

Using mobile sensing data to assess stress: Associations with perceived and lifetime stress, mental health, sleep, and inflammation

Digital Health
Volume 7: 1–11
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076211037227
journals.sagepub.com/home/dhj


Michelle L Byrne^{1,2} , Monika N Lind¹, Sarah R Horn¹, Kathryn L Mills¹, Benjamin W Nelson³, Melissa L Barnes¹ , George M Slavich⁴  and Nicholas B Allen^{1,5}

Abstract

Background: Although stress is a risk factor for mental and physical health problems, it can be difficult to assess, especially on a continual, non-invasive basis. Mobile sensing data, which are continuously collected from naturalistic smartphone use, may estimate exposure to acute and chronic stressors that have health-damaging effects. This initial validation study validated a mobile-sensing collection tool against assessments of perceived and lifetime stress, mental health, sleep duration, and inflammation.

Methods: Participants were 25 well-characterized healthy young adults ($M_{\text{age}} = 20.64$ years, $SD = 2.74$; 13 men, 12 women). We collected affective text language use with a custom smartphone keyboard. We assessed participants' perceived and lifetime stress, depression and anxiety levels, sleep duration, and basal inflammatory activity (i.e. salivary C-reactive protein and interleukin-1 β).

Results: Three measures of affective language (i.e. total positive words, total negative words, and total affective words) were strongly associated with lifetime stress exposure, and total negative words typed was related to fewer hours slept (all large effect sizes: $r = 0.50 - 0.78$). Total positive words, total negative words, and total affective words typed were also associated with higher perceived stress and lower salivary C-reactive protein levels (medium effect sizes; $r = 0.22 - 0.32$).

Conclusions: Data from this initial longitudinal validation study suggest that total and affective text use may be useful mobile sensing measures insofar as they are associated with several other stress, mental health, behavioral, and biological outcomes. This tool may thus help identify individuals at increased risk for stress-related health problems.

Keywords

Mobile sensing, stress, affective language, text, inflammation, mental health

Submission date: 22 January 2021; Acceptance date: 17 July 2021

Introduction

In recent years, the global disease burden has shifted away from premature death toward more years lived with disability.^{1,2} Non-communicable, chronic diseases are now responsible for more years lost and greater health treatment costs than any other type of illness or disease.³ It is thus important to identify modifiable risk factors for chronic diseases that can be targeted for intervention and prevention programs.

Chronic inflammation is one mechanism that appears to be relevant for a wide variety of physical and mental health

¹Department of Psychology, University of Oregon, USA

²Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Australia

³Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, USA

⁴Cousins Center for Psychoneuroimmunology and Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, USA

⁵School of Psychological Sciences, University of Melbourne, Australia

Corresponding author:

Nicholas B Allen, Department of Psychology, University of Oregon, Eugene, OR 97403-1227, USA.

Email: nallen3@uoregon.edu



problems,⁴ including heart disease,⁵ obesity,^{6,7} and depression.⁸ Inflammatory activity has been shown to be upregulated by psychological stress in addition to tissue damage and microbial threats.⁹ However, frequently assessing inflammatory levels over time, within a particular individual, is a difficult and expensive way to index stress-related disease risk, and continuous, real-time assessments are not currently possible. Other measures of stress that are known to be associated with chronic illnesses and inflammation include depression and anxiety,^{10,11} perceived stress,¹² stressful life events,¹³ and poor sleep.¹⁴ However, many of these factors are either biased (e.g. if self-reported) or also difficult to repeatedly measure over long periods of time.

Recent technological advances have enabled the ongoing assessment of stress markers that might be relevant for estimating changes in chronic disease risk. For example, mobile sensing data, which are continuously collected from a person's naturalistic smartphone use, may be able to estimate levels of both acute and chronic stressors that have health-damaging effects. Mobile sensing has the advantage of measuring objective, naturalistic behavioral data unobtrusively, while preserving ecological validity. This benefit could improve upon self-reported questionnaires, which may suffer from response- or interpretation bias, and physiological variables, which are more difficult to obtain and challenging to measure in ecologically valid contexts.

Mobile sensing tools can measure behaviors that are known to be associated with stress and mood, including language use, facial expressions, sleep, and music choice, among others.^{15–18} Some research has even shown that user location, type of sound in the environment (including voices), and being around others (based on Bluetooth data) are all associated with self-reported perceived stress levels.¹⁹ However, natural language use from the keyboard is especially interesting for two reasons. Firstly, there is accumulating evidence that natural language processing can be used to detect critical health events,²⁰ including elevated suicide risk,^{21,22} and is also related to depression,²³ heart disease,²⁴ and overall mortality (particularly less positive emotional words in the latter case).²⁵ Secondly, few mobile sensing tools capture the content of communication, opting instead to focus on communication frequency (e.g. number of texts sent and number of phone calls made²⁶). As such, focusing on the content of the language represents an important innovation in assessing stress-related signals that may be useful for non-invasively estimating stress-related health risks on a continual basis.

To accomplish this goal, we conducted a longitudinal validation study of healthy young adults in which we used a mobile sensing collection tool called Effortless Assessment of Risk States (EARS)²⁶ to validate naturalistic text collection from smartphones (using a custom keyboard) as an objective behavioral marker of both short-term stress

(i.e. perceived and contextual stress, mental health symptoms and biological stress markers) and long-term stress (i.e. cumulative lifetime stress exposure). Focusing on communication content (e.g. affective sentiment) improves upon prior studies that have measured communication frequency only. Furthermore, no study has examined associations between mobile sensing and biological (i.e. immune) markers of stress. Therefore, this study was designed to show if affective language measured noninvasively and in an ecologically valid way is associated with self-reported and biological markers of stress. We focused specifically on healthy young adults given the fact that they are less likely to have already experienced a serious inflammation-related disease. As such, any insights gained from this research could have the ability to potentially identify *pre-clinical* risk processes that could in turn be targeted to reduce the initial onset of disease.

