News Hurts: Exploring the Link Between Pandemic-Related Social Media Use and Trauma Symptoms

Claudine Tinsman¹, Siddharth Arora², Max Van Kleek³

¹ ³ University of Oxford, Department of Computer Science, ² University of Oxford, Mathematical Institute
claudine.tinsman@cs.ox.ac.uk¹, siddharth.arora@maths.ox.ac.uk², max.van.kleek@cs.ox.ac.uk³

Abstract

The link between indirect exposure to disasters via traditional media and the development of post-traumatic stress symptoms (PTSs) has long been a topic of interest in multiple fields. However, it is unclear to what degree existing findings are applicable to social media.

We used random forest regression and classification to conduct an exploratory investigation into the association between type of social media use (i.e. active and passive use) and pandemic-related PTSs. Controlling for post-traumatic stress disorder (PTSD) risk factors and exposure to the pandemic, we found an association between passive social media use and high levels of pandemic-related PTSs among individuals who had not directly been exposed to the effects of COVID-19.

Introduction

The COVID-19 pandemic led to rapid and unpredictable changes to daily life for millions of adults across the United Kingdom. News about the pandemic was disseminated rapidly through both traditional and social media, making it difficult to escape a deluge of information about death and infection statistics, stories of tragedy, and constantly changing public safety measures. Existing research in traditional media studies suggests that individuals who are only exposed indirectly to a traumatic event via media coverage can experience symptoms of post-traumatic stress. However, the degree to which such findings, primarily drawn from traditional media research (e.g. television or print newspapers), are applicable to social media is unknown.

Unlike traditional media, social media enables users to act both as producers (active users) and consumers (passive users) of content (Ha and Yun 2014). Both types of use have been associated with either decreases or increases in well-being, depending on factors such as age or platform selected (Frison and Eggermont 2016) (Lin et al. 2016) (Weinstein 2017) (Valkenburg, van Driel, and Beyens 2021) (Ha and Yun 2014). While spending more time using social media to consume COVID-19 information has been associated with high levels of post-traumatic stress symptoms (PTSs) among the general population (Chao et al. 2020), there is no research on the relationship between active or passive social media use and PTSs among a sample of the population that has not been directly exposed to the pandemic.

Given the gap in the literature, we explored the following research question:

Among adults indirectly exposed to the COVID-19 pandemic, are either active or passive social media use associated with pandemic-related post-traumatic stress symptoms (PTSs)?

To address our research question, we performed random forest regression and classification on cross-sectional survey data from 352 UK adults who had not been directly exposed to COVID-19 at the time of the study. We collected data on time spent using both traditional and social media, active and passive social media use, symptoms of PTSD, and a broad range of potential PTSD risk factors.

Background

In this section, we provide domain-specific background from multiple disciplines regarding the importance of media use, risk factors for direct exposure to traumatic events, and risk factors for indirect exposure across both traditional and social media.

Post-traumatic Stress Disorder

Post-traumatic Stress Disorder (PTSD) is a type of trauma and stressor-related disorder that manifests as psychological distress following exposure to a traumatic event (American Psychiatric Association 2013). Many sufferers experience persistent trauma-related emotions, physical ailments, and reduced social and work functioning (van Minnen et al. 2015). Initial assessment of post-traumatic stress is performed using the Post-traumatic Stress Disorder Checklist (PCL-5) (American Psychiatric Association 2013). A diagnosis of PTSD requires that the event consist of “actual or threatened death, serious injury, or sexual violence” (American Psychiatric Association 2013). Therefore, individuals who indirectly experience a traumatic event by watching or reading about cannot be diagnosed with PTSD, regardless of symptoms experienced.

The development of PTSD is influenced by a combination of objective and subjective risk factors. Principal objective
risk factors are: Identifying as a female, having a history of anxiety disorders, depression, and PTSD (Worthington, Mandavia, and Richardson-Vejlgaard 2020; Brewin, Andrews, and Valentine 2000; Kessler et al. 2005b). Young age, being an ethnic minority, low socio-economic status (SES) and low education are also possible risk factors for PTSD (Kobayashi, Sledjeski, and Delahanty 2019; Sareen 2014). Subjective risk factors depend on an individual’s perception of the event as traumatic (Weinberg, Michael and Gil 2016). Such factors include perceived proximity to the traumatic event, the subjective experience of the trauma as a threat, dissociation in the face of the threat, and personality traits of the individual (Weinberg, Michael and Gil 2016).

