ABSTRACT
While most mortality rates have decreased in the US, maternal mortality has increased and is among the highest of any OECD nation. Extensive public health research is ongoing to better understand the characteristics of communities with relatively high or low rates. In this work, we explore the role that social media language can play in providing insights into such community characteristics. Analyzing pregnancy-related tweets generated in US counties, we reveal a diverse set of latent topics including Morning Sickness, Celebrity Pregnancies, and Abortion Rights. We find that rates of mentioning these topics on Twitter predicts maternal mortality rates with higher accuracy than standard socioeconomic and risk variables such as income, race, and access to health-care, holding even after reducing the analysis to six topics chosen for their interpretability and connections to known risk factors. We then investigate psychological dimensions of community language, finding the use of less trusting, more stressed, and more negative affective language is significantly associated with higher mortality rates, while trust and negative affect also explain a significant portion of racial disparities in maternal mortality. We discuss the potential for these insights to inform actionable health interventions at the community-level.

KEYWORDS
maternal mortality, health disparities, language, topic modeling, community characteristics

1 INTRODUCTION
The United States has one of the highest maternal mortality rates of any country in the Organization for Economic Cooperation and Development group [11, 23]. Approximately 700 individuals die from pregnancy-related causes [3, 15, 68] and an estimated 60% of these deaths are suspected to be preventable [68]. While the international trend has seen a reduction in maternal mortality, despite increased budgets, rates in the US have more than doubled in the past 25 years [3].¹ Black and Latina mothers bear a disproportionate brunt of this burden: Black women are three to four times more likely to die during childbirth, even after controlling for numerous socioeconomic and risk factors [68]. These rates vary by geography: e.g., in New York City, Black women are 12 times more likely to die during childbirth than white women [64, 73].²

Public health research has examined potential causes for maternal mortality and disparities, pointing to issues such as access to insurance, bias in health-care, segregated hospitals, and inadequate post-delivery care [8, 27, 49, 50, 54]. While it is understood that each of community, health facility and system, patient, and provider all play a part, there is an overall pervasive concern that the specific causes and mechanisms for maternal mortality and disparities are not adequately understood [68]. The WHO cites a “general lack of good data – and related analysis – on maternal health outcomes” as a bottleneck for gaining insights into this issue [3].

In this work, we seek to partially address this gap, focusing on community-level factors that characterize maternal mortality as revealed through social media language. We examine whether community variables derived from social media language data can predict community maternal mortality rates and its racial disparity. While emotions and language analyzed using social media data are shown to have high-efficacy in tasks ranging from predicting allergies or life satisfaction to depression or heart disease mortality [26, 30, 33, 66, 76], the potential of social media has yet to be examined in this manner to help shed understanding on maternal mortality at the community level.

Our contributions in this work are in three-folds:

1. We show that there is a diverse set of pregnancy-related topics ranging from Morning Sickness, to Abortion Rights, to Maternal Studies. We demonstrate that these topics predict maternal mortality rates with higher accuracy than standard socioeconomic (SES), risk factors, and race.

2. This issue has garnered increases attention in part due to concentrated efforts by policy-makers, advocacy groups, and celebrities, in addition to long-standing work by community organizations [35, 36, 40, 59, 81], e.g., see collaborations between the Atlanta-based Black Mamas Matter Alliance and the Black Maternal Health Caucus.

Footnotes:
¹Both authors contributed equally to this research.
²Note, on the other hand, US infant mortality is at a historic low [13].

Proceedings of The Web Conference 2020

Quantifying Community Characteristics of Maternal Mortality Using Social Media

Rediet Abebe∗
Harvard University
rabe@fas.harvard.edu

Salvatore Giorgi∗
University of Pennsylvania
sgiorgi@seas.upenn.edu

Anna Tedijanto
Cornell University
ajt232@cornell.edu

Anneke Buffone
University of Pennsylvania
buffone.anneke@gmail.com

H. Andrew Schwartz
Stony Brook University
has@cs.stonybrook.edu

Copyright and Reuse:
ACM Reference Format:

*Both authors contributed equally to this research.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW ’20, April 20–24, 2020, Taipei, Taiwan
© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.
https://doi.org/10.1145/3366423.3380066
• We show that a select set of six topics, chosen for their interpretability and relations to known maternal health factors, hold as much predictive power as all pregnancy-related topics. Specifically, four of these topics – Maternal Studies, Teen Pregnancy, Abortion Rights, and Congratulatory Remarks – have negative associations with mortality rates.

