

# Measuring Campaign Message Exposure and Public Communication Environment Exposure: Some Implications of the Distinction in the Context of Social Media

Robert Hornik

Annenberg School for Communication, University of Pennsylvania, Philadelphia, Pennsylvania

## ABSTRACT

*The comments in this article suggest that the measurement differences associated with the distinction between assessing exposure to specific campaign messages and assessing exposure to ideas in the public communication environment may become sharper in the context of social/digital media, compared to traditional media. The argument: using social media to deliver specific campaign messages to a large targeted audience is hard, and measurement of exposure will be more demanding than with conventional media sources. In contrast, when the goal is to capture exposure to information in the public communication environment, "big data" based content analyses of social media streams may be a valuable approach; still there is much more work to do to establish the validity of these approaches and their usefulness in accounting for behavior.*

Here is a distinction that is obvious, but perhaps we do not consider its implications for measurement consistently, and particularly its implications in the context of social media. We measure exposure to communication for (at least) two different purposes. Sometimes we are interested in whether or not a communication campaign or other specific messages available through media produced individual exposure to those messages. At other times we are interested in exposure to particular topics and ideas that are available broadly in multiple media sources, traditional and digital and social media and interpersonal conversations. In that circumstance we are not focused on whether respondents saw a particular tweet, a particular advertisement, or had a specific conversation, but rather whether they are being exposed to given ideas as they rise and fall in the public communication environment (hereafter PCE) generally.

With the changes in the media environment, and increasing attention to social media, it may be that the distinction between these two cases looms larger. First for the specific message case: many campaigns now argue that traditional mass media may not reach their audiences and therefore build complementary social media strategies to achieve exposure particularly among youth and young adults. And it is true that a large majority of this audience (and other audiences as well) are heavy users of these forms of media. However it turns out to be more difficult to assure exposure to specific messages through social media.

Traditional mass media campaigns rely on passive reception—that is buying time on specific programs where audiences are known to be routinely in the audience. They do not rely so much on active seeking by the audience, but that is often required for social media; individuals must seek out that information by clicking through to a website or visiting a Facebook page. There is a quasipassive exposure path for digital/social media, when, for example, advertisers buy search terms on Google or other similar sites, and flash their messages to users who were not specifically looking for a specific site, but its effective reach and frequency has yet to be established. Still, when there is a need for active seeking of tweets, or YouTube video comments or Facebook entries, the fraction of the population directly exposed to any one message is almost always very small. While a very few of those items go viral, the proportion that do so is infinitesimal. Any specific campaign may propose a strategy to gain repeated exposure through social media, but this will be hard, and thus hardnosed measurement of exposure is required, against the same criteria of reach and frequency that would be applied to traditional media outreach. Whether measured through self-report, or Internet clicks, exposure should be compared against a denominator representing the target population.

However, when our interest is in assessing exposure for the second case—understanding what is in the public communication environment to assess the diffusion of ideas—then we might look at social media in a different light. A reasonable hypothesis (but it is only an hypothesis) is that the world of social media reflects what people are being exposed to: Can we sample the content of that environment and view that as an indicator of what ideas are getting exposure? We can sample the tweets, or the blogs, or the Facebook comments, or Google searches (as captured in Google Trends) about e-cigarettes and see how they vary in quantity, theme, or sentiment over time (Cole-Lewis et al., 2015). A media effects analysis can use those PCE

estimates to predict an outcome of interest (Depue, Southwell, Betzner, & Walsh, 2015). Does variation over time in pro-e-cigarette tweets predict increases in e-cigarette sales? We don't need to argue that any substantial fraction of the population has seen specific social media items to make inferences about the public communication environment; only we must assure ourselves that the items are a good sample of what is available and being seen, and, going one step further, also the topic of unseen personal conversations.

We cannot just assume that sampled social media content (e.g., tweets drawn from the Twitter fire hose over a specific time period) fairly represent the PCE. How would one validate a claim that over time content analysis of media sources captures the/a public communication environment? In part, assurance comes from thoughtful procedures used to sample the content stream (Ruths & Pfeffer, 2014). Evidence for high recall and precision from a search term, that selected items capture most relevant items in a source and that most selected items are in fact relevant, will help (Stryker, Wray, Hornik, & Yanovitzky, 2006). But that will provide only limited assurance, since even good sampling procedures produce a valid picture only if the content stream itself represents the PCE. Claims of PCE relevance might require some alternative measure of the PCE against which a specific content analysis can be compared. Here are two such alternative strategies: one could gather survey data over multiple weeks and ask respondents who use a particular source to recall how often they had seen the specific type of content of interest in the relevant time period in that source (Kelly, Niederdeppe, & Hornik, 2009). For example, does variation in quantity of news coverage of e-cigarettes over a 52-week period predict weekly survey self-reports of exposure to e-cigarette information from the media? This construct validity approach assumes that over time variation in content is reflected in respondents' likelihood of recalling exposure. The second validity approach would compare the over time variation in content from one source with the over time variation in content from other sources. Does week-by-week coverage of e-cigarettes in the Associated Press (AP) wire match the frequency of mentions of e-cigarettes in tweets? This approach would be easier to implement but makes an even more demanding assumption: that the PCE is equally reflected in multiple sources over time. Can one assume that both the Twitter stream and the AP stream both capture the same PCE and with the same time lags, so both will vary concomitantly over time?

These brief comments suggest that the measurement differences associated with the distinction between assessing exposure to specific messages and assessing exposure to ideas in the public communication environment may become sharper in the context of social/digital media. The arguments: using social media to deliver specific campaign messages is hard, and measurement of exposure will be more demanding than with conventional media sources; in contrast, "big data" based content analysis of social media streams may be a valuable approach to capturing exposure to information in the PCE; but there is much more work to do in establishing the validity of our approaches and their usefulness in accounting for behavior.

## Acknowledgments

These comments reflect ongoing discussions among colleagues at the Annenberg School/Penn. Research reported in this publication was supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) and FDA Center for Tobacco Products (CTP) under Award Number P50CA179546. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH or the Food and Drug Administration (FDA).

## References

- Cole-Lewis, H., Varghese, A., Sanders, A., Schwarz, M., Pugatch, J., & Augustson, E. (2015). Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. *Journal of Medical Internet Research*, 16(8). doi:10.2196/jmir.4392
- Depue, J. B., Southwell, B. G., Betzner, A. E., & Walsh, B.M. (2015). Encoded exposure to tobacco use in social media predicts subsequent smoking behavior. *American Journal of Health Promotion*, 29(4), 259–261. doi:10.4278/ajhp.130214-ARB-69
- Kelly, B. J., Niederdeppe, J., & Hornik, R. C. (2009). Validating measures of scanned information exposure in the context of cancer prevention and screening behaviors. *Journal of Health Communication*, 14(8), 721–740. doi:10.1080/10810730903295559
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213), 1063–1064. doi:10.1126/science.346.6213.1063

Stryker, J. E., Wray, R. J., Hornik, R. C., & Yanovitzky, I. (2006). Validation of database search terms for content analysis: The case of cancer news coverage. *Journalism & Mass Communication Quarterly*, 83(2), 413–430. doi:[10.1177/107769900608300212](https://doi.org/10.1177/107769900608300212)