Random forests (RF) are an ensemble method comprising decision trees (Breiman 2001) used for both classification and regression tasks that have been commonly used to predict PTSD. They consist of multiple decision trees, where each tree uses a random subset of predictor variables at each node of the tree (Speiser et al. 2019). The prediction is generated by averaging scores across all trees (i.e., the entire forest), which allows for high predictive accuracy while avoiding over-fitting (Breiman 2001).

Random Forests have been used to predict PTSD in a broad range of contexts, demographics, and sample sizes. Augsburger and Galatzer-Levy used RF classification to predict the severity of PTSD among 94 individuals who had been hospitalised following a traumatic experience (2020). Using RF classification in conjunction with clustering techniques, Siegel et al. identified military-related PTSD subtypes (2021). Random Forest classification has been used to identify people at risk of PTSD based on smartphone-based self-reported patient surveys (Wshah, Skalka, and Price 2019). Prout et al. (2020) employed regression trees to identify key predictors of distress during the COVID-19 pandemic, which included anxiety, depression and post-traumatic stress. However, the study did not account for direct/indirect exposure to the pandemic (2020).

**Media Exposure to Traumatic Events**

Events such as war (Silver et al. 2013), natural disasters (Hall et al. 2019), terrorism (Pfefferbaum et al. 2003), and pandemics (Zhao and Zhou 2020) have been treated as disasters in the context of empirical research. Controlling for direct exposure, time spent watching television coverage of terrorist attacks or natural disasters has positively correlated with symptoms of PTSD for periods spanning one month to several years after initially viewing the images (Silver et al. 2013; Schlenger et al. 2002; Garfin et al. 2018; Hall et al. 2019). Specific risk factors, both objective and subjective, have been identified for indirect exposure to disasters via the media (Pfefferbaum et al. 2014; Chao et al. 2020).

**Age**  Research on age as a risk factor for PTSs associated with disaster media use has yielded mixed results: Younger adults were found to be more at risk of PTSs where the disaster was a bombing (Jones et al. 2016), and older adults were at higher risk of developing PTSs in cases of bioterrorism (Dougall, Hayward, and Baum 2005).

**Gender**  Associations between gender, PTSs, and media use are mixed and vary depending on cultural context (Yeung et al. 2018). Gender was not associated with disaster media-related PTSs in cases of indirect exposure to 9/11 coverage via television news coverage (Lengua, Long, and Meltzoff 2006). Yet, across American and Israeli population samples for multiple disasters, women reported more PTSs associated with disaster media use than men (Keinan, Sadeh, and Rosen 2003; Pfefferbaum et al. 2003). Conversely, higher levels of exposure to traditional media coverage of Typhoon Hato were associated with greater levels of PTSs in men than women (Hall et al. 2019).

**Pre-existing Mental Health Conditions**  Research has found that individuals who experienced symptoms of anxiety and trauma in the past were more likely to view more television coverage of the event than those who had not (Ahern et al. 2004). In a study of 898 young adults conducted during the early days of the COVID-19 pandemic (April-May 2020), those who reported either confirmed or suspected pre-existing mental health condition were significantly more likely to score above the clinical threshold for PTSD compared to those with no diagnosis (Liu et al. 2020). The authors did not control for direct and indirect exposure to COVID-19. Controlling for the number of hours of event coverage consumed, a survey of adolescents who had not directly experienced the Boston Marathon bombings found that prior experience of trauma was a significant predictor of more severe PTSs (Busso, McLaughlin, and Sheridan 2014).

**Proximity**  Geographic proximity has been shown to play a key mediating role between media use and PTSs: In the wake of 9/11, Galea et al. found living close to the World Trade Centre at the time of the attack was a risk factor for PTSs both among individuals who had witnessed the event, and those who hadn’t (2002). Similarly, a survey conducted two years after the Oklahoma City Bombing on school children living within one hundred miles of the disaster found that participants experienced symptoms of PTSs related to print and television disaster content, even if they had not been exposed to the disaster, or directly knew anyone who had (Pfefferbaum et al. 2003).

