Fear and anxiety about COVID-19 have swept across the globe. Understanding the factors that contribute to increased emotional distress regarding the pandemic is paramount—especially as experts warn about rising cases. Despite large amounts of data, it remains unclear which variables are essential for predicting who will be most affected by the distress of future waves. We collected cross-sectional data on a multitude of socio-psychological variables from a sample of 948 United States participants during the early stages of the pandemic. Using a cross-validated hybrid stepwise procedure, we developed a descriptive model of COVID-19 emotional distress. Results reveal that trait anxiety, gender, and social (but not government) media consumption were the strongest predictors of increasing emotional distress. In contrast, commonly associated variables, such as age and political ideology, exhibited much less unique explanatory power. Together, these results can help public health officials identify which populations will be especially vulnerable to experiencing COVID-19-related emotional distress.
Introduction

COVID-19 has killed close to three million people and infected more than 100 million people worldwide, creating a public health, economic, and social crisis that affects the well-being of all segments of the population. Alongside these sobering statistics, the number of reported psychiatric cases has risen sharply in numerous countries since the start of the outbreak, illustrating a global mental health problem (Neelam et al., 2021; Pan et al., 2021). In just the first few months of the pandemic in the United States, there was a significant increase in anxiety (17% rise) and depression (18% rise; Center for Disease Control and National Center for Health Statistics, 2020), which is estimated to cause an economic loss of $1.6 trillion in 2021 alone (Cutler and Summers, 2020). The pandemic is not showing any signs of waning and experts are warning of future deadly waves of cases (Center for Disease Control, 2020b; Chavez et al., 2020; Maragakis, 2020). To help mitigate the psychological toll of the pandemic, it is essential to develop a clear picture of the factors that contribute to increasing COVID-19 distress. To answer this question, we measured a wide range of socio-psychological factors frequently associated with increased emotional distress, as well as factors highlighted by news outlets and policymakers covering COVID-19-related issues in the early days of the pandemic. To assess which of these variables have the most explanatory power, we leveraged a data-driven approach in which we built a psychological profile to predict COVID-19-related emotional distress.

To date, an extraordinary amount of COVID-19 data has been compiled, including information about attitudes, beliefs, and preventive behaviors that can stymie the spread of the virus. This research has understandably focused on documenting the behaviors—and compliance with those behaviors—that are most likely to prevent the transmission of COVID-19 (e.g., hand washing, mask-wearing, etc.). For instance, in the United States, women are more likely to wear cloth face coverings than men (Czeisler et al., 2020), younger respondents are less likely to engage in handwashing (Haston et al., 2020), and young men are less likely to have knowledge about how the virus spreads (Alsan et al., 2020). Moreover, in some countries, such as the United States, COVID-19 has become a politicized issue, such that political ideology and conservative media consumption are both linked to less physical distancing, despite active ‘stay-at-home’ orders (Gollwitzer et al., 2020).

A separate, but related, branch of research has focused on the emotional reactions (Heffner et al., 2021) and mental health consequences of COVID-19 (Vindegaard and Benros, 2020), which together paint an emerging picture of the factors associated with virus-related distress. For example, individuals who align with conservative political perspectives are less likely to report fearing the virus (Conway et al., 2020), and comprise a greater proportion of COVID-19 cases, hospitalizations, and deaths worldwide (Purdie et al., 2021). Research also reveals that older people are less worried about contracting the virus, yet are paradoxically more likely to partake in preventative measures (Andrade et al., 2020; Barber and Kim, 2020). Some factors, such as age, are not consistently linked to COVID-19 fear (Soraci et al., 2020). Other factors, however, such as gender, appear to be a reliable predictor of COVID-19 distress. For instance, females report more mental health issues (Di Crosta et al., 2020), greater perceived stress (Flesia et al., 2020), and an increase in psychiatric symptoms following the transmission of COVID-19 (Mazza et al., 2020). Although this data is informative, it remains challenging to determine which factors are the most important and hold the most explanatory power in predicting COVID-19 distress. Furthermore, since many factors have internal dependence and the relationships between them can collectively influence how a person feels or behaves, separately testing each factor makes it difficult to establish its unique predictive power. In other words, given the litany of factors now associated with COVID-19, which factors in particular are independently critical for predicting increasing emotional distress?