• We examine variables associated with racial disparities in maternal mortality (i.e., the difference between rates for Black women and other races), finding that language-based scores for trust and affect hold explanatory power for the county-level relationship between race and maternal mortality, even after controlling for standard SES and risk-factors.

2 BACKGROUND AND RELATED WORK

Maternal Mortality Background. Public health research has sought better measurements of maternal mortality rates and their causes and consequences [3, 23, 68]. There is a long line of work exploring what community, patient, hospital, provider, or systemic-level factors may contribute to high rates of mortality and disparities in the US [34, 48, 56, 57]. At the patient-level, cardiovascular conditions, which are related to stress, cause about one third of all pregnancy-related deaths [68]. At the community and systemic-level, studies have shown that delivery site, segregation, and discrimination in maternity care during visits all play a role [8, 27, 49, 50]. At the systemic-level, sociological and economic research have shown racial disparities in mortality and life-expectancy [18, 55]. In line with such studies, there are numerous calls to use a data-driven approach to better grasp the role and causes of maternal mortality related to each of the above main categories [68].

Social Media Data for Health. Twitter data and more generally social media data has been a popular source for exploring community-level health measurements [67]. Examples include excessive alcohol consumption [26], depression [29, 62], heart disease, [33], and more generally population health and well-being [25, 38, 76]. In addition to measuring community-level insights, these data sources have been used to study health information seeking and sharing [31] and individual-level predictions [30]. In recent years, there has also been interest in understanding the societal and ethical implications and limitations around the use of social media data for health studies and roles for computing as a diagnostic of social problems [1, 4, 16, 17, 20].

Maternal Health. An emerging topic of interest has been the use of language-driven analysis to understand pregnancy and maternal experiences. For instance, De Choudhury et al. [28] studied Twitter posts to understand changes in emotions for mothers; Antoniak et al. [5] looked at narrative paths in individuals sharing childbirth stories on an online forum. Focusing on support, Costa Figueiredo et al. [22], Gui et al. [42], Vydiswaran et al. [80] looked at how online peer support and information exchange for pregnant individuals, their caregivers, and individuals experiencing fertility issues. Abebe et al. [2] looked at information seeking for pregnancy and breastfeeding related to HIV. To our knowledge, ours is the first work to employ a language-driven study to understand maternal mortality in the US.

3 DATA

We used three sets of data sets for this study, described below:

3.1 Twitter Data and Seed-Words

To generate our pregnancy data set, we started with a random 10% sample of the entire Twitter stream collected between 2009 and 2015 [70]. We then used this data set to build two subsets: (1) pregnancy-related tweets and (2) tweets geo-located to US counties.

Pregnancy-Related Tweets. The first data set consisted of tweets related to pregnancy and birth. Tweets were pulled from the main data set if they contained the following seed-words: pregnancy, pregnant, infant, fetus, miscarriage, prenatal, trimester, complications, pregnancy, birth, childbirth, pregnancies, baby, children, pregnancy, mother, newborn, child, as well as their plural form, hashtags such as #pregnancy, and capitalizations such as Pregnancy. These seed-words were selected by examining nearest neighbors from word2vec for words related to ‘pregnancy’ and ‘pregnant.’

We then manually examined a random sample of 1,000 tweets from the data set to test for relevance to pregnancy. Tweets that were deemed off-topic, such as those containing phrases like ‘miscarriage of justice’ were used to generate phrases for further data cleaning. We also randomly sampled tweets for specific seed-words and if a substantial (i.e., more than 20%) of the tweets were unrelated to pregnancy, all tweets were removed from the data set, reducing the seed-set. After these cleaning steps, we kept 74.40% of the data set, and validated in fresh sample of 1,000 tweets that over 95% of them are related to pregnancy.

U.S. County Tweets. The second data set consisted of tweets geo-located to U.S. counties. For this we used the County Tweet Lexical Bank [39]. This data set was geo-located using self-reported location information (from the user description field) and latitude / longitude coordinates [76]. The data were then filtered to contain only English tweets [58]. We then limited our data set to Twitter users with at least 30 posts and U.S. counties with at least 100 such users. The final Twitter data set consisted of 2,041 U.S. counties.

