

PUBLIC OPINION AND SOCIAL MEDIA

Potential and Pitfalls

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Michael C. Smith, The George Washington University, mikesmith@gwu.edu
David A. Broniatowski, The George Washington University, broniatowski@gwu.edu

Abstract

- Social media are a promising new data source for surveying public opinion.
- Despite clear advantages, analyses of social media data face some challenges.
- We seek to elucidate these challenges and draw relevant lessons from more traditional survey techniques.
- We argue that social media studies should carefully consider elements of study design
 - We plan to focus on issues of research validity:
 - Internal validity (present a pilot study here)
 - External validity
 - Construct validity
 - Statistical conclusion validity
 - Examples throughout relate to vaccine hesitancy and refusal
- Common pitfalls and techniques to avoid these are discussed.

Introduction, Background

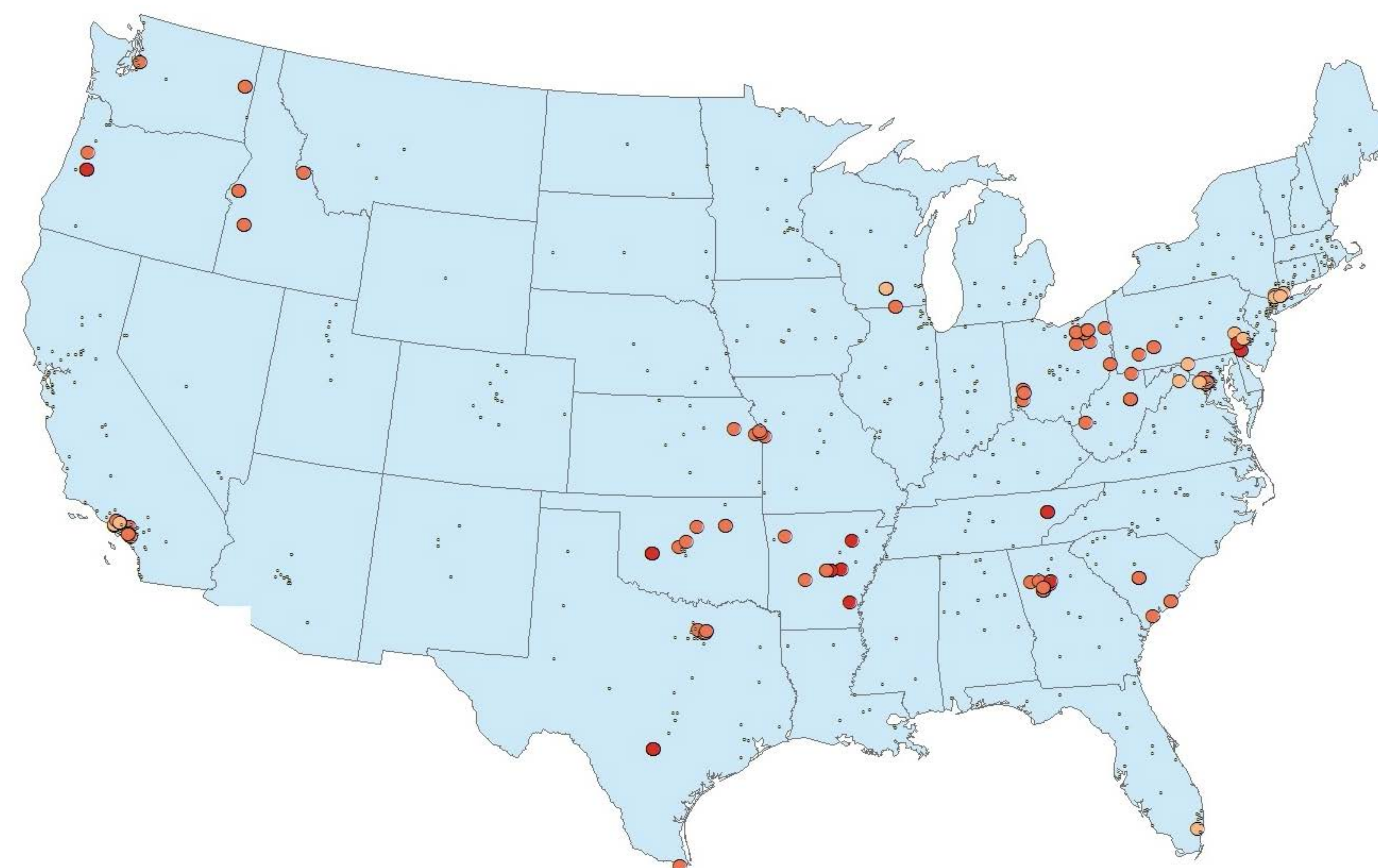
- Social Media are a promising source of data for surveillance of public opinion
 - e.g., disaster response, public health, political views [1, 2, 3]
- **Advantages:** breadth, depth, lack of response bias, access to minority viewpoints
- **Disadvantages:** should be addressed if we are to leverage potential
- Framework of internal, external, construct, and statistical conclusion validity [4]
 - Organic data strong in external validity, vulnerable to other three types
 - Structure research: compare through these lenses
- Draw lessons from traditional surveys in context of validity structure
 - To strengthen social media's and surveys' synergy for measuring opinion

