Tweet Sixteen and Pregnant: Missing Links in the Causal Chain from Reality TV to Fertility.

A replication study of Kearney & Levine (*American Economic Review*, 2015)

David A. Jaeger* Theodore J. Joyce† Robert Kaestner‡

International Journal for Re-Views in Empirical Economics, Volume 3, 2019-1, DOI: 10.18718/81781.10

JEL: J13, L82, C81
Keywords: social media, fertility, birth control, abortion, *16 and Pregnant*, event study, replication study

Data Availability: The data we used in this paper come mainly from the replication files made available by Kearney and Levine (KL) on the American Economic Review website. The zip-file can be downloaded at aeaweb.org. In addition, we use national-level data from Google from the file googletrends.dta, which was obtained directly from Phil Levine on 28 August 2016. This file is referenced in KL's replication files on the AER website, but not included in the downloadable zip file. The downloadable zip-file, Stata do and log files, and detailed instructions for the use of all data are available from the website of the journal www.iree.eu.

Please Cite As: Jaeger, David A. , Theodore J. Joyce, and Robert Kaestner (2019). Tweet Sixteen and Pregnant: Missing Links in the Causal Chain from Reality TV to Fertility. A replication study of Kearney & Levine (American Economic Review, 2015). *International Journal for Re-Views in Empirical Economics*, Vol 3(2019-1). DOI: 10.18718/81781.10

Abstract
We replicate and extend the analysis of the positive association between social media (Google searches and tweets) and the MTV program *16 and Pregnant* recently published by Melissa Kearney and Phillip Levine (2015). We find that the relationship disappears or even turns negative when we include in the analysis periods when new episodes of *16 and Pregnant* were not being broadcast. The results are also sensitive to the use of weights. Our results cast substantial doubt on social media as a link in the causal chain between reality television and fertility.

*CUNY Graduate Center, 365 Fifth Ave, New York, NY 10016 and University of Cologne, IZA, and NBER, e-mail: dj@djaeger.org

†Baruch College, CUNY, 55 Lexington Ave, New York, NY 10010 and NBER, e-mail: theodore.joyce@baruch.cuny.edu

‡University of Illinois, 815 West Van Buren Street, Suite 525, Chicago, IL 60607 and NBER, e-mail: kaestner@uic.edu.

Acknowledgements: We thank Phil Levine for providing data used in this analysis, and Nils Braakman and an anonymous referee for helpful comments.

Received August 8, 2018; Revised December 31, 2018; Accepted January 2, 2019; Published January 29, 2019.

1 Introduction

In a recent study of teen pregnancy in the *American Economic Review*, Melissa Kearney and Philip Levine (2015, henceforth KL) found that the MTV show *16 and Pregnant* was associated with a 4.3 percent decline in teen birth rates between July 2009 and December 2010. We have recently cast doubt on KL's interpretation of this association as causal (Jaeger, Joyce, and Kaestner 2018). In this paper we re-examine KL's evidence using data from Twitter and Google Trends in which they link the timing of broadcasts of *16 and Pregnant* to increased use of phrases related to birth control and abortion on these platforms, in an effort to establish a causal link between the show and fertility behavior.[1] The use of social media to support traditional econometric analyses is a major focus of their study and part of a growing trend in social science research (see, for example, Choi and Varian 2012 and Stephens-Davidowitz 2017).[2]

We first replicate KL's analysis of Google and Twitter data, which is limited to periods around the first broadcasts of episodes of *16 and Pregnant*. We show that these results are unaffected by correcting for minor mistakes KL made in the coding of broadcast dates. We then use the full sample of tweets available in the data KL posted on the *American Economic Review* website and re-evaluate their evidence. We continue to show that, not surprisingly, broadcasts of new episodes of *16 and Pregnant* are associated with increased tweets about and Google searches for the phrase "16 and Pregnant." In not limiting the periods in the analysis, however, we find that broadcasts of new episodes *did not* increase tweets about or Google searchers for terms related to birth control and abortion relative to the pre-*16 and Pregnant* period or relative to periods in which new episodes of the show were not being broadcast. In addition, we show that KL's results are sensitive to their choice of weights. We conclude that KL's social media analyses do not support the causal link to fertility behaviors that they claim.

2 Data

2.1 KL's Original Data

KL use data from Google Trends and Twitter to assess interest in *16 and Pregnant*, birth control, the contraceptive pill, and abortion. KL regress tweet rates and Google Trend indices on indicators for broadcasts of *16 and Pregnant* as well as the Google Trend index and tweets about *16 and Pregnant* in an attempt to establish a causal link from the show to potential behavioral changes that would plausibly explain the decline in teen birth rates that they claim was related to the show's broadcast. Most of the data used in our analysis was included in the files made available by the *American Economic Review* when KL was published. KL's "do" files refer to one data file that was not included in the replication package and we obtained this file directly from Phillip Levine.[3]

[1] KL write, "In all of these approaches using high frequency data, we believe that the results plausibly provide causal estimates of the impact of the show" (p. 3621).