3.2 Mortality Rates

The World Health Organization (WHO) defines maternal mortality as “the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes” with the Centers for Disease Control and Prevention (CDC) expanding this time period to 1 year [24, 82]. Data for maternal mortality was collected from the CDC WONDER online database [14]. We collected rates from 2009-2017, so as to match the time-span of our Twitter sample in addition to more recent years (2016 and 2017) since these rates are on the rise [68]. Mortality rates are listed under the following International Classification of Diseases, Tenth Revision (ICD-10) categories: O00-O07 (pregnancy with abortive outcome) and O10-O99 (other complications of pregnancy, childbirth and the puerperium). The CDC suppresses data if a county experiences less than 10 deaths in a given time period for privacy reasons. Of the 2,041 counties in our Twitter set only 197 also had mortality rates (i.e., counties experiencing 10 or more deaths).
Since the CDC does not report age-adjusted rates for counties with low mortality numbers, we took the crude rate as reported and created our own age-adjusted rate. To do this, we built a model using median age of females (American Community Survey, 2014; 5-year estimates) and predicted maternal mortality, taking the residuals as our new “age-adjusted maternal mortality rate.” This age-adjusted value is used throughout the paper.

3.3 Socioeconomic Measures and Risk Factors
In addition to mortality, we collected additional county-level variables on socioeconomics, risk factors, and race. Socioeconomics included unemployment rate, median income, and education (percentage of people with Bachelor’s degrees and High School graduate percentage). For risk factors, we included insurance rates and access to health-care (the ratio of population to number primary care providers). Finally, we also explored the relationship between language and maternal mortality with respect to percentage of Black individuals in each county. As discussed previously, the disparity in mortality rates for Black women is large and providing evidence toward the factors at play for such a disparity is a key application for our analyses. Additionally, to account for overall rates of birth, all analysis included a birth rate covariate (the rate per 1,000 women, aged 15-50, with births in the past 12 months).

The birth rate, race, SES variables, and insurance rates were collected from the 2014 American Community Survey (5 year estimates), whereas the primary care providers was collected from the 2017 County Health Rankings (as reported by the Area Health Resource File/American Medical Association, 2014). We were able to obtain these values for each of the counties which met the Twitter and mortality inclusion criteria above.

Overall, we obtained data for 197 U.S. counties and county equivalents that met each of the data requirements above and conducted our study on these counties. The full list of these counties is included in the project page.5

4 TOPICS AND THEORETICAL LINGUISTIC FEATURES
We used three sets of features that will characterize maternal mortality through language. First, we created a set of automatically-derived topics built over the pregnancy-related tweets. These topics reveal a diversity of themes in discussions around pregnancy on the platform. Next, we used a small set of theoretically-driven language features – (affect, depression, stress, and trust) – in order to access psychological traits of a community and their relations to maternal mortality. Finally, we use a large, general set of topics (non-pregnancy related) to identify broader language patterns.

4.1 Pregnancy-Related Topics
We start with our data set of over 5 million pregnancy-related tweets described in Section 3. We automatically extracted topics using Latent Dirichlet Allocation (LDA) [12]. LDA is a generative statistical model which assumes that each document (in our case tweet) contains a distribution of topics, which in turn, are a distribution of words. We use the Mallet software package [60], which estimates the latent variable of the topics using Gibbs sampling [37]. All default Mallet settings were used, except $\alpha$, which is a prior on the expected topics per document. We set $\alpha = 2$ since tweets are shorter than the typical length of documents. The number of topics is a free parameter and we chose 50 topics.3

<table>
<thead>
<tr>
<th>Topic Label</th>
<th>Top Weighted Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teen</td>
<td>teen, rate, rates, teenage, highest, mortality, low, states, teens, higher, number, 20, country, american, united, education, lowest, population</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>(1.34%)</td>
</tr>
<tr>
<td>Morning</td>
<td>morning, sickness, purpose, symptoms, lives, wanted, williamson, tv, experience, cure, bra, marianne, thinking, signs, oral, teenagers, simon</td>
</tr>
<tr>
<td>Sickness</td>
<td>(0.54%)</td>
</tr>
<tr>
<td>Celebrity</td>
<td>kim, kardashian, kayne, amber, rose, beyonce, years, west, harry, finish, swear, north, who’s, kayne’s, taylor, sets, louis, wiz</td>
</tr>
<tr>
<td>Pregnancies</td>
<td>(1.42%)</td>
</tr>
<tr>
<td>Rights</td>
<td>women, abortion, care, health, abortions, bill, mortality, #prolife, rights, law, gift, support, circumstances, crisis, irrelevant, #prochoice, forced</td>
</tr>
<tr>
<td>Maternal</td>
<td>risk, defects, study, health, weight, linked, flu, cancer, early, diet, drinking, smoking, blood, safe, alcohol, diabetes, autism, acid, disease, drug</td>
</tr>
<tr>
<td>Studies</td>
<td>(2.56%)</td>
</tr>
<tr>
<td>Congratulatory Remarks</td>
<td>congrats, congratulations, :, boy, happy, love, daughter, son, &lt;3, wait, sister, late, healthy, cousin, xx, amazing, #meet, proud</td>
</tr>
<tr>
<td>Remarks</td>
<td>(3.06%)</td>
</tr>
</tbody>
</table>

