Both Mirror and Complement: A Comparison of Social Media Data and Survey Data about Flu Vaccination

David Broniatowski, Ph.D., George Washington University; Mark Dredze, Ph.D., Johns Hopkins University; Karen Hilyard, Ph.D., University of Georgia; Maeghan Dessecker, M.A., M.P.H., University of Georgia; Amelia Jamison, M.A., M.P.H., University of Maryland; Michael Paul, Ph.D., University of Colorado, Boulder; Michael Smith, M.S., George Washington University; Sandra Cruise Quinn, Ph.D., University of Maryland

Background

Social media offers researchers opportunities to assess public knowledge, attitudes, and health behaviors in real-time. Researchers have relied on survey methods to provide reliable and accurate data, but are time consuming. Survey research struggles to capture the attitudes of young people, minority groups, and uninsured: all groups well represented on social media. However, social media brings unique challenges including huge amounts of data and difficulty in accurately assessing demographic information about social media users.

Research Questions

The goal of this study is to compare Twitter data to published CDC survey data on influenza vaccination. Our hypothesis is that Twitter data will complement these findings published by CDC data.

Measures & Methods

Traditional Survey Data

CDC’s ‘FluView®’ data from nationally representative surveys including the National Immunization Survey (NIS-Flu), the Behavioral Risk Factor Surveillance System (BRFSS), and the National Health Interview Survey (NHIS)

Vaccination rates are estimated by age, gender, geographic location (NHIS Regional) and for each month during the flu season.

Measures & Methods

TWEET DATA

Since 2012, collected tweets that contain health-related keywords, using the Twitter search API 2

A subset of “flu-related” tweets were identified as having one flu-related keyword (RF, influenza) and one vaccine-related keyword (shot, vaccines, vaccination). Retweets and Non-English tweets were excluded.

Twitter data has been collected over the last 3 years.

Direction for Future Research

Develop methods to infer additional demographics including age and race/ethnicity.

We move beyond existing CDC statistics to collect data not already captured in existing research.

Conclusions

Twitter data is promising means of tracking flu vaccine behavior.

Acknowledgments


Citations


