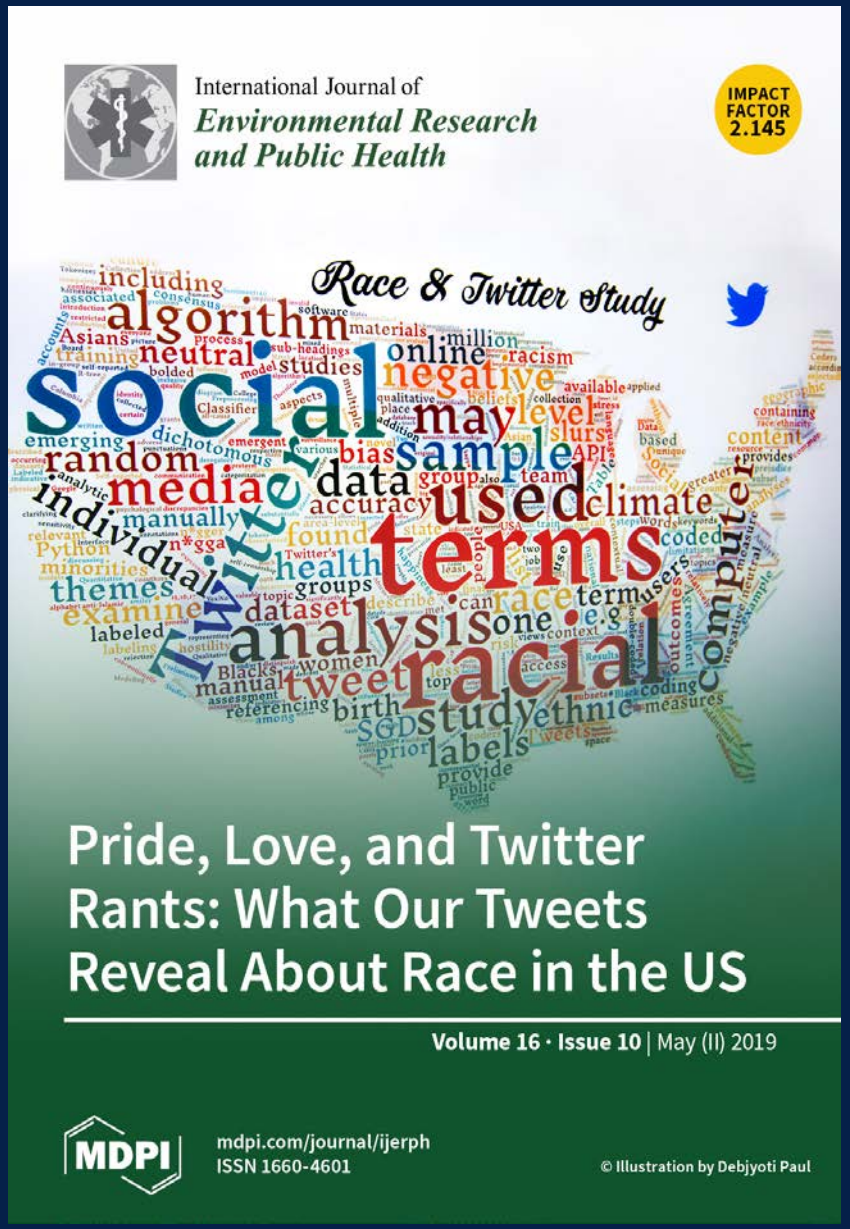
UCSF School of Medicine

State-level racial attitudes and adverse birth outcomes: applying natural language processing to Twitter data to quantify state context for pregnant women

Presented at the 2019 PRISM Health Symposium, San Francisco, CA. 6 Dec 2019

Thu Nguyen, ScD MSPH
Department of Family & Community Medicine
University of California, San Francisco


International Journal of
Environmental Research and Public Health
IMPACT FACTOR 2.145
Race & Twitter Study
Pride, Love, and Twitter Rants: What Our Tweets Reveal About Race in the US
Volume 16 · Issue 10 | May (II) 2019
MDPI
mdpi.com/journal/ijerph
ISSN 1660-4601
© Illustration by Debjyoti Paul

Place-level racial attitudes, bias, and birth outcomes

Significance and Research Gap

- Traditional data collection
  - relies primarily on self-report
  - underestimate of the impact of racism on health

- Racial disparities in birth outcomes persist and widening
  - 2017: Preterm NH White: 9.05%; NH Black: 13.93%. LBW is 2x+ for Black vs White infants for 2006-2016

- Racial bias is an important risk factor for adverse birth outcomes with stress as a potential mediator

NIH K99/R00 to create novel indicator of place-level racial attitudes and bias

- Develop state and county level indicator of racial attitudes and bias from Twitter data
- Examine the associations between place-level racial attitudes and adverse birth outcomes
- Research represents a new way to measure area-level racial attitudes and bias

