



Published in final edited form as:

J Racial Ethn Health Disparities. 2020 October ; 7(5): 888–900. doi:10.1007/s40615-020-00712-y.

Twitter-characterized sentiment towards racial/ethnic minorities and Cardiovascular Disease (CVD) outcomes

Dina Huang¹, Yuru Huang¹, Nikki Adams², Thu T. Nguyen³, Quynh C. Nguyen^{1,*}

¹Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD 20742, USA

²Center for Substance Abuse Research, University of Maryland, College Park, MD 20742, USA

³Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA 94158, USA

Abstract

Sentiments towards racial/ethnic minorities may impact cardiovascular disease (CVD) through direct and indirect pathways. In this study, we assessed the association between Twitter-derived sentiments towards racial/ethnic minorities at state level and individual level CVD-related outcomes from the 2017 Behavioral Risk Factor Surveillance System (BRFSS). Outcomes included hypertension, diabetes, obesity, stroke, myocardial infarction (MI), coronary heart disease (CHD), and any CVD from BRFSS 2017 (N=433,434 to 433,680 across outcomes). A total of 30 million race-related tweets were collected using Twitter Streaming Application Programming Interface (API) from 2015 to 2018. Prevalence of negative and positive sentiment towards racial/ethnic minorities were constructed at state level and merged with CVD outcomes. Poisson regression was used, and all the models adjusted for individual level demographics as well as state level demographics. Individuals living in states with the highest level of negative sentiment towards racial/ethnic minorities had 11% higher prevalence of hypertension (PR:1.11, 95% CI: 1.08, 1.14), 15% higher prevalence of diabetes (PR:1.15, 95% CI: 1.08, 1.22), 14% higher prevalence of obesity (PR:1.14, 95% CI: 1.10, 1.18), 30% higher prevalence of stroke (PR: 1.30, 95% CI: 1.16, 1.46), 14% higher prevalence of MI (PR:1.14, 95% CI: 1.03, 1.25), 9% higher prevalence of CHD (PR: 1.09, 95% CI: 1.00, 1.19) and 16% higher prevalence of any CVD outcomes (PR: 1.16, 95% CI: 1.09, 1.24). Conversely, Twitter-derived positive sentiment towards

*Correspondence: qtnguyen@umd.edu; Tel.: 301-405-6425.

Author Contributions:

Conceptualization: Quynh C. Nguyen; Data Curation: Nikki Adams, Thu T. Nguyen, Dina Huang; Investigation: Nikki Adams, Thu T. Nguyen; Methodology: Quynh C. Nguyen, Dina Huang; Formal analysis: Dina Huang; Writing—original draft preparation: Dina Huang, Yuru Huang; Writing—review and editing: Dina Huang, Yuru Huang, Thu T. Nguyen, Nikki Adams, Quynh C. Nguyen; Visualization: Dina Huang; Funding acquisition: Quynh C. Nguyen

Publisher's Disclaimer: This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Ethnic Statement:

The study is original research and was not published in any form elsewhere. The study has been approved by University of Maryland Institutional Review Board. Informed consent was exempt because this study utilized publically available Twitter data and BRFSS data for secondary data analyses.

Conflict of Interest:

All the authors declare that they have no conflict of interest.

racial/ethnic minorities was associated with a lower prevalence of CVD outcomes. Programs and policies that promote racially inclusive environments may improve population health.

Keywords

Racial sentiment; racial/ethnic racial/ethnic minorities; Twitter; CVD outcomes

Introduction

Cardiovascular disease (CVD) is a leading cause of death in the US for all race/ethnicities. Racial/ethnic minorities experienced a higher risk of CVD than whites and the risk of CVD varies greatly across race/ethnic groups. Blacks experience earlier onset of hypertension and have a higher prevalence of hypertension, while Hispanics have a lower prevalence of hypertension, which is similar to whites [1, 2]. Different social context including residential segregation, discrimination, acculturation and environmental stressors may explain the differential rates of CVD across racial/ethnic groups in the US [3, 4]. Chronic exposure to racial discrimination contributes to CVD [5]. Previous studies suggest two hypothesized pathways linking negative sentiment towards racial/ethnic minorities and CVD. Negative sentiment towards racial/ethnic minorities from discrimination might directly increase cortisol (stress hormone), alter autonomic control, induce elevated inflammation, blunt β -adrenergic receptor responsiveness, and consequently contribute to CVD [6–8]. Additionally, negative sentiment can be linked to CVD indirectly via health behaviors. Perceived discrimination is associated with increased risk of unhealthy behaviors including unhealthy food consumption (e.g., fried foods, higher caloric foods), alcohol consumption and substance abuse [9–11].

Williams developed a framework that described the relationship between race and health outcomes [12]. Fundamental causes, such as racism, influence other social determinants of health outcomes and facilitate the biological processes leading to health disparities. Numerous research has explored physiological effects of racism or discrimination on heart rate and blood pressure [13]. Armstead examined the relationship of racial stressors, anger, and blood pressure responses and found that racist stimuli but not anger induced significant increases in blood pressure [14], which may partly explain the etiology of racism causing hypertension. Szanton et. al confirmed the link between self-reported racial discrimination and red blood cell oxidative stress, which identified a potential cellular pathway by which racial discrimination increase cardiovascular disease risks [15]. Racial discrimination may have both acute and prolonged effects on cardiac functions [16]. Brondolo and colleagues conducted an review of the empirical research investigating the effects of three levels of racism individual, internalized, and institutional racism and concluded that all levels of racism may influence the risk of hypertension [17].

Existing evidence supports the notion that perceived discrimination is associated with CVD risk factors and adverse CVD events [18, 19] and some of the studies suggest the association is more pronounced among African Americans [20, 19, 21, 22]. Everyday discrimination at baseline was associated with 33% (HR: 1.33, 95% CI: 1.11, 1.64) greater incidence of

metabolic syndrome after 14 years of follow up in the Study of Women's Health Across the Nation (SWAN) and the association was most evident in Black, Hispanic and Japanese Women [23]. An experimental study also found perceived racial discrimination to be associated with increased vascular responses and that African Americans had greater responses than white Americans [24]. Some studies purely focused on African American population due to the high vulnerability to CVD for this population. However, findings are mixed with some findings no association among African Americans [25, 26] and some finding an association [27]. An analyses of Multi-Ethnic Study of Atherosclerosis (MESA) data explored modification effects by racial/ethnic groups. However, investigators did not find differences in the association between perceived discrimination and CVD risk by race/ethnicity [28]. Subgroup analyses beyond comparisons between whites and African Americans comparison are also needed.

Additionally, gender plays a unique role in racial discrimination experiences. Numerous studies have demonstrated that there are gender differences in the prevalence, onset, and management of CVD [29, 30]. Socioeconomic and psychosocial factors, such as lack of social support and isolation, seem to have a more significant impact on CVD outcomes in women compared to men [30]. Gender specific psychological responses to discrimination may play a role in the pathology of CVD events [28]. Gender differences in the components of metabolic syndrome might be an explanation for different CVD risks [29, 31, 32]. Abdominal obesity is the most common condition of metabolic syndrome among women while the most prevalence component of metabolic syndrome varied among men in the US [33]. Modification of effects by gender has been understudied. Hence, it is crucial to examine the role of gender in potentially modifying the relationship between sentiment towards racial/ethnic minorities and CVD outcomes.

