

# Tracking Mental Health Risks and Coping Strategies in Healthcare Workers' Online Conversations Across the COVID-19 Pandemic

Molly E. Ireland, Kiki Adams, and Sean A. Farrell

Receptiviti

{mireland, kiki, sfarrell}@receptiviti.com

## Abstract

The mental health risks of the COVID-19 pandemic are magnified for medical professionals, such as doctors and nurses. To track conversational markers of psychological distress and coping strategies, we analyzed 67.25 million words written by self-identified healthcare workers ( $N = 5,409$ ; 60.5% nurses, 39.5% physicians) on Reddit beginning in June 2019. Dictionary-based measures revealed increasing emotionality (including more positive and negative emotion and more swearing), social withdrawal (less affiliation and empathy, more "they" pronouns), and self-distancing (fewer "I" pronouns) over time. Several effects were strongest for conversations that were *least* health-focused and self-relevant, suggesting that long-term changes in social and emotional behavior are general and not limited to personal or work-related experiences. Understanding protective and risky coping strategies used by healthcare workers during the pandemic is fundamental for maintaining mental health among front-line workers during periods of chronic stress, such as the COVID-19 pandemic.

## 1 Introduction

The COVID-19 pandemic has magnified existing mental health disparities globally. Relative to people working in other fields, healthcare workers have experienced greater exposure to COVID-19 and, consequently, higher risk of death and illness as well as more time spent apart from loved ones during quarantine (Walton et al., 2020). An estimated 150,000-200,000 healthcare workers have died globally from COVID-19 since the start of the pandemic, with higher rates of infection for nurses, women, and workers involved in COVID-19 screening, and higher mortality rates among doctors (Chutiyaami et al., 2021; WHO, 2022). Deaths and illnesses among healthcare workers have led to severe understaffing in the hardest-hit areas, causing widespread overwork and burnout in the

healthcare field (Andel et al., 2022). Healthcare workers experienced higher rates of depression and suicide than many other professions before the pandemic (Kalmoe et al., 2019), and suicidality, depression, and anxiety disorders have increased among healthcare workers in the last 2 years (Spoorthy et al., 2020; Young et al., 2021).

Beyond pandemic-related social isolation, personal health risks, and overwork, healthcare workers additionally cope with feeling responsible for the deaths and symptoms they witness firsthand in their patients (Zhang et al., 2020)—experiences exacerbated early in the pandemic by the fact that healthcare workers were often the only people permitted to be physically present in patients' final hours (Rabow et al., 2021). For many, the stress of the pandemic has been aggravated by widespread skepticism of vaccines and the medical system (Schneider et al., 2021). Others have reported survivors' guilt related to having early vaccine access, feelings of powerlessness with respect to limited COVID-19 patient treatment options, and the chronic stress of having insufficient personal protective equipment while working, particularly early in the pandemic (Rabow et al., 2021).

Dealing with chronic stress at the front line of an epidemic or pandemic requires extraordinary coping and emotion regulation skills—and, at the same time, likely compromises the mental health of even the most resilient nurses and doctors. In this project, we followed the linguistic trajectories of healthcare workers' risky and protective coping strategies over the course of the pandemic. The following sections first review past research on risk factors and resilience evident in language use following community traumas. We then describe a longitudinal study tracking social and emotional language used by several thousand self-labeled nurses and doctors on Reddit, a popular online social discussion platform, over a baseline period followed by roughly 2 years of the pandemic. Analyses focused on main

effects over time and moderator models exploring how language trajectories varied as a function of health-relevance, self-relevance, and role (nurse or physician). Finally, we explore the ethical, theoretical, and practical implications of the findings for clinical psychology and mental health technology.

### 1.1 Coping with Shared Trauma over Time

Tracking naturalistic language use on the internet is an effective method of measuring how people cope with trauma and experience emotions over time (Vine et al., 2020). Research has, for example, used both dictionary-based and open-vocabulary analyses of online language use (including social media, online forums, and search engine activity) to understand how individuals anticipate and then cope with traumatic events such as suicide attempts (De Choudhury et al., 2016; Ophir et al., 2020; Roy et al., 2020), relationship dissolution (Seraj et al., 2021), illnesses such as breast cancer (Verberne et al., 2019) and autoimmune disease (Jordan et al., 2019), and mental health conditions such as anxiety (Ireland and Iserman, 2018) and depression (Eichstaedt et al., 2018).

Several studies of community coping with shared traumas—such as the September 11<sup>th</sup> attacks and natural disasters—have found evidence of both distress and coping in naturalistic conversational language. Results show a common pattern of increasing affiliative and emotional language in the immediate 1-2 weeks after a traumatic event followed by a refractory period during which such communal coping indicators drop below baseline, theoretically reflecting social withdrawal (Cohn et al., 2004; Pennebaker and Harber, 1993; Stone and Pennebaker, 2002). For acute traumas, language typically returns to baseline after 4-6 weeks (Pennebaker and Chung, 2005).

Analyses of social media language use surrounding epidemics (e.g., Zika, Ebola) and sociopolitical movements (e.g., the Arab Spring) have focused primarily on the transmission of information about symptoms or events rather than psychological dimensions of messages (Hassan Zadeh et al., 2019; Howard et al., 2011). Previous analyses of psychological language use during epidemics or disease outbreaks have typically focused on tracking markers of distress over short spans of time. For example, Tausczik et al. (2012) tracked anxiety language in tweets about the H1N1 epidemic, revealing that fears about H1N1 were intense but short-lived, de-

clining within weeks of the initial news about the disease.

At least one study has used dictionary-based measures to track coping across the first months of the COVID-19 pandemic. Based on a large Reddit sample of people posting in major U. S. city forums, Ashokkumar and Pennebaker (2021) found that anxious language spiked and positive emotional, angry, and analytic language dropped in March 2020. People also referred less to friends and more to family early in the pandemic. After roughly 6 weeks, these language categories plateaued but remained distinct from pre-pandemic levels in the previous year. It is unclear whether these patterns vary as a function of individuals' life stressors or will continue to shift over time.

### 1.2 Linguistic Markers of Distress

Overwork compromises mental health and has downstream consequences for the quality of individuals' close relationships and job performance. There are several potential indicators of burnout and work stress that may carry over from the workplace to online conversations. The clearest linguistic markers of distress and vulnerability to mental health conditions tend to be self-references (*I, me, my*) and negative emotional language, alone and particularly in combination (Baddeley et al., 2013; Coppersmith et al., 2015a; Tackman et al., 2019).

Work-related stress has disrupted healthcare workers' relationships throughout the COVID-19 pandemic. Long-term quarantining away from romantic partners and family members due to frequent exposure to the disease increases loneliness and relationship conflict (Murata et al., 2021). Relationship problems are closely linked with mental health; for example, breakups and relationship conflict are common triggers of suicide attempts (Bagge et al., 2013) and depressive episodes (Monroe et al., 1999). Thus, in tracking healthcare workers' conversational language use over the pandemic, it is critical to target linguistic markers of affiliation and social behavior.