[2] Kearney and Levine's description of the social media analysis as "secondary and suggestive" (Kearney and Levine 2016) and "peripheral" (Kearney and Levine 2018) is surprising, given that fully half of the data analysis in KL is devoted to social media. In Kearney and Levine (2014), which garnered much attention from the national press, they argue that the analysis of social media and its effect on attitudes was a primary contribution.

[3] We use the data files topsy-trend-daily.dta, googletrends-state.dta, topsy-state-daily.dta, topsy-state.dta from aeaweb.org. We acquired the file googletrends.dta directly from Phillip Levine on 28 August 2016. All files are included in our replication files along with the "do" file to generate all of the results in this paper. DOI: notyet.

KL purchased Twitter data from Topsy Labs, a social media company that was founded in 2007 and acquired by Apple in 2013 before closing in 2015. These data contain information about tweets regarding *16 and Pregnant*, birth control, abortion, and adoption, as well as the total number of tweets, on each day from 1 January 2009 to 31 December 2012. KL create daily rates (per 1 million overall tweets) mentioning “abortion,” “birth control,” or “16 and Pregnant.” They also provide data from Topsy that gathered the same information disaggregated by state.

KL use data from Google Trends to measure the relative frequency for searches for “16 and Pregnant,” “how get birth control,” “how get abortion,” and “how get birth control pill” at a weekly frequency over 209 weeks from January 2009 to December 2012. Google Trends creates an index of query shares, and queries can be specific to a time-period, geographic region or both. The index for a search of a specific query is constructed by dividing the number of searches for the specific term by the total number of searches conducted over the time-period or geography and then assigning a value to 100 to the time-period-geography with the largest share. All other period-geography units are normalized using the largest value. Thus, an index value of 0.5 for a period-geography search of “how to get abortion” is half as large as the share of the query that generated the largest share in that period and geographic unit.[4]

2.2 Corrections to KL’s Data

KL miscoded 11 of the 53 days on which a new episode of *16 and Pregnant* was broadcast.[5] Each miscoding was off by one day, and these obviously affect the lagged indicators for broadcast days as well as the characterization of periods as “in season” (i.e. the period between broadcast of the first show in the season and the last).[6] In addition, in the analysis where KL focus only on the days that are “in season,” they incorrectly programmed the variable indicating the day after a new episode of *16 and Pregnant* was broadcast.[7] Neither of these mistakes has a large impact on the qualitative interpretation of KL’s results, and we note for the record that we were able to replicate all of KL’s results using their (incorrect) dates and variable coding.[8]

[4]See KL pp. 3617-3619 for a more detailed description of their social media data.

[5]We checked broadcast dates from the official 16 and Pregnant website (mtv.com/shows/16-and-pregnant/episode-guide, last seen 7 August 2018) as well as the TV Guide sites for each season of the show that is relevant to KL’s analysis (tvguide.com/tvshows/16-pregnant/episodes-season-1/304110/, tvguide.com/tvshows/16-pregnant/episodes-season-2/304110/, and www.tvguide.com/tvshows/16-pregnant/episodes-season-3/304110/, for the first three seasons, respectively; last seen 7 August 2018).

[6]The corrected “in-season” periods are 11 June 2009 to 30 July 2009 (Season 1), 16 February 2010 to 20 April 2010 (Season 2 part 1), 26 October 2010 to 4 January 2011 (Season 2 part 2), 19 April 2011 to 28 June 2011 (Season 3), and 27 March 2012 to 5 June 2012 (Season 4). The first episode of season 5 of *16 and Pregnant* was not broadcast until 14 April 2014, outside of our analysis period.

[7]KL first dropped the observations that were not “in season” and then used the x[_n-1] construct in Stata to create the lagged variable. This works when each observation is temporally contiguous to the previous observation, but not when there are breaks in the time series.

[8]Because KL did not make available daily-level data across states for the state-level Twitter analysis in Table 4 below, we are not able to evaluate how the miscoding of the “in-season” periods affects the results. We use their dating of the 11 periods in this analysis.

3 Replication and Extensions

3.1 National Twitter Analysis

KL define tweet rates as the total number of tweets with the specified terms each day per 1 million total tweets made that day. They graphically show (KL Figure 8) that the daily time series of the tweet rate for "16 and Pregnant" has clear spikes on the day after a new episode of the show was broadcast, as well as elevated tweet activity during the weeks in which new episodes of *16 and Pregnant* was airing (KL Figure 7). In Figure 1, we provide a similar graph of daily tweet rates for "16 and Pregnant" in 2010, but also include the tweet rates for "birth control" (Panel A) and "abortion" (Panel B). Other years exhibit similar properties and are available from the authors by request. In both panels the vertical lines represent the broadcast dates of *16 and Pregnant*. While there are clear spikes in the tweet rate for "16 and Pregnant," we find no similar pattern for the tweet rate of "birth control" or "abortion."

