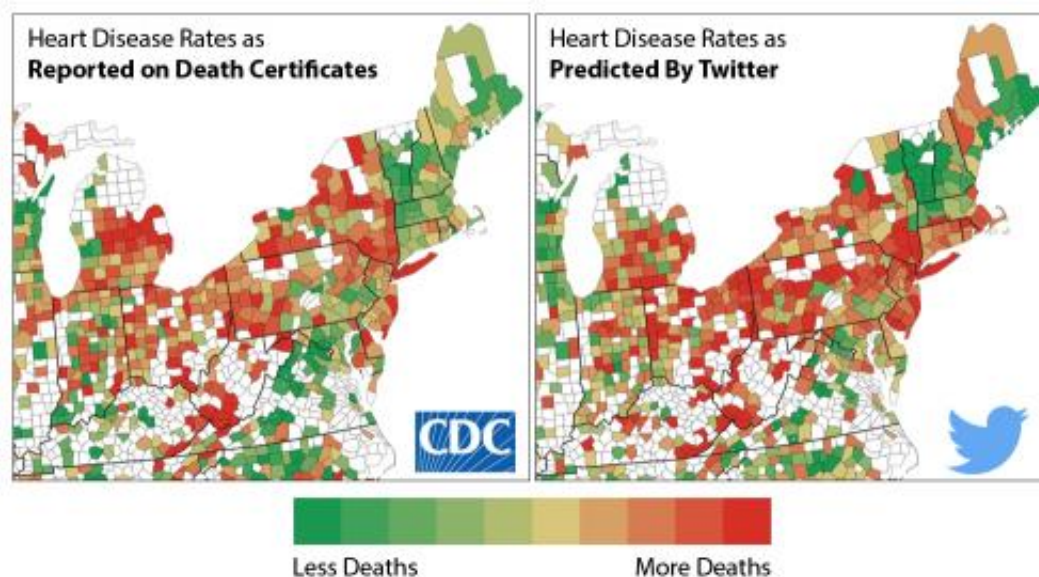# Twitter can predict rates of coronary heart disease, research says

January 21 2015

Penn researchers compared the language of tweets and CDC data on coronary heart disease deaths on a county-by-county level. On one hand, they found that expressions of negative emotions such as anger, stress and fatigue in a county's tweets were associated with higher heart disease risk. On the other hand, positive emotions like excitement and optimism were associated with lower risk. Because the people tweeting these words were not the ones dying, the study suggests that Twitter language is capturing some greater aspect of community life in those counties that has implications for heart disease risk. Credit: University of Pennsylvania

Twitter has broken news stories, launched and ended careers, started

social movements and toppled governments, all by being an easy, direct and immediate way for people to share what's on their minds. Researchers from the University of Pennsylvania have now shown that the social media platform has another use: Twitter can serve as a dashboard indicator of a community's psychological well being and can predict rates of heart disease.

Previous studies have identified many factors that contribute to the risk of heart disease: traditional ones, like low income or smoking but also psychological ones, like stress. The Penn researchers demonstrated that Twitter can capture more information about heart disease risk than many traditional factors combined, as it also characterizes the psychological atmosphere of a community.

They found that expressions of negative emotions such as anger, stress and fatigue in a county's tweets were associated with higher heart disease risk. On the other hand, positive emotions like excitement and optimism were associated with lower risk.

Researchers have long assumed that the psychological well being of communities is important for physical health, but is hard to measure. Using Twitter as a window into a community's collective mental state may provide a useful tool in epidemiology and for measuring the effectiveness of public-health interventions.

The study, published in the journal *Psychological Science*, was led by Johannes Eichstaedt, a graduate student in the School of Arts & Science's Department of Psychology, and included H. Andrew Schwartz, a visiting assistant professor in the School of Engineering and Applied Science's Department of Computer and Information Science; Margaret Kern, an assistant professor at the University of Melbourne, Australia; Gregory Park, a postdoctoral fellow in the School of Arts and Science's Department of Psychology; and director Martin Seligman, both of the

Positive Psychology Center, as well as Lyle Ungar, a professor of computer and information science.

With billions of users writing daily about their daily experiences, thoughts and feelings, the world of social media represents a new frontier for psychological research. Such data could be an invaluable public health tool if able to be tied to real-world outcomes. With this in mind, the researchers from the World Well-Being Project have long been studying the degree to which the language people use online represents their inner thoughts and feelings.

As there is no way to directly measure peoples' inner emotional lives, the team drew on traditions in psychological research that glean this information from the words people use when speaking or writing. Earlier research from the group has shown that such linguistic analysis can work as well as traditional questionnaires in assessing an individual's personality.

"Getting this data through surveys is expensive and time consuming, but, more important, you're limited by the questions included on the survey," Eichstaedt said. "You'll never get the psychological richness that comes with the infinite variables of what language people choose to use."

Having seen correlations between language and emotional states, the researchers went on to see if they could show connections between those emotional states and physical outcomes rooted in them. They had an ideal candidate in coronary heart disease, the leading cause of death worldwide.

"Psychological states have long been thought to have an effect on coronary heart disease," said Kern. "For example, hostility and depression have been linked with heart disease at the individual level through biological effects. But negative emotions can also trigger

behavioral and social responses; you are also more likely to drink, eat poorly and be isolated from other people which can indirectly lead to heart disease."

As a common cause of early mortality, public health officials carefully count when heart disease is identified as the underlying cause on death certificates. They also collect meticulous data about possible risk factors, such as rates of smoking, obesity, hypertension and lack of exercise. This data is available on a county-by-county level in the United States, so the research team aimed to match this physical epidemiology with their digital Twitter version.

Drawing on a set of public tweets made between 2009 and 2010, the researchers used established emotional dictionaries, as well as automatically generated clusters of words reflecting behaviors and attitudes, to analyze a random sample of tweets from individuals who had made their locations available. There were enough tweets and health data from about 1,300 counties, which contain 88 percent of the country's population.

They found that negative emotional language and topics, such as words like "hate" or expletives, remained strongly correlated with heart disease mortality, even after variables like income and education were taken into account. Positive emotional language showed the opposite correlation, suggesting that optimism and positive experiences, words like "wonderful" or "friends," may be protective against heart disease.

"The relationship between language and mortality is particularly surprising," Schwartz said, "since the people tweeting angry words and topics are in general not the ones dying of heart disease. But that means if many of your neighbors are angry, you are more likely to die of heart disease."

This finding fits into existing sociological research that suggests that the combined characteristics of communities can be more predictive of physical health than the reports of any one individual.

"We believe that we are picking up more long-term characteristics of communities," Ungar said. "The language may represent the 'drying out of the wood' rather than the 'spark' that immediately leads to mortality. We can't predict the number of heart attacks a county will have in a given timeframe, but the language may reveal places to intervene."

Other caveats to the method's predictive power include the social factors that influence what kinds of messages people choose to share on Twitter.

"If everyone is a little more positive on Twitter than they are in real life, however, we would still see variation from location to location, which is what we're most interested in," Schwartz said.

This variation could be used to marshal evidence of the effectiveness of public-health interventions on the community level, rather than on an individual level. The team's findings show that these tweets are aggregating information about people that can't be readily accessed in other ways.

"Twitter seems to capture a lot of the same information that you get from health and demographic indicators," Park said, "but it also adds something extra. So predictions from Twitter can actually be more accurate than using a set of traditional variables."

Provided by University of Pennsylvania