While racial bias is increasingly recognized as an important contributor to health disparities, previous research has been hampered by two major challenges—the measurement of area-level racial bias and quantification of its impact on health. Traditionally, racial bias has been measured at the individual level. Self-reported racial attitudes and beliefs are subject to a number of limitations including social desirability bias and self-censorship [34, 35], risking underestimating the impact of racism [36, 21]. Estimating experiences of discrimination and discriminatory attitudes solely at the individual level underestimates the effect of racism on health. Racism operates at multiple levels including internalized, personally mediated, and institutional [37]. Until now, characterization of the social environment towards racial/ethnic minorities has rarely been studied and few studies focus on sentiment towards racial/ethnic minorities as a possible environmental level risk factor in CVD studies. Moreover, existing studies rely on survey data collected in a few small areas, making it difficult to compare across locations and inform policy changes. Leveraging existing sources of big data may increase the breadth and cost-effectiveness of research endeavors and to allow for larger and more geographically diverse studies. In the online space, people express a variety of views and beliefs, including those that are race-related. In addition, research suggests that the sense of anonymity provided by online spaces emboldens people to express views they may not display during in-person interactions [38]. These aspects make social media an attractive source for capturing sensitive topics such as race-related discussions.

Study aims and hypotheses

In this study, we characterize sentiment towards racial/ethnic minorities across the contiguous US using Twitter. Twitter as a data source provides a large volume of data. Information from Twitter may be less affected by the type of reporting bias that affects survey responses. We assess the associations between social environment towards racial/ethnic minorities and CVD outcomes using Behavioral Risk Factor Surveillance System (BRFSS) 2017. We hypothesize that negative sentiment towards racial/ethnic minorities would be associated with higher prevalence of CVD outcomes and positive emotions towards racial/ethnic minorities would be associated with lower prevalence of CVD outcomes. We investigate race/ethnicity and sex as potential modifiers of the association between sentiment towards racial/ethnic minorities and CVD outcomes where non-Hispanic Blacks and females may experience higher prevalence of CVD outcomes when exposed to higher levels of negative sentiment.

Methods

Twitter variables

From 2015 to 2018, we collected a 1% random general tweet samples using Twitter Streaming Application Programming Interface (API) across 48 states in the contiguous US and the District of Columbia. Tweets with geotags (latitude and longitude coordinates) and tweets without geotags were collected. Tweets without geotags had “place” information (usually city and state). The place information associated with tweets was used to determine the state location of where the tweet was sent. These tweets were filtered by a race term list which consisted of 518 terms of race/ethnic groups as well as race/ethnic slurs. The race term list was constructed based on racial and ethnic terms from US Census, racial slurs from an online database [39] and from prior studies on race-related online conversation [40, 41].

To prevent undue influence of a small number of very frequent users and to exclude potential Twitter bots, we excluded users who had more than 1,000 race-related tweets/a year in the data set, representing 3–4% of all tweets [42]. We also removed duplicated tweets and dropped irrelevant tweets such as those mentioning “Indian Shores” and “egg whites.” After exclusion, a total of 1,320,647 tweets with geotags and 30,977,747 of all tweets were included in analyses. We pre-processed the tweets to remove stop words, emojis, urls, punctuations as well as hashtags. Each tweet was tokenized and term frequency inverse document frequency (TFIDF) was used to assign inverse weights that are proportional to word frequency.

We derived both positive and negative sentiments of race-related tweets using Python 3.7. Positive and negative sentiments are modeled separately as they may have independent effects [43]. We used manually labeled training data from Sentiment140 (n=498) [44], Sanders (n=5,113) [45], Kaggle (n=7,086) [46], and a sample of tweets filtered by race-terms labeled by our research team (n=6,481). Sentiment 140, Sanders, and Kaggle datasets are all publicly available training datasets specifically labelled for sentiment analysis. Support Vector Machine (SVM) was used to classify tweets to “negative” versus “not negative.” The “not negative” category included positive and neutral tweets. SVM is a non-

probabilistic classifier to separate data points to different hyperplanes utilizing maximum margins. Each data point is assigned to the hyperplane with the maximum distance. We used five-fold cross validation to assess the model performance and reached a high accuracy (90.63%) as well as a high F1 score (83.81%). F1 score is a commonly used model performance measure in analyses of binary classification that balances model precision (positive predictive value) and recall (sensitivity). A high F1 score (close to 1) is generally an indicator of excellent model performance especially when the data are unbalanced.

Similarly, we used SVM to classify tweets to “positive” versus “not positive”. We obtained 1,659 tweets labeled by our research team in addition to the manually labeled training data mentioned above for classifying sad tweets. The “not positive” category included negative and neutral tweets. We performed five-fold cross validation and reached a high accuracy (88.91%) and a high F1 score (80.59%). After assigning sentiment to each tweet, the percent of negative tweets and positive tweets were calculated separately at state level and were then categorized into tertiles. State level percent of negative tweets and percent of positive tweets were merged to individual BRFSS health outcomes data (i.e., individuals living in the same state would have the same Twitter-derived characteristics).

Study population, Outcomes, and Covariates

The Behavioral Risk Factor Surveillance System (BRFSS) is a telephone-based survey that focuses on chronic health conditions and health behaviors of adults across 50 states of US and District of Columbia [47]. BRFSS includes about 400,000 interviews each year, providing extensive data source on self-reported health-related conditions. The BRFSS survey data are collected annually through a random-digit-dialed telephone, for both landline and cellular phone samples. In 2017, the mean combined response rate was 44.9%, ranging from 30.6% in Illinois to 64.1% in Wyoming [48].

BRFSS Health Measures.—In this analysis, we examined CVD risk factors including hypertension, diabetes and obesity as well as CVD events such as stroke, myocardial infarction (MI), coronary heart disease (CHD) and “any CVD”. For hypertension, diabetes (excluding gestational diabetes), stroke, MI and CHD, the questions were asked as “Has a doctor, nurse or other health professional ever told you that you had ...?” Each outcome was asked separately. Self-reported body mass index (BMI) greater than or equal to 30 kg/m² was defined as obesity. We defined “any CVD” as individual who had one or more of the three CVD events stated above. From BRFSS 2017, a total of 450,016 participants were included in analyses. We excluded participants with missing age (N=6,101) or missing sex (N=287). The number of participants varied by outcomes with the following totals: 443,415 participants for hypertension, 443,020 participants for diabetes, 443,718 participants for obesity, 443,624 participants for stroke, 443,615 participants for MI, 443,550 participants for CHD and 443,718 for any CVD.

Covariates.—We included the following individual level covariates in models: age, sex, education, race/ethnicity, and marital status. Age groups were defined using five-year intervals from 18 to 80, and adults aged over 80 were grouped together. Education was categorized as “Did not graduate high school”, “Graduated high school”, “Attended college

or Technical School” and “Graduated from college or Technical School”. Race/ethnicity was categorized as “non-Hispanic White”, “non-Hispanic Black”, “Hispanic” and “Others”. Sex and race/ethnicity were investigated as possible effect-modifiers.

American Community Survey.—We also included state level demographics including percent of non-Hispanic White, percent of non-Hispanic Black, percent of Hispanic, and median household income from American Community Survey (ACS) 2017 1-year estimates [49].

Statistical Methods

Stata MP15 (StataCorp LP, College Station, TX) was used for statistical analyses and R software [50] was used for mapping. Poisson regression was used to estimate the prevalence ratio of each outcome. We implemented separate analyses for geotagged tweets (tweets with latitude and longitude coordinates) and all tweets. All the analyses accounted for survey design and survey weights. Models were adjusted for individual level demographics including age, race/ethnicity, sex, marital status as well as state level demographics including percent non-Hispanic white, percent non-Hispanic Black, percent Hispanic and median household income. To examine whether the estimated effects of sentiment on CVD outcomes vary by sex and race/ethnicity, interactions terms between sentiment and sex and sentiment and racial/ethnic categories were included. Wald test was used to assess the significance of interaction terms. We then performed stratified analyses by sex and by racial/ethnic groups to obtain sex- and racial-specific prevalence ratios.