KL formally test for an association between *16 and Pregnant* and tweets for "birth control" and "abortion," by regressing the natural logarithm of the tweet rate for each term on indicators for the day a new episode of *16 and Pregnant* was broadcast. KL limit their analysis to the 336 days during weeks in which a new episode of *16 and Pregnant* was broadcast, although their data include all 1,461 days between 1 January 2009 and 31 December 2012.

Limiting the periods of analysis in this way is potentially problematic for several reasons. First, re-runs of *16 and Pregnant* are shown throughout the year and at numerous times during the day, leading to potentially constant exposure to the show's messages since its inception. Second, the available Twitter data includes five months prior to any broadcasts of *16 and Pregnant*. KL omit these data from their analysis, even though they provide a useful baseline of tweeting of "birth control" and "abortion" in the months leading up to the show's debut. Lastly, including tweets from the periods when new episodes of *16 and Pregnant* are not broadcast is more comparable to their Twitter analyses across states and time, as well as to the Google Trend analyses that use all available weekly data from January 2009 to December 2012. KL provide little justification for difference in focus from the full period (using Google data and state$\times$time variation with tweets) to only the periods that are "in season" (using national-level Twitter data).[9] We feel that the more appropriate approach, and the one that KL themselves use when analyzing Google searches and tweets across space, is to include all periods and contrast the prevalence of tweets and searches about "birth control," "abortion," and the name of the show itself between those periods when new episodes are being broadcast and those periods prior to any broadcast of *16 and Pregnant* or when only re-runs are being shown.

We present our re-analysis of the association between *16 and Pregnant* and tweets about birth control and abortion in Table 1. In column (1) we show estimates of the association between the broadcast of new episodes on the tweet rate for "16 and Pregnant" during the whole period from 1

[9] KL state (p. 3620): "When we use Google Trends data, we consider the entire time period between January 2009, the beginning of the year in which the show began, and December 2012, and focus on weekly variation, distinguishing between the weeks in which a new episode was in season relative to other weeks of the year.... When we use Twitter data, we restrict our attention just to those weeks in which the show is 'in season' (which we listed previously) and take advantage of daily variation in outcomes."

January 2009 to 31 December 2012.[10] As in KL's model, we include indicators for days when a new episode was broadcast and the subsequent days. We also include an indicator for the pre-*16 and Pregnant* period and the days after *16 and Pregnant* began broadcasting that are "out of season," except for the day immediately succeeding a broadcast of a new episode. The reference category is therefore "in season" excluding days on which a new episode is broadcast and the day after. The bottom part of the table shows the difference between the "Day of" and "Day after" coefficients and the "Pre-*16 and Pregnant*" and "Out of Season" coefficients. Following KL, we also include a quadratic time trend. The regression is unweighted and we present Newey-West (1987) standard errors with one lag. As demonstrated in Figure 1, there is a huge increase in the "16 and Pregnant" tweet rate on the day of, but especially the day after, the broadcast of a new episode, relative to the pre-*16 and Pregnant* period as well as relative to the "out of season" periods.

In column (2) we replicate KL's results from their Table 3, column (4), using the tweet rate for "birth control" as the dependent variable, but correcting the coding of indicators for the date of a new episode and its lag. The differences between these results and KL's published results are trivial and we find a significant and positive relationship between new episodes and tweets about birth control.[11] In column (3) we estimate the same model but we correct for KL's miscoding of some days the show was broadcast. Correcting the dates of broadcasts reduces the "in-season" periods by 3 days, and the coefficient on the "day of" indicator is no longer statistically significantly different from zero. In both columns, following KL, we weight the regressions using the total number of tweets occurring on that day. In column (4) we do not weight, but the results are similar to those in column (3).

KL use "in season" days that are not the day of or day after the broadcast of a new episode as the reference category. But there are two other, potentially better, "control" periods available in the data that can be used as the baseline for tweets about birth control and abortion, namely the period before *16 and Pregnant* was first broadcast (the period between 1 January 2009 and 10 June 2009, inclusive) and the "out of season" periods after 11 June 2009 during which new episodes of the show were not being broadcast. In column (5), we estimate the same model as in column (1), with the log tweet rate for birth control as the dependent variable, using all of available data from 1 January 2009 to 31 December 2012. As in column (1), we include the "Pre-*16 and Pregnant*," and "Out of Season" indicators, maintaining the same reference category, "in season" days on which a new episode is broadcast or the day after, as in column (3) and KL's analysis. All four variables ("Day of New Episode," "Day after New Episode," "Pre-*16 and Pregnant*," and "Out of Season") are mutually exclusive and along with the reference category cover all days in the analysis.

We find in column (5) that tweets about birth control actually declined by about 45 percent relative to the pre-*16 and Pregnant* period.[12] Similarly, relative to the "out of season" days, tweeting about birth control declined by about 18 percent on the days a new episode was broadcast. We see no reason (and KL provide no justification) for preferring "in-season non-broadcast days" as the baseline for tweets about birth control and abortion. Recalling that the model includes a

[10] Days with zero tweets are dropped. These occur predominantly in the period before *16 and Pregnant* began broadcasting.

