Social Media for Hypothesis Testing

David A. Broniatowski
Department of Engineering Management and Systems Engineering
The George Washington University

This work supported in part by grant number R01GM114771 to D A Broniatowski and S C Quinn
Scientific value: Hypothesis testing

- Social media and survey data complement one another, in terms of:
  - Demographics
  - Clock speed
  - Immediacy
  - Analytic rigor

- Social media analysts are on the cusp of developing the research norms and practices that characterize high quality survey research
  - We can test hypotheses rapidly with very large samples

Study setting: Disneyland measles outbreak

• Began December 2014 at Disneyland in California
• Led to 111 cases in seven states (as well as Canada and Mexico)
• Cases began among unvaccinated people
• Called attention to the issue of herd immunity
• Led to proposals to curtail vaccine refusal through legislative means

Statistics, Stories…. or gist?

• Ongoing debate: Statistics vs. stories (“either-or”)
  • Does including a story lead to more effective communications than presenting “just the facts” (i.e., statistical data)?
  • Hesitance to include stories because of concerns of appearing biased

• Fuzzy Trace Theory: Gist and verbatim encoded in parallel
  • Verbatim representation (statistical details)
    • ”Measles can lead to pneumonia, deafness, lifelong brain damage, and even death, and almost 1/3 of children with measles have to be hospitalized”
  • Gist: Communicates bottom-line meaning
    • ”Taking any risk that your child could get the measles and suffer serious complications isn’t worth it. Vaccination is the best way to protect your child”
  • Stories are effective because they communicate a gist.
    • Also cue motivationally relevant moral and social principles

Analysis of measles media coverage: no response bias

- Measured shares on Facebook
- Used M-Turk to categorize article content:
  - 1) statistics about viruses or vaccines
  - 2) "gist", or bottom line meaning
    - Positive or negative summary opinion about endorsing or opposing vaccination
  - 3) Other expected covariates based on prior literature

What led to article shares:

- Results are consistent with Fuzzy Trace Theory
  - Significant effects of gist and verbatim, but NOT stories
- Stories are effective to the extent that they communicate gist
- Among articles with gists shared at least once (n=257) articles expressing positive opinions about those endorsing vaccination AND those opposing vaccination were 57.8 times more often

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE β</th>
<th>Z-value</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>-.000556</td>
<td>.0000893</td>
<td>-6.22*</td>
<td>1.00</td>
</tr>
<tr>
<td>Readability</td>
<td>-.000723</td>
<td>.00149</td>
<td>-0.49</td>
<td>1.00</td>
</tr>
<tr>
<td>Image</td>
<td>0.59</td>
<td>0.09</td>
<td>6.91*</td>
<td>1.80</td>
</tr>
<tr>
<td>Stories</td>
<td>0.34</td>
<td>0.19</td>
<td>1.82</td>
<td>1.41</td>
</tr>
<tr>
<td>Statistics</td>
<td>0.29</td>
<td>0.08</td>
<td>3.48*</td>
<td>1.33</td>
</tr>
<tr>
<td>Gist</td>
<td>0.82</td>
<td>0.15</td>
<td>5.36*</td>
<td>2.27</td>
</tr>
<tr>
<td>Stories x Statistics</td>
<td>0.05</td>
<td>0.22</td>
<td>0.24</td>
<td>1.05</td>
</tr>
<tr>
<td>Stories x Gist</td>
<td>0.25</td>
<td>0.32</td>
<td>0.80</td>
<td>1.29</td>
</tr>
<tr>
<td>Statistics x Gist</td>
<td>-0.17</td>
<td>0.20</td>
<td>-0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Stories x statistics x Gist</td>
<td>-.35</td>
<td>0.40</td>
<td>-0.89</td>
<td>0.70</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-1.08</td>
<td>0.12</td>
<td>-8.91*</td>
<td></td>
</tr>
</tbody>
</table>

Note. * = p < 0.001. β = logistic regression coefficient; SE β = standard error of β; OR = Odds Ratio

Tracking Pseudoscientific Claims

- 85% of users previously tweeted about vaccines in 2015
- At least 57% of users previously tweeted an anti-vaccine message

As if bots weren’t bad enough already, now they’re anti-vaccine

A dose of the measles-mumps-rubella vaccine. (Damian Dovarganes/AP)
Hypotheses

• Do bots post about vaccines more than the average Twitter user?
• Are bots more likely to be pro-vaccine or anti-vaccine?

Methods
• Identified public lists of different types of bot and troll accounts
• Searched through vaccine stream for their tweets
• Examined relative proportions of neutral, pro-vaccine, and anti-vaccine tweets
• Used Bot-o-Meter API to measure the “bot scores” of 9,994 randomly sampled tweets
Russian Trolls and Sophisticated Bots Tweet About Vaccines More Frequently than Average

*** = p<0.001, ** = p<0.01, * = p<0.05. All results remained significant after controlling for multiple comparisons using Holm-Bonferroni procedure.
“Unidentified” Accounts and Content Polluters are Most Likely to be Anti-Vaccine

<table>
<thead>
<tr>
<th></th>
<th>Assorted Users</th>
<th>Known Bots and Trolls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Polarized (%)</td>
<td>Anti-Vaccine (%)</td>
</tr>
<tr>
<td>Bot Score between 20% and 80%</td>
<td>39(^**)</td>
<td>60(^***)</td>
</tr>
<tr>
<td>Bot Score &gt; 80%</td>
<td>26</td>
<td>49(^*),(^x)</td>
</tr>
<tr>
<td>Bot Score unknown</td>
<td>37(^*),(^x)</td>
<td>62(^***)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Online organized misinformation and/or disinformation campaigns can undermine public health

• In partnership with our collaborators, we are developing new techniques to assess how compelling and influential messages might be
  • Based on empirically validated theory: Fuzzy Trace Theory
  • Complementary to existing survey techniques

• Future directions: Better understanding the drivers of coherent gist communications in online messages
  • How these vary across sociodemographic groups (e.g., different interpretations of emergency use authorization versus standard vaccination)