Results

The prevalence of negative tweets across the US was 39.9%, and the prevalence of positive tweets across the US was 6.4% for all race-related tweets. Figs 1–2 map the state level prevalence of negative and positive sentiment of all race-related tweets. Mississippi (46.6%), Louisiana (46.2%) and Michigan (44.9%) had the highest prevalence of negative sentiment towards racial/ethnic minorities. The southeastern US had higher prevalence of negative sentiment compared to western US. Utah had the highest prevalence of positive sentiment towards racial/ethnic minorities (8.1%), followed by North Dakota (7.5%) and California (7.4%). The western US generally had higher prevalence of positive sentiment compared to the eastern US. For race-related tweets with geotags (eFigures 1–2), the national trends were similar where the western US had more positive sentiment and the southeastern US had higher prevalence of negative sentiment.

Descriptive statistics of BRFSS participants are presented in Table 1. Half of participants were males (49%). The majority of participants were non-Hispanic white (62.6%); 11.8% of the participants were non-Hispanic black and 17.1% of the participants were Hispanic. Moreover, 7.1% of the participants were non-Hispanic “other race,” and 1.4% were non-Hispanic multiracial.

Utilizing all race-related tweets, we observed an association between negative sentiment and higher prevalence of CVD outcomes, while positive sentiment towards racial/ethnic minorities was associated with lower prevalence of CVD outcomes (Table 2). People living

in states with the highest level of negative tweets towards racial/ethnic minorities had 11% higher prevalence of hypertension (PR:1.11, 95% CI: 1.08, 1.14), 15% higher prevalence of diabetes (PR:1.15, 95% CI: 1.08, 1.22), 14% higher prevalence of obesity (PR:1.14, 95% CI: 1.10, 1.18), 30% higher prevalence of stroke (PR: 1.30, 95% CI: 1.16, 1.46), 14% higher prevalence of MI (PR:1.14, 95% CI: 1.03, 1.25), 9% higher prevalence of CHD (PR: 1.09, 95% CI: 1.00, 1.19) and 16% higher prevalence of any CVD (PR: 1.16, 95% CI: 1.09, 1.24). Conversely, people living in states with the highest tertile of positive tweets towards racial/ethnic minorities had 3% lower prevalence of hypertension (PR:0.97, 95% CI: 0.94, 1.00), 6% lower prevalence of diabetes (PR:0.94, 95% CI: 0.90, 0.99), 3% lower prevalence of obesity (PR:0.97, 95% CI: 0.94, 1.00), 11% lower prevalence of stroke (PR: 0.89, 95% CI: 0.80, 0.98), 9% lower prevalence of MI (PR:0.91, 95% CI: 0.83, 0.98), 6% lower prevalence of CHD (PR: 0.94, 95% CI: 0.86, 1.02) and 10% lower prevalence of any CVD (PR: 0.90, 95% CI: 0.86, 0.95).

We tested effect-modification by sex and racial/ethnic groups for all the CVD outcomes (Table 3 and Table 4). Females were more affected by negative sentiment towards racial/ethnic minorities compared to males in terms of hypertension, stroke, CHD and any CVD. The effects of negative sentiment towards racial/ethnic minorities on diabetes and obesity were more evident in males (Table 3). The prevalence ratio of hypertension was 1.13 (95% CI: 1.08, 1.17) for females and was 1.10 (95% CI: 1.06, 1.15) for males, comparing people living in states with the highest level of negative tweets to the lowest. The prevalence ratio of CHD was 1.16 (95% CI: 1.00, 1.34) for females and was 1.04 (95% CI: 0.93, 1.17) for males comparing people living in states with the highest level of negative tweets to the lowest. For any CVD outcomes, the prevalence ratio was 1.20 (95% CI: 1.09, 1.32) for females and was 1.13 (95% CI: 1.04, 1.23) for males comparing people living in states with the highest level of negative tweets to the lowest.

We tested for effect modification by racial/ethnicity with Whites as the referent. Examining negative sentiment and hypertension, significant interactions were detected for Blacks and Hispanics. For MI, there was also a significant interaction detected for Hispanics. Examining positive sentiment and our health outcomes, significant interactions were detected for Hispanics for hypertension, stroke, and any CVD. We present stratified analyses in Table 4. The associations between negative sentiment with hypertension, MI and any CVD were stronger for non-Hispanic whites and non-Hispanic blacks than other race/ethnicity groups. However, the association between negative sentiment with diabetes, obesity and stroke was stronger in Hispanics than any other racial/ethnic groups. For CVD risk factors such as hypertension, diabetes and obesity, positive sentiment towards racial/ethnic minorities was more protective in non-Hispanic blacks than non-Hispanic whites.

We conducted sensitivity analyses for the subset of tweets with latitude and longitude information or geotags. Descriptive results showing the prevalence of positive and negative sentiment for geotagged tweets are presented in eFigures 1–2. Sensitivity analyses restricting to tweets with geotags followed a similar pattern as compared to analyses with all tweets (eTable 3).

Discussion

In this study, we constructed a new measure of racial sentiment using Twitter data and assessed the association between negative and positive sentiment with CVD outcomes in a nationally representative sample. Our results suggested that negative sentiment towards racial/ethnic minorities was significantly associated with higher prevalence of CVD risk factors including hypertension, diabetes and obesity as well as CVD events including stroke, MI, CHD, and any CVD. Conversely, positive sentiment towards racial/ethnic minorities was associated with lower prevalence of hypertension, diabetes, obesity, stroke, MI, and any CVD. The associations between negative sentiment with hypertension, MI and any CVD were also stronger for non-Hispanic whites and non-Hispanic blacks than other race/ethnicity groups.

In this study, we used negative sentiment and positive sentiment as two separate variables. Negative emotions may be linked to CVD outcomes through inducing direct biological changes including autonomic control and elevated inflammation [8], and through indirect pathway via unhealthy behaviors [51, 9]. Positive emotions have been associated with reduced blood pressure [52] and decreased risk of mortality [53, 54]. Both negative and positive emotion may share common pathways to CVD outcomes, however, the correlation between positive and negative emotions was generally low and in previous studies the two affective states were associated with different external factors [43], thus providing evidence for assessing positive and negative sentiment as separate factors.

Our study found the relationship between negative sentiment and CVD outcomes tended to be stronger among females than males (except for risk factors like obesity and diabetes). Males and females are different in metabolism, psychosocial attributes, social roles and social behaviors [30]. CVD outcomes of females may be more affected by a negative social environment towards racial/ethnic minorities. Males have higher risk of CVD, however, socioeconomic and psychosocial factors may have a greater impact on CVD in women compared to men [30]. The results of previous studies suggest the complexity of gender in the association between discrimination and CVD. A laboratory study exposed 92 African American college students to socially stressful stimuli and male college students showed more pronounced elevation in blood pressure than females [55]. A cross-sectional study found self-reported discrimination was associated with Interleukin-6 (IL-6) in women but not in men and was not associated with C-Reactive Protein (CRP) in both gender groups [56].