[11] Using KL's incorrect coding of this variable, we were able to exactly replicate their results: the coefficient on the "Day of" indicator is 0.1204 (standard error of 0.0468) and the coefficient on the "Day after" indicator is 0.2287 (standard error of 0.0579)

[12] $\exp(0.369) - 1 = 0.446$.

quadratic time trend, we argue that the cleanest reference categories are the pre-*16 and Pregnant* period followed by those periods in which new episodes of *16 and Pregnant* were not being broadcast. There is no evidence that broadcasts of *16 and Pregnant* increased twitter activity about birth control relative to the pre-*16 and Pregnant* period. We find essentially identical results in column (6) when we estimate the model without weighting by the total number of tweets but calculate the standard errors using the Newey-West procedure with one lag.

We repeat the analysis of columns (2) through (6) using the log tweet rate for abortion as the dependent variable in columns (7) through (11). As in column (2), we are able to replicate closely KL's result from Table 3, column (5).[13] The results here are quite analogous to those using the log tweet rate for birth control. In columns (10) and (11), we again find that, relative to both the pre-*16 and Pregnant* period and the "out of season" periods, tweeting about abortion decreases in response to broadcasts of the show.

It is worth noting here that although KL include audience viewing of *Teen Mom* and *Teen Mom 2* in defining the treatment in their analysis of the impact *16 and Pregnant* on birth rates, they include none of the broadcast dates in these shows in their social media analysis. Exclusion of these shows potentially induces measurement error in our (and KL's) analyses that use the full period of data available. This measurement error would be expected to bias the coefficient of the impact the broadcast of *16 and Pregnant* towards zero, however. Our finding of a statistically significant and negative coefficient relative to both the pre-*16 and Pregnant* period and the "out of season" periods is therefore all the more striking.

KL also correlate the tweet rates for "birth control" and "abortion" with the tweet rate for the phrase "16 and Pregnant," again limiting their analysis to the "in season" periods. We replicate their estimates in columns (1) and (5) of Table 2.[14] In both cases the coefficient on the tweet rate for "16 and Pregnant" is nearly identical to KL's estimates and positively associated with the tweet rate for "birth control" and "abortion." In columns (2) and (6), we re-estimate these models without weighting by the total number of tweets. The coefficients are smaller by a factor of three or more and are no longer statistically significant. KL (footnote 25) suggest that weighting serves to correct for heteroskedasticity, although they also employ robust standard errors to address the issue. Angrist and Pischke (2009) have raised questions about heteroskedasticity as a rationale for weighting when the form of heteroscedasticity is unknown. Solon, Haider, and Wooldridge (2015) have recently suggested that researchers compare weighted and unweighted estimates, and the differences between the weighted and unweighted results here suggests misspecification or heterogeneous effects. Weighting in this context seems dubious, at best.

Twitter became available for public use in July 2006 and subsequent growth was exponential. In Figure 2 we plot the weighting variable that KL use, the total number of tweets per day. The shaded areas indicate the "in-season" periods that KL analyze (and that correspond to columns (1), (2), (5), and (6) in Table 2. Differences between the weighted and unweighted results come from

[13] If we use their incorrectly-coded lagged day of broadcast indicator, we can exactly replicate the coefficients on "Day of New Episode" (0.142 with a standard error of 0.036) and "Day After New Episode" (0.212 with a standard error of 0.046).

[14] As in Table 1, we can exactly replicate the coefficients on log tweets about *16 and Pregnant* in the birth control regression (0.077 with a standard error of 0.034) and the abortion regression (0.064 with a standard error of 0.025) if we use KL's incorrectly-coded lagged broadcast indicator.

a 12-fold increase in the average daily volume of tweets from the first "in-season" period in 2009 to the last "in-season" period in 2012.[15] To the extent that KL find an effect in the weighted results, it is largely driven by the last two "in-season" periods in 2011 and 2012, the latter being after the period of their analysis of relationship between *16 and Pregnant* and fertility.

In columns (3) and (7), we re-estimate KL's model but include all available data between the show's inception on 11 June 11 2009 and 31 December 2012.[16] The coefficient on the tweet rate for "16 and Pregnant" in the birth control regression in column (3) is 60 percent smaller and marginally statistically significant and the coefficient on the tweet rate for "16 and Pregnant" in the abortion regression in column (7) becomes negative and statistically insignificant. In columns (4) and (8), we re-estimate the models from columns (3) and (6) without using weights and using the Newey-West (1987) procedure to correct for serial correlation and heteroskedasticity in the error terms. In neither case do we find a statistically significant association with the tweet rate for "16 and Pregnant." These results are consistent with the lack of a visual association between tweeting about *16 and Pregnant* and tweeting about birth control and abortion in Figure 1.