Table 1: Sample pregnancy topics with representative words

We find that our data reveals a rich set of themes related to pregnancy and birth. In Table 5, we show a sample of six topics, which are hand-selected to demonstrate the breadth of topics in the data set.5 The first column provides the topic label, which were hand-generated by the authors, and the frequency with which the topic occurs in the data set.4 The last column corresponds to the top 10 most representative words for the topic.

These above topics show that pregnancy-related discussions on Twitter can range from personal-health disclosure such as in Morning Sickness, to political conversations related to Abortion Rights, and light topics such as Congratulatory Remarks. Topics that were not included in manuscript due to length constraints include Royal Baby, Food Cravings, and Pregnancy Timeline. Each of these topics shows varying levels of popularity across the counties.

4.2 Theoretical Features
We also explore a set of theoretically-driven language features: affect, depression, trust, and stress. We downloaded pre-existing models to derive county-level language features including:
- affect – positive and negative emotional valence trained over Facebook posts [71].
- depression – degree of depressive personality (a facet of the big five personality test) fit over social media users’ language [75].

---

3Before running the rest of our analysis, we ran LDA using 10, 20, 50, 100, and 200 topics. We selected 50 topics based on manual inspection of coherence and interpretability of the topics.

4Note, since there are 50 topics, the average value is 2%. Furthermore, since some themes, such as celebrity pregnancy, occur in more than one topic, the overall frequency of this theme in the data set is higher than the corresponding value in this table.
• **trust** – degree of trustfulness (how much one tends to trust persons or entities that they do not personally know) fit over social media users’ language [83].
• **stress** – amount of stress fit over social media users’ language and Cohen’s Stress scale [19, 43].

### 4.3 General Topics

Finally, we use a larger set of LDA topics built over a more general data set. By doing this in tandem with the pregnancy-related topics, we can zoom in on pregnancy-related themes while also exploring a larger set of language correlates, which might help in characterizing communities suffering from higher or lower rates of mortality. To this end, we downloaded a set of 2,000 topic posteriors that were automatically-derived over the MyPersonality data set [77]. These topics have been used over a large class of problems and have been found to be robust both in terms of interpretability and predictive power [33, 51, 65, 69], so they form a point of comparison for our domain-specific topics.

### 5 METHODS

To understand the relationship between community level language and maternal mortality, we perform three types of statistical analyses: (1) **prediction** – can language be used to predict mortality rates in a cross-sectional cross validation setup? (2) **differential language analysis** – can we gain insights into communities which suffer from higher or lower maternal mortality through language? and (3) **mediating language analysis** – can language be used to understand the mechanisms through which Black communities experience increased rates of maternal mortality? All data processing, feature extraction and statistical analysis are performed using the open source Python package DLATK [78].

#### 5.1 Prediction

We use two types of predictive models, depending on the type of independent variables. All non-language variables (i.e., SES and risk factors) are modeled with an ordinary least squares (OLS) regression, whereas language features use an $\ell_2$ regularized (Ridge) regression [47]. In addition to regularization, we also use a feature selection pipeline in all language based models, since the number of features can be larger than the number of observations ($N=197$ counties). The pipeline first removes all low variance features and then features that were not correlated with our outcome. Finally, we applied Principal Component Analysis (PCA) to further reduce the number of features. All models are evaluated in a 10-fold cross validation setup, with the Ridge regularization parameter $\alpha$ tuned on the training set within each fold. Predictive accuracy is measured in terms of a single Pearson correlation between the actual values and the predicted values, whereas standard errors are calculated across all 10 folds.

