Did an Ebola Outbreak Influence the 2014 U.S. Federal Elections? (Hint: Only if you ignore autocorrelation).

Leonid Tiokhin[1,2] ,Daniel Hruschka[1,2]

[1] Arizona State University School of Human Evolution and Social Change

[2] Arizona State University Center for Evolution & Medicine

ABSTRACT

In a recent paper, Beall, Hofer, and Schaller (2016) use observational time series data to test the hypothesis that the 2014 Ebola outbreak influenced the 2014 U.S. Federal Elections. They find substantial associations between online search volume for Ebola and people's tendency to vote Republican, an effect observed primarily in states with norms favoring Republican candidates. However, the analyses do not deal with the well-known problem of temporal autocorrelation in time series. We show that all variables analyzed exhibit extremely high levels of temporal autocorrelation (i.e. similarity in data-point values across time). After appropriately removing first-order autocorrelation, the observed relationships are attenuated and non-significant. This suggests that either no real associations exist, or that existing data are insufficiently powered to test the proposed hypotheses. We conclude by highlighting other pitfalls of observational data analysis, and draw attention to analytical strategies developed in related disciplines for avoiding these errors.

BODY

In a recent paper, (Beall, Hofer, & Schaller, 2016, BHS) use observational time series data to test the hypothesis that the 2014 Ebola outbreak influenced the 2014 U.S. Federal Elections. This represents one example of a returning interest in using observational data: 1) to assess long-term temporal predictions of psychological theories in naturalistic settings (Jebb, Tay, Wang, & Huang, 2015) and 2) to examine how psychological theories can predict cross-population variation in attitudes and behavior (Eppig, Fincher, & Thornhill, 2010; Fincher & Thornhill, 2012; Gelfand et al., 2011; Murray, Schaller, & Suedfeld, 2013; Schaller & Murray, 2008). While such non-experimental designs hold considerable promise, they also introduce analytic challenges that can lead to spurious inferences if left unaddressed (Hackman & Hruschka, 2013; Hruschka & Hackman, 2014; Hruschka & Henrich, 2013; Jebb et al., 2015; Pollet, Tybur, Frankenhuis, & Rickard, 2014). Such pitfalls include artificially deflated standard errors from pseudoreplication of observational units (Hruschka & Henrich, 2013), spurious associations in time series data (Yule, 1926), hidden confounding from inappropriately disaggregated data (Hackman & Hruschka, 2013), and extending inferences from group-level analyses to individual behavior (Pollet et al., 2014). Fortunately, these pitfalls are well-studied in economics, sociology, and anthropology, and researchers have developed numerous techniques to attenuate their threats to valid conclusions (Naroll, 1961; Yule, 1926). Over two decades ago in psychology, there was also some discussion of these issues related to time series data (Gergen & Gergen, 1987; McCleary & Welsh, 2015).

Here, we use the BHS analyses to illustrate how using observational data without attention to these issues can lead to spurious inferences. BHS use the coincidence of the Ebola epidemic and the 2014 U.S. elections to assess two hypotheses derived from theories of the behavioral immune

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2165v2 | CC BY 4.0 Open Access | rec: 28 Jun 2016, publ: 28 Jun 2016

system (Schaller & Murray 2008). First, they hypothesize that the perceived threat of disease should increase political conservatism**.** Second, they hypothesize that disease threats may cause increased conformism—a bandwagon effect, where "voters show an increased inclination to support whichever political candidate is leading in recent polls" (pg. 2). BHS assess these hypotheses by correlating time series of: 1) online searches of Ebola and 2) daily polling data for U.S. congressional elections, a month before and a month after the CDC's announcement of the first Ebola Case (September 30, 2014). They find strong correlations between Ebola search volume and support for Republican candidates at both the national and state level, interpreting this as support for the first hypothesis. They also find that correlations between Ebola and Republican support are strongest in states that started off with greater support of Republican candidates and with longstanding Republican voting norms. They interpret this as support for the bandwagon effect.

It has been known for nearly a century that correlations between time series without appropriate controls are prone to being spuriously high (Yule, 1926), because temporally adjacent observations are often highly correlated. A simple method for dealing with such threats is to analyze the changes between time points rather than the absolute values of the time series. This process of detrending removes first-order autocorrelation, and is often the first step in time series analysis (Jebb et al., 2015).

Here we apply this simple procedure and reanalyze the BHS time series (see Supplementary Materials, available online, for further details). We find exceedingly high levels of temporal autocorrelation in the time series variables ($r > 0.90$). This indicates detrending is a necessary first step (see Table S1 in Supplementary Materials). After detrending the data, our analyses show no empirical support for either of the two hypotheses (Table 1). At both national and state levels, there are no longer strong or significant associations between Ebola search volume and preference for Republican candidates. Moreover, there is no support for a moderating bandwagon effect: states leaning Republican in either current or past elections did not show correlations greater than zero or greater than the correlations observed in Democratic states (Table 1). These results are robust to the composition of the sample (including/excluding outliers; excluding/including six states with insufficient data on daily changes; see Supplementary Materials).

Table 1 – Comparison of results across three approaches: (1) original analyses, (2) analyses removing first-order autocorrelation, (3) original analyses with subsample used for first-order autocorrelation.