3.2 State-Level Twitter Analysis

KL also analyze the tweet rate for "birth control" and "abortion" at the state level over time, creating a panel dataset with 11 (unequal length) time periods from January 2009 to December 2012. These periods consist of the pre-*16 and Pregnant* period followed by alternating "in season" and "out of season" periods, as discussed in footnote 8. We present our exact replication of KL's results in Table 3. Our columns (1) and (3) replicate KL's results from their Table 4, Panel B, columns (4) and (5). When conditioning on period and time fixed effects, KL find a positive and statistically significant association between tweets about *16 and Pregnant* and birth control, but a negative and not statistically significant association with tweets about abortion.

As in all of their analyses using Twitter data, KL weight the observations by the total number of tweets in a state×period. In columns (2) and (4) of Table 3 we re-estimate the models from columns (1) and (3) without using weights. We find that the significant coefficient in the birth control regression is now 64 percent smaller and no longer statistically significant while the coefficient in the abortion regression remains not statistically significantly different from zero. As with the national-level analysis, the rationale for weighting seems unjustified, and may also exacerbate measurement error. KL (footnote 41) acknowledge that assigning a geographic location to a tweet is "a work in progress. . . and it is prudent to interpret our reported results using geographic Twitter data with some caution." Given the possibility of measurement error with respect to the location of tweets and the lack of any covariates besides the two-way fixed effects, the lack of robustness to weighting is not surprising. We conclude that there is little to infer about the causal link between *16 and Pregnant* and fertility behaviors from these results.

[15] The average volume of daily tweets in the five "in-season" periods from 2009 to 2012 are 5.08 million, 10.47 million, 12.56 million, 23.88 million, and 60.66 million, respectively.

[16] There are 1,294 days for this period but two days with a tweet rate of zero for "16 and Pregnant" are dropped.

3.3 National Google Analysis

We present our results using KL's national Google data in Table 4. In all columns we follow KL and present conventional standard errors in parentheses and do not use weights.[17] Although the Google search indices for "16 and Pregnant," "How get birth control", "How get birth control pill," and "How get abortion" are uninterrupted time series, KL do not correct for serial correlation in the disturbances. We present Newey-West (1987) standard errors (employing one lag) in square brackets. As above with our analysis of national Twitter trends, in column (1) we first show that new episodes of *16 and Pregnant* strongly predict Google searches for "16 and Pregnant." A week in which a new episode is broadcast is associated with a 38-point increase in the Google Search index for "16 and Pregnant" relative to weeks without a new episode.

The remaining columns replicate results from KL's Tables 3 and 4. Results on the association between weeks with a new episode in columns (2), (4), and (6) replicate KL's Table 3, columns (1) through (3), respectively, and results on the association between searches for "16 and Pregnant" in columns (3), (5), and (7) replicate KL's Table 4, Panel A, columns (1) through (3), respectively. We are able to replicate their results exactly. Both new broadcasts of *16 and Pregnant* and searches for "16 and Pregnant" are unrelated to searches for "how get the birth control," how get birth control pill," or "how to get abortion." In each case the Newey-West standard errors are larger than the unadjusted ones resulting in t-ratios that are less than one for each coefficient. Like our results using Twitter, including the full time series of information from January 2009 to December 2012 yields no association between Google search activity on behavioral terms ("birth control" and "abortion") and measures of interest in *16 and Pregnant*.[18]

4 Conclusion

KL's analyses of social media data played an important role in their argument that *16 and Pregnant* lowered teen birth rates. By linking phrases associated with preventing or terminating pregnancy with the timing of broadcasts of *16 and Pregnant*, KL attempted to provide evidence for a potential causal change from reality TV to fertility. This evidence was crucial for their argument because KL were evaluating a point-in-time national policy change (see footnote 2). We have shown elsewhere (Jaeger, Joyce, and Kaestner 2018), however, that a causal interpretation of KL's results based solely on their identification strategy is implausible.

Our reassessment of KL's social media analysis shows that these results, too, are extremely fragile and, at best, inconclusive. Although we were able to replicate their published results exactly, and although coding and data gathering mistakes by KL had relatively minor impacts on their re-

[17] Despite reporting robust standard errors in the Twitter analysis, KL report conventional standard errors when analyzing national Google search indices.

[18] We were also able to replicate KL's state-level Google search results in their Table 4, Panel B, columns (1) and (3). We view these results as being uninformative about the relationship between Google searches for "16 and Pregnant" and those for "How to get birth control" and "How to get abortion". The regressions are rudimentary: a pre-/post-treatment analysis with 12 or 15 states, two broad time periods (January 2005–May 2009 and June 2009–December 2010), and no comparison group during a period of rapid growth in Internet activity. We note for the record, however, that although KL's equation (9) includes the state- and period-specific unemployment rate, their replication programs did not include this variable in either the state-level Twitter or Google analyses, and we were able to replicate their results exactly without including the unemployment rate in any analysis.

sults, their choice of reference period for the Twitter results is questionable. When we expand their analysis to cover all of the available data, it is clear that there is little or no positive association between *16 and Pregnant* and tweets about birth control and abortion. KL's original research showed little evidence that Google searches related to birth control and abortion are related to measures of exposure to the or searching for "16 and Pregnant" and we confirm those results.

