Building a Framework for Social Media Data Reuse in Public Health Research

Lorinette S. Wirth, MPH

"Pay close attention to the content of social media access, production, and education...shift from narrowly focusing on the brightest signals to highlight those that are obscured...make a case for the scientific need of social media data access in the public health and social welfare. We will need industry- and university-based researchers to unite." – Gray M.

BACKGROUND

- Social media generates massive amounts of patient-reported data.
- Unique challenges arise when working with social media data, as is common with other sources of secondary data.
- Scientists must adopt innovative and multidisciplinary approaches to utilize this data for improved healthcare quality.
- Consensus as to where and how social media data utilization fits within public health research design is needed.
- Theoretical frameworks can serve as a roadmap for the reuse of patient-generated data from social media for conducting rigorous, evidence-based research.
- Leads to better formulated research questions and hypotheses, more appropriate selection of research methods, and conclusions.
- Recent attention has been placed on online health communities (OHCs), a subset of social media where patients and their caregivers gather to learn about a shared illness, seek and offer support, and connect with others in similar situations (e.g., “CancerSurvivorsNetwork”).

OBJECTIVE: to develop a framework with recommendations for the reuse of patient-generated data from OHCs in public health research

- Insights gained are expected to be mostly generalizable to other social media sources and will serve as a foundation for future studies

METHODS

- Inclusion criteria included articles that 1) had available peer-reviewed abstracts that were published after 2005, 2) reported analysis of social media data within (bio)medical or public health domains, 3) were written in English, and 4) were original research.
- We reviewed abstracts (n=300) and excluded studies that described disease surveillance, health interventions, health messaging strategies, and evaluations, and use of OHCs for research purposes.
- Recommendations and guidelines from the the Research Data Alliance (RDA) working groups were incorporated into framework.

RESULTS

Table 1. Description of a sample of reviewed articles

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dillies &amp; Newbold</td>
<td>Hypothesis driven/Thematic/Content working and interest groups were incorporated into framework.</td>
</tr>
<tr>
<td>Song, Mei et al.</td>
<td>Recommendations and guidelines from the Research Data Alliance (RDA) working groups were incorporated into framework.</td>
</tr>
<tr>
<td>Thompson, Vanacker-Otten et al.</td>
<td>Social media data reuse has the potential to guide public health research and policy, if used in a methodologically sound manner.</td>
</tr>
<tr>
<td>Lu, Yingjie et al.</td>
<td>Social media data reuse has the potential to guide public health research and policy, if used in a methodologically sound manner.</td>
</tr>
</tbody>
</table>

Table of Contents:

- CONCLUSIONS
- ACKNOWLEDGEMENTS

CONCLUSIONS

- Social media data reuse has the potential to guide public health research and policy, if used in a methodologically sound manner.
- Learn about patient’s experiences, what works and doesn’t work to improve healthcare quality.
- Not meant to replace other data source, but compliment existing data and research.
- While there is possibility for noisy or inaccurate data, the significance of the availability of fast, large-scale, low-cost, and unadulterated data collection via social media data rise is undeniable.
- Social media data reuse has the potential to guide public health research and policy, if used in a methodologically sound manner.
- The framework that is presented in this current project is the first of its kind, addressing the gaps bought by the scale and nature of this type of data. It should be used and expanded upon by researchers wanting to conduct rigorous studies using patient-generated data from social media.
- Future studies should expand upon this framework from an interdisciplinary perspective and design quality control methods such as scoring algorithms to assess validity of social media data and research based on the issues raised in this framework.

ACKNOWLEDGEMENTS

This project was sponsored in part through grants from the Alfred P. Sloan Foundation #IG-2014-13746 and the National Science Foundation NSF ACI1349002 via the RDA US Data Share Fellows program. I would like to thank my RDA mentor, Dr. Leslie McIntosh for her guidance and support towards this project.

REFERENCES

- Zhang, S. 2016. Computational Approaches to Characterizing Online Health Communities. 
- Gray M. Putting Social Media in Its Place: A Curatorial Theory for Media’s Social Risks. DOI: 10.1177/2056305115578863
- Gray M. Putting Social Media in Its Place: A Curatorial Theory for Media’s Social Risks. DOI: 10.1177/2056305115578863
- Gray M. Putting Social Media in Its Place: A Curatorial Theory for Media’s Social Risks. DOI: 10.1177/2056305115578863
- Gray M. Putting Social Media in Its Place: A Curatorial Theory for Media’s Social Risks. DOI: 10.1177/2056305115578863
- Gray M. Putting Social Media in Its Place: A Curatorial Theory for Media’s Social Risks. DOI: 10.1177/2056305115578863