Given the well documented racial/ethnic disparities in CVD health and limited number of cohorts with diverse study populations, this study adds to the limited literature [24–27]. Our study suggests that social environment towards racial/ethnic minorities is associated with CVD in both racial/ethnic minorities and non-Hispanic whites. Our study highlights the potential impact of racial sentiment on health outcomes. Online racial sentiment can be perceived as one form of cultural racism which works as a risk factor for cardiovascular diseases and act as a barrier to health promotion. David Williams and colleagues defined cultural racism as the infusion of the ideology of inferiority in the values, language, imagery, symbols, and unstated assumptions of the larger society. It is displayed through media,

stereotyping, and norms within society and its institutions [57]. Importantly, cultural racism produces an environment where institutional and individual-level discrimination can thrive. Views and activities described online and via social media can help shape perceived norms, beliefs and subsequently behaviors of people. Social climates that are less welcoming may encourage racial discrimination or tolerance of discrimination. Using social media data, this study sought to measure the social context related to race in order to quantify its influence on cardiovascular outcomes. Our results identify online racial sentiment as one source of cultural racism that is risky for both racial/ethnic minorities and white Americans and open new avenues through which to address the burden of cardiovascular disease in historically conflicted communities. For example, intervention programs can be designed specifically for neighborhoods that are found to be less accepting of racial/ethnic minorities. Creating an environment more inclusive of all racial groups can improve cardiovascular health outcomes for the total population.

Strengths and limitations

This study has some notable strengths. To the best of our knowledge, this is the first study leveraging Twitter data to examine racial sentiment and cardiovascular disease outcomes. We utilized a nationally representative sample of BRFSS participants that is geographically and ethnically diverse and allowed for the examination of various CVD risk factors and CVD events including hypertension, diabetes, obesity, stroke, MI and CHD. We utilized extensive racial/ethnic information from BRFSS for stratified analyses, allowing comparison of associations across different racial/ethnic groups. We constructed new measures of racial sentiment from Twitter data and examined the links between these social environment measures and individual level health outcomes. Results with geotagged tweets were similar to analyses incorporating all race-related tweets. The social environment towards racial/ethnic minorities was constructed at state level and the results of our study may inform state level policies towards racial/ethnic minorities and subsequently promote population health via policy changes. Our study highlights the potential influence of social and structural factors on health.

Our study has some limitations. The sociocultural model presents the following interrelated factors affecting health: individual, relationship, community and societal characteristics. Overlapping ellipses represent how one factor at one level influences other factors at another level. Our study focused on sentiment towards racial/ethnic minorities as a social environmental factor. It is important to note that both individual level and contextual level racial sentiment are risk factors for CVD [58]. We focused on the social context at state level because study results may inform future state level policy towards racial/ethnic minorities and promote population health through policy changes. In this study, we do not have data on individual level stress measures and perceived discrimination, which are possible mediators of the relationship between a hostile social environment and health outcomes. Future studies may consider structural equation modeling to assess the overall effect of discrimination and CVD and to assess the role of individual level stress, individual perceived discrimination, and social context on CVD as well as investigate the interplay of factors across levels.

Furthermore, we collected tweets from 2015 to 2018 to construct social environment measures and thus there is some temporal mismatch in the Twitter variables and the outcome data which came from the BRFSS 2017. A three-year Twitter collection period was used in order to ensure sufficient number of geotagged tweets for a somewhat rare topic of race/ethnicity. Some states had fewer tweets than others and a multiple year collection period helped ensure reliable estimates for all states. Additionally, a longer period of time in social environment exposure may be preferred because cardiovascular outcomes are chronic conditions which may take years to develop. Negative sentiment towards racial/ethnic minorities in the social environment could be underestimated since Twitter's hateful conduct policy prohibits accounts that "promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease" [59]. Thus accounts which espouse racist sentiment may have been removed and therefore not included in our analysis, resulting in an underestimation of negative racial sentiment. Moreover, we utilized machine learning to determine the racial sentiment of tweets, however, for some tweets misclassification may have occurred because these algorithms may not recognize nuanced language features such as humor or sarcasm.

Also, given Twitter does not collect demographic information of Twitter users, we were unable to directly compare user differences. However, studies are currently underway to predict characteristics such as gender, race and age of Twitter users leveraging information contained in their tweets and profile [60]. We did sensitivity analyses restricted to tweets with geotags. Although the prevalences of negative and positive sentiment were quite different comparing geotagged tweets to all tweets, the geographic distribution of the sentiment of the tweets did not change substantially; we consistently observed more negative sentiment in the southeastern US and more positive sentiment in the western US. Analyses examining associations between sentiment and health outcomes also produced similar results for geotagged tweets and all tweets. Twitter location service is off by default [61]. Users can choose to turn on or turn off the location at any time and users can also delete past locations [62]. Differences in prevalence of sentiment variables among geotagged tweets and all tweets may be due to differences in Twitter users who choose to enable location tracking versus those who do not. It is also possible that the sample of geotagged tweets is too small (4%) to be representative of all tweets, leading to differences in sentiment towards racial/ethnic minorities. Another limitation of our analyses is that we only included tweets. Nonetheless, millions of tweets are sent daily, and 90% of Twitter users make their profile public [63]. Twitter readily makes available an API for researchers to download publicly available tweets. Facebook and Instagram do not have an API for that purpose and more profiles are private on those platforms.

Conclusion

Our research leveraging racially charged comments in social media advances a long-standing challenge in measuring the social context likely encountered by racial/ethnic minorities and provide new opportunities to investigate the role of the social context in shaping health and health disparities. This research has the potential to lead to greater tracking of area-level racial bias and to provide essential information needed to the develop

interventions to reduce the impact of racial bias on health and health disparities. In this study, we measured sentiment towards racial/ethnic minorities at state level using Twitter and assessed the influence of negative and positive sentiment towards racial/ethnic minorities with CVD outcomes. Our results suggest that negative sentiment towards racial/ethnic minorities was associated with higher prevalence of CVD outcomes and positive sentiment was associated with lower prevalence of CVD outcomes. The novel use of Twitter to measure social environments may provide a possible direction for racial discrimination and health research. The study results may inform state level policies towards racial/ethnic minorities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding:

This study was supported by grants from the National Institutes of Health's Big Data to Knowledge Initiative (BD2K) (award number: 5K01ES025433) and the National Library of Medicine (award number: R01LM012849 (Dr. Nguyen, Q.C., PI).

Collection and storage of the Twitter data used in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number U01DA038360, awarded to Center for Substance Abuse Research (CESAR) at the University of Maryland, College Park. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