The fragility of KL's results raises general questions about using social media data. The potential for data mining and selective presentation of results in such a rich data environment is great, and clear justification should be given for selecting a limited number of search phrases and subjective periods to search. Pre-analysis plans should be used to minimize the appearance of "cherry picking" results that support authors' claims.

References

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics.* Princeton: Princeton University Press.

Choi, Hyunyoung and Hal Varian (2012). "Predicting the Present with Google Trends." *Economic Record* 88(S1): 2–9.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner (2018). "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?" *Journal of Business and Economic Statistics*, forthcoming. DOI: 10.1080/07350015.2018.1497510.

Kearney, Melissa S., and Phillip B. Levine (2014). "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." National Bureau of Economic Research. Working Paper 19795. nber.org/papers/w19795.

Kearney, Melissa S. and Phillip. B. Levine (2015). "Media Influences on Social Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing." *American Economic Review* 105(12), 3597–3632. DOI: DOI:10.1257/aer.20140012.

Kearney, Melissa S. and Phillip. B. Levine (2016). "Does Reality TV Induce Real Effects? A Response to Jaeger, Joyce, and Kaestner (2016)." IZA. Discussion Paper 10318. iza.org/dp10318.pdf.

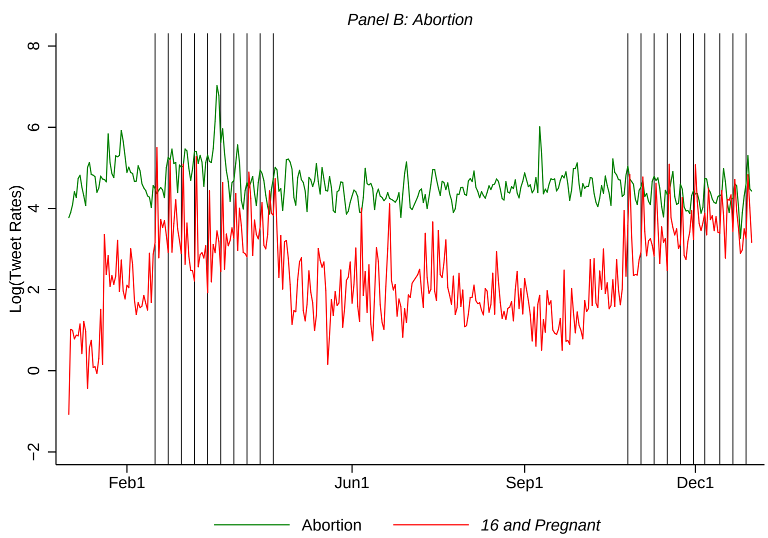
Kearney, Melissa S. and Phillip B. Levine (2018). "A Comment on Jaeger, Joyce, and Kaestner (2018) *Did Reality TV Really Cause a Decline in Teenage Childbearing?*" Mimeo, NBER. nber.org/data-appendix/w19795/KL_Response_to_JJK-JBES-July_2018_FINAL.pdf.

Newey, Whitney K. and West, Kenneth D. (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703–708. DOI: 10.2307/1913610.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge (2015). "What Are We Weighting For?" *Journal of Human Resources* 50(2): 301–316. DOI: 10.3368/jhr.50.2.301.

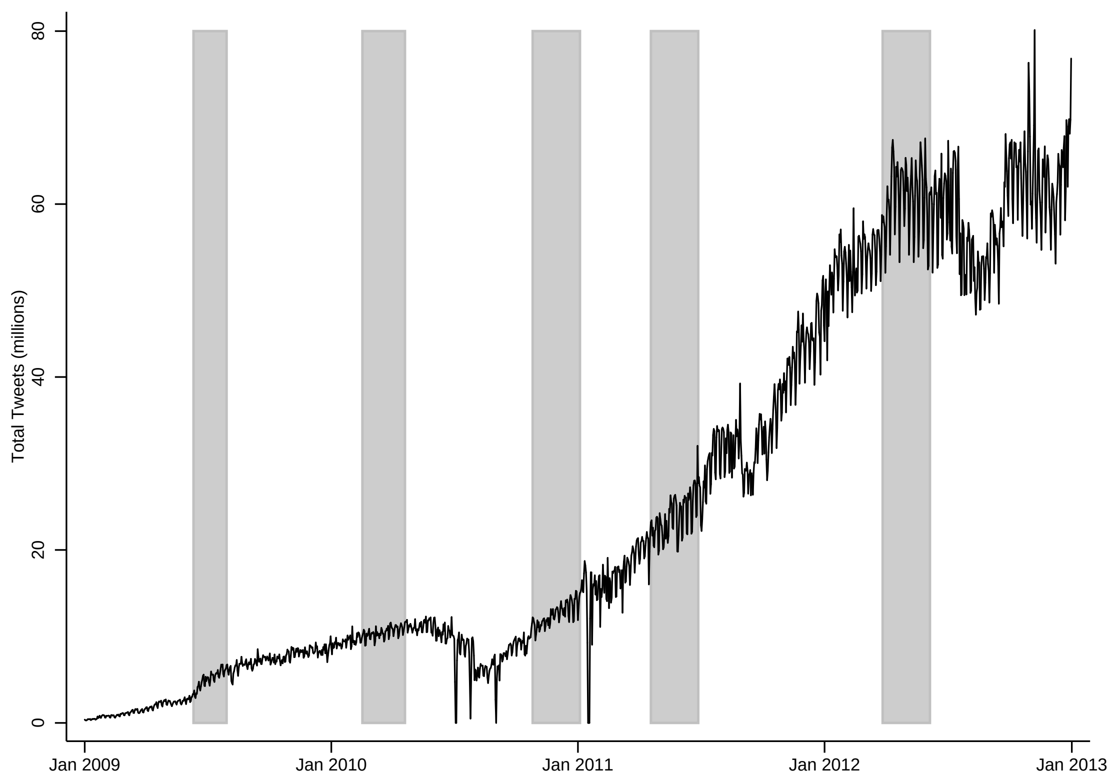
Stephens-Davidowitz, Seth (2017) *Everybody Lies.* New York: Harper Collins.