**COVID-19 and PTSs**  There is evidence that the COVID-19 pandemic is associated with an increase in mental health concerns amongst the global population: In a meta-review of seventy-one papers published during the first six months of 2020, Liu et al. found that anxiety, depression, insomnia, and post-traumatic stress disorder were prevalent mental health disorders across nine national populations (Liu et al. 2021).

In the UK specifically, a survey of 3074 adults conducted during the first six weeks of lockdown (March-May 2020) found that women, young people (18–29 years), those from more socially disadvantaged backgrounds and those with pre-existing mental health problems were at risk of experiencing worse mental health outcomes during the
pandemic (O’Connor et al. 2021). The shift from working outside of the home to working from home (Goyinas and Ozdamar 2021), perceived vulnerability to the virus (Boyraz, Legros, and Tigershtrom 2020), and living in isolation (Hendriksen et al. 2021) have been identified as risk factors for developing pandemic-related PTSs and a decline in mental well-being among the population. These findings suggest that individual differences play a significant role in mental health outcomes relating to a pandemic, both among directly and indirectly exposed individuals.

**Active and Passive Social Media Use**

HCI and media studies researchers frequently categorise social media use as either active or passive (Verduyn et al. 2015; Burke, Marlow, and Lento 2010). **Passive social media use** refers to patterns of behaviour that consist of monitoring the lives of others by viewing their social media content (Verduyn et al. 2017) (Frison and Eggermont 2016). **Active social media use** refers to patterns of behaviour that enable direct exchanges between users” (Verduyn et al. 2017). Such behaviours include liking, commenting, sending messages, and otherwise engaging with other users in a private (e.g. direct messaging) or public manner (e.g. posting a status update) (Verduyn et al. 2015).

Research has often found active social media use to be associated with improvement in well-being, and passive use associated with passive social media use (Valkenburg, van Driel, and Beyens 2021). However, there is no consensus on the matter (Valkenburg, van Driel, and Beyens 2021; Frison and Eggermont 2016; Weinstein 2017): Frison et al. found that chatting with other users on Facebook (i.e. active use) predicted positive affect amongst female adolescents but not male adolescents (2016). The authors also found that passive use predicted depressive symptoms among girls who passively use Facebook and boys who actively used it (Frison and Eggermont 2016). Yet, in another study of adolescents, Lin & Utz found that passive Facebook use was associated more strongly with positive emotions than negative emotions (2015). These differing conclusions suggest that individual differences play an important role in predicting whether a user is likely to experience emotional and psychological harm from their social media use behaviours.

The absence of consensus may arise from a lack of consistent metrics to measure decreases in mental well-being, a pervasive problem in traditional media research (Valkenburg and Peter 2013; Trifiro and Gerson 2019). Furthermore, studies on active and passive use have been conducted on individual platforms and on narrow age demographics (Valkenburg, van Driel, and Beyens 2021; Frison and Eggermont 2016) (Lin et al. 2016; Weinstein 2017). Social media platforms have different types of users and features that evolve over time and are not uniform (Kross et al. 2021), calling into question the validity of the active-passive use dichotomy (Valkenburg, van Driel, and Beyens 2021).

**Social Media Use and Post-traumatic Stress Symptoms**

The association between time spent using social media and PTSs has not been well-researched, but evidence suggests that a similar association may exist between time spent using social media and PTSs: In the wake of the Paris terrorist attacks in 2015, a survey of 1760 French participants found that those who had consumed moderate to high levels of social media content about the attack reported higher levels of PTSs than those who had not (Robert et al. 2021). Recent evidence suggests exposure to pandemic-related news in the initial stages of the outbreak was associated with negative affect, anxiety, depression and stress both among individuals who had knowingly and directly been exposed to the virus and those who had not (Chao et al. 2020). Li et al. found that nurses not involved in caring for COVID-19 patients and members of the general public not directly exposed to the virus experienced higher levels of PTSD symptoms than front-line nurses (2020). The authors theorised that, due to China’s strict isolation policy during the early stages of the pandemic epidemic, non-front-line nurses and the general public felt compelled to consume more COVID-19 media to keep informed about the pandemic, which, in turn, was