#### 5.2 Differential Language Analysis

Differential Language Analysis (DLA) is used to identify language characterizing maternal mortality [52, 77]. Here we individually regress each of our language variables (i.e., pregnancy related topics and theoretical features) using an OLS regression, adding in access to health-care, birth rates, socioeconomics and risk factors as covariates. We adjust for multiple comparisons by applying a Benjamini–Hochberg false discovery rate correction to the significance threshold ($p < .05$) [10]. For LDA topics we visualize topics significant correlations as word clouds. The word clouds display the top 15 most prevalent words within a topic sized according to their posterior likelihood.

### 5.3 Mediating Language Analysis

We explore the relationship between maternal mortality and the percentage of Black individuals within a county, as expressed through the county’s language. Language based mediation analysis has been used in the past to explore the relationship between socioeconomics and excessive drinking [26]. For this analysis, we residualize the crude maternal mortality rate, as reported by the CDC, on median age of female, birth rates, all socioeconomic variables (income, education and unemployment), insurance rates and rates of primary care providers.

For each language variable, both the pregnancy related LDA topics and theoretical language features, we consider the mediating relationship between the topic (mediator), percentage Black (independent variable) and residualized maternal mortality rates (dependent variable). We follow the standard three-step, Baron and Kenny approach [9]. Step 1: we regress our independent ($x$ and dependent variables ($y$; path $c$) in a standard OLS regression. Step 2: we regress the independent variable ($x$) and mediator ($m$; path $a$). Finally, in Step 3 we create a multi-variate model and regress both the mediator ($m$; topic) and independent variable ($x$; percentage Black) with maternal mortality ($y$; path $c’$). The three models are as follows:

$$y = cx + \beta_1 + \epsilon_1,$$

(1)

$$m = ax + \beta_2 + \epsilon_2,$$

(2)

$$y = c’x + m + \beta_3 + \epsilon_3.$$  

(3)

The mediation effect size ($c – c’$) is taken as the reduction in the effect size between the direct relationship (i.e., percentage Black and maternal mortality) and the mediated relationship. To test for significance, we use a Sobel $p$ [79] and correct all $p$ values for false discoveries via a Benjamini–Hochberg procedure.

### 6 RESULTS

We begin by looking at correlations between maternal mortality and various socioeconomics and risk factors. Table 2 shows the set of correlation coefficients. These results state that the percentage of the population that is Black and unemployment rate were positively correlated with maternal mortality rate and insurance access, income, and education were negatively correlated with maternal mortality rate. Additionally, birth rates were not significantly correlated with maternal mortality. Note that in this paper, we only consider 197 counties in the US due to constraints around Twitter and county-mapped data as discussed in Section 3. While the correlation values do not exactly match correlations for all US counties, the general direction of relationship between maternal mortality rates and these SES and risk-factors was the same, with those the strongest associations – such as percent Black – also matching.

We next look at the predictive accuracy of our 50 topics, the 2000 general topics, and the above SES and risk-factors as well as percent Black values. For this, note that we used linear regression
with maternal mortality values as the outcome variable and the aforementioned language variables as the explanatory variables. Figure 1 shows that the 2000 general Facebook topics had the highest predictive power with a Pearson $r = .72 \ [.65, .79]$*** while risk factors (PCP access and insurance rate) were the lowest with a Pearson $r = .21 \ [.05, .38]$**. Overall SES factors, risk factors, and race, had significantly less predictive accuracy (using a paired t-test) than the 50 pregnancy-related topics from the Twitter data ($t = -4.63, p < .001$) and the 2000 general topics ($t = -4.74, p < .001$).

For the Differential Language Analysis (DLA), we selected the 6 topics of interest. We ran a multi-linear regression, treating the maternal mortality rate as an outcome variable and the prevalence of these topics in the counties as the explanatory variable with birth rates, race, risk factors, and socioeconomics as covariates. We found that five of the 6 topics, shown in Figure 2 had significant associations with maternal mortality rates. Maternal studies had the most negative association – i.e., counties where there are relatively more tweets related to this topic had lower rates of mortality. Note that each of the four topics in the figure – Maternal Studies, Teen Pregnancies, Congratulatory Remarks, and Abortion Rights – all show negative associations with maternal mortality rates. Celebrity Births, not shown, is positively associated ($r = .20 \ [.07, .33]$*) with higher mortality.