Figure 1: Twitter Activity for *16 and Pregnant*, Birth Control, and Abortion
Season 2 (2010)



Note: These graphs plot the log tweet rate for *16 and Pregnant*, birth control, and abortion for the second season (2010) of *16 and Pregnant*. The vertical lines represent the days of broadcasts of *16 and Pregnant*. Note that the vertical scale is different in each panel.

Figure 2: Total Number of Tweets
2009–2012



Note: This graph shows the total number of daily tweets from 2008 to 2012, which KL use as weights in the national-level Twitter regressions shown in Table 2. The shaded areas indicate ”in season” periods for *16 and Pregnant*, using the corrected dates as discussed in the text.

Table 1: Association between New Episodes of 16 and Pregnant and Twitter Activity about Birth Control and Abortion

Dependent Variable:	Log Tweet Rate: 16 and Pregnant	Log Tweet Rate: Birth Control					Log Tweet Rate: Abortion				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Coefficients											
"In Season" Excluding Day of and Day After New Episode	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Day of New Episode	−0.233**	0.140***	0.042	0.037	0.044	0.040	0.159***	0.090**	0.168***	0.095**	0.172***
	(0.112)	(0.049)	(0.057)	(0.045)	(0.057)	(0.045)	(0.039)	(0.040)	(0.048)	(0.048)	(0.053)
Day After New Episode	1.253***	0.240***	0.202***	0.155***	0.197***	0.158***	0.216***	0.140***	0.170***	0.158***	0.184***
	(0.148)	(0.062)	(0.062)	(0.051)	(0.057)	(0.051)	(0.049)	(0.047)	(0.058)	(0.049)	(0.057)
Pre-16 and Pregnant (before 11 June 2009)	−1.016***				0.414***	0.309***				0.326***	0.418***
	(0.284)				(0.079)	(0.065)				(0.088)	(0.100)
"Out of Season" Excluding Day After New Episode and Pre-*16 and Pregnant* Period	−1.268***				0.207***	0.137***				0.197***	0.153***
	(0.080)				(0.035)	(0.034)				(0.030)	(0.041)
Differences in Coeffcients											
Day of New Episode - Pre-*16 and Pregnant*	0.783***				−0.369***	−0.269***				−0.232**	−0.246**
	(0.286)				(0.090)	(0.070)				(0.094)	(0.101)
Day after New Episode - Pre-*16 and Pregnant*	2.269***				−0.217**	−0.151**				−0.168**	−0.234**
	(0.313)				(0.089)	(0.077)				(0.095)	(0.104)
Day of New Episode - "Out of Season"	1.035***				−0.163***	−0.097**				−0.102**	0.019
	(0.104)				(0.055)	(0.041)				(0.044)	(0.047)
Day after New Episode "Out of Season"	2.521***				−0.010	0.021				−0.039	0.031
	(0.146)				(0.054)	(0.047)				(0.045)	(0.052)
Include Only "In-Season" Days		x	x	x			x	x	x		
Weighted by Total Number of Tweets		x	x		x		x	x		x	
Number of Days	1,322	336	333	333	1,455	1,455	336	333	333	1,455	1,455