### 1.3 Linguistic Markers of Coping

Just as self-directed negativity is a common indicator of psychological distress, the opposite pattern tends to reflect efforts to gain emotional distance from personal problems—a tactic that provides relief in the moment but may be risky long term. Research on self-talk and expressive writing has found that people tend to naturally self-distance, using

less “I” and more “you,” when recalling negative events or while discussing stressful events, with stronger effects for more distressing topics (Dolcos and Albarracin, 2014; Kross and Ayduk, 2017). The same strategy is effective experimentally as well, with people experiencing less distress when asked to write about negative life experiences or personal concerns using self-distancing (e.g., writing *you* instead of *I*). Psychological distance theoretically provides an emotional buffer, allowing people to consider the events that are causing them distress from the more objective perspective of an outside observer or friend. Thus, lower first-person singular pronoun usage may be a healthy emotion regulation strategy, especially when experiencing acute distress.

Despite the well-established body of work showing that self-distancing can help with emotional control and distress, decreased first-person singular pronoun is not an unambiguous sign of effective coping. In contexts where self-references indicate self-disclosure or self-reflection, using more “I”—or alternating between “I” and other personal pronouns—may be healthier. For example, people with ambiguous sexual self-concepts who used less first-person singular when discussing their sexuality were more likely to report drinking alcohol to cope with personal problems (Hancock et al., 2018). In expressive writing, where people repeatedly privately write about their deepest thoughts and feelings on a distressing topic, individuals tend to have better long-term mental and physical health after the writing intervention if their language indicates a perspective shift (moving from high to low self-references, or vice versa) across sessions (Pennebaker and Chung, 2007; Seih et al., 2008).

Separate research on compassion has found that discussing others’ suffering in a less emotional, more socially distant way is associated with better mental health and greater likelihood of taking proactive steps to help the people who are suffering or need assistance (Buechel et al., 2018; Ministero et al., 2018). That is, people may be better able to provide assistance if they feel others’ pain less acutely. These findings dovetail with research and practice regarding healthcare workers’ bedside manner, where the goal is to show humanistic compassion for patients while maintaining enough distance to carry out complex and often risky and painful tasks (Weissmann et al., 2006).

| Word category                | Examples                       |
|------------------------------|--------------------------------|
| <i>Function Words</i>        |                                |
| First-person singular (“I”)  | <i>I, me, my</i>               |
| Third-person plural (“they”) | <i>they, them, their</i>       |
| Negations                    | <i>no, not, never</i>          |
| <i>Affect</i>                |                                |
| Positive emotion             | <i>lucky, love, happy</i>      |
| Amusement                    | <i>haha, lol, funny</i>        |
| Admiration                   | <i>cool, amazing, best</i>     |
| Negative emotion             | <i>hate, worry, sad</i>        |
| Disgust                      | <i>creepy, vomit, ugh</i>      |
| <i>Social</i>                |                                |
| Affiliation                  | <i>call, party, together</i>   |
| High empathy                 | <i>ally, rescue, we</i>        |
| Low empathy                  | <i>yourself, asshat, waste</i> |
| Prosocial                    | <i>help, support, thanks</i>   |
| Swear words                  | <i>dang, fuck, douche</i>      |

Table 1: Social and emotional language categories showing significant linear or curvilinear effects over time. Linguistic categories, affiliation, swear words, and prosocial are from LIWC-22 (Pennebaker et al., 2022). Affect categories are from SALLEE (Adams, 2022). High and low empathy are novel lexica adapted from Sedoc et al. (2020).

#### 1.4 Hypotheses & Analytic Strategy

The current project took a quasi-exploratory approach, modeling the trajectories of a wide range of language variables that are theoretically relevant to risky and protective emotions and social behaviors (see Table 1). The main predictions were that healthcare workers would show signs of increasing distress (more negativity, less positivity), social detachment or isolation (more *I* and *they*, fewer social references, less empathetic language), and social problems (increased conflict and swearing, and decreased prosocial and polite language) over time. Linear, quadratic, and cubic associations were tested for all models. Finally, we tested three moderators for each model: professional role (nurse or doctor) and two aspects of linguistic context (first-person singular pronouns and references to health, e.g., *medicine, symptom, vaccine*).

## 2 Method

To obtain the initial sample, we first scraped a large sample of comments and submissions from medical-themed forums, or subreddits (r/medicine, 312,357 posts; r/nurses, 14,927 posts; r/emergencymedicine, 46,019 posts; r/AskDocs, 1,617,327 posts; r/StudentNurse, 191,525 posts), that appeared to be moderated by healthcare professionals and included “flair” indicating users’ real-life qualifications or specializations. Initially, 2,182,155 messages posted between October 2018 and January 2021 were scraped using the Pushshift

| Rank                | Content  | $M$ ( $SD$ ) | Function | $M$ ( $SD$ ) |
|---------------------|----------|--------------|----------|--------------|
| <i>High Empathy</i> |          |              |          |              |
| 1                   | love     | 0.13 (0.98)  | and      | 2.03 (1.78)  |
| 2                   | great    | 0.1 (0.68)   | are      | 0.57 (1.14)  |
| 3                   | feel     | 0.1 (0.45)   | your     | 0.4 (1.02)   |
| 4                   | thank    | 0.09 (0.91)  | we       | 0.33 (0.84)  |
| 5                   | help     | 0.09 (0.67)  | how      | 0.26 (0.88)  |
| 6                   | patients | 0.08 (0.36)  | her      | 0.15 (0.63)  |
| 7                   | hospital | 0.07 (0.41)  | our      | 0.13 (0.56)  |
| 8                   | home     | 0.07 (0.4)   | us       | 0.1 (0.51)   |
| 9                   | patient  | 0.07 (0.37)  | through  | 0.07 (0.34)  |
| 10                  | life     | 0.06 (0.49)  | omg      | 0.01 (0.62)  |
| <i>Low Empathy</i>  |          |              |          |              |
| 1                   | time     | 0.22 (0.69)  | the      | 3.29 (2.77)  |
| 2                   | think    | 0.18 (0.56)  | in       | 1.31 (1.7)   |
| 3                   | new      | 0.14 (0.8)   | that     | 0.97 (1.35)  |
| 4                   | going    | 0.12 (0.46)  | but      | 0.62 (0.93)  |
| 5                   | same     | 0.11 (0.95)  | be       | 0.54 (0.97)  |
| 6                   | better   | 0.09 (0.53)  | not      | 0.53 (1.08)  |
| 7                   | thing    | 0.08 (0.41)  | if       | 0.45 (0.78)  |
| 8                   | say      | 0.08 (0.39)  | like     | 0.41 (1.06)  |
| 9                   | lol      | 0.07 (0.87)  | one      | 0.29 (0.97)  |
| 10                  | long     | 0.07 (0.47)  | up       | 0.27 (0.77)  |

Table 2: Frequency ranks and descriptive statistics for content and function words in the high and low empathy lexica. All numbers are % of total words per document (concatenated messages per user per month).