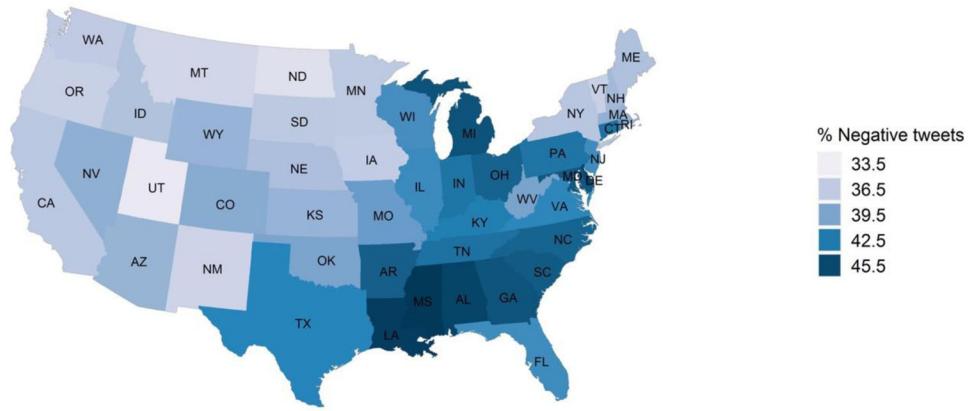
References

1. Lackland DT. Racial differences in hypertension: implications for high blood pressure management. *The American journal of the medical sciences*. 2014;348(2):135–8. doi:10.1097/maj.0000000000000308. [PubMed: 24983758]
2. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S et al. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. *Circulation*. 2018;137(12):e67–e492. doi:10.1161/cir.0000000000000558. [PubMed: 29386200]
3. Geronimus AT, Hicken M, Keene D, Bound J. “Weathering” and age patterns of allostatic load scores among blacks and whites in the United States. *American journal of public health*. 2006;96(5):826–33. doi:10.2105/ajph.2004.060749. [PubMed: 16380565]
4. James SA, Keenan NL, Strogatz DS, Browning SR, Garrett JM. Socioeconomic status, John Henryism, and blood pressure in black adults. The Pitt County Study. *Am J Epidemiol*. 1992;135(1):59–67. [PubMed: 1736661]
5. Brewer LC, Cooper LA. Race, discrimination, and cardiovascular disease. *The virtual mentor : VM*. 2014;16(6):270–4. doi:10.1001/virtualmentor.2014.16.06.stas2-1406.
6. Hill L, Sherwood A, McNeilly M, Anderson NB, Blumenthal JA, Hinderliter AL. Impact of Racial Discrimination and Hostility on Adrenergic Receptor Responsiveness in African American Adults. *Psychosomatic medicine*. 2018;80(2):208–15. [PubMed: 29206724]
7. Townsend SSM, Major B, Gangi CE, Mendes WB. From “In the Air” to “Under the Skin”: Cortisol Responses to Social Identity Threat. *Personality and Social Psychology Bulletin*. 2011;37(2):151–64. doi:10.1177/0146167210392384. [PubMed: 21239591]
8. Saban KL, Mathews HL, Bryant FB, Tell D, Joyce C, DeVon HA et al. Perceived discrimination is associated with the inflammatory response to acute laboratory stress in women at risk for cardiovascular disease. *Brain, behavior, and immunity*. 2018;73:625–32.

9. Pascoe EA, Richman LS. Effect of discrimination on food decisions. *Self and Identity*. 2011;10(3):396–406.
10. Richman LS, Boynton MH, Costanzo P, Banas K. Interactive effects of discrimination and racial identity on alcohol-related thoughts and use. *Basic and Applied Social Psychology*. 2013;35(4):396–407.
11. Gibbons FX, Etcheverry PE, Stock ML, Gerrard M, Weng C-Y, Kiviniemi M et al. Exploring the link between racial discrimination and substance use: What mediates? What buffers? *Journal of personality and social psychology*. 2010;99(5):785. [PubMed: 20677890]
12. Williams DR. Race and health: basic questions, emerging directions. *Annals of epidemiology*. 1997;7(5):322–33. [PubMed: 9250627]
13. Harrell JP, Hall S, Taliaferro J. Physiological Responses to Racism and Discrimination: An Assessment of the Evidence. *American journal of public health*. 2003;93(2):243–8. doi:10.2105/ajph.93.2.243. [PubMed: 12554577]
14. Armstead CA, Lawler KA, Gorden G, Cross J, Gibbons J. Relationship of racial stressors to blood pressure responses and anger expression in black college students. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*. 1989;8(5):541–56.
15. Szanton SL, Rifkind JM, Mohanty JG, Miller ER 3rd, Thorpe RJ, Nagababu E et al. Racial discrimination is associated with a measure of red blood cell oxidative stress: a potential pathway for racial health disparities. *International journal of behavioral medicine*. 2012;19(4):489–95. doi:10.1007/s12529-011-9188-z. [PubMed: 21913047]
16. Hoggard LS, Hill LK, Gray DL, Sellers RM. Capturing the cardiac effects of racial discrimination: Do the effects “keep going”? *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*. 2015;97(2):163–70. doi:10.1016/j.ijpsycho.2015.04.015. [PubMed: 25931114]
17. Brondolo E, Love EE, Pencille M, Schoenthaler A, Ogedegbe G. Racism and hypertension: a review of the empirical evidence and implications for clinical practice. *American journal of hypertension*. 2011;24(5):518–29. doi:10.1038/ajh.2011.9. [PubMed: 21331054]
18. Dolezsar CM, McGrath JJ, Herzig AJM, Miller SB. Perceived racial discrimination and hypertension: a comprehensive systematic review. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*. 2014;33(1):20–34. doi:10.1037/a0033718.
19. Brondolo E, Rieppi R, Kelly KP, Gerin W. Perceived racism and blood pressure: a review of the literature and conceptual and methodological critique. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*. 2003;25(1):55–65. doi:10.1207/s15324796abm2501_08. [PubMed: 12581937]
20. Williams DR, Neighbors H. Racism, discrimination and hypertension: evidence and needed research. *Ethn Dis* 2001;11(4):800–16. [PubMed: 11763305]
21. Krieger N, Sidney S. Racial discrimination and blood pressure: the CARDIA Study of young black and white adults. *Am J Public Health*. 1996;86(10):1370–8. [PubMed: 8876504]
22. Lewis TT, Barnes LL, Bienias JL, Lackland DT, Evans DA, Mendes de Leon CF. Perceived discrimination and blood pressure in older African American and white adults. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2009;64(9):1002–8. doi:10.1093/gerona/glp062.
23. Beatty DM, Chang Y, Brown C, Bromberger JT, Matthews KA. Everyday Discrimination and Metabolic Syndrome Incidence in a Racially/Ethnically Diverse Sample: Study of Women’s Health Across the Nation. *Psychosomatic medicine*. 2018;80(1):114–21. [PubMed: 28787363]
24. Thomas KS, Nelesen RA, Malcarne VL, Ziegler MG, Dimsdale JE. Ethnicity, perceived discrimination, and vascular reactivity to phenylephrine. *Psychosom Med* 2006;68(5):692–7. doi:10.1097/01.psy.0000238214.80871.e6. [PubMed: 17012522]
25. Dunlay SM, Lippmann SJ, Greiner MA, O’Brien EC, Chamberlain AM, Mentz RJ et al., editors. *Perceived discrimination and cardiovascular outcomes in older African Americans: insights from the Jackson Heart Study Mayo Clinic Proceedings*; 2017: Elsevier.

