

Psychological Language on Twitter Predicts County-Level Heart Disease Mortality

[Johannes C. Eichstaedt](#), [Hansen Andrew Schwartz](#), [Margaret L. Kern](#), ...

First Published January 20, 2015 [Article](#)

Abstract

Hostility and chronic stress are known risk factors for heart disease, but they are costly to assess on a large scale. We used language expressed on Twitter to characterize community-level psychological correlates of age-adjusted mortality from atherosclerotic heart disease (AHD). Language patterns reflecting negative social relationships, disengagement, and negative emotions—especially anger—emerged as risk factors; positive emotions and psychological engagement emerged as protective factors. Most correlations remained significant after controlling for income and education. A cross-sectional regression model based only on Twitter language predicted AHD mortality significantly better than did a model that combined 10 common demographic, socioeconomic, and health risk factors, including smoking, diabetes, hypertension, and obesity. Capturing community psychological characteristics through social media is feasible, and these characteristics are strong markers of cardiovascular mortality at the community level.

References

Auchincloss A. H., Gebreab S. Y., Mair C., Diez Roux A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual Review of Public Health*, 33, 107–122. doi:10.1146/annurev-publhealth-031811-124655 [Google Scholar](#) [CrossRef](#), [Medline](#)

Boehm J. K., Kubzansky L. D. (2012). The heart's content: The association between positive psychological well-being and cardiovascular health. *Psychological Bulletin*, 138, 655–691. doi:10.1037/a0027448 [Google Scholar](#) [CrossRef](#), [Medline](#)

Boehm J. K., Kubzansky L. D. (in press). Positive psychological well-being and cardiovascular disease. In Kop W., Katzel L., Waldstein S. (Eds.), *Cardiovascular behavioral medicine*. New York, NY: Springer. [Google Scholar](#)

Centers for Disease Control and Prevention. (2009). Behavioral Risk Factor Surveillance System annual survey data. Retrieved from http://www.cdc.gov/brfss/annual_data/annual_2009.htm [Google Scholar](#)

Centers for Disease Control and Prevention. (2010a). Behavioral Risk Factor Surveillance System annual survey data. Retrieved from http://www.cdc.gov/brfss/annual_data/annual_2010.htm [Google Scholar](#)

Centers for Disease Control and Prevention. (2010b). Underlying cause of death 1999-

2010. CDC WONDER online database [Data set]. Retrieved from <http://wonder.cdc.gov/ucd-icd10.html> [Google Scholar](#)

Chaix B., Merlo J., Evans D., Leal C., Havard S. (2009). Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine*, 69, 1306–1310. doi:10.1016/j.socscimed.2009.07.018 [Google Scholar](#) [CrossRef](#), [Medline](#)

Chew C., Eysenbach G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE*, 5(11), Article e14118. Retrieved from <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0014118> [Google Scholar](#)

Chida Y., Steptoe A. (2008). Positive psychological well-being and mortality: A quantitative review of prospective observational studies. *Psychosomatic Medicine*, 70, 741–756. doi:10.1097/PSY.0b013e31818105ba [Google Scholar](#) [CrossRef](#), [Medline](#)

Chida Y., Steptoe A. (2009). The association of anger and hostility with future coronary heart disease: A meta-analytic view of prospective evidence. *Journal of the American College of Cardiology*, 53, 936–946. doi:10.1016/j.jacc.2008.11.044 [Google Scholar](#) [CrossRef](#), [Medline](#)

Cohen D. A., Farley T. A., Mason K. (2003). Why is poverty unhealthy? Social and physical mediators. *Social Science & Medicine*, 57, 1631–1641. doi:10.1016/S0277-9536(03)00015-7 [Google Scholar](#) [CrossRef](#), [Medline](#)

De Choudhury M., Counts S., Horvitz E. (2013). Social media as a measurement tool of depression in populations. In *WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference* (pp. 47–56). doi:10.1145/2464464.2464480 [Google Scholar](#) [CrossRef](#)

de Quincey E., Kostkova P. (2009). Early warning and outbreak detection using social networking websites: The potential of Twitter. In Kostkova P. (Ed.), *Electronic Healthcare: Second International ICST Conference, eHealth 2009, Istanbul, Turkey, September 23–25, 2009* (pp. 21–24). doi:10.1007/978-3-642-11745-9_4 [Google Scholar](#) [CrossRef](#)

Diez Roux A. V., Mair C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186, 125–145. doi:10.1111/j.1749-6632.2009.05333.x [Google Scholar](#) [CrossRef](#), [Medline](#)

Eysenbach G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, 11(1), Article e11. Retrieved from <http://www.jmir.org/2009/1/e11/>. [Google Scholar](#)