Working with Tweets

- Twitter data collection
  - June 2015 - December 2017
  - English language, U.S., used one or more of race-related keywords
  - N=26,027,740 tweets from 2,498,717 Twitter users

- Compiling and refining keyword list
  - 518 race-related keywords
  - Group into 5 main racial/ethnic categories: Blacks, Hispanics, Asians, Whites, and Middle Easterners

- Sentiment analyses(train and predict sentiment)
  - Machine Learning , Support Vector Machines (SVM)

- Construct area level predictors

Birth outcomes

- Low birth weight: $\leq$ 2499 g
- Preterm birth: <37 weeks of gestation

- Birth outcomes data come from the 2015-2017 restricted Natality File with geographical identifiers

- Restricted to singleton births with no congenital abnormalities with available data on gestational age (N= 8,369,697) and birth weight (N= 8,367,143).

Sentiment Analysis

- Support Vector Machines (SVM), a supervised machine learning model to label the tweets

- Training data: Sentiment140(n=498), Kaggle(n=7,086), Sanders(n=5,113) and n=6,481 tweets from our research group.

- 5-fold cross validation to assess the model performance
  - Negative & non-negative tweet classification: Accuracy = 91%, F1 score = 84%.
  - Positive & not positive tweet classification: Accuracy = 89%, F1 score =81%

Statistical Analysis

- For each year, state-level derived sentiment towards racial and ethnic minorities were merged with data on births during that year.

- Estimated **prevalence ratios** using log binomial models controlling for individual-level maternal characteristics and state demographic characteristics.
  - **Individual-level** maternal characteristics: age, sex, race, ethnicity, foreign birth, education, marital status, smoking, body mass index, first birth status, and prenatal care.
  - **state-level:** % non-Hispanic Black, % Hispanic, and economic disadvantage (factor score of 5 variables: percent unemployed; percent with some college, percent with high school diploma, percent children in poverty, percent single parent households, and percent median household income)

UCSF

Table. State negative sentiment towards race/ethnic minorities and individual level birth outcomes

	Low Birth Weight	Preterm birth
State level Twitter-derived variables	Prevalence Ratio (95% CI)	Prevalence Ratio (95% CI)
Proportion of race-related tweets that are negative		
Total sample		
3rd tertile vs 1st (lowest)	1.08 (1.02, 1.13)	1.10 (1.00, 1.21)
N	8,367,143	8,369,697
Among Minorities		
3rd tertile vs 1st (lowest)	1.10 (1.06, 1.15)	1.11 (1.06, 1.18)
N	3,905,417	3,906,659
Among Whites		
3rd tertile vs 1st (lowest)	1.08 (1.00, 1.18)	1.10 (0.96, 1.26)
N	4,749,843.00	4,751,254.00

Table. State negative sentiment towards race/ethnic minorities and individual level birth outcomes among racial/ethnic subgroups

	Low Birth Weight	Preterm birth
State level Twitter-derived variables	Prevalence Ratio (95% CI)	Prevalence Ratio (95% CI)
Middle Easterners (among Minorities)		
Proportion of tweets about Middle Easterners including Muslims that are negative		
3rd tertile vs 1st (lowest)	1.04 (1.01, 1.07)	1.06 (1.01, 1.11)
N	3,905,417	3,906,659
Among Blacks		
Proportion of tweets about Blacks that are negative		
3rd tertile vs 1st (lowest)	1.09 (1.05, 1.14)	1.11 (1.05, 1.06)
N	1,278,401	1,278,978
Among Hispanics		
Proportion of tweets about Hispanics that are negative		
3rd tertile vs 1st (lowest)	1.00 (0.93, 1.07)	0.92 (0.88, 0.97)
N	1,649,486	1,649,859
Among Asians		
Proportion of tweets about Asians that are negative		
3rd tertile vs 1st (lowest)	1.11 (1.04, 1.18)	1.15 (1.03, 1.29)
N	444,561	444,733
Among Whites		
Proportion of tweets about Whites that are negative		
3rd tertile vs 1st (lowest)	0.98 (0.93, 1.03)	0.93 (0.89, 0.98)
N	4,749,843	4,751,254

UCSF

Qualitative Research

Mixed Methods

- Content analysis of tweets using race-related terms

International Journal of
Environmental Research and Public Health

MDPI

Article

Pride, Love, and Twitter Rants: Combining Machine Learning and Qualitative Techniques to Understand What Our Tweets Reveal about Race in the US

Thu T. Nguyen [1,*], Shaniece Criss [2], Amani M. Allen [3], M. Maria Glymour [1,4], Lynn Phan [5], Ryan Trevino [6], Shrikha Dasari [1] and Quynh C. Nguyen [7]

- Focus groups in Berkeley, CA and Greenville, SC in Summer 2019 to to understand people's experiences with engaging in discussions related to race and ethnicity on Twitter and their response to these online discussions