Based on the research summarized above, we hypothesized that more negative sentiment words typed (i.e. more positive sentiment score, more total negative words typed, and less positive to total affective words typed) would be associated with: (a) higher self-reported perceived stress, as measured by the Perceived Stress Scale (PSS); (b) higher self-reported depression and anxiety, as measured by the Depression, Anxiety, and Stress Scales (DASS); (c) greater lifetime stress exposure, as measured by the Stress and Adversity Inventory for Adults (STRAIN); (d) higher basal levels of inflammation, as indexed by salivary C-reactive protein (sCRP) and salivary levels of the key pro-inflammatory cytokine, salivary interleukin-1 β (sIL-1 β); and (e) fewer self-reported hours of total sleep, as measured by the Pittsburgh Sleep Quality Index (PSQI).

Methods and materials

Participants and recruitment

We conducted the Effortless Assessment of Stressful Experiences (EASE) study (University of Oregon Institutional Review Board (IRB) protocol number 07212016.019) in 2016–2017. We assessed 25 healthy young adults (13 male, 12 female; 13 men, 12 women) over fall and winter terms of 2016 and 2017, who were recruited from the university community ($M_{\text{age}}=20.64$ years old, $SD=2.74$). Participants provided written informed consent to take part in the study. The inclusion criteria were owning an Android phone and not having an immunological medical condition. Participants self-reported their race/ethnicity as 64% White, 12% Asian, 12% multiracial, and 12% Hispanic. The average yearly total pre-tax income of the participants' parents was \$85,080 (range: \$0–\$300,000), and parents' education was 20% high school, 16% some college, 8% associate's degree, 36% bachelor's degree, 12% master's degree, and 8% PhD.

Procedure

Data were collected twice during the school term: once during a period that was expected to be relatively less academically stressful and then again during a final exam period that was expected to be more stressful. The baseline assessment occurred 3–7 weeks before the first final exam of the term, thus avoiding weeks when participants might have had a midterm or major project due. The follow-up assessment occurred during the week prior to the last final exam. Therefore, each participant had two measurements of each variable (except for our measure of cumulative lifetime stress exposure, which was measured only once, at baseline). For continuous text data, we either averaged a continuous score of sentiment value or totaled by number of affective, positive, and negative words for Bing analysis (see later). We administered self-report questionnaires of perceived and contextual stress, mental health symptoms, and sleep duration on the last day of each week of data collection, and asked about mental health symptoms over the past week. To assess participants' inflammatory levels, we collected saliva samples once per week on the last day of each week.

Measures

Perceived stress. We assessed perceived stress with the 14-item PSS.²⁷ This version of the PSS consistently has Cronbach's alpha values over 0.70 across versions used in multiple cultures and countries.²⁸

Lifetime stress exposure. We measured lifetime stress exposure using the STRAIN, which is an online stress assessment system that quantifies an individual's exposure to acute and chronic stressors occurring over the entire life course²⁹ (see <https://www.strainsetup.com>). The STRAIN uses intelligent logic to shape the interview based on participants' responses, and it assesses the severity, frequency, timing, and duration of each stressor reported. The STRAIN has been extensively validated in relation to many different psychological, biological, and health outcomes (e.g. Cazassa et al.³⁰, Mayer et al.³⁰, and Sturmhuber et al.³²). For analyses, we focused on three of the STRAIN's main outcomes: total count of all stressors, all acute stressors, and all chronic stressors occurring over the life course. In the STRAIN system, acute stressors are episodic stressors that typically last a day or so and include situations such as getting into an accident or hearing bad news. In contrast, chronic stressors last a month or longer and include situations such as persistent marital, housing, or financial problems.

Mental health symptoms. We assessed self-reported symptoms of perceived stress, depression, and anxiety with the 42-item DASS.³³ Internal consistency of this version of the

DASS is high across clinical and community samples, with Cronbach's alpha coefficients for the Depression, Anxiety, and Stress scales of 0.97, 0.92, and 0.95, respectively.³⁴

Sleep duration. We asked participants about their sleep behaviors over the past week using the PSQI,³⁵ and calculated an average subjective total sleep time in hours per night as a measure of sleep duration.

Inflammation. We assessed two markers of basal inflammation involved in the acute phase response³⁶ that are known to be associated with psychosocial stress and mental health: the acute-phase protein sCRP and the cytokine sIL-1 β (for a review, see Howren et al.⁸). CRP is a general inflammatory and somatic maintenance marker³⁷ that activates the complement system, which is responsible for the body's systemic inflammatory response to pathogens and tissue damage.³⁸ Elevated levels have been shown to be associated with various psychosocial stressors^{39–41} as well as a variety of chronic disease conditions.⁴ IL-1 β is a key pro-inflammatory cytokine produced by immune cells⁴² that is responsive to psychological stress^{43,44} and is associated with a variety of mental health conditions, including depression.⁴⁵

We assessed these inflammatory markers in saliva to see whether inflammation measured under basal conditions (i.e. not in response to stress) was associated with self-reported and passive sensing measures of stress. In this way, the inflammatory measures were considered to be measures of inflammatory health rather than stress-related biological reactivity. We sampled saliva because levels of localized inflammatory markers from the oral compartment have been shown to be sensitive to social stress^{46,47} and are associated with psychological health.^{48,49} Salivary markers of inflammation also have been related to activity in brain regions implicated in depression and social stress.^{47,50} This may partly be due to a process whereby the brain receives information from the local trigeminal nerve during acute oral inflammation.⁵¹ In addition, sCRP has been shown to be associated with CRP levels in blood at medium-to-large effect sizes,^{52–54} although one study found blood and sCRP were not related,⁵⁵ and the correlation between salivary and blood levels of IL-1 β is not yet known.

A research assistant collected 2 mL of whole, unstimulated saliva into a polystyrene transport tube (Corning Axygen) from each participant, using the passive drool method. The research assistant then placed the sample immediately into an -80°C freezer for storage until it was shipped overnight on dry ice for analysis. The research assistant also recorded participants' body temperature on the forehead with an Exergen temporal artery digital thermometer; however, no participants measured over $100.4^{\circ}\text{F}/38^{\circ}\text{C}$; so, none were excluded based on fever.