Ford E. S., Capewell S. (2011). Proportion of the decline in cardiovascular mortality disease due to prevention versus treatments: Public health versus clinical care. *Annual Review of Public Health*, 32, 5–22. doi:10.1146/annurev-publhealth-031210-101211 [Google Scholar](#) [CrossRef](#), [Medline](#)

Fox S., Zickurh K., Smith A. (2009). Twitter and status updating, fall 2009. Retrieved from Pew Research Internet Project Web site: <http://www.pewinternet.org/2009/10/21/twitter-and-status-updating-fall-2009> [Google Scholar](#)

Fredrickson B. L., Mancuso R. A., Branigan C., Tugade M. M. (2000). The undoing effects of positive emotions. *Motivation and Emotion*, 24, 237–258. doi:10.1023/A:1010796329158 [Google Scholar](#) [CrossRef](#), [Medline](#)

Friedman H. S., Kern M. L. (2014). Personality, well-being, and health. *Annual Review of Psychology*, 65, 719–742. doi:10.1146/annurev-psych-010213-115123 [Google Scholar](#) [CrossRef](#), [Medline](#)

- Ginsberg J., Mohebbi M. H., Patel R. S., Brammer L., Smolinski M. S., Brilliant L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014. doi:10.1038/nature07634 [Google Scholar](#) [CrossRef](#), [Medline](#)
- Grimmer J., Stewart B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21, 267–297. doi:10.1093/pan/mps028 [Google Scholar](#) [CrossRef](#)
- Howell R., Kern M. L., Lyubomirsky S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes. *Health Psychology Review*, 1, 83–136. doi:10.1080/17437190701492486 [Google Scholar](#) [CrossRef](#)
- Labarthe D. R. (2010). *Epidemiology and prevention of cardiovascular disease: A global challenge*. Sudbury, MA: Jones & Bartlett. [Google Scholar](#)
- Lett H. S., Blumenthal J. A., Babyak M. A., Sherwood A., Strauman T., Robins C., Newman M. F. (2004). Depression as a risk factor for coronary artery disease: Evidence, mechanisms, and treatment. *Psychosomatic Medicine*, 66, 305–315. doi:10.1097/01.psy.0000126207.43307.c0 [Google Scholar](#) [CrossRef](#), [Medline](#)
- Leyland A. H. (2005). Socioeconomic gradients in the prevalence of cardiovascular disease in Scotland: The roles of composition and context. *Journal of Epidemiology & Community Health*, 59, 799–803. doi:10.1136/jech.2005.034017 [Google Scholar](#) [CrossRef](#), [Medline](#)
- Lloyd-Jones D. M., Hong Y., Labarthe D., Mozaffarian D., Appel L. J., Van Horn L., . . . Rosamond W. D. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: The American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*, 121, 585–613. doi:10.1161/CIRCULATIONAHA.109.192703 [Google Scholar](#) [CrossRef](#)
- Lochner K. A., Kawachi I., Brennan R. T., Buka S. L. (2003). Social capital and neighborhood mortality rates in Chicago. *Social Science & Medicine*, 56, 1797–1805. doi:10.1016/S0277-9536(02)00177-6 [Google Scholar](#) [CrossRef](#), [Medline](#)
- Menezes A. R., Lavie C. J., Milani R. V., O'Keefe J., Lavie T. J. (2011). Psychological risk factors and cardiovascular disease: Is it all in your head? *Postgraduate Medicine*, 123, 165–176. doi:10.3810/pgm.2011.09.2472 [Google Scholar](#) [CrossRef](#), [Medline](#)
- Mislove A., Lehmann S., Ahn Y.-Y., Onnela J. P., Rosenquist J. N. (2011). Understanding the demographics of Twitter users. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (pp. 554–557). Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816/3234> [Google Scholar](#)
- Paul M. J., Dredze M. (2011a). A model for mining public health topics from Twitter. Retrieved from the Johns Hopkins University Web site: http://www.cs.jhu.edu/~mpaul/files/2011.tech.twitter_health.pdf [Google Scholar](#)
- Paul M. J., Dredze M. (2011b). You are what you tweet: Analyzing Twitter for public health. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (pp. 265–272). Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880/3264> [Google Scholar](#)
- Pennebaker J. W., Chung C. K., Ireland M., Gonzales A., Booth R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net. [Google Scholar](#)
- Pennebaker J. W., Mehl M. R., Niederhoffer K. G. (2003). *Psychological aspects of natural*

language use: Our words, our selves. *Annual Review of Psychology*, 54, 547–577. doi:10.1146/annurev.psych.54.101601.145041 [Google Scholar](#) [CrossRef](#), [Medline](#)