To examine which factors are most predictive of COVID-19 emotional distress, we amassed a set of 30 diverse variables which fall into seven broad categories: COVID-19 media consumption, demographics (i.e., age, gender, political ideology, socio-economic status, education), mental health, personality traits, emotional regulation abilities, and general COVID-19 knowledge and behavior. These variables were selected because of their prior implications in predicting people’s attitudes and behaviors during previous pandemics, or because they were repeatedly mentioned by the news (Aronson, 2020) and public-facing government websites in relation to COVID-19 (Center for Disease Control, 2020a). For example, historical pandemic research revealed that during the 2009–2010 H1N1 influenza outbreak, both the type of media source and the degree to which people consumed news was associated with behavioral compliance meant to attenuate the spread of the virus (Lin et al., 2014). Moreover, specific demographic variables, such as being older, female, and more educated, were also linked to a higher rate of adopting preventative behaviors during the H1N1 pandemic (Bish and Michie, 2010). Even changes in dispositional traits, such as the ability to successfully regulate one’s emotions, have been associated with prior viral outbreaks (Cisler et al., 2010; Zinbarg et al., 2016), and more recently, COVID-19 (Roma et al., 2020). Finally, with the flourish of COVID-19 research on psychological wellbeing, it is becoming increasingly clear that mental health outcomes are compromised during global pandemics (Salari et al., 2020).

Our goal in this work was simple: in order to document a well-characterized psychological profile of COVID-19 emotional distress, we measured all 30 factors simultaneously, aiming to capture each variable’s inter-dependence on one another, as well as their unique predictive power. This approach allows us to take a data-driven approach to identify which factors are the most critical in predicting COVID-19 distress. A large online convenience sample representing the demographics of the United States population completed a battery of well-validated inventories previously shown to reliably assess the variables outlined above. We used this data to build a cross-validated predictive model of emotional distress to COVID-19 (assessed on a 15-item scale). This distress scale was designed to measure the severity of negative emotional experiences (e.g., fear, worry, stress) people felt and ruminated about during the early days of the pandemic.

Methods

Participants. 1000 participants (506 females; mean age = 44.75, SD ± 15.86) were recruited using the online participant platform Prolific between March 24 and 26 of 2020. Prolific uses a “representative sampling approach” which stratifies across three demographics: Age, sex, and ethnicity according to census data (Prolific Team, 2019). Of these 1000 participants, 45 were excluded based on preregistered checks and seven were removed from the analysis because they did not indicate their gender, resulting in a final sample of N = 948 (see Table 1 for demographic information). We preregistered (https://osf.io/2uj6) both our sample size and exclusion criterion (i.e., an attention check asking subjects to place their cursor in the middle of the screen), and our cross-validation approach minimizes the potential for false discovery. Participants completed a battery of questionnaires and were paid $3.50 for a 20-min study. The experiment was approved by the local ethics committee and participants filled out a consent form before beginning the experiment.
Table 1 Participant demographics.

|                | N     |
|----------------|-------|
| Age            | 948   |
| Gender (%)     |       |
| Female         | 491 (51.8) |
| Male           | 457 (48.2) |
| Race (%)       |       |
| White          | 692 (73.0) |
| Black          | 126 (13.3) |
| East Asian     | 42 (4.4) |
| Hispanic Latinx| 37 (3.9) |
| South Asian    | 20 (2.1) |
| Mixed          | 15 (1.6) |
| Other          | 9 (0.9) |
| Native American| 4 (0.4) |
| Middle Eastern | 3 (0.3) |
| Personal income (%) |       |
| <$20,000       | 183 (19.3) |
| $20,000-$34,999| 161 (17.0) |
| $35,000-$49,999| 156 (16.5) |
| $50,000-$74,999| 179 (18.9) |
| $75,000-$99,999| 117 (12.3) |
| $100,000-$149,999| 88 (9.3) |
| $150,000-$199,999| 35 (3.7) |
| >$200,000      | 29 (3.1) |
| Education      |       |
| Mean (SD)      | 15.48 (2.59) |
| Total          | 948   |