Table 2: Correlations with risk factors, socioeconomics, and race. All non-birth rate correlations controlled for birth rates. Reported standardized $\beta$ with 95% confidence intervals in square brackets; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, after Benjamini–Hochberg correction.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Birth Rates</th>
<th>Race</th>
<th>Risk Factors</th>
<th>Socioeconomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate per 1,000 women</td>
<td>.10 [-.04, .24]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black (percent)</td>
<td>.49 [.36, .61]***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Care Providers</td>
<td>-.23 [-.38, -.09]**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninsured (percent)</td>
<td>.27 [.12, .41]***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income (log median)</td>
<td>-.42 [-.55, -.29]***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment (percent)</td>
<td>-.14 [-.28, .01]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School or more (percent)</td>
<td>-.38 [-.52, -.23]***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s Degree (percent)</td>
<td>.26 [.12, .39]***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Differential Language Analysis of theoretically relevant features. Reported standardized $\beta$ with 95% confidence intervals in square brackets; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, after Benjamini–Hochberg correction.

Figure 2: Differential Language Analysis using 6 pregnancy related LDA topics, controlled for race, risk factors and socioeconomics. Reported standardized $\beta$ with 95% confidence intervals in square brackets; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, after Benjamini–Hochberg correction.

We also used 4 theoretical features within the DLA framework: affect, depression, stress and trust. Results are presented in Table 3. We see higher rates of maternal mortality associated with higher distrust, higher stress, higher depression, and with less affect.

Figure 1: Prediction accuracy for non-language variables (red), pregnancy related LDA topics (purple) and general set of LDA topics (blue). Reported Pearson $r$ from 10-fold cross validation, errors bars are 95% CI.

Finally, we explore disparities by race at the population level. The county-level health disparity itself can be seen simply from the strong correlation between the two variables: communities that are more Black, have greater maternal mortality. We turn to Twitter-based community characteristics as mediators (i.e. explainers) of this race-mortality relationship. The idea behind mediation analysis, is that if included a 3rd variable (i.e. a Twitter measurement) in the linear analysis reduces the relationship of the first 2 (i.e. race and...
Table 4: Mediating Language Analysis: Analysis seeks to explain the correlation, $c = .36$, between percent Black and residualized maternal mortality through differences in language. $\alpha$: correlation between the theoretical factor and percent Black; $\beta$: correlation between the theoretical factor and residualized maternal mortality. Reported Pearson $r$ with 95% confidence intervals in square brackets; ***$p < 0.001$, **$p < 0.01$, *$p < .05$, after Benjamini–Hochberg correction. The $c - c'$ column uses a Sobel $\hat{p}$ for significance [79].

<table>
<thead>
<tr>
<th></th>
<th>$c - c'$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>-0.11**</td>
<td>-.40 [-.53,-.27]**</td>
<td>-.27 [-.41,-.13]***</td>
</tr>
<tr>
<td>Depression</td>
<td>-0.04</td>
<td>-.26 [-.39,-.12]***</td>
<td>.14 [.01,.28]*</td>
</tr>
<tr>
<td>Stress</td>
<td>-.01</td>
<td>-.06 [-.29,.08]</td>
<td>.15 [-.02,.28]*</td>
</tr>
<tr>
<td>Trust</td>
<td>-0.14**</td>
<td>-.51 [-.63,-.39]**</td>
<td>-.27 [-.42,-.12]***</td>
</tr>
</tbody>
</table>

The results shown in this work demonstrate the efficacy of social care and racial disparities from public health research: for instance, failure by hospitals, and facilities to provide unbiased and nondiscriminatory care has already been shown to result in lower follow-up visits by Black and Latina women, which is believed to drive higher mortality rates. Trust in physicians and medical institutions has been extensively studied [44–46, 61], with multiple studies focusing on racial and ethnic differences in levels of trust [6, 7, 32, 41]. Findings repeatedly show ethnic and racial differences in trust towards health-care systems, in addition to showing that distrust is associated with racial disparities in use of preventive services [63]. The affect result is also related to the Congratulatory Remarks topic, indicating that communities with both more positive language and more positive discussions around pregnancy and birth may also be experiencing lower maternal mortality rates and disparities. These observations, along with existing discussions, provide potential actionable insights for policies at the community level.

