Table S1 *Variable Sources and Transformation*

Variable			Description of variable	Unit	Years covered	Source
Included variable	**Transformation**	**Categories**				
Atherosclerotic Heart Disease (AHD) mortality	averaged across years		International Classification of Disease (ICD) 10 code I25.1 recorded as underlying cause of death on death certificates, prepared for the county level and age-adjusted through the CDC (using year 2000 population estimates)	per 100,000 population	2009-2010	CDC Wonder, Underlying Cause of Death (CDC, 2010)
Income	log-transformed		Median household income	2010 inflation-adjusted US dollars	2008-2010	American Community Survey (ACS, 2010) 3-Year Estimates (Table DP03)
Educational Attainment Index	Independently standardized and then averaged	High school grad	Attainment of high school graduation or higher	% of population	2008-2010	ACS 3-Year Estimates (Table DP02) (ACS, 2010)
		College grad	Attainment of bachelor's degree or higher			
Diabetes			Adults (age 20+) diagnosed with diabetes	% of population	2008-2010	County-level estimates based on CDC's Behavioral Risk Factor Surveillance System (BRFSS) data (2009-2010), obtained through 2013 County Health Rankings (CHR; 2010) (see note).
Obesity			Body Mass Index >= 30, based on self-reported height and weight			
Smoking			Current adult smokers who have smoked >= 100 cigarettes in their lifetime		2005-2011	
Hypertension	averaged	male	Male adults (age 30+) who self-reported systolic BP of at least 140mm Hg and/or self-reported taking medication	% of population	2009	County-level estimates prepared through the Institute for Health Metrics and Evaluation (IHME; 2009) on the basis of CDC BRFSS data (see note).
		female	Female adults (age 30+) who self-reported systolic BP of at least 140mm Hg and/or self-reported taking medication			
% Black			Population of one race - Black or African American alone	% of population	2010	U.S. Census, Demographic Profile Data (Table DP01) (U.S. Census Bureau, 2010)
% Hispanic			Hispanic or Latino			
% Female			Female			
% Married	averaged	male	Male adults (age 15+) now married (not separated)	% of population	2008-2010	ACS 3-Year Estimates (Table DP02) (ACS, 2010)
		female	Female adults (age 15+) now married (not separated)			

Note on sources used for selected variables:

Diabetes and Obesity: County Health Rankings (CHR; 2010) used data from the National Center for Chronic Disease Prevention and Health Promotion's Division of Diabetes Translation (part of the CDC), which provides the Diabetes Public Health Resource (DPHR; 2010). DPHR used data from the CDC's Behavioral Risk Factor Surveillance System (BRFSS; 2009-2010), an ongoing national survey. DPHR developed county-level estimates from state-level BRFSS data using small area estimation techniques, including Bayesian multilevel modeling, multilevel logistic regression models, and a Markov Chain Monte Carlo simulation method.

Smoking: County-level estimates (based on BRFSS state-level data) were calculated for CHR by CDC staff.

Hypertension: The Institute for Health Metrics and Evaluation (IHME; 2009) used National Health Examination and Nutrition Survey data (1999-2008) to characterize the relationship between self-reported and physical measurements for various health factors. They used the resulting model to predict physical measurements for 2009 BRFSS participants (who supplied self-reported measures) and employed small area estimation techniques to estimate hypertension prevalence at the county-level.

Table S2

Dictionary Evaluation

	Dictionary	Top Ten Dictionary Words by Frequency	Two Rater Agreement	Accuracy
Risk Factors	Anger	shit f*** hate damn b*tch hell f***ing mad stupid b*tches	70.0%	60.0%
	Negative Relationships	hate alone jealous blame evil rude lonely independent hated ban	86.0%	75.5%
	Negative Emotion	sorry mad sad scared p*ssed crying horrible afraid terrible upset	87.0%	79.5%
	Disengagement	tired bored sleepy lazy blah meh exhausted yawn distracted boredom	91.0%	88.0%
	Anxiety	crazy pressure worry scared awkward scary fear doubt horrible afraid	81.5%	55.0%
Protective Factors	Positive Relationships	love home friends friend team social welcome together kind dear	75.0%	55.5%
	Positive Emotion	great happy cool awesome amazing glad excited super enjoy wonderful	93.0%	88.5%
	Engagement	learn interesting awake interested alive learning creative alert involved careful	74.5%	79.0%

Note. Each dictionary was evaluated by two independent raters. 200 random instances of tweets containing words from the dictionary in question were extracted, and the expert raters determined whether the word expressed the associated dictionary concept within the tweet. On average, the raters agreed 81.5% of the time, and a third rater was brought in to break ties. Accuracy refers to the percentage of tweets that expressed the associated dictionary concept, out of the 200 random instances sampled for every dictionary.