API (<https://github.com/pushshift/api>). Doctors and nurses were categorized via regular expression searches over the flair text of these messages, searching for commonly used phrases and acronyms used by medical professions (e.g., MD, M.D., MBBS for doctors; Nurse, PCCN, Nursing, NP, LPN, CAN, RN, R.N., BSN for nurses). A total of 2,585 doctors and 4,138 nurses were identified. Next, we downloaded all available comments and posts from the 6,723 self-labeled doctors or nurses on Reddit, totaling over 1.25 million texts, beginning in June 2019. The start date was selected in order to establish baseline norms for the sample, providing roughly 6 months of data from before the virus began spreading globally and 9 months before the WHO declared a pandemic.

Texts were concatenated by user and then by month, excluding months containing fewer than 100 words. We also excluded months for which fewer than 50% of the words were recognized by our dictionaries; given that over half of conversational language typically consists of function words (“stop words” such as pronouns and articles), texts containing half or more words that were not captured by our lexica are unlikely to be conversational English. Finally, we excluded months in which all punctuation made up 50% or more of the text (indicating, e.g., ASCII art). The final dataset included 5,409 unique users ( $n = 3,271$  or 60.5% nurses;  $n$

$= 2,138$  or 39.5% medical doctors) and 67,247,147 words ( $M = 1,090$ ,  $SD = 2,355$ , median = 434 words per user per month).

For mixed-effects regression modeling, we regressed language markers on time (linear, quadratic, and cubic effects), including random slopes for time, nested within authors, and specifying an autocorrelation structure of order 1 (corAR1) to account for the non-independence of adjacent (lag-1) months; all models used the nlme package (Pinheiro et al., 2021) and were plotted with splot (Iserman, 2022) in R version 4.2.0 (R Core Team, 2022). To simplify the time variable and make the regression coefficients more interpretable, we transformed months into quarters and then assigned each quarter a sequential number, starting with Q3 2019 as sequence 0 and ending with Q1 2022 as sequence 11. We then squared and cubed the sequence variable to create the polynomial predictors.

All references to statistical significance below use an adjusted p-value threshold rather than the traditional .05 in order to partly account for inflated false discovery rates, or Type I errors, due to multiple comparisons. Mixed-effects regression models tested effects for 30 language variables, each of which were explored in mixed-effects regression models including six tests (three linear and polynomial effects and three moderators). Thus, the corrected  $\alpha$  level is .00028 using the Bonferroni method, a conservative but intuitive correction that is suitable for exploratory analyses in large samples (VanderWeele and Mathur, 2019).

Data collection methods and analytic strategies were approved by internal ethical review at Receptiviti, Inc. and meet federal guidelines for exempt research under the U. S. Department of Health and Human Services’ (2017) revised Common Rule. Consistent with the Reddit API’s User Agreement, all quantitative data are available online, and we will not profit from the use of these data.

Deidentified data, the high and low empathy lexica we developed, and R code for downloading individuals’ messages can be accessed via the project’s Open Science Framework (OSF) page.<sup>1</sup>

## 2.1 Language Measures

**LIWC and SALLEE.** Texts were analyzed using the latest version of the Linguistic Inquiry and

<sup>1</sup>[https://osf.io/scmb7/?view\\_only=53e8bd3359b3460a907d19f5cb5a0ef6](https://osf.io/scmb7/?view_only=53e8bd3359b3460a907d19f5cb5a0ef6)

| Word category         | Linear $\beta$ | Quadratic $\beta$ | Cubic $\beta$ |
|-----------------------|----------------|-------------------|---------------|
| <i>Function Words</i> |                |                   |               |
| I                     | -0.09          | -0.18             | <b>0.18</b>   |
| They                  | -0.03          | <b>0.42</b>       | <b>-0.30</b>  |
| Negations             | 0.01           | 0.25              | <b>-0.20</b>  |
| <i>Affect</i>         |                |                   |               |
| Emotionality          | -0.1           | <b>0.55</b>       | <b>-0.39</b>  |
| Positive emo.         | -0.12          | <b>0.58</b>       | <b>-0.42</b>  |
| Amusement             | -0.05          | <b>0.41</b>       | <b>-0.28</b>  |
| Admiration            | -0.14          | <b>0.59</b>       | <b>-0.43</b>  |
| Negative emo.         | 0.01           | 0.13              | -0.09         |
| Disgust               | -0.09          | <b>0.41</b>       | <b>-0.27</b>  |
| <i>Social</i>         |                |                   |               |
| Affiliation           | <b>0.18</b>    | <b>-0.63</b>      | <b>0.42</b>   |
| High empathy          | <b>0.22</b>    | <b>-0.66</b>      | <b>0.42</b>   |
| Low empathy           | -0.06          | <b>0.47</b>       | <b>-0.34</b>  |
| Prosocial             | 0.15           | <b>-0.45</b>      | <b>0.27</b>   |
| Swear words           | -0.07          | <b>0.47</b>       | <b>-0.32</b>  |
| Question marks        | 0.01           | <b>-0.45</b>      | <b>0.34</b>   |

Table 3: Standardized  $\beta$  from the polynomial mixed-effects regression model controlling for role (doctor or nurse) and including linear, quadratic, and cubic sequence (time) effects, random slopes for time within authors, and random intercepts for authors. *I* = first-person singular pronouns, *they* = third-person plural pronouns, *emo.* = emotion. Bold = two-tailed  $p \leq .0003$ .

Word Count, LIWC-22 (Pennebaker et al., 2022; Boyd et al., 2022) and a sentiment analysis framework, SALLEE (Syntax-Aware Lexical Emotion Engine; Adams 2022). LIWC is a widely used and well-validated dictionary-based text analysis tool that outputs the percentage of words in a given text that fall into one or more of several dozen grammatical (e.g., pronouns, articles), psychological (e.g., emotions, drives), and topical (e.g., work, health) categories. SALLEE is dictionary-based as well, providing fine-grained measures of specific emotions (e.g., curiosity, surprise, disgust) and summary affective states (e.g., emotionality, positive emotion) in addition to using syntax-based logic allowing words adjacent to emotion terms (e.g., swear words, negations, and intensifiers) to influence category weights (Adams, 2022).

**Empathy lexica.** The high and low-empathy lexica were both adapted from the data-driven empathy dictionary developed by Sedoc et al. (2020), which was initially trained on a gold-standard empathic reaction corpus (Buechel et al., 2018). For the revised dictionaries, we first took words in the highest and lowest-weighted quartiles of Sedoc et al.’s (2020) empathy dictionary. We then removed person and place names (e.g., *Abuja*, *Charles*; excepting names used synonymously with low or high empathy, such as *Bundy* and *Gandhi*, respectively), low-frequency misspellings (e.g., *entrapreneurship*), numerals, and other words that appeared

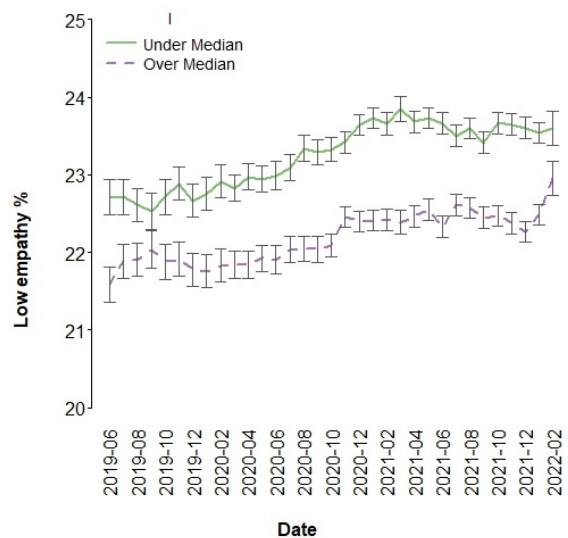


Figure 1: Low empathy language (% of total words) as a function of first-person singular pronoun ("I") usage. Error bars are standard errors.

to be highly contextual or time-specific. Removal judgments were made by the authors, with disagreements resolved through discussion. Wildcards were added sparingly to capture additional word variants where it was safe to do so (e.g., *ambulance\**), and missing British English spellings (e.g., *analyse*) were added. Finally, we separated content and function words in order to explore whether effects were robust across both types of words. The final revised dictionaries included 4,059 words (2,105 low empathy, 1,954 high empathy).