The results here are not without limitations: as with other studies heavily relying on social media data, there are inherent issues of selection bias in who is on the platform and which users meet the inclusion thresholds we set for the pregnancy-related and county-mapping data sets. There is also selection bias in tweets that are geo-located as well as language use by the individuals on Twitter compared to other platforms. It is imperative to not take these data sets as being representative of the U.S., the counties we study, or even individuals that maybe included in the data sets.

Furthermore, we do not control for linguistic differences across different parts of the U.S. and some topics, as a result, may show significant spatial and geographic associations. Likewise, we set the seed-words for constructing the pregnancy-related data set using word2vec, which may also suffer bias issues: e.g., certain words which may be commonly used to discuss pregnancy and birth by certain groups of under-represented individuals may not pass this analysis. While we attempt to control for this by having a relatively large number of seed-words and instead relying on data cleaning, this remains a notable limitation.

We were hindered by the availability of outcome data: a lot of the relevant data is available only at the county-level and crucial data like disparities by race were entirely unavailable. While we believe that studies like ours will provide additional data-sources, models, and measurements to further our understanding of maternal mortality and disparities, availability of ground truth data presents a significant bottleneck. The availability of ground truth data about mortality and disparities, including data regarding mortality rates for groups of individuals belonging to marginalized communities, as well as disaggregated data by different demographics such as race, age, education, income, and other demographics would allow for more fine-grained analysis.

8 ETHICS STATEMENT

This study was reviewed by the University of Pennsylvania institutional review board and, due to lack of individual human subjects, found to be exempt. All data used in this study are publicly available. While the county-level language estimates are publicly available and will be posted on the project page, the original tweets, which are also publicly available, are unable to be redistributed by the authors due to Twitter’s Terms of Service. For additional privacy protection, we automatically replace any Twitter user names with <user> in our analysis and presentation in this paper.

---

5All data available: https://github.com/wwbp/maternal_mortality
REFERENCES


APPENDIX
We include further details on results and discussions from the main text below:

A SAMPLE TWEETS
The last column shows a sample of three tweets for the topic. To find these representative tweets, we extract topic loadings over a random set of 500,000 pregnancy-related tweets. We then order the tweets by topic loadings and hand-select three tweets (out of the top ten) that best describe the topic, ignoring noisy or uninformative tweets. For example, a tweet “teen rates!!!” would load extremely high in our first topic, but it doesn’t capture any additional information over the list of the highest-weighted words within the topic. Note that all typos and emoticons in the tweets are included unchanged.

<table>
<thead>
<tr>
<th>Topic Label</th>
<th>Sample Tweets                                                                misión</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teen Pregnancy (1.34%)</td>
<td>Decreasing infant mortality around the world &lt;URL&gt; #BecauseOfYolo #blah</td>
</tr>
<tr>
<td>Morning Sickness (0.54%)</td>
<td>The purpose of our lives is to give birth to the best which is within us Marianne Williamson #spirituality Ecotopic pregnancy diagnosis symptoms and complications &lt;URL&gt; Getting a sickness that isn’t morning sickness while #pregnant #sucks #cough #throat hurts #stuffy nose #blah</td>
</tr>
<tr>
<td>Celebrity Pregnancies (1.42%)</td>
<td>Amber rose is pregnant ? #dammwiz hopefully kim k’s pregnancy doesn’t last 72 days Taylor swift pregnant by harry</td>
</tr>
<tr>
<td>Abortion Rights (2.15%)</td>
<td>Lawmakers ban shackling of pregnant inmates &lt;URL&gt; #SouthAfrica to care for all #HIV positive infants &lt;URL&gt; #worldaidsday #womensrights #children Nebraska governor rejects prenatal care funding for illegal immigrants &lt;URL&gt;</td>
</tr>
<tr>
<td>Maternal Studies (2.56%)</td>
<td>Lower autism risk with folic acid supplements in pregnancy &lt;URL&gt; Postpartum cardiovascular risk linked to glucose intolerance during pregnancy &lt;URL&gt; Increased autism risk linked to hospital-diagnosed maternal infections &lt;URL&gt;</td>
</tr>
<tr>
<td>Congratulatory Remarks (3.06%)</td>
<td>Congrats to &lt;USER&gt;- and &lt;USER&gt; on the birth of their baby #5yearsago i gave birth to my wonderful daughter &lt;3 &lt;3 &lt;3 Awwwwwww my nephew’s wife is pregnant &lt;3 congrats!</td>
</tr>
</tbody>
</table>

Table 5: Sample topics with sample tweets

B THEORETICAL MODELS
We present high-level details for each of the four models used in this paper. Detailed descriptions and evaluations can be found in the corresponding papers. Note that none of the models described below were developed for this paper.