Table S3

Cross-Correlations between Dictionaries and Topics

	Anger	Negative Relationships	Negative Emotion	Disengagement	Anxiety	Positive Relationships†	Positive Emotion	Engagement
Anger	1	.76 [.73, .78]	.60 [.57, .64]	.72 [.69, .74]	.29 [.24, .34]	.18 [.26, .36]	-.33 [-.38, -.28]	-.30 [-.35, -.25]
Negative Relationships			.70 [.68, .73]	.67 [.64, .70]	.37 [.32, .41]	.42 [.50, .58]	-.04 [-.09, .01]	-.09 [-.14, -.04]
Negative Emotion				.55 [.51, .59]	.43 [.38, .47]	.45 [.50, .58]	.19 [.14, .24]	.04 [-.02, .09]
Disengagement					.29 [.24, .34]	.28 [.37, .46]	-.16 [-.21, -.11]	-.27 [-.32, -.22]
Anxiety						.38 [.29, .39]	.23 [.18, .28]	.16 [.11, .21]
Positive Relationships							.48 [.43, .52]	.23 [.18, .28]
Positive Emotion								.61 [.58, .64]

Topics	Included Word	Anger	Negative Relationships	Negative Emotion	Disengagement	Anxiety	Positive Relationships†	Positive Emotion	Engagement
Hostility, Aggression	bullsh*t	.94	.58	.43	.62	.19	-.03	-.45	-.40
	a**hole	.93	.62	.48	.61	.19	.00	-.41	-.39
	retarded	.81	.65	.56	.54	.21	.06	-.26	-.30
Hate, Inter-personal Tensions	hating	.88	.74	.54	.68	.23	.13	-.33	-.36
	drama	.87	.67	.53	.66	.26	.18	-.28	-.29
	passion	.67	.84	.66	.60	.33	.37	.02	-.08
Boredom, Fatigue	bored	.70	.60	.47	.87	.20	.16	-.26	-.35
	tired	.69	.70	.62	.87	.31	.32	-.04	-.21
	bed	.50	.61	.56	.69	.30	.41	.08	-.12
Skilled Occupations	management	-.42	-.32	-.23	-.41	.03	.29	.38	.69
	service	-.41	-.28	-.17	-.39	.08	.33	.51	.63
	conference	-.45	-.28	-.16	-.42	.11	.34	.56	.65
Positive Experiences	experience	-.30	-.12	-.01	-.26	.15	.42	.57	.76
	company	-.30	-.12	.11	-.21	.18	.54	.78	.67
	weekend	-.35	-.11	.09	-.22	.14	.55	.89	.62
Optimism, Resilience	opportunities	-.33	-.20	-.12	-.31	.10	.35	.41	.69
	achieve	-.21	-.07	.00	-.22	.17	.36	.39	.68
	strength	-.14	.06	.04	-.08	.29	.55	.48	.68

Note. Dictionary cross-correlations (Pearson *r*) are given, with 95% confidence intervals in brackets. To ease inspection, topic-dictionary correlations are color formatted, ranging from dark red (strongly negative) to dark green (strongly positive). Particularly strong correlations between topic clusters and dictionaries are emphasized with bolder boxes. Topics correspond to the topics shown in Figure 1, in the same order. The "included words" are dominant unique words in each cloud, which help identify the topic.
† The word "love" was removed from the dictionary, as it accounted for more than a third of all word occurrences in the dictionary, and distorted the results (see discussion).

Table S4

Performance of Regression Models Predicting AHD Mortality on the Basis of Different Sets of Predictors

Model	Demographic	SES	Health	Twitter	Accuracy of County-Level AHD Prediction
1	X				.14 [.09, .19] ⎤ ***
2	X			X	.42 [.38, .45] ⎦
3		X			.23 [.18, .28] ⎤ ***
4		X		X	.41 [.38, .45] ⎦
5			X		.27 [.20, .34] ⎤ ***
6			X	X	.42 [.38, .46] ⎦
7	X	X			.32 [.27, .37] ⎤ ***
8	X	X		X	.41 [.38, .45] ⎦
9	X		X		.33 [.26, .40] ⎤ ***
10	X		X	X	.42 [.38, .46] ⎦
11		X	X		.29 [.23, .35] ⎤ ***
12		X	X	X	.42 [.38, .46] ⎦
13	X	X	X		.36 [.29, .43] ⎤ * ⎤ *
14	X	X	X	X	.42 [.38, .46] ⎦
15				X	.42 [.38, .45] ⎦