Changes to the original empathy dictionary (Sedoc et al., 2020) were not intended not to improve measurement accuracy; rather, we aimed to increase interpretability and generalizability, with the long-term goal of making the lexica accessible to clinicians and mental health care providers. As in the original dictionary, high empathy words in the revised lexica focused primarily on suffering (e.g., *ravaged*, *hurt*, *lost*) using expressive (e.g., *emotions*, *feel*), prosocial language (e.g., *provide*, *reunite*), whereas low empathy language included unemotional or technical words (e.g., *acknowledge*, *result*) and disagreeable or insensitive language (e.g., *idgaf*, *trashy*). For examples used in the current sample, see Table 2. The two dictionaries were moderately negatively correlated,  $r = -.278$ .

### 3 Results

Regression results were consistent with our hypotheses with a few notable exceptions. Both effect sizes and AIC comparisons indicated that cubic

models were the best fits for the data, and quadratic models were always a better fit than linear models, based on the  $\Delta AIC > 2$  criterion. The standard pattern was an approximately flat line at baseline followed by relatively sharp changes over the first year of the pandemic followed by another plateau or period of more gradual change in the same direction (see Table 3), similar to the overall patterns found by Ashokkumar and Pennebaker (2021).

For the social language categories, doctors and nurses both showed increasing rates of low-empathy words (a pattern that was strongest for less self-referential language; see Figure 1), swearing (Figure 2), and social detachment (more "they" pronouns, Figure 3) over the course of the pandemic. In parallel, healthcare workers showed decreasing rates of words reflecting or referring to social harmony and social engagement (high-empathy, prosocial, affiliation, and question marks) over time.<sup>2</sup>

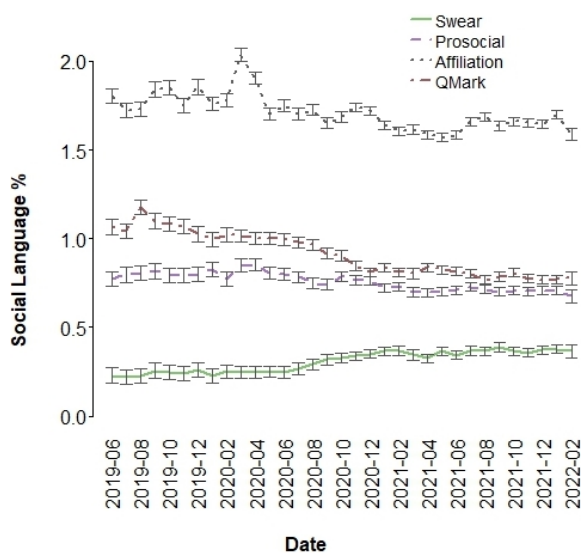


Figure 2: LIWC social categories (affiliation, prosocial, question marks, swear words) with significant polynomial effects over time.

Contrary to our predictions, first-person singular pronouns (e.g., *I*, *me*, *my*) decreased over the first year of the pandemic and then plateaued at a relatively low level (Figure 4). Nurses in particular used markedly less "I" (5.2% to 4.2%) from baseline to early 2022. Doctors' first-person singular usage was lower than nurses' at baseline (3.9%), perhaps reflecting physicians' relatively higher status (Kacewicz et al., 2014).

<sup>2</sup>For high empathy, effects were parallel and the conclusions of hypothesis tests were identical when function words were removed from the lexicon. For low empathy, results were not significant after removing function words, all  $t < |2|$ .

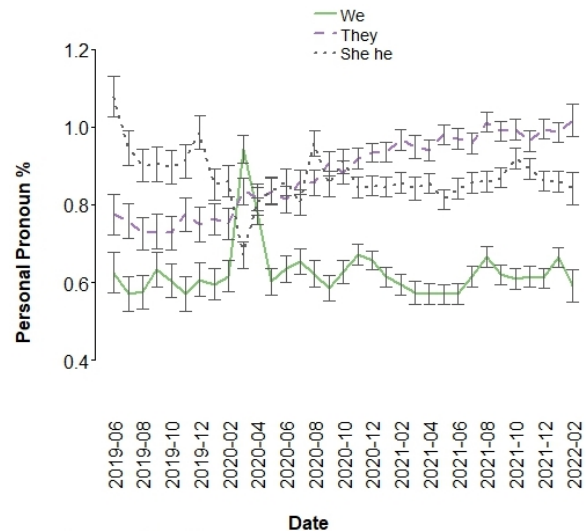


Figure 3: Other-focused pronouns over time. We = first-person plural, they = third-person plural, she/he = third-person singular.

The emotional language results were partly consistent with our predictions. As expected, emotionality and some negative emotions (namely disgust) increased over time. However, most negative emotion categories did not change significantly over time (e.g., sadness, fear). More surprisingly, overall positive emotional language increased over time, with amusement and admiration showing the strongest effects for specific emotions (Figure 5). Amusement is a low-frequency category ( $M = 0.71$ ,  $SD = 1.62$ ; 56.6% of months had 0% amusement) but showed robust quadratic and cubic effects.

Results for words referring to politeness and conflict from LIWC-22 were nonsignificant, despite showing the predicted trends (increasing conflict and decreasing politeness over time), both  $ps > .10$ ,  $ts < |3|$ . Those categories' low base rates ( $M = 0.32\%$  and  $0.24\%$ , respectively) may have limited our ability to detect subtle shifts over time.

### 3.1 Moderation by Health and Self-Relevance

For most variables, effects were not moderated by whether the conversations focused on health. There were a few exceptions: for swearing, positive emotions, and disgust, effects were strongest for conversations that were *not* about health. That is, changes in healthcare workers' language over time do not appear to be driven by online discussions of COVID-19 or challenges in their jobs as nurses and doctors; rather, linguistic changes were most evident in casual conversations about interests or hobbies, suggesting that the coping strate-

gies that people have developed in response to the exigencies of the pandemic are carrying over into everyday conversations.

Moderation by self-referential language (*I*, *me* and *my* usage) was mixed. The overall pattern was for effects to be stronger for negative categories (negations, negative emotions, low empathy) when people were *not* talking about their own experiences; conversely, effects were strongest for positive or prosocial categories (affiliation, positive emotion, and high empathy) when people *were* talking about themselves. Such patterns are consistent with the self-protective tendency to distance oneself from negativity (Ayduk and Kross, 2010). People may feel more comfortable venting (e.g., expressing disgust) when not talking about themselves.