UCSF

Content Analysis: negative sentiment themes

Themes	Example Tweets
Negative Sentiment	
Innocuous	▪ Can't Watch The (professional basketball team) Play. These N*ggas Boring AF
Complaints	▪ N*ggas don't bring sh*t but headaches
Insults using derogatory language	▪ You bend over backwards for that n*gga you a f*ggot ▪ Some girls really need to dye their fkn hair and maintain that weave looking good #sh*tsghetto ▪ Stupid a*s ho*s and n*ggas bruh
Generalizations, use of racial slurs in derogatory ways	▪ I work like a freaking Mexican ▪ I seriously hate when I hear about Desi or Arab mothers preferring lighter skinned brides for their sons. Stop this stupid mentality. ▪ Middle Eastern/Arabic accents piss me off more than most things. ▪ You can't use big words around hood niggas ▪ That gook got so mad he was steamed rice
Hostile tweets, some mentioning violence	▪ N*gga gon break his Kneecaps [url] ▪ And if they are carrying a Mexican flag in Az. they need to be arrested.

UCSF

Content Analysis: positive sentiment themes

Positive Sentiment	
Cultural pride	▪ asian at heart, forever ▪ mexican anything is the best way to my heart tbh ▪ #AskFluffy Why do you love being Mexican?
Food	▪ God bless Mexicans and their delicious food <33,343 [url] ▪ Mexican food and family can definitely turn a day around ▪ We are getting Chinese food for lunch. I repeat. We are getting Chinese food for lunch.
Loyalty within friendships	▪ Stil down with my day one n*ggas! ▪ Don't worry bout my n*ggas cuz I gottem ▪ I'll never leave my n*ggas hanging
Denying stereotypes	▪ I don't think I'm ghetto at all..... ▪ N*ggas that are locked up are prolly some of the smartest most innovative people on earth ▪ Everybody wanna be Haitian now wasn't like that bout 15 years ago Lol

Feedback from the group

- How to more rigorously investigate how these place-level measures are associated with birth outcomes? Potential for confounding by other state-level characteristics (policies, sociodemographics). We controlled for some but others potential confounders remain.

- Other suggested analyses? In addition to sentiment analysis, how could we also capture racial attitudes and bias using Twitter data?

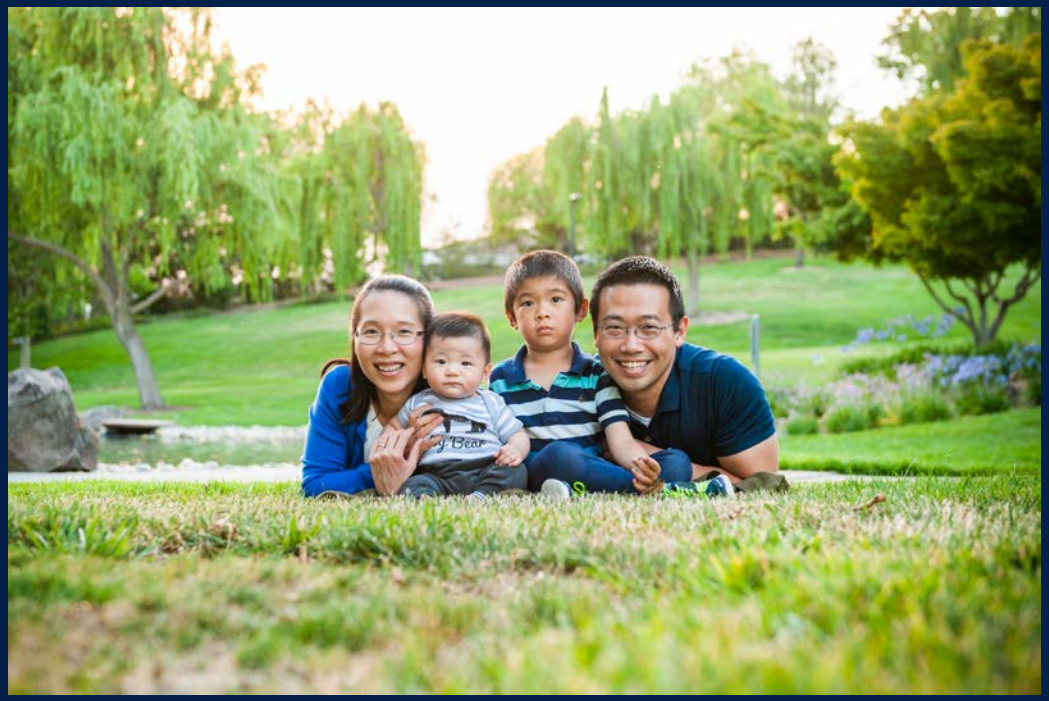
Thank You

- Amani Allen
- M. Maria Glymour

Leal Hsiao

Milo & Sebastian

- HEARTs research group
- HRS research group
- K scholars
- Funding: R00MD012615



- Contact:
- thu.nguyen@ucsf.edu

UCSF