26. Okhomina VI, Glover L, Taylor H, Sims M. Dimensions of and responses to perceived discrimination and subclinical disease among African-Americans in the Jackson heart study. *Journal of racial and ethnic health disparities*. 2018;5(5):1084–92. [PubMed: 29313298]
27. Chae DH, Lincoln KD, Adler NE, Syme SL. Do experiences of racial discrimination predict cardiovascular disease among African American men? The moderating role of internalized negative racial group attitudes. *Social science & medicine (1982)*. 2010;71(6):1182–8. doi:10.1016/j.socscimed.2010.05.045. [PubMed: 20659782]
28. Everson-Rose SA, Lutsey PL, Roetker NS, Lewis TT, Kershaw KN, Alonso A et al. Perceived discrimination and incident cardiovascular events: The Multi-Ethnic Study of Atherosclerosis. *American journal of epidemiology*. 2015;182(3):225–34. [PubMed: 26085044]
29. Regitz-Zagrosek V, Lehmkuhl E, Weickert MO. Gender differences in the metabolic syndrome and their role for cardiovascular disease. *Clinical research in cardiology : official journal of the German Cardiac Society*. 2006;95(3):136–47. doi:10.1007/s00392-006-0351-5. [PubMed: 16598526]
30. Moller-Leimkuhler AM. Gender differences in cardiovascular disease and comorbid depression. *Dialogues in clinical neuroscience*. 2007;9(1):71–83. [PubMed: 17506227]
31. Beigh SH, Jain S. Prevalence of metabolic syndrome and gender differences. *Bioinformatics*. 2012;8(13):613–6. doi:10.6026/97320630008613. [PubMed: 22829741]
32. Lee S, Ko Y, Kwak C, Yim E-S. Gender differences in metabolic syndrome components among the Korean 66-year-old population with metabolic syndrome. *BMC Geriatr* 2016;16:27-. doi:10.1186/s12877-016-0202-9. [PubMed: 26803469]
33. Pradhan AD. Sex Differences in the Metabolic Syndrome: Implications for Cardiovascular Health in Women. *Clinical Chemistry*. 2014;60(1):44. doi:10.1373/clinchem.2013.202549. [PubMed: 24255079]
34. Stocké V. Determinants and Consequences of Survey Respondents' Social Desirability Beliefs about Racial Attitudes. *Methodology*. 2007;3(3):125–38. doi:10.1027/1614-2241.3.3.125.
35. An BP. The role of social desirability bias and racial/ethnic composition on the relation between education and attitude toward immigration restrictionism. *The Social Science Journal*. 2015;52(4):459–67. doi:10.1016/j.soscij.2014.09.005.
36. Nuru-Jeter AM, Michaels EK, Thomas MD, Reeves AN, Thorpe RJ Jr, LaVeist TA. Relative roles of race versus socioeconomic position in studies of health inequalities: a matter of interpretation. *Annual review of public health*. 2018;39:169–88.
37. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. *American journal of public health*. 2000;90(8):1212. [PubMed: 10936998]
38. Suler J The online disinhibition effect. *Cyberpsychology & behavior*. 2004;7(3):321–6. [PubMed: 15257832]
39. The Racial Slur Database. 2018 <http://www.rsd.org/>. Accessed May 21st 2019.
40. Anderson M, Hitlin P. Social media conversations about race. 2016 <https://www.pewinternet.org/2016/08/15/social-media-conversations-about-race/>. Accessed May 21st 2019.
41. Bartlett J, Reffin J, Rumball N, Williamson S. Anti-Social Media. London 2014 https://www.demos.co.uk/files/DEMOS_Anti-social_Media.pdf. Accessed May 21st 2019.
42. Lansley G, Longley PA. The geography of Twitter topics in London. *Computers, Environment and Urban Systems*. 2016;58:85–96. doi:10.1016/j.compenvurbsys.2016.04.002.
43. Diener E, Emmons RA. The independence of positive and negative affect. *J Pers Soc Psychol* 1984;47(5):1105–17. [PubMed: 6520704]
44. Iglehart JK. Expanding Coverage for Children — The Democrats' Power and SCHIP Reauthorization. *New England Journal of Medicine*. 2009;360(9):855–7. doi:10.1056/NEJMp0900461. [PubMed: 19196665]
45. <measles in england and wales analysis of factors underlying seasonal patterns.pdf>.
46. Kaggle UMich SI650 - Sentiment Classification. 2011 <https://www.kaggle.com/c/si650winter11>. Accessed Nov. 30 2018.
47. Behavioral Risk Factor Surveillance System. Centers for Disease Control and Prevention. 2019 <https://www.cdc.gov/brfss/index.html>. Accessed April 24 2019.

48. The Behavioral Risk Factor Surveillance System 2017 Summary Data Quality Report. Center for Disease Control and Prevention 2018.
49. American Community Survey (ACS). 2019 <https://www.census.gov/programs-surveys/acs/news/data-releases/2017/release.html>. Accessed April 19 2019.
50. R Core Team, R: A Language and Environment for Statistical Computing In: R Foundation for Statistical Computing, Vienna, Austria 2013 <http://www.R-project.org/>. Accessed Nov. 21 2018.
51. Sims M, Diez-Roux AV, Gebreab SY, Brenner A, Dubbert P, Wyatt S et al. Perceived discrimination is associated with health behaviours among African-Americans in the Jackson Heart Study. *J Epidemiol Community Health*. 2016;70(2):187–94. [PubMed: 26417003]
52. Ho M-W. Happiness is a heartbeat away. *Science in Society Archive*. 2007 <http://www.i-sis.org.uk/happinessIsAHeartbeatAway.php>. Accessed April 23 2019.
53. Danner DD, Snowdon DA, Friesen WV. Positive emotions in early life and longevity: findings from the nun study. *J Pers Soc Psychol* 2001;80(5):804–13. [PubMed: 11374751]
54. Chida Y, Steptoe A. Positive psychological well-being and mortality: a quantitative review of prospective observational studies. *Psychosom Med* 2008;70(7):741–56. doi:10.1097/PSY.0b013e31818105ba. [PubMed: 18725425]
55. Morris-Prather CE, Harrell JP, Collins R, Leonard KL, Boss M, Lee JW. Gender differences in mood and cardiovascular responses to socially stressful stimuli. *Ethnicity & disease*. 1996;6(1–2):123–31. [PubMed: 8882841]
56. Kershaw KN, Lewis TT, Roux AVD, Jenny NS, Liu K, Penedo FJ et al. Self-reported experiences of discrimination and inflammation among men and women: The multi-ethnic study of atherosclerosis. *Health Psychology*. 2016;35(4):343. [PubMed: 27018725]
57. Williams DR, Lawrence JA, Davis BA. Racism and Health: Evidence and Needed Research. *Annu Rev Public Health*. 2019;40:105–25. doi:10.1146/annurev-publhealth-040218-043750. [PubMed: 30601726]
58. Ahmed AT, Mohammed SA, Williams DR. Racial discrimination & health: pathways & evidence. *Indian J Med Res* 2007;126(4):318–27. [PubMed: 18032807]
59. <Gay 2004 theory of measles elimination implication for design of elimination strategies.pdf>.
60. Cesare N, Grant C, Hawkins JB, Brownstein JS, Nsoesie EO. Demographics in social media data for public health research: does it matter? arXiv preprint arXiv:171011048. 2017.
61. <Hill 1911 epidemiology of diphtheria scarlet fever and measles.pdf>.
62. <butler 1947 whooping cough and measles.pdf>.
63. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist JN, editors. Understanding the demographics of twitter users. Fifth international AAAI conference on weblogs and social media; 2011.



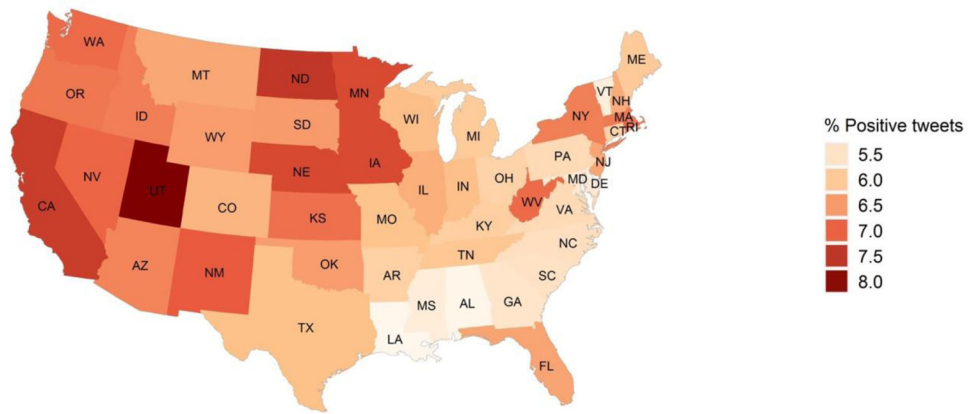


Fig 2. Prevalence of Twitter-derived positive sentiment towards racial/ethnic minorities at state level

Table 1.