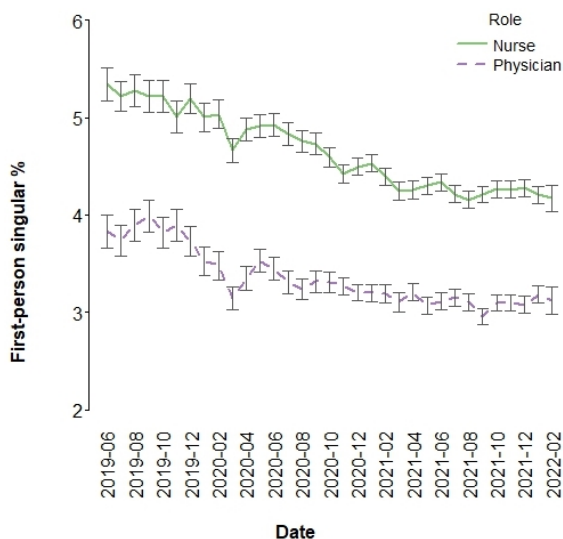


Figure 4: LIWC first-person singular pronoun usage (% of total words) as a function of role.

### 3.2 Additional Analyses

**Early Pandemic Spikes.** Many of the plots show deviations at the start of the pandemic followed by linear or flat patterns. First-person singular pronouns dipped sharply in March 2020 followed by a return to near baseline and then a gradual decrease over time. Affiliation language spiked in the first month of the pandemic, followed by a slow linear decline. Although there was no overall linear or curvilinear effect for first-person plural pronouns, there is a clear spike at the start of the pandemic where "we" increases and other pronouns drop before quickly returning to near baseline (Figure 3). Sadness and fear spiked in the same month, declined, and then increased gradually in the following months.

**Word-Level Analyses.** To better understand the results from the most data-driven (and thus least immediately intuitive) dictionaries, high and low empathy, we examined word-level frequencies. Table 2 shows that the most frequently used low-empathy content words are not rude or callous per se, but seem to reflect a degree of detachment (e.g., *lol*, *things*, *week*, *think*). Low empathy function words had some overlap with LIWC's composite analytic language category, including an article (*the*), impersonal pronouns (*that*, *there*), and prepositions (*up*, *in*)—all of which reflect more formal, categorical thinking—as well as negations (*no*, *not*, *never*).

**New Case Rates.** Monthly global new case rates (cases per million; Hannah Ritchie and Roser 2020) were largely uncorrelated with the language variables of interest in this study. In mixed-effects models regressing new case rates onto language variables, none met a  $p < .001$  cut-off. The strongest effect was for first-person singular pronouns, quadratic effect  $\beta = .013$ , 95% CI [0.005,0.021],  $SE = .004$ ,  $p = .002$ . However, controlling for new case rates as a covariate did not affect the conclusions for any models involving changes in first-person singular over time.

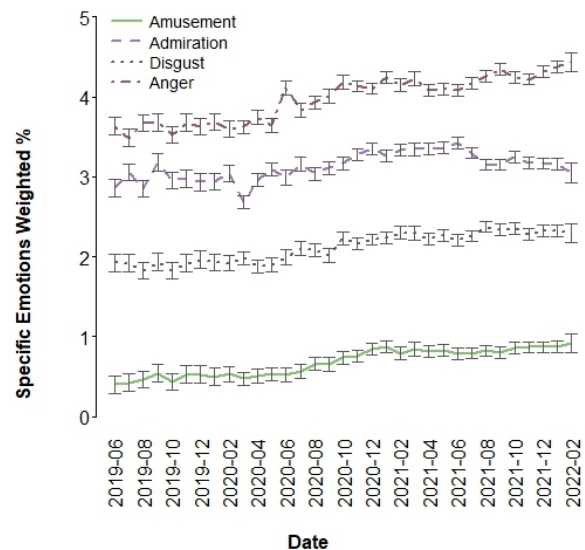


Figure 5: SALLEE emotions (amusement, admiration, disgust, and anger) that increased over time. All cubic effects except for anger are significant,  $p < .001$ ; anger showed a nonsignificant but positive trend.

## 4 Discussion

The online conversational language of doctors and nurses over the course of the pandemic shows a coherent picture of people coping with chronic stress by self-distancing (fewer first-person singular pro-

nouns) and adopting a more socially detached perspective (less empathic and affiliative language). At the same time, healthcare workers did not seem to be eschewing emotions; rather, emotional language increased over time, including more references to disgust and positive emotions in general.

The emotional effects should be qualified by the standard caveats of any language-based sentiment analysis: Affective words, when categorized correctly, indicate that a person is attending to and talking about an emotion—which sometimes but not always correlates with their emotional state at the time of speaking or writing (Sun et al., 2020; Eichstaedt et al., 2021). Thus, increases in positive emotional language may reflect emotion regulation attempts or coping strategies more than improvements in well-being or mood. What is most striking is not that positive emotional language increased near the end of our sample—which could be explained by decreasing case rates and a slow return of pre-pandemic freedom in much of the world—but that positive emotionality only dropped notably during the first month of the pandemic and did not decrease again during later spikes in global case or mortality rates (Figure 6). Indeed, post hoc analyses show that positive emotional language correlated weakly with global new case rates per million,  $r = .015$ . That pattern may support the supposition that positive language shifts reflected coping strategies (such as positive reframing) rather than overall well-being (Robbins et al., 2019).

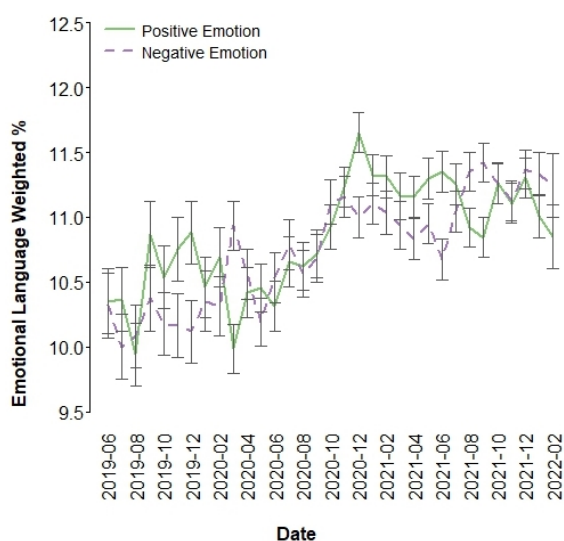


Figure 6: SALLEE composite positive and negative emotions. Only positive emotion showed significant linear, quadratic, and cubic effects; negative emotion is shown for context.