### Table 1: Excerpt of Pearson’s r Correlations for Media Use

<table>
<thead>
<tr>
<th></th>
<th>COVID-19 Traditional Media (hrs/day)</th>
<th>COVID-19 Social Media (hrs/day)</th>
<th>Active Social Media Use</th>
<th>Passive Social Media Use</th>
<th>White</th>
<th>Black / African / Caribbean /Black British</th>
<th>East Midlands, England</th>
<th>London</th>
<th>Secondary Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCL-5 Sum Score</td>
<td>PCL-5 Binary Score</td>
<td>Active Social Media Use</td>
<td>Passive Social Media Use</td>
<td>White</td>
<td>Black / African / Caribbean /Black British</td>
<td>East Midlands, England</td>
<td>London</td>
<td>Secondary Education</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PCL-5 Sum Score</th>
<th>PCL-5 Binary Score</th>
<th>Active Social Media Use</th>
<th>Passive Social Media Use</th>
<th>White</th>
<th>Black / African / Caribbean /Black British</th>
<th>East Midlands, England</th>
<th>London</th>
<th>Secondary Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.30***</td>
<td>0.27***</td>
<td>0.24***</td>
<td>0.22***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.14***</td>
<td>0.18***</td>
<td>0.15**</td>
<td>0.15**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.23***</td>
<td>0.21***</td>
<td>0.19***</td>
<td>0.23***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.34***</td>
<td>0.31***</td>
<td>0.30***</td>
<td>0.42***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45***</td>
<td>0.44***</td>
<td>0.41***</td>
<td>0.56***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.35***</td>
<td>0.33***</td>
<td>0.32***</td>
<td>0.50***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.36***</td>
<td>0.37***</td>
<td>0.37***</td>
<td>0.55***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.37***</td>
<td>0.37***</td>
<td>0.40***</td>
<td>0.60***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$**p < .05,** $ $***p < .01,** $ $***p < .001**$
associated with PTSs (Li et al. 2020).

**Methods**

We conducted an internet-based survey using JISC Online Surveys, with recruitment conducted by Prolific Academic. The collection took place over two days in October 2020. Participants were selected as a stratified sample representative of the UK population in terms of age, sex, and ethnicity (Prolific 2022). The online survey did not allow participants to leave questions blank. However, participants were given the option to select a 'prefer not to respond' option for every question, which was treated as missing data.

**Demographics**

Total Household Income, Age, and Education were measured as interval variables and adapted from a recent study by (Zhao and Zhou 2020). Gender was adapted from (Bauer et al. 2017), and ethnicity was measured as a categorical variable adopted from the Harmonised Ethnicity Standard by the UK Government Statistics Service (United Kingdom Government 2020).

**Feelings of Proximity to COVID-19**

Location was determined following the Nomenclature of Territorial Units for Statistics (NUTS) hierarchical classification of administrative areas for the United Kingdom (Office for National Statistics 2021).

**Exclusion of Direct Experience to COVID-19**

Given that our aim was to examine only the secondary effects of exposure to COVID-19, we removed cases that corresponded to direct exposure to COVID-19 as defined by Criterion A (Exposure) for PTSD according to the DSM-V (American Psychiatric Association 2013).

Participants were asked if they had 1) been infected by COVID-19, 2) knew someone close to them who was infected with COVID-19, 3) knew someone who died as a result of contracting COVID-19, or 4) been exposed to it due to their profession (i.e. first responders). Responses were ‘no’ (coded as 0) or ‘yes’ (coded as 1) for each item. Any participant who responded ‘yes’ or ‘decline to respond’ to at least one item was excluded from the analysis.

**Environment**

A single item accounted for the change in working environment. "Are you currently working from home because of the COVID-19 pandemic?" Response options were 1) Yes. I currently only work from home. 2) Yes, I mostly work from home but still leave the house for work sometimes. 3) No. I mostly/only work at a workplace different to my home. 4) No. I am currently unemployed or furloughed. 5) No. I worked from home before the pandemic and still do.

**COVID-19 Stressors**

One dichotomous item, adapted from the checklist for pandemic-related stressors (Main et al. 2011), measured isolation caused by the pandemic with the statement, "I lived alone for a long time due to the COVID-19 pandemic.” Response options were Yes/No.

One dichotomous item measured perceived vulnerability to COVID-19 with the statement, "I consider myself to be part of a group that is at higher risk of developing severe COVID-19 symptoms.” Response options were Yes/No.

**Prior Diagnosis of Mental Health Condition**

Prior diagnoses of anxiety, depression, and PTSD were assessed using dichotomous items questions adapted from (National Center for Health Statistics 2021). Response options were Yes/No.

