Commentary

Statistics in Medicine

(wileyonlinelibrary.com) DOI: 10.1002/sim.5698 Published online in Wiley Online Library

Rejoinder to commentaries on social contagion theory

We thank the commentators for the care and attention they have given to this topic and for the new insights they are contributing to what we all regard as an important and engaging area, namely the study of the structure and function of human social networks.

Most of the commentators note the same limitations and promise that we, in our prior papers, have also noted with the use of generalized estimating equation (GEE) models implemented in longitudinal networks. Like us, they cannot see any superior alternative to analyzing such data, especially in networks of the size we analyze, given the statistical tools currently at our disposal.

As a scientific principle, we take it to be the case that some observation of the world is better than no observation. The fact that we can imagine the existence of a space-based telescope that would be superior to the terrestrial variety does not mean that the latter, with all the limitations imposed by light pollution and atmospheric interference, offers no value, offers no information about the natural world. We think that the current state of network statistics is like an earth-based telescope: the methods are not perfect, not free of all biases or assumptions. But still, they are much better than nothing.

We are grateful for Wasserman's deft highlighting of this fundamental point. Wasserman has been a pioneer in network statistics, and he knows how hard it is to reap a harvest in this terrain. Like him, we see no point in packing our bags and going home, because network phenomena are so important.

Like virtually all social scientists, we have, in many of our papers, made use of observational data, and so we must cope with limitations in extant methods. But we also have published a number of experiments, exploring both the structure and function of networks ranging in size from thousands to tens of millions of subjects [1–3]. And we note with interest the experimental work of others [4–6]. It is worth noting, however, that experiments have their own limitations, including that they are thinned-out versions of reality. What one gains in robust causal inference, one loses in verisimilitude. So, we think that both approaches will play an important role in network science in the coming years. We just need to invent better statistical tools, a sentiment in which we are joined by all four commentators.

Thomas, as usual, thinks deeply about the limitations of current network models. Yet, he does not suggest anything specific that one could do differently than we did, given our data and the current state of statistical knowledge. And, in many cases, he identifies limitations in current methods that we, and others, have previously pointed out. The challenge is what to do about them.

In his discussion, Thomas appears to overlook the fact that we ourselves, for most of our papers, have published results with both dichotomous and continuous versions of the same outcome – for example, dichotomous obesity and continuous body mass index [7, 8], dichotomous smoking and continuous number of cigarettes smoked per day [9], dichotomous and continuous versions of a happiness index [10], and dichotomous and continuous versions of a lonely days per week [11]. In all cases, the two types of models have led to the same conclusions, a fact we have noted in our papers. Moreover, we have also previously explored double-lagged models of exactly the sort he (and others) have suggested, where alter's change in status from t-2 to t-1 is used as a predictor for ego's change in status from t-1 to t, a fact mentioned in the supplements to our papers. And those analyses also confirmed our main results. This has been previously noted by VanderWeele [12].

Moreover, as Thomas notes, the use of other sorts of approaches, whether propensity score matching, for example, as implemented by Aral [13] (and with which we are quite familiar [14]), or SIENA models [15], do not by any means solve the generic critique that Shalizi and Thomas have offered in a prior paper regarding the difficulty of causal inference with observational network data [16]. This is true even with the implementation of further lags, although further lags do increase our confidence in causal interpretations and solve certain other statistical problems [12].

The simulation study involving network geometry that Thomas reports is a clever contribution. We are glad to see that Thomas availed himself of the public-use version of the FHS-Net data for this purpose, and we encourage others to explore that dataset. Still, it is unclear how to modify current models to account for this theoretical concern. And other analyses based on real data have confirmed that the models we used perform very well, despite various theoretical issues being raised here and elsewhere, as noted in our paper [17–19]. Incidentally, we are glad to see that Thomas confirms that our ‘symmetry test works exactly as advertised for the continuous model’ because we already had reported such results in our prior papers. Thomas’ concern that the sparseness of FHS-Net may contribute to some bias is mitigated by the fact that we and others have replicated our results in several other datasets, as we note in our paper.

In his featured article, O’Malley makes numerous points directly relevant to our paper. He provides a comprehensive overview of network statistics, correctly noting all the ways that inference regarding peer effects is so much more challenging within the context of network data. He correctly points out that transitivity among peers can potentially lead to bias in estimation, even if one implements, as we did, separate models for each type of peer (friends, spouses, siblings, etc.). However, the extent of this bias is likely to be small, and, to our knowledge, there are no extant models that can correct for this issue.

Some critical comments have centered around the need, in network analysis, to deploy assumptions that are (in any case) the same as in all other causal estimation with observational data; as O’Malley notes, this requires ‘alternative identification strategies’, such as the need to ‘make a multivariate parametric assumption, assume the non-existence of unmeasured confounding variables, or assume that an instrumental variable is valid’. We agree entirely with O’Malley that more work is needed to address the robustness of the currently available models to misspecification, especially because some misspecification is unavoidable, here as elsewhere.