Weighted demographics for individuals in BRFSS 2017

Demographics	n^a	%
Sex		
Male	198,725	48.7%
Female	251,007	51.3%
Education		
Did not graduate high school	32,687	13.6%
Graduated high school	122,577	28.0%
Attended college or Technical School	124,655	31.0%
Graduated from college or Technical School	168,390	27.3%
Race		
Non-Hispanic white	337,166	62.6%
Non-Hispanic black	35,765	11.8%
Hispanic	37,078	17.1%
Other	22,076	7.1%
Multiracial, Non-Hispanic	9,176	1.4%
Marital Status		
Married	232,891	50.4%
Not married	213,979	49.6%
CVD risk factors		
Hypertension	181,098	32.5%
Diabetes	60,444	10.9%
Obesity	128,641	27.3%
CVD events		
Stroke	18,956	3.2%
MI	26,004	4.2%
Coronary heart disease	25,389	4.0%
Any CVD	51,562	8.5%
Twitter-derived sentiment using all race-related tweets		
	Mean	SD
% positive sentiment	6.4%	0.7%
% negative sentiment	39.9%	3.4%
Twitter-derived sentiment using race-related tweets, geotagged		
% positive sentiment	13.9%	2.6%
% negative sentiment	13.9%	5.1%

^a n displays the unweighted counts (numerator) of the listed characteristic. Any CVD captures individuals with any of the examined CVD events (stroke, MI, coronary heart disease).

Table 2.

Twitter-derived sentiment with CVD outcomes, weighted^a

	Hypertension	Diabetes	Obesity	Stroke	MI	CHD	Any CVD
<i>Twitter Sentiment^c</i>	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b
Positive tweets							
3rd tertile (highest)	0.97 (0.94, 1.00)	0.94 (0.90, 0.99)	0.97 (0.94, 1.00)	0.89 (0.80, 0.98)	0.91 (0.83, 0.98)	0.94 (0.86, 1.02)	0.90 (0.86, 0.95)
2nd tertile	0.97 (0.94, 1.00)	0.95 (0.90, 0.99)	0.97 (0.95, 1.00)	0.88 (0.81, 0.96)	0.93 (0.86, 1.00)	0.97 (0.90, 1.05)	0.90 (0.86, 0.95)
Negative tweets							
3rd tertile (highest)	1.11 (1.08, 1.14)	1.15 (1.08, 1.22)	1.14 (1.10, 1.18)	1.30 (1.16, 1.46)	1.14 (1.03, 1.25)	1.09 (1.00, 1.19)	1.16 (1.09, 1.24)
2nd tertile	1.06 (1.04, 1.08)	1.03 (0.94, 1.07)	1.05 (1.02, 1.08)	1.13 (1.04, 1.22)	1.08 (1.00, 1.15)	1.06 (0.99, 1.13)	1.07 (1.03, 1.13)
N	433,434	433,066	433,680	433,605	433,597	433,540	433,680

^aData source for health outcome: BRFSS 2017

^bAdjusted Poisson regression models were run for each outcome separately. Models controlled for individual level age, race, sex, marriage as well as state level percent of non-Hispanic white, percent of non-Hispanic black, percent of Hispanic, and median household income. Twitter-derived sentiment were categorized into tertiles, with the lowest tertile serving as the referent group. Analyses accounted for survey weights and complex survey design to produce nationally representative estimates.

^cTwitter sentiment including % positive tweets and % negative tweets were constructed at state level with all the race-related tweets. We categorized % positive tweets and % negative tweets into tertiles with the lowest levels as the references.

Table 3.

Twitter-derived sentiment with CVD outcomes stratified by sex, weighted^a

CVD outcomes	Twitter Sentiment ^f	Prevalence Ratio (95% CI) ^b	
		Male	Female
Hypertension	Positive tweets		
	3rd tertile (highest)	0.96 (0.92, 0.99)	0.98 (0.94, 1.01)
	2nd tertile	0.97 (0.94, 1.00)	0.98 (0.95, 1.01)
	Negative tweets		
	3rd tertile (highest)	1.10 (1.06, 1.15)	1.13 (1.08, 1.17)
	2nd tertile	1.05 (1.02, 1.08)	1.07 (1.04, 1.11)
N		188,772	238,587
Diabetes	Positive tweets		
	3rd tertile (highest)	0.91 (0.84, 0.98)	0.98 (0.91, 1.05)
	2nd tertile	0.95 (0.89, 1.01)	0.94 (0.89, 1.01)
	Negative tweets		
	3rd tertile (highest)	1.17 (1.07, 1.27)	1.13 (1.04, 1.23)
	2nd tertile	1.03 (0.97, 1.09)	1.03 (0.97, 1.10)
N		188,586	238,405
Obesity	Positive tweets		
	3rd tertile (highest)	0.97 (0.93, 1.01)	0.97 (0.93, 1.01)
	2nd tertile	1.01 (0.97, 1.05)	0.94 (0.91, 0.98)
	Negative tweets		
	3rd tertile (highest)	1.16 (1.11, 1.22)	1.12 (1.06, 1.17)
	2nd tertile	1.07 (1.03, 1.11)	1.03 (0.99, 1.07)
N		188,898	238,707
Stroke	Positive tweets		
	3rd tertile (highest)	0.92 (0.79, 1.07)	0.86 (0.77, 0.97)
	2nd tertile	0.89 (0.79, 1.02)	0.86 (0.75, 0.99)
	Negative tweets		

CVD outcomes	Twitter Sentiment ^c	Prevalence Ratio (95% CI) ^b	
		Male	Female
N	3rd tertile (highest)	1.27 (1.08, 1.50)	1.33 (1.14, 1.55)
	2nd tertile	1.12 (1.00, 1.26)	1.13 (1.01, 1.27)
		188,853	238,677
MI	Positive tweets		
	3rd tertile (highest)	0.89 (0.80, 0.99)	0.92 (0.81, 1.06)
	2nd tertile	0.93 (0.84, 1.02)	0.93 (0.82, 1.05)
	Negative tweets		
	3rd tertile (highest)	1.14 (1.01, 1.28)	1.14 (0.98, 1.33)
	2nd tertile	1.06 (0.97, 1.16)	1.10 (0.99, 1.23)
		188,850	238,672
CHD	Positive tweets		
	3rd tertile (highest)	0.95 (0.85, 1.05)	0.94 (0.82, 1.07)
	2nd tertile	0.93 (0.84, 1.04)	1.02 (0.91, 1.15)
	Negative tweets		
	3rd tertile (highest)	1.04 (0.93, 1.17)	1.16 (1.00, 1.34)
	2nd tertile	1.03 (0.94, 1.12)	1.11 (1.00, 1.24)
		188,819	238,646
Any CVD	Positive tweets		
	3rd tertile (highest)	0.90 (0.83, 0.97)	0.90 (0.82, 0.98)
	2nd tertile	0.89 (0.84, 0.96)	0.91 (0.85, 0.99)
	Negative tweets		
	3rd tertile (highest)	1.13 (1.04, 1.23)	1.20 (1.09, 1.32)
	2nd tertile	1.06 (1.00, 1.12)	1.10 (1.02, 1.18)
		188,898	238,707

^aData source for health outcome: BRFSS 2017

^bAdjusted Poisson regression models were run for each outcome separately. Models controlled for individual level age, race, marriage as well as state level percent of non-Hispanic white, percent of non-Hispanic black, percent of Hispanic, and median household income. Twitter-derived sentiment were categorized into tertiles, with the lowest tertile serving as the referent group. Analyses accounted for survey weights and complex survey design to produce nationally representative estimates.