**Measures of Post-Traumatic Stress**

The PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (PCL-5) was used to assess the severity of PTSs. The 20-item checklist was measured using a 5-point Likert scale, ranging from 0 (not at all) to 4 (extremely) and includes four sub-scales that map the PTSD symptom clusters of the DSM-5 (American Psychiatric Association 2013). The internal consistency of the scale was evaluated with Cronbach’s alpha. Item scores were summed into a total PTSD symptom severity score ranging from 0–80. A score of 31 or above was treated as indicative of probable PTSD (American Psychiatric Association 2013).

**COVID-19 Social Media Time**

This construct measures the general amount of social media use to obtain information about COVID-19. An assessment tool adapted from a prior study by Lin & Utz (2015) was modified to measure time spent on social media accessing COVID-19 information.

The construct of social media use was measured using five items in a grid question format: ‘On average, how many hours per day have you spent accessing pandemic-related information, using the following media, since the lockdown began on March 23, 2020?’ 1) Facebook 2) Instagram 3) YouTube 4) Twitter 5) WhatsApp 6) TikTok 7) Snapchat 8) LinkedIn. Options once again ranged from "less than 1 hour per day” to "more than 12 hours per day.” The scores of the five items were averaged to form the social media use index, and the Cronbach’s alpha score was calculated to verify internal consistency. The items were averaged to form the COVID-19 Social Media (hrs/day) index.

**COVID-19 Traditional Media Use**

The construct was measured using five items in a grid question format: ‘On average, how many hours per day have you spent accessing pandemic-related information, using the following media, since lockdown began on March 23, 2020?’ 1) Television, 2) Newspapers, 3) Radio, 4) Webpages (excluding social media) 5) Other. Options once again ranged from ‘less than 1 hour per day’ to ‘more than 12 hours per day’. The score of the five items was averaged to form the Traditional Media Use Index, and its internal consistency was verified using the Cronbach’s alpha score. The items were averaged to form the COVID-19 Traditional Media (hrs/day) index.
Active and Passive Social Media Use

We adapted the questions from the Multidimensional Scale of Facebook Use to measure both general and pandemic-related social media use across multiple social media platforms (Frison and Eggermont 2016). All questions from the original questionnaire loaded onto three factors: Active public Facebook Use, Active private Facebook use, and Passive Facebook Use on a seven-point Likert scale. We selected this particular scale because it has demonstrated good validity and reliability in prior research, and it is easier to adapt to multiple social media platforms than other comparable scales (Gerson, Plagnol, and Corr 2017). Questions like: "How often do you visit a Facebook profile of someone that does not belong to your friends list?" was modified to "How many times per day do you look at social media content posted by someone you are not friends with, are not following, or being followed by?" One of the two original items pertaining to active private use was dropped due to redundancy when applied to multiple social media platforms. Furthermore, the scale was adapted to ask participants about their COVID-19 social media active and passive use (e.g. 'How many times per day do you chat privately with someone on social media about COVID-19?'). The item was dropped from the general social media use patterns questionnaire and the COVID-19 social media use patterns questionnaire.

Due to the changes made to the scale for general social media use and COVID-19 social media use, a principal component analysis (PCA) was conducted to examine if and how the structure differed from the original scale applied to Facebook. A variance/covariance matrix was employed for the PCA.

Random Forest Regression and Classification

A Pearson’s r correlation analysis was run on all study variables, including the categorical and ordinal variables re-coded as dummies, resulting in 63 variables. Variance inflation was also calculated: Both analyses indicated collinearity between several predictive variables. Given that the aim of our study was to identify key predictors of PTSs and the presence of mixed variable types, we chose to retain all of the variables rather than perform feature selection. We chose to analyse the data using Random Forest regression and classification, as they are suitable for nonlinear and nonparametric modelling frameworks.

We built an RF predictive model for both regression and classification, including all variables as predictors using the randomForest package in R (Liaw and Wiener 2002). We analysed the ranking of predictor variables, which we use to identify the relative importance of our media variables are key predictors of PTSs from our pool of explanatory variables. This methodology presents an advantage over univariate techniques where problems are highly dimensional and variables are highly correlated, which is often the case with observational data collected from human subjects (Han, Guo, and Yu 2016).