VanderWeele’s clever contribution builds on his earlier work that shows our method is a valid test of the no-contagion hypothesis [12]. Here, he highlights the subtleties in assessing social ripples beyond one degree of separation. One of the basic horizons that social network data open up is that, if there is one degree of influence in a social process, then it may be possible to find evidence for influence at two and three (or more) degrees. It is not surprising that the effect decays with social distance (as we have argued before, this can be understood as a kind of ‘social friction’), but still, if one person affects another, then the latter might go on and pass the effect along. VanderWeele points out that the burgeoning literature on causal mediation may help us to deal with the additional challenges that arise when testing influence across multiple degrees of separation. Developing ways to assess the magnitude and meaning of such effects in an increasingly interconnected world is a topic of substantial interest to us and others [20, 21].

We thank Wasserman, Thomas, and O’Malley for recognizing that the development of the network data alone, using the FHS, was both difficult and valuable. We also thank all the commentators for recognizing the utility and novelty of the directionality test, which we hope will prove to be a fruitful area for investigation by statisticians, because it might provide a lever to have more confidence in causal estimates with observational data. Prior work by Shalizi and Thomas has already identified certain other assumptions this test may involve [16], and we hope that further exploration of the utility of this approach will be forthcoming. O’Malley and VanderWeele are also quite right that improved versions of the permutation test for topological clustering that we implemented would be extremely useful; these scholars are, in our judgment, among the most creative in contributing new methodological approaches to network data. In short, there is a pressing need for still more, and better, methods in the area of network statistics.

NICHOLAS A. CHRISTAKIS
Department of Sociology, Harvard University
180 Longwood Avenue, Boston, MA 02115, U.S.A.
E-mail: christak@hcp.med.harvard.edu

JAMES H. FOWLER
Division of Medical Genetics
University of California San Diego
School of Medicine, 9500 Gilman Drive
La Jolla, CA 92093, U.S.A.
E-mail: jhfowler@ucsd.edu

Statistics
in Medicine

References

1. Fowler JH, Christakis NA. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences* 2010; **107**:5334–5338.
2. Rand DG, Arbesman S, Christakis NA. Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences* 2011; **108**:19193–19198.
3. Bond RM, Fariss CJ, Jones JJ, Krames AD, Marlow C, Settle JE, Fowler JH. A 61-million-person experiment in social influence and political mobilization. *Nature* 2012; **489**:295–298.
4. Aral S, Walker D. Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Management Science* 2011; **57**:1623–1639.
5. Centola D. The spread of behavior in an online social network experiment. *Science* 2010; **329**:1194–1197.
6. Valente TW. Network interventions. *Science*; **337**:49–53.
7. Fowler JH, Christakis NA. Estimating peer effects on health in social networks: a response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. *Journal of Health Economics* 2008; **27**:1400–1405.
8. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 2007; **357**:370–379.
9. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *New England Journal of Medicine* 2008; **358**:2249–2258.
10. Fowler JH, Christakis N. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham heart study. *British Medical Journal* 2008; **337**:a2338.
11. Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology* 2009; **97**:977–991.
12. VanderWeele TJ, Ogburn EL, Tchetgen EJ. Why and when "flawed" social network analyses still yield valid tests of no contagion. *Statistics, Politics, and Policy* 2012; **3**:1–11.
13. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* 2009; **106**:21544–21549.
14. Christakis NA, Iwashyna TJ. The health impact on families of health care: a matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Social Science and Medicine* 2003; **57**:465–475.
15. Steglich C, Snijders TAB, Pearson M. Dynamic networks and behavior: separating selection from influence. *Sociological Methodology* 2010; **40**:329–393.
16. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research* 2011; **40**:211–239.
17. Iwashyna T, Gebremariam A, Hutchins M, Lee J. Can longitudinal GEE models distinguish network influence and homophily? An agent-based modeling approach to measurement characteristics, 2012. Under Review.
18. VanderWeele TJ. Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research* 2011; **40**:240–255.
19. Ver Steeg G, Galstyan A. Statistical tests for contagion in observational social network studies. http://arxiv.org/lab/1211.4889.
20. Christakis NA, Fowler JH. *Connected: The Surprising Power of our Social Networks and How They Shape Our Lives*. Little Brown and Co.: New York, 2009.
21. Landon BE, Keating NL, Barnett ML, Onnela JP, Paul S, O'Malley AJ, Keegan T, Christakis NA. Variation in patient-sharing networks of physicians across the United States. *Journal of the American Medical Association* 2012; **308**:265–273.