Twitter sentiment including % positive tweets and % negative tweets were constructed at state level with all the race-related tweets. We categorized % positive tweets and % negative tweets into tertiles with the lowest levels as the references.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Twitter-derived sentiment with CVD outcomes stratified by race/ethnicity, weighted^d

CVD outcomes	Twitter Sentiment ^e	White, Non-Hispanic				Hispanic	Other race, Non-Hispanic		Multiracial, Non-Hispanic
		Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b		Prevalence Ratio (95% CI) ^b	Prevalence Ratio (95% CI) ^b	
	Positive tweets								
	3rd tertile (highest)	0.98 (0.96, 1.01)	0.89 (0.80, 0.98)	0.97 (0.83, 1.13)	1.02 (0.85, 1.22)	1.05 (0.87, 1.27)			
	2nd tertile	0.98 (0.95, 1.00)	0.93 (0.87, 0.99)	0.99 (0.88, 1.11)	1.02 (0.87, 1.18)	1.13 (0.95, 1.34)			
	Negative tweets								
	3rd tertile (highest)	1.14 (1.11, 1.18)	1.12 (1.04, 1.21)	0.97 (0.87, 1.08)	1.10 (0.89, 1.35)	1.05 (0.88, 1.25)			
	2nd tertile	1.09 (1.06, 1.11)	1.01 (0.95, 1.08)	0.96 (0.89, 1.03)	1.07 (0.95, 1.22)	1.08 (0.94, 1.25)			
N		331,344	34,811	31,901	20,481	8,822			
	Positive tweets								
	3rd tertile (highest)	1.00 (0.94, 1.05)	0.79 (0.63, 0.98)	0.84 (0.67, 1.06)	0.97 (0.70, 1.34)	1.13 (0.80, 1.58)			
	2nd tertile	0.98 (0.93, 1.03)	0.89 (0.78, 1.03)	0.93 (0.78, 1.11)	1.09 (0.85, 1.41)	1.05 (0.74, 1.49)			
	Negative tweets								
	3rd tertile (highest)	1.15 (1.08, 1.23)	1.16 (0.97, 1.38)	1.28 (1.08, 1.51)	0.91 (0.62, 1.34)	0.62 (0.43, 0.90)			
	2nd tertile	1.06 (1.02, 1.11)	1.04 (0.90, 1.20)	1.04 (0.92, 1.16)	1.02 (0.83, 1.26)	0.74 (0.56, 0.97)			
N		331,169	34,764	31,831	20,421	8,806			
	Positive tweets								
	3rd tertile (highest)	1.00 (0.97, 1.03)	0.86 (0.76, 0.97)	0.87 (0.75, 1.00)	1.10 (0.88, 1.24)	1.14 (0.91, 1.42)			
	2nd tertile	1.00 (0.96, 1.03)	0.93 (0.85, 1.01)	0.97 (0.86, 1.08)	1.02 (0.84, 1.24)	0.90 (0.73, 1.11)			
	Negative tweets								
	3rd tertile (highest)	1.11 (1.07, 1.15)	1.10 (1.00, 1.21)	1.24 (1.11, 1.37)	1.24 (0.95, 1.62)	1.04 (0.82, 1.33)			
	2nd tertile	1.04 (1.01, 1.07)	1.06 (0.97, 1.15)	1.08 (1.01, 1.17)	1.16 (0.98, 1.37)	0.96 (0.82, 1.13)			
N		331,539	34,834	31,914	20,493	8,825			
	Positive tweets								
	3rd tertile (highest)	0.91 (0.82, 1.02)	0.99 (0.66, 1.48)	0.55 (0.31, 0.99)	1.06 (0.55, 2.04)	1.35 (0.76, 2.39)			
	2nd tertile	0.87 (0.79, 0.96)	0.82 (0.62, 1.09)	0.60 (0.40, 0.91)	1.02 (0.54, 1.93)	1.07 (0.65, 1.78)			

CVD outcomes	Twitter Sentiment ^c	White, Non-Hispanic		Black, Non-Hispanic		Hispanic		Other race, Non-Hispanic		Multiracial, Non-Hispanic	
		Prevalence Ratio (95% CI) ^b	N	Prevalence Ratio (95% CI) ^b	N	Prevalence Ratio (95% CI) ^b	N	Prevalence Ratio (95% CI) ^b	N	Prevalence Ratio (95% CI) ^b	N
	Negative tweets										
	3rd tertile (highest)	1.29 (1.14, 1.46)		1.20 (0.87, 1.65)		1.65 (1.11, 2.45)		0.90 (0.53, 1.53)		1.01 (0.60, 1.71)	
	2nd tertile	1.11 (1.02, 1.20)		1.08 (0.82, 1.42)		1.05 (0.77, 1.42)		1.03 (0.69, 1.54)		1.21 (0.78, 1.87)	
N		331,486	34,830	31,908	20,483	8,823					
	Positive tweets										
	3rd tertile (highest)	0.90 (0.83, 0.98)		0.76 (0.49, 1.18)		1.29 (0.78, 2.15)		1.12 (0.72, 1.74)		1.13 (0.66, 1.96)	
	2nd tertile	0.89 (0.82, 0.96)		0.97 (0.74, 1.28)		1.16 (0.78, 1.72)		0.95 (0.61, 1.47)		1.37 (0.87, 2.18)	
MI	Negative tweets										
	3rd tertile (highest)	1.21 (1.10, 1.34)		1.38 (0.93, 2.02)		0.73 (0.52, 1.02) ^f		0.89 (0.55, 1.45)		0.96 (0.54, 1.72)	
	2nd tertile	1.08 (1.01, 1.16)		1.28 (0.93, 1.76)		0.96 (0.78, 1.20)		0.85 (0.60, 1.21)		0.87 (0.57, 1.72)	
N		331,476	34,828	31,906	20,488	8,824					
	Positive tweets										
	3rd tertile (highest)	0.92 (0.84, 1.00)		0.94 (0.60, 1.49)		1.09 (0.62, 1.90)		1.57 (0.93, 2.64)		1.05 (0.59, 1.87)	
	2nd tertile	0.90 (0.83, 0.98)		1.26 (0.96, 1.65)		1.13 (0.7, 1.73)		1.40 (0.86, 2.28)		0.85 (0.50, 1.45)	
CHD	Negative tweets										
	3rd tertile (highest)	1.15 (1.05, 1.27)		0.95 (0.65, 1.40)		0.84 (0.58, 1.22)		0.74 (0.42, 1.29)		1.27 (0.78, 2.06)	
	2nd tertile	1.05 (0.98, 1.12)		1.02 (0.74, 1.42)		1.06 (0.82, 1.36)		0.95 (0.65, 1.39)		1.01 (0.67, 1.54)	
N		331,430	34,822	31,905	20,486	8,822					
	Positive tweets										
	3rd tertile (highest)	0.89 (0.84, 0.94)		0.87 (0.67, 1.15)		0.92 (0.65, 1.31)		1.07 (0.73, 1.57)		1.12 (0.76, 1.66)	
	2nd tertile	0.86 (0.82, 0.91)		0.95 (0.79, 1.14)		0.94 (0.72, 1.23)		1.05 (0.73, 1.50)		1.07 (0.77, 1.50)	
Any CVD	Negative tweets										
	3rd tertile (highest)	1.22 (1.14, 1.30)		1.16 (0.93, 1.46)		0.98 (0.78, 1.25)		0.92 (0.63, 1.32)		1.06 (0.73, 1.55)	
	2nd tertile	1.07 (1.03, 1.13)		1.12 (0.93, 1.36)		1.01 (0.86, 1.19)		0.97 (0.75, 1.26)		0.96 (0.73, 1.28)	
N		331,539	34,834	31,914	20,493	8,825					

^dData source for health outcome: BRFSS 2017

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Adjusted Poisson regression models were run for each outcome separately. Models controlled for individual level age, sex, marriage as well as state level percent of non-Hispanic white, percent of non-Hispanic black, percent of Hispanic, and median household income. Twitter-derived sentiment were categorized into tertiles, with the lowest tertile serving as the referent group. Analyses accounted for survey weights and complex survey design to produce nationally representative estimates.

Twitter sentiment including % positive tweets and % negative tweets were constructed at state level with all the race-related tweets. We categorized % positive tweets and % negative tweets into tertiles with the lowest levels as the references.