The Stress Physiology Investigative Team (SPIT) lab at Iowa State University conducted the immunological assays using Salimetrics enzyme-linked immunosorbent assay

Table 1. Descriptive variables.

Measure	Mean \pm SD (range)
PSS total score	24.52 \pm 8.65 (6.5–38.5)
STRAIN total lifetime count of all stressors ^a	22.39 \pm 15.82 (9–68)
STRAIN total lifetime count of acute stressors ^a	10.57 \pm 9.99 (2–37)
STRAIN total lifetime count of chronic stressors ^a	11.83 \pm 6.30 (5–31)
DASS anxiety	6.30 \pm 4.71 (0–16)
DASS depression	6.66 \pm 5.32 (0–17)
DASS stress	10.98 \pm 7.78 (0.5–29)
PSQI average hours sleep per night	6.97 \pm 1.32 (3.5–9)
Raw sIL-1 β (pg/ml)	23.51 \pm 17.77 (9.56–81.45)
Raw sCRP ^b (pg/ml)	868.46 \pm 890.98 (12.78–2764.14)
Ln(sIL-1 β)	2.92 \pm 0.62 (1.50–4.20)
AFINN total score ^c	0.60 \pm 0.78 (–1.57–2)
Bing total negative words ^c	10.83 \pm 10.37 (0–42.5)
Bing total positive words ^c	14.77 \pm 14.77 (0–56)
Bing total affective words ^c	25.60 \pm 24.45 (0–98.5)
Bing positive to total affective words ratio ^c	0.57 \pm 0.21 (0–1)
Total words typed ^c	725.06 \pm 623.32 (1.5–2549)
Bing affective to total words ratio ^c	0.03 \pm 0.01 (0–0.06)

DASS: Depression, Anxiety, and Stress Scale; PSQI: Pittsburgh Sleep Quality Index; PSS: Perceived Stress Scale; sCRP: salivary C-reactive protein; sIL-1 β : salivary interleukin-1 β ; STRAIN: Stress and Adversity Inventory. AFINN is not an acronym.

^aThe STRAIN was only administered once during the study, as it assesses lifetime stress exposure, so the resulting scores are not averaged.

^bTransforming the CRP values did not improve skew; so, the raw values were used for analyses.

^cThree participants did not have week 1 text data recorded, so only week 2 is reported (instead of an average across weeks) for those participants.

(ELISA) kits (Carlsbad, CA). The inter-assay coefficients of variations (CVs) were calculated from optical density and were 3.82% for sCRP and 7.39% for sIL-1 β ; the intra-assay CVs, in turn, were for 3.17% CRP and 1.47% for IL-1 β . For assay values that were out of range (either too high, i.e. right-censored, or low, i.e. left-censored, for the assay to detect), we winsorized these values in the same way that we winsorized outliers that were >3 SDs. There were no left-censored immune values and no right-censored sIL-1 β values, but there were five right-censored sCRP values. The sCRP skew was 1.06 and kurtosis was 2.58, and the sIL-1 β skew was 1.50 and kurtosis was 5.03; so, both were log transformed. Whereas this transformation improved the sIL-1 β skew statistics, it did not improve sCRP; therefore, only sIL-1 β values were transformed for analyses.

Affective language use. The first version of the EARS tool included a custom keyboard installed on the user's smartphone that logged every third word they typed into the phone across all apps, not including passwords or credit card information. For this study, we captured only every third word to err on the side of protecting the privacy and confidentiality of participants' information and communications. This conservative approach limited our unit of analysis to the unigram (i.e. single words). Although phrase- and vector-based approaches tend to outperform unigram-based approaches, there is a rich literature of robust unigram-based findings through using the Linguistic Inquiry and Word Count,⁵⁶ including research showing that unigram-based approaches can detect differences in stress (e.g. Coppersmith et al.⁵⁷ and Guntuku et al.⁵⁸). For more information on the EARS tool and changes to newer versions, refer to the paper describing the tool.²⁶

We analyzed text sentiment (i.e. unigrams) using the R package `tidytext`⁵⁹ with two different lexicons: "AFINN"⁶⁰ and "bing."⁶¹ For words in the dataset that exist in each affective dictionary (i.e. only including words that are marked as affective, not neutral words), AFINN gives a continuous scale value based on sentiment value of each word, whereas bing categorizes each word as either positive or negative sentiment and then totals the number in each category. We calculated five affective scores for each participant: (a) the average AFINN score, which is a number between -5 and 5 , with negative numbers representing more negative sentiment; (b) the total number of negative words typed over the week using the bing lexicon; (c) the total number of positive words typed over the week using the bing lexicon; (d) the total number affective (negative + positive words) typed over the week using the bing lexicon; and (e) the ratio of total positive words to total affective (positive + negative) words using the bing lexicon (i.e. the proportion of affective words that are positive). Additionally, we calculated the total words typed post-hoc, regardless of affective

sentiment, to see whether any effects were simply due to the amount typed rather than to affect specifically, and a ratio of total affective words to total (any) words typed. The word totals reported are for the entire week but they represent only about a third of the total number of words actually typed because we only collected every third word.

Data analysis

Due to the exploratory nature of this study, effect sizes are used to describe all strength of associations. To assess whether there was within-person variability across stressful contexts, we first examined differences by week for each variable. We first examined differences in weeks for each variable using a two-tailed paired *t*-test, but the only variable that significantly changed across weeks was the DASS anxiety subscale (mean week 1 = 5.24, mean week 2 = 7.67, $t(23) = -2.14$, 95%CI [-4.35, -0.07], $p = 0.04$; remaining statistics for null changes not shown). Therefore, we ultimately averaged each variable across the weeks and only used one (averaged) measure per person in subsequent analyses. Due to the exploratory and cross-sectional nature of the study, we then calculated Pearson's correlations (i.e. did not include any covariates) between all mobile sensing variables (AFINN score, Bing total negative words, Bing total positive words, Bing total affective words, and Bing positive-to-total affective words ratio) and traditional markers of stress exposure and impact [perceived stress, total lifetime stressors (i.e. acute + chronic), total lifetime acute stressors, total lifetime chronic stressors, depression, stress, anxiety, average hours of sleep, sCRP, and sIL-1 β].