Measuring COVID-19 emotional distress. To assess the emotional experiences relating to the COVID-19 pandemic, we asked participants to report on a 0 (does not apply at all) to 100 (strongly applies) scale how generally nervous, calm (reverse coded), worried, and stressed they currently felt. On a second scale, we specifically asked how much time an individual “spent thinking about the coronavirus”, and the “amount of stress, panic, and worry the coronavirus created in their lives” (see supplement for a full list of questions). These two scales tapped into a similar factor (see results below), and were thus analyzed as a single emotional distress index.

Measuring variables related to COVID-19 emotional distress. All participants completed a series of questionnaires in a pseudorandom order designed to capture variables assumed to be related to COVID-19 emotional distress. For the mental health category, we used standard, well-validated, and reliable clinical measures of depression, which include asking questions such as ‘I felt depressed’ (Center for Epidemiologic Studies Depression scale; Radloff, 1977), anxiety (e.g., ‘How often have you been bothered by worrying too much about different things?’, Generalized Anxiety Disorder-7 item scale; Spitzer et al., 2006), and alexithymia (e.g., ‘I am often confused about what emotion I am feeling’ (Toronto Alexithymia Scale-20 items; Bagby et al., 1994). To assess stable personality traits, we relied on the extraversion and neuroticism subscales of the Big 5 Inventory-2-S (e.g., ‘I am someone who tends to be quiet; I am someone who is temperamental, gets emotional easily’ (Soto and John, 2017), and the often used Intolerance of Uncertainty Scale (e.g., ‘Unforeseen events upset me greatly’ (Carleton et al., 2007))—both of which are predictors of emotional distress following stressful events (Oglesby et al., 2016; Ulaşzek et al., 2010). Emotion regulation—an ability known to have explicit ramifications on an individual’s emotional distress levels (Côté et al., 2010)—was assessed with one of the most common emotion regulation questionnaires (e.g., ‘I control my emotions by not expressing them’ (Gross and John, 2003), as well as a measure that indexes how social support systems can scaffold the ability to regulate emotions (e.g., ‘When something bad happens, my first impulse is to seek out the company of others’ (the Interpersonal Regulation Questionnaire; Williams et al., 2018).

To determine a participant’s knowledge of COVID-19, we created a quiz that tapped into how well-informed a participant was given the data available at the time (e.g., ‘Which is NOT a common symptom reported by those who have become infected?’ and ‘If you come into contact with the coronavirus and have to self-quarantine, how long would you have to isolate yourself?’; see supplement for full questionnaire). We probed media consumption and general COVID-19 behaviors by asking participants which source (e.g., Facebook, Twitter, Government sources, television, etc.) they used to get their COVID-19 news, and how often they consumed news (1 = not at all to 5 = a great deal), and what types of preventative behaviors they were partaking in (e.g., handwashing frequency). Finally, we collected typical demographic variables including age, political ideology, gender, socioeconomic status, and education level. Political ideology was assessed on a 100-point sliding scale, indicating how strongly individuals identified with liberal or conservative ideology in the United States (Dodd et al., 2012). Taken together, these assessments capture a diverse, but not exhaustive, set of variables related to the emotional distress of COVID-19.