Random Forest Regression

For RF regression, the final prediction was calculated by averaging the prediction from individual trees. We used 500 decision trees in the ensemble. For split-point selection, we employed the square root of the total number of predictors when constructing regression trees and one-third of the total predictors for the classification trees (Breiman 2001). Predictor importance from regression trees was obtained based on increases in node purity (important features would be associated with a larger increase in node purity), whereas predictor importance from classification trees was based on the mean decrease in the Gini index, which measures how important a variable is to estimating the outcome variable across the entire variable: A higher mean decrease in the Gini index signifies a more important variable.

Random Forest Classification For the RF classification model, all PCL-5 scores were converted to binary scores. Per clinical guidance, a score above a cutoff of 31 can be treated as 'probable PTSD' (Bovin et al. 2016; Weathers et al. 2020). The classification model was weighted to correct for a severe class imbalance (see Results section). Both classes (low PTSs/high PTSs) were allocated a class weight that was inversely proportional to the number of observations in the class.

For RF classification, the prediction from each tree was a class membership (e.g., Low COVID-19 PTSs/High COVID-19 PTSs), and the final prediction was calculated using a majority voting scheme (whereby the predicted class is the same as the majority class).

In this instance, we used (1) Regression trees to predict the severity of PTSs, and (2) Classification trees to identify if a participant experienced PTSs above the clinical threshold ('high PTSs').

Results

Response Sample

Our initial study sample population of individuals who had not been directly exposed to COVID-19 consisted of 366 cases and constituted a representative sample of the UK population in terms of age, gender, ethnicity, and location (United Kingdom Government 2020). Fourteen of these cases were missing more than 10% of their response data and were eliminated from the analysis.

Social Media Active and Passive Use Scales

A principal component analysis (table 2) was conducted on the ten items from the scales for COVID-19 Active Use, Active Use, COVID-19 Passive Use, and Passive Use. A scree plot and parallel analysis indicated a two-factor solution explaining 61.34% of the total variance. Due to the assumptions of covariance, oblique rotation was applied to the components.

All items pertaining to passive use and private chatting loaded onto one factor, yielding a scale with good internal consistency (Nunnally 1978) of $\alpha = 0.82$, $M = 2.00$, and $\sigma = 0.83$. All items pertaining to active and public use, both general and pandemic-related, loaded onto the other, yielding a scale with acceptable internal consistency (Nunnally 1978) $\alpha = 0.78$, $M = 1.23$, $\sigma = 0.49$. 

$\sigma$.
The mean of the pandemic-related social media time index was averaged to form the PCL-5 index. It yielded a high internal consistency of (\(\alpha = 0.94\)). The mean of the index was \(M = 0.67\), with a standard deviation of \(\sigma = 0.7\). By definition, participants could not meet DSM-5 Criterion A for direct exposure to trauma and no assumption or diagnosis of probable PTSD was made for individuals who scored above the clinical threshold of 31. Instead, the new variable ‘PCL-5 Binary Score’ was created by dichotomising the PCL-5 Sum scores above 31 to ‘High PTSs’ (Bovin et al. 2016; Weathers et al. 2020). The cumulative distribution function of PCL-5 scores indicated that over 90% of the data fell below the threshold score of 31, indicating a strong class imbalance for the PCL-5 Binary Score.

### Media Use Scales

The mean of the pandemic-related social media time index was \(M = 1.35\), \(SD = 1.08\) with a Cronbach’s \(\alpha = 0.912\). A larger mean indicates more average hours of daily use. The mean of the pandemic-related traditional media use index was \(M = 1.77\), \(SD = 1.32\) and Cronbach’s \(\alpha = 0.81\). The scores on five types of traditional media use were averaged to form the traditional media use index. A higher number indicates more hours of daily use of pandemic-related traditional media use. All Cronbach’s alpha values cleared the 0.70 threshold (Nunnally 1978), indicating high internal reliability.

A Pearson’s \(r\) correlation analysis was run on all study variables, including the categorical and ordinal variables recoded as dummies, resulting in 63 variables. Due to the table’s substantial length, only the significant correlations for all media variables are included in this paper in table 1. Daily Traditional COVID-19 media use, Daily Social COVID-19 Media Use, Passive Social Media Use, and Active Social Media Use were significantly and positively correlated with each other, as well as the PCL-5 sum score for PTSD and the PTSD Binary Score.