Results

Descriptive statistics and correlations

Table 1 shows the values and SDs of all of the main study variables, averaged across weeks. Three participants did not have week 1 text data recorded, so only week 2 is reported instead of an average across weeks for those participants. Cronbach's alpha statistics for the questionnaires with scales were 0.89 for DASS stress, 0.79 for DASS anxiety, 0.89 for DASS depression, and 0.80 for the PSS. Time of saliva collection ranged from 7:45 a.m. to 3:20 p.m. ($M = 10:57$ a.m.), and these times were not related to levels of sCRP ($r = -0.06$, $p = 0.70$) or sIL-1 β ($r = 0.12$, $p = 0.44$). Figure 1 shows the correlation effect sizes across all of the main study variables.

Visualization of effects

Figure 2 highlights medium and large effect sizes of correlations for measures of different types (i.e. effect sizes of correlations between self-report questionnaires, or between

text measures, or between the two immune markers, are not illustrated). In addition to the correlations depicted in Figure 2, most of the self-reported questionnaires were also correlated with one another. Moreover, higher scores on the PSS, DASS anxiety, DASS stress, and STRAIN total, STRAIN acute, and STRAIN chronic stressor count were all associated with fewer average hours of sleep per night.

Summary of hypothesized results

Overall, some of our naturalistic measures of affective language use were associated with several markers of biological and self-reported stress (for exact correlation effect sizes, see Figure 1). The total number of negative words typed was moderately associated with higher scores on the PSS and DASS stress scales, and strongly associated with experiencing more lifetime stressors, acute stressors, and chronic stressors, as assessed by the STRAIN. Total number of negative words typed was also strongly related to fewer hours of sleep. Results were similar for total affective (i.e. positive and negative) words. A more negative affective language sentiment (AFINN) score was moderately associated with higher levels of sCRP. The ratio of positive as compared to total affective words was associated with higher scores on the DASS depression scale, but no other measure of affective language use was associated with DASS depression.

Post-hoc analysis

We also examined post-hoc if similar associations were evident for total words (i.e. any content) typed, rather than for affective words typed (refer to Figure 1 for effect sizes). The number of total words typed was moderately associated with higher scores on the PSS and DASS stress scales, and with fewer hours sleep, and was strongly related to experiencing more total lifetime stressors, acute stressors, and chronic stressors as assessed by the STRAIN. Therefore, we conducted another post-hoc analysis to investigate whether any of these variables were related to a ratio of total affective words to total words typed. However, there were no medium-to-large effect sizes (see Figure 1 for the exact correlations).

Discussion

Although stress is a central construct in most major models of mental and physical health, it can be challenging to assess, especially in a continuous, non-intrusive manner. To address this issue, we conducted a preliminary study examining how text collected using mobile sensing is associated with self-report and biological markers of stress exposure and impact. Overall, we found evidence that the mobile sensing data were associated with other measures of stress. More specifically, as hypothesized, more total