Note. Performance of regression models predicting atherosclerotic heart disease (AHD) mortality from demographic variables (percentage of Blacks, Hispanics, married, and female residents), socioeconomic variables (income and education), health variables (incidence of diabetes, obesity, smoking, and hypertension), Twitter language, and all combinations of these sets of predictors. Accuracy refers to the Pearson r correlation between the set of predictors and CDC reported AHD. Brackets give 95% confidence intervals. The models are trained on one part of the data ("training set") and evaluated on another ("hold-out set"), to avoid distortion through chance. A model combining Twitter and all predictors (Model #14) significantly outpredicted the model with all predictors (Model 13), suggesting that Twitter has incremental predictive validity. Twitter language by itself significantly outpredicted a model with all SES, demographic and health predictors (Model 15 compared to Model 13). Predictive performance between two models was compared through paired t-tests, comparing the sizes of standardized residuals of county-level predictions from each model. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

Table S5

Varimax-rotated Factor Structure of the County-level Frequencies of the 20 most Frequent Words in the Positive Relationship Dictionary

Words	Partnership factor	Social factor
love	**.65**	.39
home	.11	.35
friends	.47	**.53**
friend	.43	**.48**
team	-.07	.30
social	-.32	.13
welcome	-.09	.43
together	.40	.34
kind	-.23	.50
dear	.11	.41
agree	-.30	**.51**
loved	.03	**.51**
relationship	**.73**	.05
liked	.02	.12
loving	.18	.33
boyfriend	**.72**	.10
appreciate	.06	.27
girlfriend	**.66**	.06
helping	-.25	.38
united	-.27	.09

County-level correlations		
Socioeconomic Status (SES)†	-.43 [-.47, -.38]	.14 [.08, .19]
Atherosclerotic Heart Disease	.18 [.13, 23]	-.02 [-.07, .04]

Note. Examination of the eigenvalues and the Scree test revealed a clear two factor structure. Words are ordered in descending frequency of occurrence. Factor scores were imputed through regression (random factors, Thompson's method). Pearson correlations (*r*) are given with 95% confidence intervals in brackets. The 20 words shown account for 89.1% of all word occurrences of the positive relationship dictionary.
† SES index combining standardized high school and college graduation rates, and median income.

Table S6

Top Ten Dictionary Words by Frequency and Their Correlations with Atherosclerotic Heart Disease (AHD)

Anger Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
shit	.12 [.06, .17]	.07 [.02, .13]	2,178,219
fuck	.20 [.15, .25]	.17 [.11, .22]	1,551,388
hate	.23 [.18, .28]	.19 [.13, .24]	1,307,810
damn	.03 [-.02, .09]	-.03 [-.08, .03]	1,252,834
bitch	.13 [.07, .18]	.06 [.01, .12]	864,810
hell	.01 [-.04, .07]	-.05 [-.11, .00]	781,102
fucking	.28 [.23, .33]	.29 [.24, .34]	651,694
mad	.13 [.08, .19]	.09 [.03, .14]	514,694
stupid	.11 [.06, .16]	.06 [.00, .11]	410,894
bitches	.13 [.08, .18]	.09 [.03, .14]	305,033

Negative Relationships Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
hate	.23 [.18, .28]	.19 [.13, .24]	1,307,810
alone	.13 [.08, .18]	.09 [.03, .14]	292,621
jealous	.05 [-.01, .10]	.04 [-.02, .09]	177,374
blame	-.01 [-.07, .04]	-.01 [-.06, .04]	100,930
evil	-.07 [-.13, -.02]	-.07 [-.13, -.02]	94,161
rude	.04 [-.01, .10]	.02 [-.03, .08]	78,552
lonely	.05 [-.01, .10]	.01 [-.05, .06]	70,916
independent	-.04 [-.09, .01]	-.02 [-.08, .03]	39,313
hated	.10 [.05, .15]	.09 [.04, .14]	39,251
ban	-.05 [-.10, .00]	-.02 [-.07, .03]	36,417

Negative Emotions Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
sorry	.04 [-.02, .09]	.04 [-.01, .09]	757,751
mad	.13 [.08, .19]	.09 [.03, .14]	514,694
sad	.00 [-.05, .06]	.00 [-.05, .05]	428,082
scared	.09 [.03, .14]	.03 [-.03, .08]	168,420
pissed	.19 [.14, .24]	.15 [.10, .20]	140,696
crying	.11 [.06, .17]	.09 [.04, .14]	123,994
horrible	.07 [.02, .12]	.08 [.02, .13]	113,522
afraid	.05 [-.01, .10]	.04 [-.02, .09]	104,582
terrible	.03 [-.03, .08]	.06 [.00, .11]	104,195
upset	.10 [.05, .15]	.08 [.02, .13]	93,648