### Prediction of Post-Traumatic Stress Symptoms

#### Random Forest Regression

The calculated average difference between the predicted and observed PCL-5 scores on an 80-point scale was \(RMSE = 5.04\) and the mean absolute error \(MAE = 3.61\), with predicted PCL-5 sum scores ranging from 1.42 to 48.48. The features that most contributed to the regression model’s predictive power were, in order of decreasing importance: (1) Prior Anxiety Diagnosis, Daily COVID-19 Traditional Media Use, Prior Depression Diagnosis, Passive Social Media Use, Daily COVID-19 Social Media Time, Active Social Media Use, Age: 25-29, Age: 18-24, and Prior Diagnosis of PTSD.

#### Random Forest Classification

The RFC model treated PCL-5 scores above 31 (high PTSs) as a positive class and PCL-5 scores below 31 as the negative class (low PTSs). Predictive features were ranked by mean decrease in Gini index. The RFC yielded a model with an accuracy of 0.99, a sensitivity of 1, and a specificity of 0.992. The top features contributing to the model were, in order of decreasing importance: (1) Prior Anxiety Diagnosis, (2) Prior PTSD Diagnosis, (3) Daily COVID-19 Social Media Time, (4) Active Social Media Use, (5) Passive Social Media Use, (6) Daily COVID-19 Traditional Media Use, (7) Prior Diagnosis of Anxiety, (8) Location: West Midlands UK, Ethnicity: Asian/Asian British, (9) Age: 30-34, Income: £30,000-£39,000/yr, (10) Daily COVID-19 Social Media Time, Age 65 yo and Above, Working from Home since before the pandemic, and (11) Prior Diagnosis of Depression.

### Discussion

Per the random forest regression (fig. 1), the top eight features in the RF regression produced a considerably more significant increase in node purity than the other features. Per the RF classification (fig. 2), the top eleven features of the classification produced a greater mean decrease in Gini coefficient, after which values increased noticeably, signifying a drop in predictive importance. In this section, we discuss the placement of media variables in the feature rankings, as well as the role of non-media variables that are important in both models.

#### Passive and Active Social Media Scales

A PCA of the scales for COVID-19 Active Use, Active Use, COVID-19 Passive Use, and Passive Use yielded a two-component scale. This suggests that active and passive use behaviours are stable across both non-disaster and disaster content. Those who tend to use social media...
when interacting with non-disaster social media will follow similar patterns when dealing with COVID-19 content. While further research is necessary to determine if such behaviours apply to other types of disaster content or if the effect is specific to COVID-19, this finding implies that digital mental health interventions that target type of social media use may help reduce distress related to the use of disaster media content on social media, as well as others.

Importance of Features in Predicting Post-Traumatic Stress Symptoms

**Media Use Variables** The presence of active use as an important feature in the regression model, but not (or at least less) for the classification model, suggests that its impact is likely more strongly mediated or moderated by individual differences than gender for predictions of high PTSs. Future investigation into the role of individual differences in the active use of social media disaster content, such as gender (Frison and Eggermont 2016), is necessary to draw more concrete conclusions (Valkenburg and Peter 2013).

In the regression, passive social media use, active social media use, and COVID-19 social media time were all ranked closely together in the model. This was not the case for the classification model, in which passive use was the most important predictor of PTSs above the clinical threshold for PTSD above any other media variable or risk factor. Our findings are consistent with both the majority of research on passive use, which associates it with negative mental health outcomes (Valkenburg and Peter 2011; Frison and Eggermont 2016) (Burke, Marlow, and Lento 2010) and research on traditional media use, which is by default passive (Ha and Yun 2014). This conclusion is further supported by the relevance of time spent consuming traditional media in both models, consistent with existing research on television and print disaster media use (Silver et al. 2013; Pfefferbaum et al. 2003). The paramount importance of passive social media use suggests that the type of media use is the feature most relevant to predicting high levels of PTSs, regardless of the presence or absence of other risk factors, which may have important implications for future research on digital well-being.

**Prior Mental Health Diagnoses** Prior diagnosis of an anxiety disorder is the most important feature in the RF regression and the third most important feature in the RF classification. Prior diagnosis of depression was also an important predictive feature for both models. These findings are consistent with its importance as a subjective risk factor for PTSD (Worthington, Mandavia, and Richardson-Vejlgaard 2020; Weinberg, Michael and Gil 2016) and disaster media-related trauma (Ahern et al. 2004; Liu et al. 2020), suggesting that it is an essential individual risk factor for developing PTSs through both indirect and direct exposure and disaster media-related trauma symptoms. This continuity of findings suggests that prior diagnoses of anxiety and depression are strong predictors of PTSs.