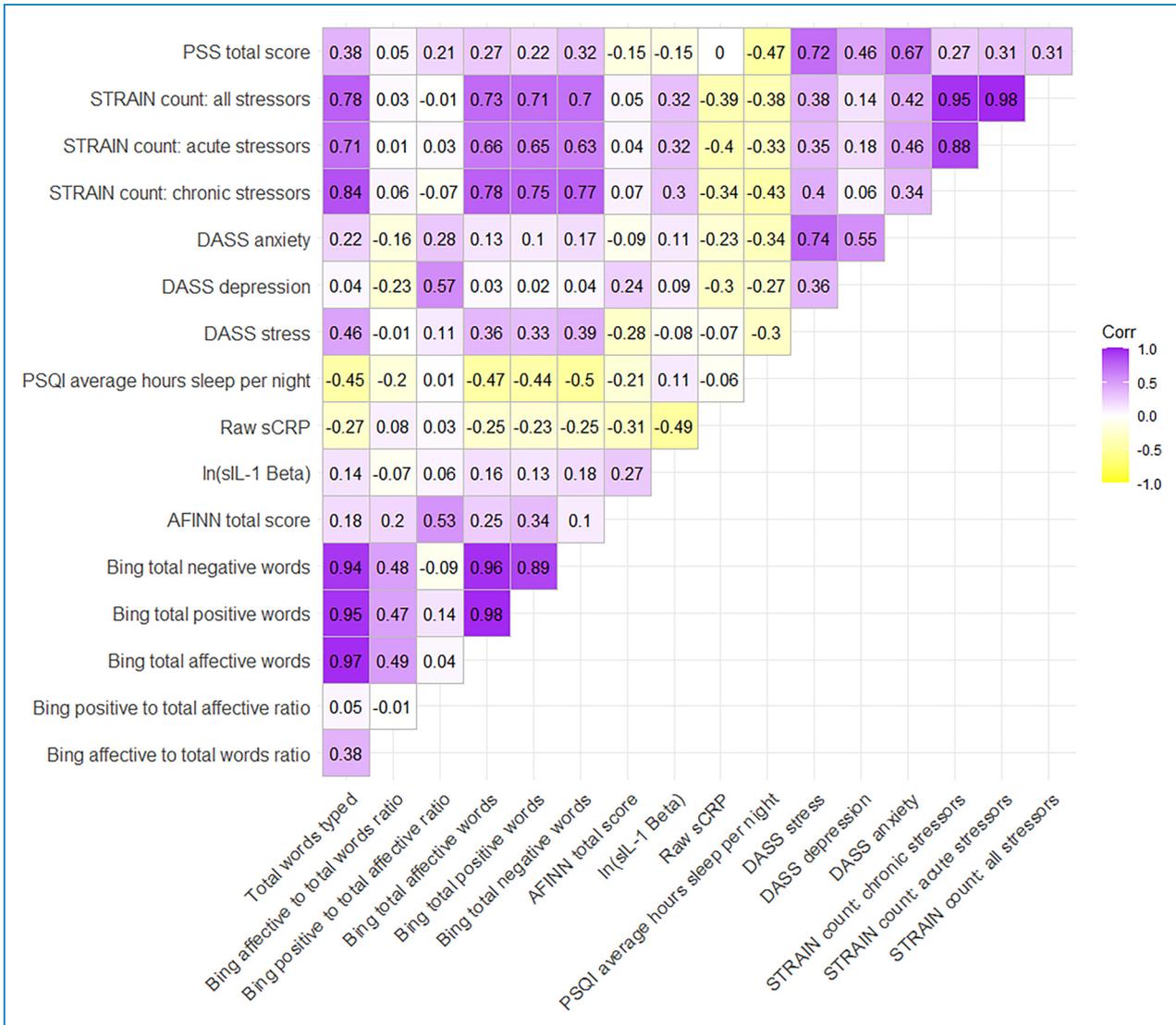


Figure 1. Effect sizes (Pearson’s correlations) between all variables. DASS: Depression, Anxiety, and Stress Scale; PSS: Perceived Stress Scale; sCRP: salivary C-reactive protein; sIL-1β: salivary interleukin-1β; STRAIN: Stress and Adversity Inventory. AFINN is not an acronym.

negative sentiment words typed was strongly associated with the number of total, acute, and chronic stressors experienced over the life course, as well as with fewer hours slept. Typing more negative sentiment words was also moderately related to higher levels of perceived stress. In addition, a more negative sentiment AFINN score was moderately associated with higher levels of inflammation, as indexed by sCRP. Although additional research with larger sample sizes is needed to further validate these associations, based on these initial data, we suggest that naturalistic text collected unobtrusively from smartphones could potentially be a promising objective marker of stress.

Of note, total affective words (i.e. both positive and negative)—and also total words typed regardless of affective sentiment—were both associated with the outcomes

assessed. Total words are likely more important than the affective sentiment, at least when measuring unigrams, and this is reflected in our post-hoc analysis showing the ratio of affective to total words was not associated with psychosocial and biological stress variables. Although our proportion of affective words to total words (about 3%) seems low, it is consistent with other studies of affective words using a different dictionary.⁶² A count of total words typed may represent how much the person is typing and using their phone, potentially due to circumstances such as how busy they are. This possibility is further supported by the negative correlation observed between total words typed and hours of sleep. Future research (a) should consider individual changes in total number of words typed as a possible marker of stress, regardless of affective

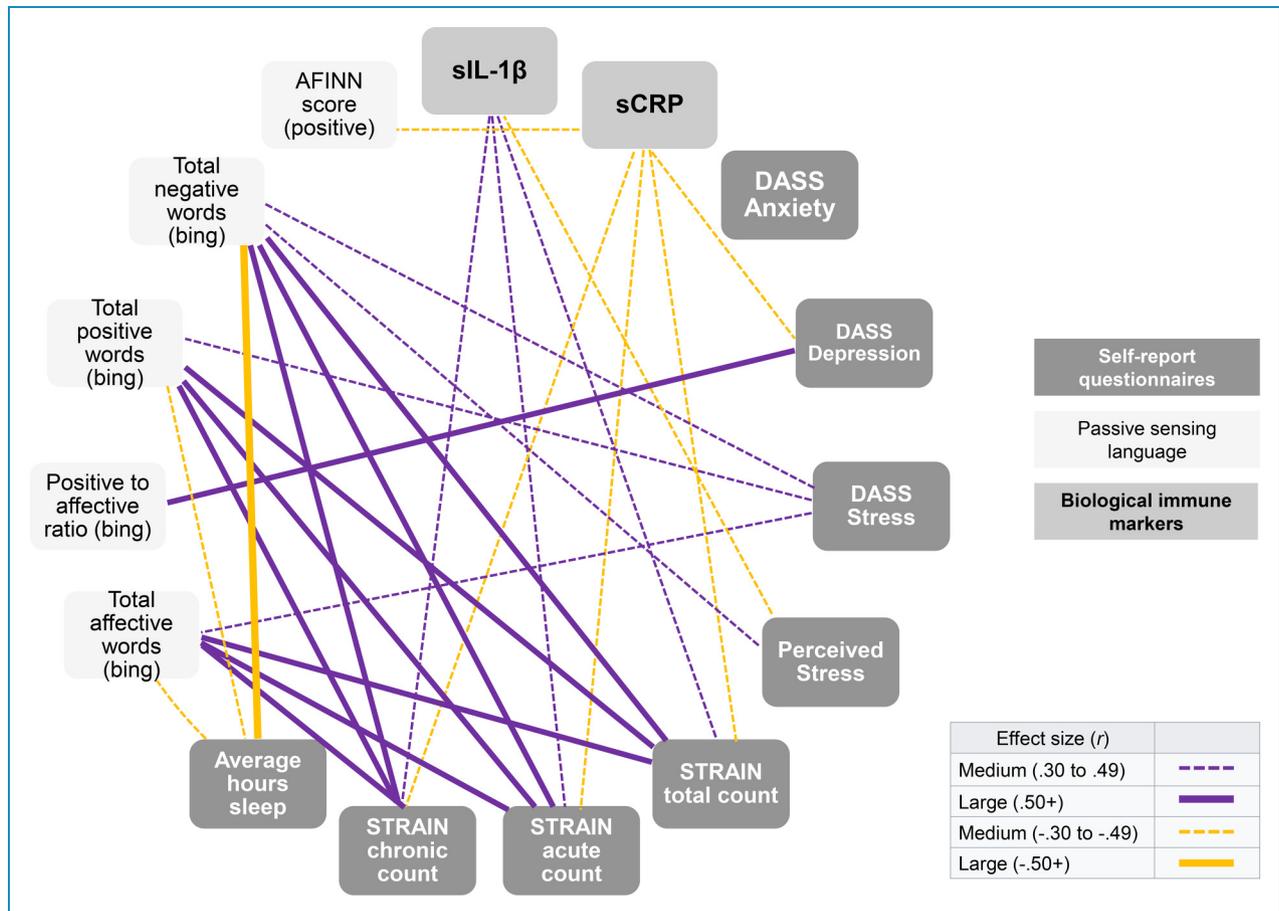


Figure 2. Medium and large effect sizes for mobile sensing variables.