Affect. An affect model was built using a set of 2,895 annotated Facebook posts. Each post was rated by two psychologists on a nine-point ordinal scale, based on the affective circumplex model introduced by Russell [74]. A $\ell_2$ penalized linear (ridge) regression was built using 1–2grams extracted from each message. Using a 10-fold cross-validation setup, the ngram model resulted in a prediction accuracy (Pearson $r$) of 0.65. Full details can be found in Preoțiuc-Pietro et al. [71].

Depression. The MyPersonality data set [53], which consisted of approximately 154,000 consenting users who shared Facebook statuses and completed a 100-item personality questionnaire was used. The personality questionnaire is based on the International Personality Item Pool proxy for the NEO Personality Inventory [21]. This work then takes the average response to the seven depression-facet items (located within the larger Neuroticism scale) to estimate user-level degree of depression. A ridge-penalized regression model was built [47] using a set of 2,000 LDA topics and 1-3grams extracted over 27,749 individuals and tested on 1,000 random individuals who used at least 1,000 words across all of their statuses. This resulted in a final prediction accuracy (Pearson $r$) of 0.39. Full details can be found in Schwartz et al. [75].

Trust. Similar to the depression model, the trust model was built using the MyPersonality Facebook data set [53]. Consenting individuals were asked to share their Facebook statuses and answer a Big-Five personality questionnaire. The average of three of the ten trust-facet items from the agreeableness domain – (1) “I believe that others have good intentions,” (2) “I trust what people say,” and (3) “I suspect hidden motives in others” (reverse-coded) – was used as a measure of trust. A predictive model was built on 26,243 users who answered the above question and also shared Facebook statuses, where each answered at least 1,000 words across all of their statuses. This resulted in a final prediction accuracy (Pearson $r$) of 0.35. Full details can be found in Zamani et al. [83].

Stress. Participants were recruited through Qualtrics (an online survey platform, similar to Amazon Mechanical Turk), where each participant answered a series of demographic questions, the 10-item Cohen’s Stress scale [19] and consented to share their Facebook statuses. The analysis was then limited to those who self-reported age and gender (female/male) and who posted at least 500 words across all Facebook statuses, resulting in a final set of 2,749 participants. A set of 2,000 Facebook topics were used as features in a ridge penalized regression model [47]. This resulted in a prediction accuracy of 0.32 (Pearson $r$), using a 10-fold cross validation setup. Full details can be found in Guntuku et al. [43].
C  DOMAIN TRANSFER: APPLYING FACEBOOK MODELS TO TWITTER DATA

All four of our theoretical models were trained and evaluated on Facebook data in their original papers, whereas we applied the models to Twitter data. Some of the models have been shown to work in other domains (i.e., stress on Facebook vs Twitter; Guntuku 2019). Additionally, previous work has found is that effect sizes tend to vanish without correcting for the domain transfer [72], which we argue makes our prediction task harder. Additionally, Riemann et al. [72] showed that user-level Facebook models applied to county-level Twitter data are stable in terms of direction of effect sizes.

D  SPATIAL DISTRIBUTIONS

Figure 3 shows the relationship between maternal mortality rates (residualized on race, median age of females, socioeconomics and risk factors) and the topic loadings for the Congratulatory Remarks topic. Markers in the scatter plot are colored according to U.S. Census regions (Midwest, Northeast, South and West). We see that lower usage of this topic is associated with high mortality rates. We also see spatial clustering across the regions. For example, the West tends to have lower rates of mortality but large variance in topic usage. The South has the most variation in mortality in addition to the largest outliers in topic usage. Figure 4 includes a similar set of plots for the theoretically-relevant features, showing significant associations between affect and trust and maternal mortality.

Figure 3: Maternal mortality rate (residualized) vs the Congratulatory Remarks topic loading. Dots are colored by which U.S. Census Region the county resides in: Midwest, Northeast, South and West.

Figure 4: Maternal mortality rate (residualized) vs theoretically relevant features. Dots are colored by which U.S. Census Region the county resides in: Midwest, Northeast, South and West.