Interestingly, a prior diagnosis of PTSD was not a high-ranking predictive feature in the RF classification and was a low-ranking predictive feature in the RF regression, which does not concur with most research on PTSD risk factors (Weinberg, Michael and Gil 2016) or traditional media studies (Busso, McLaughlin, and Sheridan 2014) but is consistent with research specific to COVID-19 and PTSs: Ashby et al. found that prior trauma was predictive of PTSs related to COVID-19 only among Asian-American respondents (2022). Given that 83.3% of the participants in the final sample identified as White (which hews closely to the actual proportion in the UK population of 86%), our findings are consistent.

**Implications** Our research contributes to HCI scholarship on the link between social media use and well-being by exploring the connection between forms of social media activity, disaster-related content, and post-traumatic stress, using the COVID-19 pandemic as a relevant context. Our research also has and has broader implications: The UK Online Safety Bill is a piece of landmark legislation which, if passed, would impose a duty of care on social networking service providers to protect users from
exposure to legal but harmful content (Great Britain and DCMS 2021). However, disaster-related content (outside of misinformation or disinformation) or type of social media use are not within scope. Understanding how such content can pose a threat to the health and well-being of the population is critical for guiding both platform policies and legislative efforts to make online spaces safer for users.

**Limitations**

There are considerations regarding the interpretation of this analysis. Random Forests produce predictive models with degrees of accuracy at the cost of interpretability. Their ‘black box’ nature limits our understanding of specific interaction terms used to build the model (Worthington, Mandavia, and Richardson-Vejlgaard 2020). Moreover, the observed associations between the selected features are not evidence of causality.

It should also be stressed that symptoms of depression and anxiety overlap with some symptoms of PTSD and are co-morbid with the disorder (Kessler et al. 2005a; Sveen, Bondjers, and Willebrand 2016). The degree to which our findings explore the relationship between potential risk factors and symptoms of post-traumatic stress, as opposed to other mental health conditions, is therefore unclear. For greater clarity, future quantitative work should incorporate validated scales measuring anxiety and depression to control for those conditions.

Another limitation relates to timing: this study was performed in October 2020, between the UK’s first and second national lockdown. However, several local lockdowns occurred during the months since the initial lockdown ended. Furthermore, the study was performed when cases of death and infection were rapidly rising: Contextual factors may have caused individuals to report the symptoms they were experiencing at the time of the survey rather than their longer-term recollection, creating the potential for recency bias (Vallar 2001).

**Ethics Statement**

This work was approved by the University of Oxford’s Departmental Ethics Review Committee and partially funded by the Information Commissioner’s Office Grant Program.

**Broader Implications** The data collected for this study is meant to be used by other researchers to further the understanding of the relationship between media use and mental health. However, we are mindful that factual information about sensitive characteristics, such as region of residence, ethnicity, gender identity, and health history, was collected as part of this survey. To minimise the risk of re-identification, each participant was automatically assigned a unique participant ID by Prolific, reducing the probability of re-identification.

**Conclusion**

We conducted an investigative study exploring the role of social media use in predicting symptoms of post-traumatic stress in an adult population indirectly exposed to COVID-19. We found that passive social media use is a more important predictor of pandemic-related social media PTSs symptoms than other variables previously associated with disaster media-related trauma.

The significant role of passive social media use in predicting high levels of pandemic-related PTSs suggests that social media use should be further examined in future digital well-being research.

**References**


Patterns: Research Note Regarding the Lack of Universal

Trifiro, B. M.; and Gerson, J. 2019. Social Media Usage European Journal of Psychotraumatology

Sveen, J.; Bondjers, K.; and Willebrand, M. 2016. applications for Classification Prediction Modeling. Psychological Science

McIntosh, D. N.; and Gil-Rivas, V. 2013. Mental- and their biological correlates. Translational Psychiatry


Weinberg, Michael; and Gil, S. 2016. Trauma as an objective or subjective experience: The association between types of traumatic events, personality traits, subjective experience of the event, and posttraumatic symptoms. Journal of Loss and Trauma, 21(2): 137–146.