Results

We first confirmed that all items on the COVID-19 emotional distress scale showed high internal consistency (Cronbach α = 0.93), and exploratory factor analysis using a hierarchical bifactor model (McDonald, 1999) confirmed that it taps into a single underlying factor (ωh = 0.8, ωbifactor = 0.95). Therefore, we aggregated these responses to form a single COVID-19 emotional distress index per participant. We then examined whether we could replicate existing research by examining the simple correlations between the average emotional distress to COVID-19 ratings for the 30 psychological, social, and demographic variables. Our results showed that 24 out of the 30 variables significantly predicted increasing COVID-19 emotional distress, including the following top three in order of importance: anxiety, depression, and neuroticism (Fig. 1). We then developed a multifactorial predictive model of COVID-19 emotional distress using a hybrid stepwise linear regression procedure to elucidate which variables among all 30 are the most important predictors. We randomly selected 25% of this data set to be held out for model testing, which resulted in a 75–25 train-test split. Next, using the training dataset, we conducted a 10-fold cross-validation procedure which folds the training set into 10 equally sized data sets. Using a combination of forward (adding variables) and backward (removing variables) selection using the Akaike Information Criterion (AIC; Venables and Ripley, 2002), we choose the optimal model minimizing the cross-validation error across folds, resulting in a robust fitting model of COVID-19 emotional distress. Figure 2 depicts a network plot of the partial correlations between all variables (only correlations exceeding an absolute value of 0.2) in the best fitting model. The network was generated using a force-directed algorithm (Csardi and Nepusz, 2006) where correlated nodes are closer together and nodes that share more connections are closer to each other. This pattern complements the previous correlation analysis by providing information about the interrelations between variables and COVID-19 emotional distress.

We took the best fitting cross-validated model trained on only the training set (75%) and validated it by predicting the emotional distress of participants in the 25% held-out testing set. The model
explained 46% of the variance ($r^2 = 0.46$) in the testing dataset, indicating that the selected variables were very successful in predicting COVID-19 emotional distress (Fig. 3A; similar predictive accuracy obtained with other statistic procedures including random forest models, $r^2 = 0.48$ and lasso regression, $r^2 = 0.47$). After refitting the cross-validated model on the full dataset, results revealed that, from the 30 original variables, only 16 were needed to maximize predictive power (Table 2). The variables that survived the model, in order of importance, were: trait anxiety (Cronbach $\alpha = 0.93$), gender (female), amount of media consumption, interpersonal emotion regulation (Cronbach $\alpha = 0.93$), political ideology (liberalism), intolerance of uncertainty (Cronbach $\alpha = 0.92$), alexithymia (Cronbach $\alpha = 0.87$), COVID-19 preventive measures (e.g., hand washing, social distancing; Cronbach $\alpha = 0.62$), Facebook and Twitter-related COVID-19 posts, age, and knowledge about COVID-19 (Fig. 3B, note extraversion (Cronbach $\alpha = 0.82$) and social events are included but do not reach significance).

Discussion

COVID-19 has impacted the lives of hundreds of millions of people. Here we examine which individuals experienced the most distress in the early stages of the pandemic. We found that women who are high in anxiety, liberal-leaning, intolerant to uncertainty, and absorb large amounts of COVID-19 media from their social networks on Facebook and Twitter were most susceptible to experiencing increased emotional distress. Of these variables, trait anxiety exerted the strongest influence and was almost three times more predictive than any of the other 30 variables. Gender was the next most influential, with women reporting significantly higher levels of COVID-19 distress than men. Although US-specific media consumption was the third most predictive variable, the media platform by which information was gleaned was instrumental in influencing distress levels. For example, while government sources did not contribute to emotional distress when controlling for other media sources, social network platforms (Facebook and Twitter) and television watching were uniquely linked with higher emotional distress levels. Partaking in preventative behaviors and having more accurate knowledge of COVID-19 had a small but significant impact in increasing COVID-19 distress.

Interestingly, some factors that have been associated with COVID-19 distress when assessed on their own, did not survive our cross-validation procedure. This does not mean these past relationships are incorrect but rather they may be subsumed or contextualized when accounting for a broader set of related variables. For example, while neuroticism and depression have been associated with stress and worry about the pandemic (Faisal et al., 2021; Garbe et al., 2020; Liu et al., 2020; Somma et al., 2020), in our data they do not seem to provide additional explanatory evidence once other variables, namely anxiety, were accounted for. Given that anxiety is primarily associated with worrying about future events (Eysenck et al., 2006), tends to temporally precede depressive disorders (Starr and Davila, 2012), and depression is more closely associated with blunted emotional responding and motivation, may explain why we found anxiety to be a much stronger predictor of COVID-19 distress than depression or neuroticism.