First-person singular pronouns (e.g., *I, me, my*) decreased across the pandemic, a pattern that was starker for nurses than for doctors. "I"-words typically indicate vulnerability to psychological distress (i.e., neuroticism or trait negative affectivity; Tackman et al. 2019) and mental health concerns related to affect dysregulation, including depression (Bucur et al., 2021; Holtzman et al., 2017), anxiety (Brockmeyer et al., 2015; Shen and Rudzicz, 2017), eating disorders (Coppersmith et al., 2015a), and suicidality (Coppersmith et al., 2015b; Stirman and Pennebaker, 2001). Although the results were inconsistent with general psychological distress, a pattern of decreasing first-person singular pronoun usage is consistent with using self-distancing as a self-regulation strategy during periods of chronic stress. People tend to naturally decrease "I" pronouns as a means of distancing themselves from distress and downregulating negative emotions (Ayduk and Kross, 2010; Dolcos and Albarracín, 2014). Coupled with less empathic language over time, however, decreased "I" rates may represent an adaptation to chronic stress that helps preserve mental stability in the moment but leads to stress dysregulation and interpersonal problems in the future, after the period of severe stress has passed (Ellis et al., 2017).

#### 4.1 Potential Applications

Occupational burnout has intensified throughout the pandemic, particularly for jobs that entail regular risk of exposure to the virus that causes COVID-19. The healthcare field has been among the most affected (Alrawashdeh et al., 2021), with women in particular experiencing more intense and debilitating burnout (Sriharan et al., 2021), as in other professions, partly as a result of gender inequality in the distribution of family responsibilities and household chores while working from home (Malisch et al., 2020). Being able to unobtrusively profile work-related stress or burnout in available texts (e.g., internal chats, emails) could help employers direct mental health resources to employees at risk of mental health crises before their symptoms become severe or their work is affected.

Before translating our findings to clinical or industrial/organizational practice, it will be necessary to disentangle which long-term or acute changes in language use are helpful or harmful. Some of healthcare workers' linguistic changes over time may be beneficial in the short-term but have long-



term costs. For example, as already noted, self-distancing decreases distress in the moment (Kross and Ayduk, 2017) but may have long-term psychological costs (Hancock et al., 2018), parallel to the psychological and social toll of keeping major life secrets (Tausczik et al., 2016), refusing to discuss conflicts with romantic partners (Laursen and Hafen, 2010), or avoiding thoughts about traumatic experiences (Pennebaker, 1989, 2018). Indeed, people who use less authentic language (a composite measure that includes "I" pronouns) tend to be perceived as less likable and credible in social and entrepreneurial contexts, likely because first-person singular pronouns are a necessary part of self-disclosure and intimacy (Markowitz et al., 2022). Therefore, increasing self-distancing over time may lead to social and occupational fallout. Further research should confirm which linguistic markers of chronic stress may be harmful before implementing any language-based intervention.

## 4.2 Limitations

As with many archival samples of naturalistic conversations online, the current sample is limited by a lack of information about the users. It is not possible to verify each user's healthcare work experience, nor can we conclusively assess demographic characteristics or personality traits that may clarify or qualify our findings. Reddit users are diverse and global, but tend to skew American, young, and masculine (Gjurković et al., 2021). Although language-based models can estimate such individual differences (Eichstaedt et al., 2021), linguistic cues to mental health such as negative self-focus (Baddeley et al., 2013) are often confounded with gender, age, and culture. For example, younger people and women tend to use "I" more (Pennebaker and Stone, 2003; Tausczik and Pennebaker, 2010), and negative affect is less stigmatized in East Asian than in Western cultures (Park et al., 2020).

The results are also limited by the relatively short baseline period. Using a longer 1 or 2-year pre-pandemic sample would have more appropriately accounted for seasonality, i.e., cyclical patterns over time operating independently of but sometimes confounded with the variables of interest (Brendstrup et al., 2004).

Finally, our conclusions are limited by the relatively narrow focus on doctors and nurses. Coping strategies and emotional experiences over the course of the pandemic may differ for people in

other workplaces (e.g., restaurants, public transit) who share doctors' and nurses' experiences with high-infectivity work environments and understaffing. However, we provisionally assume that doctors' and nurses' language patterns represent a microcosm of the global pandemic response, with people in all professions potentially showing the same linguistic changes over time to the degree that their lives have been disrupted by COVID-19.

## 4.3 Ethics and Privacy

Research on social media language is fraught with ethical ambiguity. All messages we analyzed are public, and Reddit norms encourage anonymity. Yet social media users often fail to realize the degree to which others may be able to triangulate personal information from messages they have posted online (Mneimneh et al., 2021). Furthermore, people who are comfortable disclosing private thoughts and feelings in a familiar online community may be less sanguine about researchers reading and republishing their messages. That is, despite the public nature of Reddit, users may have reasonably expected relative privacy (believing only fellow subreddit subscribers would see their messages) while writing.

To respect the individuals in this sample, texts and usernames will only be shared pending ethical review of the proposed research (see Bender et al. 2020). All deidentified, quantitative data are available at the OSF link referenced above.

## 4.4 Conclusion

Dictionary-based analyses of a large naturalistic, longitudinal sample of healthcare workers' online conversations revealed psychological strengths and vulnerabilities among people working in high-risk positions on the front lines of the COVID-19 pandemic. Understanding how people cope—adaptively and otherwise—with chronic stress can help to calibrate mental health treatment for not only doctors and nurses, but also other high-risk professions (Aulisio and May, 2020). In the workplace, such treatment improvements may decrease burnout, mitigate staffing shortages, and improve healthcare quality, thus lightening the global healthcare burden (Gandi et al., 2011). In terms of both theory and practice in clinical psychology, gaining a clearer picture of everyday coping strategies offers an opportunity to check and in some cases reject inaccurate assumptions about how chronic stress affects social and emotional behavior.