Disengagement Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
tired	.16 [.11, .21]	.10 [.05, .16]	580,979
bored	.18 [.13, .23]	.11 [.05, .16]	411,358
sleepy	-.01 [-.06, .04]	-.10 [-.16, -.05]	157,043
lazy	.04 [-.02, .09]	-.01 [-.06, .04]	138,761
blah	.07 [.02, .12]	.03 [-.02, .09]	110,085
meh	-.02 [-.07, .04]	-.04 [-.09, .01]	53,376
exhausted	.06 [.01, .12]	.09 [.03, .14]	49,955
yawn	-.03 [-.09, .02]	-.03 [-.08, .02]	21,398
distracted	-.06 [-.12, -.01]	-.04 [-.10, .01]	17,998
boredom	.04 [-.01, .10]	.04 [-.02, .09]	17,150

Anxiety Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
crazy	.13 [.08, .18]	.09 [.04, .14]	696,947
pressure	.02 [-.03, .08]	.03 [-.02, .09]	193,805
worry	.05 [-.01, .10]	.02 [-.03, .08]	172,486
scared	.09 [.03, .14]	.03 [-.03, .08]	168,420
awkward	.09 [.04, .15]	.09 [.03, .14]	152,980
scary	-.02 [-.08, .03]	-.02 [-.07, .04]	121,521
fear	-.06 [-.12, -.01]	-.05 [-.10, .00]	120,542
doubt	.09 [.03, .14]	.09 [.03, .14]	115,207
horrible	.07 [.02, .12]	.08 [.02, .13]	113,522
afraid	.05 [-.01, .10]	.04 [-.02, .09]	104,582

Positive Relationships Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
love	.13 [.08, .18]	.08 [.02, .13]	5,375,835
home	.11 [.05, .16]	.10 [.04, .15]	1,907,974
friends	.10 [.05, .15]	.09 [.04, .14]	1,005,756
friend	.05 [.00, .10]	.02 [-.03, .07]	721,639
team	-.07 [-.13, -.02]	-.05 [-.10, .01]	629,910
social	-.08 [-.14, -.03]	-.03 [-.09, .02]	448,731
welcome	-.04 [-.09, .01]	-.02 [-.07, .03]	421,685
together	.00 [-.05, .06]	-.02 [-.07, .04]	398,957
kind	-.09 [-.14, -.03]	-.04 [-.10, .01]	379,906
dear	.02 [-.03, .07]	.02 [-.03, .08]	289,738

Positive Emotion Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
great	-.15 [-.21, -.10]	-.09 [-.15, -.04]	2,375,268
happy	.06 [.01, .12]	.06 [.01, .12]	1,830,533
cool	-.09 [-.14, -.04]	-.06 [-.12, -.01]	972,187
awesome	-.07 [-.12, -.01]	-.02 [-.08, .03]	971,447
amazing	.04 [-.01, .09]	.09 [.04, .15]	715,301
glad	-.07 [-.13, -.02]	-.09 [-.15, -.04]	499,789
excited	.00 [-.06, .05]	.04 [-.01, .09]	495,371
super	-.01 [-.06, .05]	.01 [-.04, .07]	473,677
enjoy	-.07 [-.12, -.01]	-.02 [-.07, .03]	381,689
wonderful	-.05 [-.10, .00]	-.04 [-.09, .02]	204,721

Engagement Dictionary

Top Ten Words	Correlation with AHD Mortality (Pearson r with 95% CIs)	Correlation with AHD Mortality Controlled for Income and Education	Overall Frequency
learn	-.08 [-.13, -.02]	-.05 [-.11, .00]	350,873
interesting	-.17 [-.22, -.12]	-.10 [-.15, -.04]	305,703
awake	.12 [.07, .17]	.11 [.05, .16]	158,400
interested	-.10 [-.15, -.05]	-.05 [-.10, .01]	137,553
alive	.07 [.01, .12]	.06 [.01, .11]	132,898
learning	-.11 [-.16, -.06]	-.07 [-.12, -.02]	118,337
creative	-.10 [-.16, -.05]	-.04 [-.10, .01]	89,367
alert	-.04 [-.09, .01]	-.02 [-.08, .03]	80,982
involved	-.09 [-.14, -.04]	-.05 [-.11, .00]	65,361
careful	-.07 [-.12, -.02]	-.09 [-.14, -.03]	63,719