Note: Each column is from a separate regression. The period of analysis is 1 January 2009 to 31 December 2012, with restrictions as noted in the table. All regressions include a quadratic trend. Estimates from columns (2) and (7) replicate those from Kearney and Levine Table 3, columns (4) and (5), respectively, except that the lagged indicator for a new episode is correctly created (see text). Estimates in columns (3)–(6) and (8)–(11) also correct Kearney and Levine's dates of broadcast for *16 and Pregnant*. Estimates in columns (1), (5), (6), (10), and (11) use all of the available data from the replication files provided by Kearney and Levine to the *American Economic Review* website. The pre-*16 and Pregnant* indicator is equal to one for all dates from 1 January 2009 to 10 June 2009. *16 and Pregnant* began broadcasting on 11 June 2009. There are 1,461 possible days, but following Kearney and Levine, we drop those where the tweet rate is zero. All models estimated by OLS. Heteroskedasticity-consistent standard errors are shown in columns (2) through (7); Newey-West standard errors with one lag are shown in columns (1), (6), and (11). * indicates statistical significance at the 10 percent level, ** indicates statistical significance at the 5 percent level, and *** indicates statistical significance at the 1 percent level.

Table 2: Association between Twitter Activity about *16 and Pregnant* and Twitter Activity about Birth Control and Abortion

Dependent Variable:	Log Tweet Rate: Birth Control				Log Tweet Rate: Abortion			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Tweet Rate about *16 and Pregnant*)	0.077**	0.028	0.034*	0.010	0.060**	0.005	−0.022	−0.019
	(0.034)	(0.026)	(0.018)	(0.011)	(0.025)	(0.030)	(0.018)	(0.016)
Include Only "In-Season" Days	x	x			x	x		
Weighted by Total Number of Tweets	x		x		x		x	
Number of Days	333	333	1,292	1,292	333	333	1,292	1,292

Note: Each column is from a separate regression. The period of analysis is 11 June 2009 to 31 December 2012, with restrictions as noted in the table. All regressions include a quadratic trend. Estimates from columns (1) and (5) replicate those from Kearney and Levine Table 4, Panel A, columns (4) and (5), respectively, except that we correct the dates that are included as "in season". Estimates in columns (3), (4), (7), and (8) use all of the available data from the replication files provided by Kearney and Levine to the *American Economic Review* website for the period after *16 and Pregnant* began broadcasting. There are 1,461 possible days, but following Kearney and Levine, five are dropped because the tweet rate is zero. All models estimated by OLS. Heteroskedasticity-consistent standard errors are shown in columns (1)–(3), and (5)–(7); Newey-West standard errors with one lag are shown in columns (4) and (8). * indicates statistical significance at the 10 percent level, ** indicates statistical significance at the 5 percent level, and *** indicates statistical significance at the 1 percent level.

Table 3: Association between Twitter Activity about *16 and Pregnant* and Twitter Activity about Birth Control and Abortion at the State Level

Dependent Variable:	Log Tweet Rate: Birth Control		Log Tweet Rate: Abortion	
	(1)	(2)	(3)	(4)
Log(Tweet Rate for *16 and Pregnant*)	0.137**	0.049	−0.087	0.011
	(0.054)	(0.048)	(0.075)	(0.041)
State Fixed Effects	x	x	x	x
Period Fixed Effects	x	x	x	x
Weight	Total Tweets	None	Total Tweets	None
N	537	537	537	537

Note: Each column is from a separate regression. The period of analysis is January 2009 to December 2012. There are eleven (unequal length) time periods corresponding to the pre-16 and Pregnant period followed by alternating "in season" and "out of season" periods; see footnote 8. Estimates from columns (1) and (3) replicate Kearney and Levine Table 4, Panel B columns (4) and (5), respectively. All models estimated by OLS. Heteroskedasticity-consistent standard errors clustered at the state level in parentheses. * indicates statistical significance at the 10 percent level, ** indicates statistical significance at the 5 percent level, and *** indicates statistical significance at the 1 percent level.

Table 4: Association between New Episodes of *16 and Pregnant* and Google Searches for Birth Control, the Contraceptive Pill, and Abortion

Dependent Variable:	Google Index: *16 and Pregnant*	Google Index: Birth Control		Google Index: Pill		Google Index: Abortion	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Week of New Episode	38.954***	0.825		2.227		−1.627	
	(1.975)	(1.216)		(2.010)		(1.730)	
	[2.458]	[1.497]		[2.385]		[2.167]	
Google Index for *16 and Pregnant*			0.012		0.069		−0.074**
			(0.026)		(0.042)		(0.036)
			[0.030]		[0.048]		[0.053]
Number of Weeks	209	209	209	209	209	209	209

Note: Each column is from a separate regression. The period of analysis is January 2009 to December 2012. All regressions include a quadratic trend. Estimates from columns (2), (4), and (6) replicate Kearney and Levine Table 3, columns (1), (2), and (3), respectively. Estimates from columns (3), (5), and (7) replicate Kearney and Levine Table 4, Panel A, columns (1), (2), and (3), respectively. All models estimated by OLS. Following Kearney and Levine, conventional standard errors are shown in parentheses. Newey-West standard errors with 1 lag are shown in square brackets. * indicates statistical significance using conventional standard errors at the 10 percent level, ** indicates statistical significance at the 5 percent level, and *** indicates statistical significance at the 1 percent level.