Pilot Study - Methods

- Pilot study leverages internal validity:
- Aim to understand nuance of vaccine-related conversations
 - Do not claim representativeness, rather study specific construct(s)
 - Collected and classified vaccine-related Twitter messages [5] starting from the Disneyland measles outbreak, 12/8/14-3/2/15
 - Keyword filters: 50 vaccine-related keywords
 - Geolocation classified [6]
 - Topics from LDA and public health experts' interpretations
 - Geospatial analysis: Getis-Ord G_i^* statistical hotspots [7] by topic

Pilot Study Results

- Maps of statistical spatial hotspots, indicating spatially varying conversations about vaccines
- Indications of areas for promising future work by spatial nuance
- Leverage internal validity to understand nuances of a specific group or idea



- Above: map of discussions of California gov't bill about personal exemptions
- Suggests interesting future work: why Cleveland? Why Appalachia?
- Nonprobability sampling powerful to study nuances
 - Compare: probability sampling to leverage representativeness

Dissertation Directions

- Evaluating and exemplifying
 - construct validity
 - Response bias
 - Measurement bias
 - statistical conclusion validity
 - Sampling error
 - Nonresponse bias
 - external validity
- Studies replicating and augmenting surveys on vaccine behavior, e.g.:
 - Vaccine exemptions
 - Demography/topic/geography
 - Perceived risk
- Broader implications:
 - Engineering successful data triangulation
 - Actionable public health implications
 - Synergistic contributions to public opinion via validity framework for both data sources

References

1. Salathe, M., Khandelwal, S.: Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. *PLOS Computational Biology* 7(10), e1002199 (Oct 2011).
2. Huang, X., Smith, M.C., Paul, M.J., Ryzhkov, D., Quinn, S.C., Broniatowski, D.A., Dredze, M.: Examining Patterns of Influenza Vaccination in Social Media. In: AAAI Joint Workshop on Health Intelligence (W3PHIAI) (2017)
3. O'Connor, B., Balasubramanian, R., Routledge, B.R., Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM* 11(122-129), 1(2 (2010)
4. Shadish, W., Cook, T.D., Campbell, D.T.: *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning (2002)
5. Dredze, M., Broniatowski, D.A., Smith, M.C., Hilyard, K.M.: Understanding Vaccine Refusal: Why We Need Social Media Now. *American Journal of Preventive Medicine* 50(4), 550 (2016)
6. Getis, A., Ord, J.K.: The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24(3), 189(206 (Jul 1992).
7. Dredze, M., Paul, M. J., Bergsma, S., & Tran, H. (2013, June). Carmen: A twitter geolocation system with applications to public health. In AAAI workshop on expanding the boundaries of health informatics using AI (HIAI) (Vol. 23, p. 45).