## References

- Kiki Adams. 2022. [SALLEE documentation](#). Online Receptiviti Inc. documentation.
- Social Security Administration, Department of Veterans Affairs, et al. 2017. Federal policy for the protection of human subjects. final rule. *Fed. Regist.*, 82:7149–7274.
- Hamzeh Mohammad Alrawashdeh, Ala'a B Al-Tammemi, Mohammad Kh Alzawahreh, Ashraf Al-Tamimi, Mohamed Elkholy, Fawaz Al Sarireh, Mohammad Abusamak, Nafisa MK Elehamer, Ahmad Malkawi, Wedad Al-Dolat, et al. 2021. Occupational burnout and job satisfaction among physicians in times of covid-19 crisis: a convergent parallel mixed-method study. *BMC Public Health*, 21(1):1–18.
- Stephanie A Andel, Archana M Tedone, Winny Shen, and Maryana L Arvan. 2022. Safety implications of different forms of understaffing among nurses during the covid-19 pandemic. *Journal of Advanced Nursing*, 78(1):121–130.
- Ashwini Ashokkumar and James W Pennebaker. 2021. Social media conversations reveal large psychological shifts caused by covid-19's onset across us cities. *Science advances*, 7(39):eabg7843.
- Mark P Aulisio and Thomas May. 2020. Why healthcare workers ought to be prioritized in asmr during the sars-cov-2 pandemic. *The American Journal of Bioethics*, 20(7):125–128.
- Özlem Ayduk and Ethan Kross. 2010. Analyzing negative experiences without ruminating: The role of self-distancing in enabling adaptive self-reflection. *Social and Personality Psychology Compass*, 4(10):841–854.
- Jenna L Baddeley, James W Pennebaker, and Christopher G Beevers. 2013. Everyday social behavior during a major depressive episode. *Social Psychological and Personality Science*, 4(4):445–452.
- Courtney L Bagge, Catherine R Glenn, and Han-Joo Lee. 2013. Quantifying the impact of recent negative life events on suicide attempts. *Journal of Abnormal Psychology*, 122(2):359.
- Emily M Bender, Dirk Hovy, and Alexandra Schofield. 2020. Integrating ethics into the nlp curriculum. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 6–9.
- Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22.
- Bjarne Brendstrup, Svend Hylleberg, Morten Ørregaard Nielsen, Lars Skipper, and Lars Stentoft. 2004. Seasonality in economic models. *Macroeconomic Dynamics*, 8(3):362–394.
- Timo Brockmeyer, Johannes Zimmermann, Dominika Kulesa, Martin Hautzinger, Hinrich Bents, Hans-Christoph Friederich, Wolfgang Herzog, and Matthias Backenstrass. 2015. Me, myself, and i: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in psychology*, 6:1564.
- Ana-Maria Bucur, Ioana R Podină, and Liviu P Dinu. 2021. A psychologically informed part-of-speech analysis of depression in social media. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP*, pages 199–207.
- Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765.
- Muhammad Chutiyami, Allen MY Cheong, Dauda Salihu, Umar Muhammad Bello, Dorothy Ndwiga, Reshin Maharaj, Kogi Naidoo, Mustapha Adam Kolo, Philomina Jacob, Navjot Chhina, et al. 2021. Covid-19 pandemic and overall mental health of healthcare professionals globally: A meta-review of systematic reviews. *Frontiers in Psychiatry*, 12.
- Michael A Cohn, Matthias R Mehl, and James W Pennebaker. 2004. Linguistic markers of psychological change surrounding september 11, 2001. *Psychological Science*, 15(10):687–693.
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015a. From adhd to sad: Analyzing the language of mental health on twitter through self-reported diagnoses. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 1–10.
- Glen Coppersmith, Ryan Leary, Eric Whyne, and Tony Wood. 2015b. Quantifying suicidal ideation via language usage on social media. In *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM*, volume 110.
- Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110.
- Sanda Dolcos and Dolores Albarracín. 2014. The inner speech of behavioral regulation: Intentions and task performance strengthen when you talk to yourself as a you. *European Journal of Social Psychology*, 44(6):636–642.
- Johannes C Eichstaedt, Margaret L Kern, David B Yaden, HA Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, et al. 2021. Closed-and

- open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, 26(4):398.
- Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preojuic-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208.
- Bruce J Ellis, Albertine J Oldehinkel, and Esther Nederhof. 2017. The adaptive calibration model of stress responsivity: An empirical test in the tracking adolescents’ individual lives survey study. *Development and Psychopathology*, 29(3):1001–1021.
- Joshua C Gandi, Paul S Wai, Haruna Karick, and Zubairu K Dagona. 2011. The role of stress and level of burnout in job performance among nurses. *Mental Health in Family Medicine*, 8(3):181.
- Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Šnajder. 2021. Pandora talks: Personality and demographics on reddit. In *Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media*, pages 138–152.
- David W Hancock, Amelia E Talley, Jennifer Bohanek, Micah D Iserman, and Molly E Ireland. 2018. Sexual orientation self-concept ambiguity and alcohol use disorder symptomology: The roles of motivated psychological distancing and drinking to cope. *Journal of Studies on Alcohol and Drugs*, 79(1):96–101.
- Lucas Rodés-Guirao Cameron Appel Charlie Gattino Esteban Ortiz-Ospina Joe Hasell Bobbie Macdonald Diana Beltekian Hannah Ritchie, Edouard Mathieu and Max Roser. 2020. Coronavirus pandemic (covid-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>.
- Amir Hassan Zadeh, Hamed M Zolbanin, Ramesh Sharda, and Dursun Delen. 2019. Social media for nowcasting flu activity: Spatio-temporal big data analysis. *Information Systems Frontiers*, 21(4):743–760.
- Nicholas S Holtzman et al. 2017. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68:63–68.
- Philip N Howard, Aiden Duffy, Deen Freelon, Muzaamil M Hussain, Will Mari, and Marwa Maziad. 2011. Opening closed regimes: what was the role of social media during the arab spring? *Project on Information Technology and Political Islam*.
- Molly E Ireland and Micah D Iserman. 2018. Within and between-person differences in language used across anxiety support and neutral reddit communities. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 182–193.
- Micah Iserman. 2022. *splot: Split Plot*. R package version 0.5.3.
- Kayla N Jordan, James W Pennebaker, Keith J Petrie, and Nicola Dalbeth. 2019. Googling gout: exploring perceptions about gout through a linguistic analysis of online search activities. *Arthritis Care & Research*, 71(3):419–426.
- Ewa Kacewicz, James W Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C Graesser. 2014. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2):125–143.
- Molly C Kalmoe, Matthew B Chapman, Jessica A Gold, and Andrea M Giedinghagen. 2019. Physician suicide: a call to action. *Missouri medicine*, 116(3):211.
- Ethan Kross and Ozlem Ayduk. 2017. Self-distancing: Theory, research, and current directions. In *Advances in Experimental Social Psychology*, volume 55, pages 81–136. Elsevier.
- Brett Laursen and Christopher A Hafen. 2010. Future directions in the study of close relationships: Conflict is bad (except when it’s not). *Social Development*, 19(4):858–872.
- Jessica L Malisch, Breanna N Harris, Shanen M Sherer, Kristy A Lewis, Stephanie L Shepherd, Puntiwitt C McCarthy, Jessica L Spott, Elizabeth P Karam, Naima Moustaid-Moussa, Jessica McCrory Calarco, et al. 2020. Opinion: In the wake of covid-19, academia needs new solutions to ensure gender equity. *Proceedings of the National Academy of Sciences*, 117(27):15378–15381.
- David M Markowitz, Maryam Kouchaki, Francesca Gino, Jeffrey T Hancock, and Ryan L Boyd. 2022. Authentic first impressions relate to interpersonal, social, and entrepreneurial success. *Social Psychological and Personality Science*, page 19485506221086138.
- Lauren M Ministero, Michael J Poulin, Anneke EK Buffone, and Shane DeLury. 2018. Empathic concern and the desire to help as separable components of compassionate responding. *Personality and Social Psychology Bulletin*, 44(4):475–491.
- Zeina Mneimneh, Josh Pasek, Lisa Singh, Rachel Best, Leticia Bode, Elizabeth Bruch, Ceren Budak, Pamela Davis-Kean, Katharine Donato, Nicole Ellison, andrew gelman, Erica Groshen, Libby Hemphill, William Hobbs, J. Jensen, George Karypis, Jonathan Ladd, Amy O’Hara, Trivellore Raghunathan, and Stefan Wojcik. 2021. *Data acquisition, sampling, and data preparation considerations for quantitative social science research using social media data*.
- Scott M Monroe, Paul Rohde, John R Seeley, and Peter M Lewinsohn. 1999. Life events and depression in adolescence: relationship loss as a prospective risk factor for first onset of major depressive disorder. *Journal of Abnormal Psychology*, 108(4):606.