DASS: Depression, Anxiety, and Stress Scale; PSS: Perceived Stress Scale; sCRP: salivary C-reactive protein; sIL-1 β : salivary interleukin-1 β ; STRAIN: Stress and Adversity Inventory. AFINN is not an acronym.

sentiment (such as rumination); (b) measure potential confounds of this association, just as social network size; and (c) collect all words instead of every third word to enable analysis of text data as phrases using more advanced natural language processing techniques.

A higher proportion of affective words that were positive was strongly related to reporting more depressive symptoms. The direction of this relation is counterintuitive, and may represent a false positive, given the exploratory nature of the study with a small sample size. Nevertheless, it may be worth further exploring in future studies whether individual changes in sentiment compared to other affective words may be a marker of stress.

In terms of the biological data collected, participants' AFINN scores were not associated with any of the self-report measures administered; however, a more positive AFINN score was related to lower levels of the inflammatory marker sCRP, suggesting that affective language use may be reflective of this immune process. Both sIL-1 β and sCRP are often assumed to be markers of biological

stress, with higher levels representing more acute inflammation. In this study, greater lifetime stress exposure as assessed by the STRAIN was related to higher levels of sIL-1 β but lower levels of sCRP. It is possible that sCRP at normal levels (i.e. as observed in this sample of physically healthy young adults) is really more a marker of somatic maintenance (and sometimes anti-inflammatory processes) rather than a marker of acute inflammation.³⁷ Yet, higher levels of sIL-1 β were also associated with less perceived stress, and so we caution against overinterpreting these results, again, given the relatively small sample size. Future studies should collect multiple measures of inflammation, since it is not entirely clear that sCRP is only pro-inflammatory at lower, non-clinical levels. Furthermore, although we collected measures of basal inflammation, salivary markers of inflammation such as IL-6 and tumor necrosis factor (TNF)- α are responsive to acute stressors.⁶³ Future research may show that measures of inflammatory reactivity, rather than basal inflammation, are associated with mobile sensing measures.

Finally, the largest effect sizes observed for the mobile sensing text data were in relation to the total number of all stressors, acute stressors, and chronic stressors experienced over the life course, as measured by the STRAIN. A lifetime history of exposure to stress and adversity is a significant risk factor for poor mental and physical health.⁶⁴ Future studies with larger sample sizes are needed to assess lifetime stress exposure as an individual difference in the association between mobile sensing data and mental health outcomes.

The main limitation of this study was the sample size. Due to its exploratory nature, we only reported effect sizes rather than significance testing, and it is possible that the effects may not replicate in larger samples. Additionally, the study only focused on keyboard text data as one mobile sensing measure. Future research should include a range of passive measures known to be associated with stress and mental health that can be easily collected from smartphones, including sleep activity and sleep problems (in addition to sleep duration),⁶⁵ geolocation,⁶⁶ facial expression,⁶⁷ and speech acoustics.⁶⁸ Furthermore, to provide an initial test of whether mobile sensing measures are associated with biological markers of stress before the emergence of disease, we conducted this study in a sample that was unlikely to have experienced inflammatory disease. Therefore, the generalizability of the present results to samples with more diverse ages, ethnicities, and socioeconomic backgrounds remains to be investigated. For example, the age of the sample and the fact that they attended a university may mean that they have (a) more daily phone use than older populations; (b) more often use their phones for school, which may influence the content and usage of the phone, and especially the text inputted; and (c) have higher socioeconomic status and access to newer phones that have more functionality, which also support more phone use. In other samples, we may need to collect more data across longer time periods, or we may find that the association observed here between stress and total words typed does not replicate. For example, other participants who do not use their phones for academic activities may not show the same associations that we observed between total words typed and stress. We believe that if mobile sensing research is to serve everyone and improve global health, it will be important for future mobile sensing research to include participants from all backgrounds, including those from marginalized populations, who are often underrepresented in mental and physical health research.

Conclusions

In conclusion, we found that mobile sensing text data are associated with other well-validated measures of stress, including higher levels of self-reported perceived stress; more total, acute, and chronic lifetime stress exposure; higher levels of salivary inflammation as indexed by sCRP; and fewer self-reported hours slept. These data thus demonstrate the potential

promise of using mobile sensing data to help identify individuals at increased risk of developing stress-related health problems. Looking forward, additional research is needed to collect phrases of text (instead of only unigrams), to use mobile sensing measures other than keyboard text, and to assess other biological and inflammatory biomarkers of stress exposure and disease risk.

Acknowledgments: The authors wish to thank the participants in this study for their time and effort, as well as research assistant Nick Livingstone for collecting the data, the Stress Physiology Investigative Team (SPIT) lab at Iowa State University for conducting the immunological assays, and Grant Shields for assisting with the STRAIN data.

Author contributions: MLB, MNL, SRH, KLM, BWN, MLB, and NBA contributed to the conception and design, and gaining ethical approval; MLB, MNL, SRH, KLM, BWN, and MLB were involved with data acquisition and analysis; MLB, MNL, GMS, and NBA interpreted the data; MLB, MNL, and NBA drafted the article; and MLB, MNL, SRH, KLM, BWN, MLB, GMS, and NBA revised and finalized the manuscript. All authors reviewed and approved the final submitted version.

Declaration of conflicting interests: The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Using mobile sensing data to assess stress: Associations with perceived and lifetime stress, mental health, sleep, and inflammation. Nicholas Allen and Michelle Byrne both hold equity interests in Ksana Health Inc., a company that has the sole commercial license for certain versions of the Effortless Assessment of Risk States (EARS) mobile phone application and some related EARS tools. The other authors have nothing else to disclose.

Ethical approval: The ethics committee of University of Oregon approved this study (IRB protocol no. 07212016.019).