Our cross-validation approach revealed other evidence that helps contextualize existing research. For example, the predictive influence of several variables previously believed to be influential in generating negative emotional responses to COVID-19 was
diminished in the larger model: an individual’s aversion to uncertainty and political-ideological stance (Gollwitzer et al., 2020) provided less predictive power when considered alongside the other factors—which suggests that variables linked to COVID-19 need to be considered together and in the context given the high collinearity between them. Moreover, we found that although extraversion and neuroticism correlated with COVID-19 distress, these personality traits did not provide additional explanatory power above and beyond the other variables. Our research also illustrates the predictive importance of a number of demographic variables, such as age, gender, and education levels—all of which have been previously linked to greater distress levels (Barber and Kim, 2020; Bish and Michie, 2010; Czeisler et al., 2020). When controlling for other psychological variables, gender in particular, appears to be a critical factor associated with increasing COVID-19 emotional distress.

Although these results provide a comprehensive psychological profile of emotional distress related to the COVID-19 pandemic, there are limitations in interpreting survey data. Self-report is notoriously influenced by response biases, and participants may have presented a more positive view of themselves when reporting preventative behaviors (e.g., hand washing, socially isolating). Furthermore, this data was only collected at the beginning of the COVID-19 pandemic (March 2020), and therefore captures a particular snapshot in time. More research is needed to examine how emotional distress related to COVID-19 has changed over the course of the pandemic. Finally, it remains unknown how well this model would generalize to other cultures, where government policies and societal responses to COVID-19 may have been different (e.g., New Zealand).

Taken together, these results can help public health officials identify which populations will be especially vulnerable to COVID-19-related emotional distress. Understanding which variables contribute to pandemic-related emotional distress is paramount for policymakers considering how to prepare for the psychological impact of COVID-19 transmission spikes. COVID-19 has disrupted many aspects of the human experience, and the current levels of uncertainty appear to be contributing to increasing mental health problems. Swiftly tackling these issues will be fundamental for curtailing a mental-health epidemic.

Data availability
Data and analysis code (R markdown) are available at https://github.com/jpheffne/covid_models.

Received: 18 November 2020; Accepted: 19 May 2021; Published online: 09 June 2021

Note
1. It should be noted that the liberal-conservative spectrum in the United States may not mirror the political-ideological range in many other Western countries (Ware, 1996).

References
Alsan M, Stantcheva S, Yang D, Cutler D (2020) Disparities in Coronavirus 2019 reported incidence, knowledge, and behavior among US adults. JAMA Netw Open. 3(6):e2012403–e2012403. https://doi.org/10.1001/jamanetworkopen.2020.12403
Andrade EF, Pereira LJ, Oliveira APLD, Orlando DR, Alves DAG, Guillarducci JDS, Castelo PM (2020). Perceived fear of COVID-19 infection according to sex, age and occupational risk using the Brazilian version of the Fear of COVID-19 Scale. Death Stud 1–10. https://doi.org/10.1080/07481117.2020.1809786
Aronson L (2020). COVID-19 kills only old people. Only? https://www.nytimes.com/2020/03/22/opinion/coronavirus-elderly.html
Bagby RM, Parker JDA, Taylor GJ (1994) The twenty-item Toronto Alexithymia scale—II. Item selection and cross-validation of the factor structure. J Psychosom Res 38(1):23–32. https://doi.org/10.1016/0022-3999(94)90055-1
Barber SJ, Kim H (2020). COVID-19 worries and behavior changes in older and younger men and women. J Gerontol Ser B https://doi.org/10.1093/geronb/gbba068
well-being. American Psychological Association. https://doi.org/10.1037/pspi0000132.

Zinbarg RE, Mineka S, Bobova L, Craake MG, Vrshek-Schallhorn S, Griffith JW, Anand D (2016) Testing a hierarchical model of neuroticism and its cognitive facets: latent structure and prospective prediction of first onsets of anxiety and unipolar mood disorders during 3 years in late adolescence. Clin Psychol Sci 4(5):805–824. https://doi.org/10.1177/2167702615618162

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-021-00816-8.

Correspondence and requests for materials should be addressed to O.F.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021