- Stephen Murata, Taylor Rezeppa, Brian Thoma, Laura Marengo, Katie Krancevich, Elizabeth Chiyka, Benjamin Hayes, Eli Goodfriend, Meredith Deal, Yongqi Zhong, et al. 2021. The psychiatric sequelae of the covid-19 pandemic in adolescents, adults, and health care workers. *Depression and Anxiety*, 38(2):233–246.
- Yaakov Ophir, Refael Tikochinski, Christa SC Asterhan, Itay Sisso, and Roi Reichart. 2020. Deep neural networks detect suicide risk from textual facebook posts. *Scientific Reports*, 10(1):1–10.
- Jiyoung Park, Shinobu Kitayama, Yuri Miyamoto, and Christopher L Coe. 2020. Feeling bad is not always unhealthy: Culture moderates the link between negative affect and diurnal cortisol profiles. *Emotion*, 20(5):721.
- James W Pennebaker. 1989. Confession, inhibition, and disease. In *Advances in experimental social psychology*, volume 22, pages 211–244. Elsevier.
- James W Pennebaker. 2018. Expressive writing in psychological science. *Perspectives on Psychological Science*, 13(2):226–229.
- James W Pennebaker, Ryan L Boyd, Roger J Booth, Ashwini Ashokkumar, and Martha E Francis. 2022. Linguistic inquiry and word count: Liwc-22.
- James W Pennebaker and Cindy K Chung. 2005. Tracking the social dynamics of responses to terrorism: Language, behavior, and the internet. *NATO Security Through Science Series E: Human and Societal Dynamics*, 3:159.
- James W Pennebaker and Cindy K Chung. 2007. Expressive writing, emotional upheavals, and health. In HS Friedman and RC Silver, editors, *Foundations of Health Psychology*. Oxford University Press, Oxford, England, UK.
- James W Pennebaker and Kent D Harber. 1993. A social stage model of collective coping: The loma prieta earthquake and the persian gulf war. *Journal of Social Issues*, 49(4):125–145.
- James W Pennebaker and Lori D Stone. 2003. Words of wisdom: language use over the life span. *Journal of Personality and Social Psychology*, 85(2):291.
- Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2021. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-153.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Michael W Rabow, Chao-Hui S Huang, Gloria E White-Hammond, and Rodney O Tucker. 2021. Witnesses and victims both: Healthcare workers and grief in the time of covid-19. *Journal of Pain and Symptom Management*, 62(3):647–656.
- Megan L Robbins, Robert C Wright, Ana María López, and Karen Weihs. 2019. Interpersonal positive reframing in the daily lives of couples coping with breast cancer. *Journal of psychosocial oncology*, 37(2):160–177.
- Arunima Roy, Katerina Nikolitch, Rachel McGinn, Safiya Jinah, William Klement, and Zachary A Kaminsky. 2020. A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digital Medicine*, 3(1):1–12.
- Kristin E Schneider, Lauren Dayton, Saba Rouhani, and Carl A Latkin. 2021. Implications of attitudes and beliefs about covid-19 vaccines for vaccination campaigns in the united states: A latent class analysis. *Preventive medicine reports*, 24:101584.
- João Sedoc, Sven Buechel, Yehonathan Nachmany, Anneke Buffone, and Lyle Ungar. 2020. Learning word ratings for empathy and distress from document-level user responses. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1664–1673.
- Yi-Tai Seih, YC Lin, CL Huang, CW Peng, and SP Huang. 2008. The benefits of psychological displacement in diary writing when using different pronouns. *British Journal of Health Psychology*, 13(1):39–41.
- Sarah Seraj, Kate G Blackburn, and James W Pennebaker. 2021. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7).
- Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, pages 58–65.
- Mamidipalli Sai Spoorthy, Sree Karthik Pratapa, and Supriya Mahant. 2020. Mental health problems faced by healthcare workers due to the covid-19 pandemic—a review. *Asian Journal of Psychiatry*, 51:102119.
- Abi Sriharan, Savithiri Ratnapalan, Andrea C Tricco, and Doina Lupea. 2021. Women in healthcare experiencing occupational stress and burnout during covid-19: a rapid review. *BMJ open*, 11(4):e048861.
- Shannon Wiltsey Stirman and James W Pennebaker. 2001. Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic medicine*, 63(4):517–522.
- Lori D Stone and James W Pennebaker. 2002. Trauma in real time: Talking and avoiding online conversations about the death of princess diana. *Basic and Applied Social Psychology*, 24(3):173–183.
- Jessie Sun, H Andrew Schwartz, Youngseo Son, Margaret L Kern, and Simine Vazire. 2020. The language

- of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*, 118(2):364.
- Allison M Tackman, David A Sbarra, Angela L Carey, M Brent Donnellan, Andrea B Horn, Nicholas S Holtzman, To'Meisha S Edwards, James W Pennebaker, and Matthias R Mehl. 2019. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology*, 116(5):817.
- Yla Tausczik, Cindy Chung, and James Pennebaker. 2016. Tracking secret-keeping in emails. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 388–397.
- Yla Tausczik, Kate Faasse, James W Pennebaker, and Keith J Petrie. 2012. Public anxiety and information seeking following the h1n1 outbreak: blogs, newspaper articles, and wikipedia visits. *Health Communication*, 27(2):179–185.
- Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.
- Tyler J VanderWeele and Maya B Mathur. 2019. Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? *American journal of epidemiology*, 188(3):617–618.
- Suzan Verberne, Anika Batenburg, Remco Sanders, Mies van Eenbergen, Enny Das, Mattijs S Lambooi, et al. 2019. Analyzing empowerment processes among cancer patients in an online community: A text mining approach. *JMIR Cancer*, 5(1):e9887.
- Vera Vine, Ryan L Boyd, and James W Pennebaker. 2020. Natural emotion vocabularies as windows on distress and well-being. *Nature Communications*, 11(1):1–9.
- Matthew Walton, Esther Murray, and Michael D Christian. 2020. Mental health care for medical staff and affiliated healthcare workers during the covid-19 pandemic. *European Heart Journal: Acute Cardiovascular Care*, 9(3):241–247.
- Peter F Weissmann, William T Branch, Catherine F Gracey, Paul Haidet, and Richard M Frankel. 2006. Role modeling humanistic behavior: learning bedside manner from the experts. *Academic Medicine*, 81(7):661–667.
- WHO. 2022. [Who covid-19 dashboard](#).
- Kevin P Young, Diana L Kolcz, David M O'Sullivan, Jennifer Ferrand, Jeremy Fried, and Kenneth Robinson. 2021. Health care workers' mental health and quality of life during covid-19: results from a mid-pandemic, national survey. *Psychiatric Services*, 72(2):122–128.
- Yan Zhang, Lili Wei, Huanting Li, Yueshuai Pan, Jingyuan Wang, Qianqian Li, Qian Wu, and Holly Wei. 2020. The psychological change process of frontline nurses caring for patients with covid-19 during its outbreak. *Issues in Mental Health Nursing*, 41(6):525–530.