen, D., Espinoza, M., & Slovic, P. (2020). Political partisanship and mobility restriction during the COVID-19 pandemic. *Public Health, 187*, 111–114. https://doi.org/10.1016/j.puhe.2020.08.009

Hswen, Y., Gopaluni, A., Brownstein, J. S., & Hawkins, J. B. (2019). Using twitter to detect psychological characteristics of self-identified persons with autism spectrum disorder: A feasibility study. *JMIR MHealth and UHealth, 7*(2), Article e12264. https://doi.org/10.2196/12264

Hswen, Y., Naslund, J. A., Brownstein, J. S., & Hawkins, J. B. (2018). Monitoring online discussions about suicide among twitter users with schizophrenia: Exploratory study. *JMIR Mental Health, 5*(4), e11483. https://doi.org/10.2196/11483

Hswen, Y., Naslund, J. A., Brownstein, J. S., & Hawkins, J. B. (2018). Online communication about depression and anxiety among twitter users with schizophrenia: Preliminary findings to inform a digital phenotype using social media. *Psychiatric Quarterly, 89*(3), 569–580. https://doi.org/10.1007/s11126-017-9559-y

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *International AAAI conference on web and social media*.

Interpreting LIWC Output. (n.d.). Retrieved July 11, 2020, from https://liwc.wpengine.com/interpreting-liwc-output/.

n.d. Kovach Computing Services. LIWC dictionary. Retrieved May 1, 2021, from https://www.kovcomp.co.uk/wordstat/LIWC.html

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788. https://doi.org/10.1073/pnas.1320040111

Lewis, K. (2020). COVID-19: Preliminary data on the impact of social distancing on loneliness and mental health. *Journal of Psychiatric Practice, 26*(5), 400–404. https://doi.org/10.1097/PRA.0000000000000488

LIWC | Linguistic Inquiry and Word Count |m Pennebaker | Conglomerates, Inc. (n.d.). Retrieved May 21, 2020, from https://liwc.wpengine.com.

Loveys, K., Crutchley, P., Wyatt, E., & Coppersmith, G. (2017). Small but mighty: Affective micropatterns for quantifying mental health from social media language. In *Proceedings of the fourth workshop on computational linguistics and clinical psychology — from linguistic signal to clinical reality* (pp. 85–95). https://doi.org/10.18653/v1/W17-3110

McIver, D. J., Hawkins, J. B., Chunara, R., Chatterjee, A. K., Bhandari, A., Fitzgerald, T. P., Jain, S. H., & Brownstein, J. S. (2015). Characterizing sleep issues using twitter. *Journal of Medical Internet Research, 17*(6), e140. https://doi.org/10.2196/jmir.4476

Nagar, R., Yuan, Q., Freifeld, C. C., Santillana, M., Nojima, A., Chunara, R., & Brownstein, J. S. (2014). A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research, 16*(10), e236. https://doi.org/10.2196/jmir.3416

Nikolaidis, A., Paksarian, D., Alexander, L., DeRosa, J., Dunn, J., Nielson, D. M., Droney, I., Kang, M., Douka, I., Bromet, E., Milham, M. P., Stringaris, A., & Merikangas, K. R. (2020). The Coronavirus Health and Impact Survey (CRISIS) reveals reproducible correlates of pandemic-related mood states across the Atlantic. *MedRxiv, 2020*. https://doi.org/10.1101/2020.08.24.20181123, 08.24.20181123.

Office of Governor Charlie Baker and Lt. Governor Karyn Polito | Mass.gov. (n.d.). Retrieved October 7, 2020, from https://www.mass.gov/orgs/office-of-the-governor.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and word count: LIWC2015 operator's manual*. Austin, Texas: Pennebaker Conglomerates.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. https://doi.org/10.15781/T29G6Z.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2017). *The development and psychometric properties of LIWC2007* (Vol. 22).

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and word count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum Associates.

Zhang, W., Gao, F., Gross, J., Shrum, L. J., & Hayne, H. (2020a). How does social distancing during COVID-19 affect negative moods and memory? PsyArXiv. https://doi.org/10.31234/osf.io/67rhf.

Perrin, A., & Anderson, M. (2019, April 10). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*. Retrieved April 28, 2021, from https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/.

Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (Covid-19) Outbreak. The white House. n.d. https://www.whitehouse.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/. Retrieved October 7, 2020, from

Python 3.7.6 | python (n.d.). Retrieved January 21, 2020, from https://www.python.org/downloads/release/python-376/.

Symptoms of Coronavirus | CDC. (n.d.). Retrieved November 7, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

Tamersoy, A., De Choudhury, M., & Chau, D. H. (2015). Characterizing smoking and drinking abstinence from social media. In *Proceedings of the 26th ACM conference on hypertext & social media* (pp. 139–148). https://doi.org/10.1145/2700171.2791247

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment, 25*(4), 1069–1078. https://doi.org/10.1037/a0033007

vaderSentiment 3.3.1 | The Python Package Index (PyPI) (n.d.). Retrieved April 1, 2020, from https://pypi.org/project/vaderSentiment/3.3.1.

Wojcik, S., & Hughes, A. (2019, April 24). How twitter users compare to the general public. Pew research center: Internet, science & tech. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

Wong, K. O., Davis, F., Zaiane, O. R., & Yasui, Y. (2016). *Sentiment analysis of breast cancer screening in the United States using twitter*. KDIR. https://doi.org/10.5220/0006047102650274

Zhang, W., Gao, F., Gross, J., Shrum, L. J., & Hayne, H. (2020a). *How does social distancing during COVID-19 affect negative Moods and memory?* PsyArXiv. https://doi.org/10.31234/osf.io/67rhf.

Zhou, L., Zhang, D., Yang, C., & Wang, Y. (2018). Harnessing social media for health information management. *Electronic Commerce Research and Applications, 27*, 139–151. https://doi.org/10.1016/j.elerap.2017.12.003