	Original analyses	**Detrended analysis**	**Original analysis with detrended sub-sample**
Correlations of Ebola Search Volume & Voter Intentions	**r**	**r**	**r**
National[a]	0.51*	0.30	NA
All States [b]	0.31*	0.04	0.19
Republican-Leading Elections [b]	0.51***	0.09	0.39*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2165v2 | CC BY 4.0 Open Access | rec: 28 Jun 2016, publ: 28 Jun 2016

Democratic-Leading Elections [b]	-0.08	-0.03	-0.08
Positive (Republican) PVI [b]	0.55***	0.13	0.43*
Negative (Democratic) PVI [b]	-0.12	-0.05	-0.12
Differences between correlations	**d**	**d**	**d**
Republican vs. Democratic-Leading Elections [b]	0.92*	0.24	0.70
Positive vs. Negative PVI states [b]	1.11**	0.37	0.85*

Table 1 lists correlations between the Ebola Search Volume Index (ESVI) and the Voter-Intention Index (VII) across all analyses. * Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$. [a] refers to Study 1. [b] refers to Study 2.

Given that the original BHS findings are not robust to basic time series controls, this strongly suggests that either: (1) these initial findings were spurious or (2) the study design used by BHS was insufficiently powered to detect any potential associations. The second is a possibility given that these time series are quite short by conventional standards (Jebb et al., 2015) and suggests that a better design is needed to test these hypotheses.

We have described one of the simpler tools—detrending to remove 1st order correlation—to deal with inferential threats that arise in the analysis of observational data. However, there are many other well-studied techniques for the analysis of observational data that are not only useful, but necessary, for avoiding common pitfalls. For time series data, these include modeling and removing higher-order trends and seasonality as well as other factors introducing temporal autocorrelation (Jebb et al., 2015). For cross-population comparisons that may be subject to pseudo-replication of units (e.g. Mississippi and Alabama may not really be independent observations in analyses across the fifty U.S. states) one can introduce controls for macro-regional variation (Hruschka & Henrich, 2013), conduct spatially autocorrelated regressions (Anselin & Bera, 1998), or remove cultural autocorrelation by looking at changes over cultural phylogenies (Mace & Holden, 2005). To deal with potentially unmeasured confounding variables that are particularly pernicious in observational data, there are fixed effects models for panel data (Allison, 2009) and instrumental variable analyses (Angrist, Imbens, & Rubin, 1996). Each of these has a rich literature that includes checks on the assumptions and appropriate implementation of these techniques to best avoid inferential threats introduced by these myriad issues.

Author Contributions

L. Tiokhin and D. Hruschka developed the study concept. L. Tiokhin and D. Hruschka contributed to the study design. L. Tiokhin conducted all data analyses. L. Tiokhin and D. Hruschka drafted, revised, and approved the final version of the manuscript.

References

Allison, P. D. (2009). *Fixed effects regression models* (Vol. 160). SAGE publications.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, *155*, 237–290.

Beall, A. T., Hofer, M. K., & Schaller, M. (2016). Infections and Elections Did an Ebola Outbreak Influence the 2014 US Federal Elections (and if so, How)? *Psychological Science*, *27*(5), 595–605.

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society of London B: Biological Sciences*, *277*(1701), 3801–3808.

Fincher, C. L., & Thornhill, R. (2012). Parasite-stress promotes in-group assortative sociality: The cases of strong family ties and heightened religiosity. *Behavioral and Brain Sciences*, *35*(02), 61–79.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., … others. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104.

Gergen, K. J., & Gergen, M. M. (1987). The self in temporal perspective. *Life-Span Perspectives and Social Psychology*, 121–137.

Hackman, J., & Hruschka, D. (2013). Fast life histories, not pathogens, account for state-level variation in homicide, child maltreatment, and family ties in the US. *Evolution and Human Behavior*, *34*(2), 118–124.

Hruschka, D. J., & Hackman, J. (2014). When are cross-group differences a product of a human behavioral immune system? *Evolutionary Behavioral Sciences*, *8*(4), 265.

Hruschka, D. J., & Henrich, J. (2013). Institutions, parasites and the persistence of in-group preferences. *PloS One*, *8*(5), e63642.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology*, *6*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4460302/

Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in Ecology & Evolution*, *20*(3), 116–121.

McCleary, R., & Welsh, W. N. (2015). Philosophical and statistical foundations of time-series experiments. *Single-Case Research Design and Analysis (Psychology Revivals): New Directions for Psychology and Education*, 41.

Murray, D. R., Schaller, M., & Suedfeld, P. (2013). Pathogens and politics: Further evidence that parasite prevalence predicts authoritarianism. *PloS One*, *8*(5), e62275.

Naroll, R. (1961). Two solutions to Galton's problem. *Philosophy of Science*, 15–39.

Pollet, T. V., Tybur, J. M., Frankenhuis, W. E., & Rickard, I. J. (2014). What can cross-cultural correlations teach us about human nature? *Human Nature*, *25*(3), 410–429.

Schaller, M., & Murray, D. R. (2008). Pathogens, personality, and culture: disease prevalence predicts worldwide variability in sociosexuality, extraversion, and openness to experience. *Journal of Personality and Social Psychology*, *95*(1), 212.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series?–a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, *89*(1), 1–63.