Funding: This study was funded by the Stress Measurement Network via a grant from the National Institute on Aging (R24AG048024). MLB was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number K01MH11951. GMS was supported by a Society in Science—Branco Weiss Fellowship, NARSAD Young Investigator grant #23958 from the Brain & Behavior Research Foundation, and National Institutes of Health grant K08 MH103443. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. A preprint of this manuscript is available at <https://psyarxiv.com/z85qj>.

Guarantor: NBA.

ORCID iDs: Michelle L. Byrne  <https://orcid.org/0000-0002-4180-8095>

Melissa L Barnes  <https://orcid.org/0000-0002-8745-0287>
 George M Slavich  <https://orcid.org/0000-0001-5710-3818>

References

- Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *Lancet* 2012; 380: 2197–2223.
- James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1789–1858.
- Forouzanfar MH, Afshin A, Alexander LT, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; 388: 1659–1724.
- Furman D, Campisi J, Verdin E, et al. Chronic inflammation in the etiology of disease across the life span. *Nat Med* 2019; 25: 1822–1832.
- Danesh J, Wheeler JG, Hirschfield GM, et al. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med* 2004; 350: 1387–1397.
- Visser M, Bouter LM, McQuillan GM, et al. Elevated C-reactive protein levels in overweight and obese adults. *JAMA* 1999; 282: 2131–2135.
- Wee CC, Mukamal KJ, Huang A, et al. Obesity and C-reactive protein levels among White, Black, and Hispanic US adults. *Obesity* 2008; 16: 875–880.
- Howren MB, Lamkin DM and Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med* 2009; 71: 171–186.
- Slavich GM and Irwin MR. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol Bull* 2014; 140: 774–815.
- Clarke DM and Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust* 2009; 190: S54–S60.
- Egede LE. Major depression in individuals with chronic medical disorders: prevalence, correlates and association with health resource utilization, lost productivity and functional disability. *Gen Hosp Psychiatry* 2007; 29: 409–416.
- Vancampfort D, Koyanagi A, Ward PB, et al. Perceived stress and its relationship with chronic medical conditions and multimorbidity among 229,293 community-dwelling adults in 44 low- and middle-income countries. *Am J Epidemiol* 2017; 186: 979–989.
- Renzaho AMN, Houg B, Oldroyd J, et al. Stressful life events and the onset of chronic diseases among Australian adults: findings from a longitudinal survey. *Eur J Public Health* 2014; 24: 57–62.
- Hyypya M. Quality of sleep and chronic illnesses. *J Clin Epidemiol* 1989; 42: 633–638.
- Girard JM, Cohn JF, Mahoor MH, et al. Nonverbal social withdrawal in depression: evidence from manual and automatic analyses. *Image Vis Comput* 2014; 32: 641–647.
- Lopresti AL, Hood SD and Drummond PD. A review of lifestyle factors that contribute to important pathways associated with major depression: diet, sleep and exercise. *J Affect Disord* 2013; 148: 12–27.
- Randall WM and Rickard NS. Reasons for personal music listening: a mobile experience sampling study of emotional outcomes. *Psychol Music* 2017; 45: 479–495.
- Rude S, Gortner E-M and Pennebaker J. Language use of depressed and depression-vulnerable college students. *Cognit Emot* 2004; 18: 1121–1133.
- Wang F, Wang Y, Wang J, et al. Assessing mental stress based on smartphone sensing data: an empirical study. In 2019 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 1031–1038. IEEE. doi:10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00200.
- Pennebaker JW, Mehl MR and Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol* 2003; 54: 547–577.
- Coppersmith G, Ngo K, Leary R, et al. Exploratory analysis of social media prior to a suicide attempt. In Proceedings of the third workshop on computational linguistics and clinical psychology, 2016, pp. 106–117. Association for Computational Linguistics. doi:10.18653/v1/W16-0311.
- Jashinsky J, Burton SH, Hanson CL, et al. Tracking suicide risk factors through Twitter in the US. *Crisis* 2014; 35: 51–59.
- Bucci W and Freedman N. The language of depression. *Bull Menninger Clin* 1981; 45: 334–358.
- Scherwitz L, Graham LE, Grandits G, et al. Self-involvement and coronary heart disease incidence in the multiple risk factor intervention trial. *Psychosom Med* 1986; 48: 187–199.
- Danner DD, Snowdon DA and Friesen WV. Positive emotions in early life and longevity: findings from the nun study. *J Pers Soc Psychol* 2001; 80: 804–813.
- Lind MN, Byrne ML, Wicks G, et al. The effortless assessment of risk states (EARS) tool: an interpersonal approach to mobile sensing. *JMIR Ment Health* 2018; 5: e10334.
- Cohen S, Kamarck T and Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983; 24: 385–396.
- Lee E-H. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res* 2012; 6: 121–127.
- Slavich GM and Shields GS. Assessing lifetime stress exposure using the stress and adversity inventory for adults (Adult STRAIN): an overview and initial validation. *Psychosom Med* 2018; 80: 17–27.
- Cazassa MJ, Oliveira MdS, Spahr CM, et al. The stress and adversity inventory for adults (Adult STRAIN) in Brazilian Portuguese: initial validation and links with executive function, sleep, and mental and physical health. *Front Psychol* 2020; 10: 3083.
- Mayer SE, Prather AA, Puterman E, et al. Cumulative lifetime stress exposure and leukocyte telomere length attrition: the