Pierce J. R., Denison A. V. (2010). Accuracy of death certificates and the implications for studying disease burdens. In Preedy V. R., Watson R. R. (Eds.), *Handbook of disease burdens and quality of life measures* (pp. 329–344). New York, NY: Springer. [Google Scholar](#) [CrossRef](#)

Pressman S. D., Cohen S. (2005). Does positive affect influence health? *Psychological Bulletin*, 131, 925–971. doi:10.1037/0033-2909.131.6.925 [Google Scholar](#) [CrossRef](#), [Medline](#)

Riva M., Gauvin L., Barnett T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology & Community Health*, 61, 853–861. doi:10.1136/jech.2006.050740 [Google Scholar](#) [CrossRef](#), [Medline](#)

Roest A. M., Martens E. J., de Jonge P., Denollet J. (2010). Anxiety and risk of incident coronary heart disease: A meta-analysis. *Journal of the American College of Cardiology*, 56, 38–46. doi:10.1016/j.jacc.2010.03.034 [Google Scholar](#) [CrossRef](#), [Medline](#)

Rosenthal R., DiMatteo M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59–82. doi:10.1146/annurev.psych.52.1.59 [Google Scholar](#) [CrossRef](#), [Medline](#)

Rost M., Barkhuus L., Cramer H., Brown B. (2013, February). Representation and communication: Challenges in interpreting large social media datasets. Paper presented at the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2013), San Antonio, TX. [Google Scholar](#) [CrossRef](#)

Rowe J. W., Kahn R. L. (1987). Human aging: Usual and successful. *Science*, 237, 143–149. doi:10.1126/science.3299702 [Google Scholar](#) [CrossRef](#), [Medline](#)

Rugulies R. (2002). Depression as a predictor for coronary heart disease: A review and meta-analysis. *American Journal of Preventive Medicine*, 23, 51–61. doi:10.1016/S0749-3797(02)00439-7 [Google Scholar](#) [CrossRef](#), [Medline](#)

Salathé M., Freifeld C. C., Mearu S. R., Tomasulo A. F., Brownstein J. S. (2013). Influenza A (H7N9) and the importance of digital epidemiology. *New England Journal of Medicine*, 369, 401–404. doi:10.1056/NEJMp1307752 [Google Scholar](#) [CrossRef](#), [Medline](#)

Schwartz H. A., Eichstaedt J. C., Kern M. L., Dziurzynski L., Lucas R. E., Agrawal M., . . . Ungar L. H. (2013). Characterizing geographic variation in well-being using tweets. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (pp. 583–591). Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6138/6398>. [Google Scholar](#)

Schwartz H. A., Eichstaedt J. C., Kern M. L., Dziurzynski L., Ramones S. M., Agrawal M., . . . Ungar L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, 8(9), Article e73791. Retrieved from <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0073791> [Google Scholar](#)

Seifter A., Schwarzwald A., Geis K., Aucott J. (2010). The utility of “Google Trends” for epidemiological research: Lyme disease as an example. *Geospatial Health*, 4, 135–137. [Google Scholar](#) [CrossRef](#), [Medline](#)

St Louis C., Zorlu G. (2012). Can Twitter predict disease outbreaks? *British Medical Journal*, 344, Article e2353. Retrieved from <http://www.bmj.com/content/344/bmj.e2353> [Google Scholar](#)

Stone P. J., Dunphy D. C., Smith M. S., Ogilvie D. M. (1966). The general inquirer: A computer approach to content analysis. Cambridge, MA: MIT Press. [Google Scholar](#)

Tay L., Tan K., Diener E., Gonzalez E. (2013). Social relations, health behaviors, and health outcomes: A survey and synthesis. *Applied Psychology: Health and Well-Being*, 5, 28–78. [Google Scholar](#) [CrossRef](#), [Medline](#)

U.S. Census Bureau. (2009). Selected social characteristics in the United States: 2009. American Community Survey 1-year estimates. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_09_1YR_DP2&prodType=table [Google Scholar](#)

U.S. Census Bureau. (2010). Profile of general population and housing characteristics: 2010 demographic profile data. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_DP_DPDP1&prodType=table [Google Scholar](#)

World Health Organization. (1992). Atherosclerotic heart disease. In *International statistical classification of diseases and related health problems (10th ed.)*. Retrieved from <http://apps.who.int/classifications/icd10/browse/2010/en#/I25.1> [Google Scholar](#)

World Health Organization. (2011). Global status report on noncommunicable diseases 2010. Description of the global burden of NCDs, their risk factors and determinants. Retrieved from http://www.who.int/nmh/publications/ncd_report2010/en/ [Google Scholar](#)