- unique role of stressor duration and exposure timing. *Psychoneuroendocrinology* 2019; 104: 210–218.
32. Sturmbauer SC, Shields GS, Hetzel E-L, et al. The stress and adversity inventory for adults (adult STRAIN) in German: an overview and initial validation. *PLOS ONE* 2019; 14: e0216419.
 33. Lovibond PF and Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* 1995; 33: 335–343.
 34. Antony MM, Bieling PJ, Cox BJ, et al. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol Assess* 1998; 10: 176–181.
 35. Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989; 28: 193–213.
 36. Jain S, Gautam V and Naseem S. Acute-phase proteins: as diagnostic tool. *J Pharm Bioallied Sci* 2011; 3: 118.
 37. Del Giudice M and Gangestad SW. Rethinking IL-6 and CRP: why they are more than inflammatory biomarkers, and why it matters. *Brain Behav Immun* 2018; 70: 61–75.
 38. Black S, Kushner I and Samols D. C-reactive protein. *J Biol Chem* 2004; 279: 48487–48490.
 39. Häfner S, Emeny RT, Lacruz ME, et al. Association between social isolation and inflammatory markers in depressed and non-depressed individuals: results from the MONICA/KORA study. *Brain Behav Immun* 2011; 25: 1701–1707.
 40. Owen N, Poulton T, Hay FC, et al. Socioeconomic status, C-reactive protein, immune factors, and responses to acute mental stress. *Brain Behav Immun* 2003; 17: 286–295.
 41. Taylor SE, Lehman BJ, Kiefe CI, et al. Relationship of early life stress and psychological functioning to adult C-reactive protein in the Coronary Artery Risk Development in Young Adults Study. *Biol Psychiatry* 2006; 60: 819–824.
 42. Pace TWW, Hu F and Miller AH. Cytokine-effects on glucocorticoid receptor function: relevance to glucocorticoid resistance and the pathophysiology and treatment of major depression. *Brain Behav Immun* 2007; 21: 9–19.
 43. Brydon L, Edwards S, Jia H, et al. Psychological stress activates interleukin-1 β gene expression in human mononuclear cells. *Brain Behav Immun* 2005; 19: 540–546.
 44. McQuaid RJ, Gabrys RL, McInnis OA, et al. Understanding the relation between early-life adversity and depression symptoms: the moderating role of sex and an interleukin-1 β gene variant. *Front Psychiatry* 2019; 10: 151.
 45. Owen BM, Eccleston D, Ferrier IN, et al. Raised levels of plasma interleukin-1 β in major and postviral depression. *ACTA Psychiatr Scand* 2001; 103: 226–228.
 46. Lucas T, Lumley MA, Flack JM, et al. A preliminary experimental examination of worldview verification, perceived racism, and stress reactivity in African Americans. *Health Psychol* 2016; 35: 366–375.
 47. Slavich GM, Way BM, Eisenberger NI, et al. Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proc Natl Acad Sci* 2010; 107: 14817–14822.
 48. Cicchetti D, Handley ED and Rogosch FA. Child maltreatment, inflammation, and internalizing symptoms: investigating the roles of C-reactive protein, gene variation, and neuroendocrine regulation. *Dev Psychopathol* 2015; 27: 553–566.
 49. Delany FM, Byrne ML, Whittle S, et al. Depression, immune function, and early adrenarche in children. *Psychoneuroendocrinology* 2016; 63: 228–234.
 50. O'Connor M-F, Irwin MR and Wellisch DK. When grief heats up: proinflammatory cytokines predict regional brain activation. *NeuroImage* 2009; 47: 891–896.
 51. Navarro VP, Iyomasa MM, Leite-Panissi CR, et al. New role of the trigeminal nerve as a neuronal pathway signaling brain in acute periodontitis: participation of local prostaglandins. *Pflugers Arch* 2006; 453: 73–82.
 52. Byrne ML, O'Brien-Simpson NM, Reynolds EC, et al. Acute phase protein and cytokine levels in serum and saliva: a comparison of detectable levels and correlations in a depressed and healthy adolescent sample. *Brain Behav Immun* 2013; 34: 164–175.
 53. Ouellet-Morin I, Danese A, Williams B, et al. Validation of a high-sensitivity assay for C-reactive protein in human saliva. *Brain Behav Immun* 2011; 25: 640–646.
 54. Out D, Hall RJ, Granger DA, et al. Assessing salivary C-reactive protein: longitudinal associations with systemic inflammation and cardiovascular disease risk in women exposed to intimate partner violence. *Brain Behav Immun* 2012; 26: 543–551.
 55. Dillon MC, Opris DC, Kopanczyk R, et al. Detection of homocysteine and C-reactive protein in the saliva of healthy adults: comparison with blood levels. *Biomark Insights* 2010; 2010: 57–61.
 56. Tausczik YR and Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010; 29: 24–54.
 57. Coppersmith G, Harman C and Dredze M. Measuring post traumatic stress disorder in twitter. In Proceedings of the eighth international AAAI conference on weblogs and social media, 2014.
 58. Guntuku SC, Buffone A, Jaidka K, et al. Understanding and Measuring Psychological Stress using Social Media. *ArXiv181107430 Cs*, 2019, <http://arxiv.org/abs/1811.07430>.
 59. Silge J and Robinson D. Tidytext: text mining and analysis using tidy data principles in R. *J Open Source Softw* 2016; 1: 37.
 60. Nielsen FÅ. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *ArXiv11032903 Cs*, 2011, <http://arxiv.org/abs/1103.2903>.
 61. Liu B. *Sentiment analysis and opinion mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.
 62. Kross E, Verduyn P, Boyer M, et al. Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion* 2019; 19: 97–107.
 63. Szabo YZ, Slavish DC and Graham-Engeland JE. The effect of acute stress on salivary markers of inflammation: a systematic review and meta-analysis. *Brain Behav Immun* 2020; 88: 887–900.
 64. Shields GS, Moons WG and Slavich GM. Inflammation, self-regulation, and health: an immunologic model of self-regulatory failure. *Perspect Psychol Sci* 2017; 12: 588–612.
 65. Aledavood T, Torous J, Triana Hoyos AM, et al. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. *Curr Psychiatry Rep* 2019; 21: 49.

-
66. Fraccaro P, Beukenhorst A, Sperrin M, et al. Digital biomarkers from geolocation data in bipolar disorder and schizophrenia: a systematic review. *J Am Med Inform Assoc* 2019; 26: 1412–1420.
 67. Suk M. and Prabhakaran B. Real-time facial expression recognition on smartphones. In 2015 IEEE winter conference on applications of computer vision, 2015, pp. 1054–1059. IEEE. DOI:10.1109/WACV.2015.145.
 68. Slavich GM, Taylor S and Picard RW. Stress measurement using speech: recent advancements, validation issues, and ethical and privacy considerations. *Stress* 2019; 22: 408–413.
-