DOES REALITY TV INDUCE REAL EFFECTS? ON THE QUESTIONABLE ASSOCIATION BETWEEN *16 AND PREGNANT* AND TEENAGE CHILDBEARING

David A. Jaeger
Theodore J. Joyce
Robert Kaestner

November 2016

Acknowledgements: The authors thank Phil Levine for his generous assistance in replicating results, and Bill Evans, Gary Solon, and seminar participants at Columbia University, the CUNY Graduate Center, the CUNY Institute for Demographic Research, the Guttmacher Institute, Princeton University, and the University of Michigan for helpful comments. Onur Altindag provided exemplary research assistance. Joyce and Kaestner acknowledge support from the National Institute of Child Health and Human Development Grant R01 HD082133 and from the Health Economics Program of the National Bureau of Economic Research. Nielsen ratings data are proprietary and cannot be shared. Birth data from the National Center for Health Statistics (NCHS) have confidential elements and must be obtained directly from NCHS. Institutional Review Board (IRB) approval was obtained from the National Bureau of Economic Research.

Does Reality TV Induce Real Effects?
On The Questionable Association between
16 And Pregnant and Teenage Childbearing
November 2016
JEL No. J13, L82

ABSTRACT

We reassess evidence that the MTV program *16 and Pregnant* played a major role in reducing teen birth rates in the U.S. (Kearney and Levine, *American Economic Review* 2015). We find Kearney and Levine's identification strategy to be problematic. Through a series of placebo and other tests, we show that the assumptions necessary for a causal interpretation of their results are not met. We also reassess their evidence from social media and show that it is fragile and highly sensitive. We conclude that Kearney and Levine's results are uninformative about the effect of *16 and Pregnant* on teen birth rates.

David A. Jaeger
Ph.D. Program in Economics
CUNY Graduate Center
365 Fifth Ave
New York, NY 10016
and University of Cologne
and IZA
and NBER
djaeger@gc.cuny.edu

Robert Kaestner
Department of Economics
University of California, Riverside
900 University Ave
Riverside, CA 92521
and NBER
robert.kaestner@ucr.edu

Ted Joyce
Baruch College, CUNY
Department of Economics and Finance
55 Lexington Ave
New York, NY 10010
and NBER
theodore.joyce@baruch.cuny.edu

The birth rate among 15 to 19 year olds in the United States was 61.1 per 1,000 teens in 1991, the highest rate since 1971. The rate among black teens was 114.8 in 1991. Many policy makers viewed such high rates as a national crisis (National Research Council 1987). Since 1991, however, the overall teen birth rate has fallen almost continuously, declining by more than half to 24.2 in 2014. Over the same period, the birth rate among African American teens has fallen by nearly two thirds, to 35.1. While these are substantial changes, the teen birth rate in the U.S. is still higher than that in Canada and much of Western Europe, and a consensus as to the causes of the decline remains elusive.[1] In an effort to understand which policies and interventions might prove effective at helping teens avoid an unintended pregnancy, in 2010 the U.S. Office of Adolescent Health awarded 100 million dollars in grants to fund 102 Teen Pregnancy Prevention (TPP) projects. Each project was required to have a rigorous evaluation plan and over 90 percent used randomized designs. A recently released summary of results from these evaluations revealed that few of the interventions were effective at changing teen behavior with respect to sexual activity, contraceptive use, and pregnancy.[2]

The widely-held belief that teen pregnancy has adverse consequences for young mothers and the failure of the TPP initiative to demonstrate a clearly

---

[1] There are a variety of possible explanations (Boonstra 2014), including more effective contraception (John S. Santelli, *et al.* 2007; Laura Lindberg, John Santelli and Shelia Desai forthcoming), welfare reform (Robert Kaestner, Sanders Korenman and June O'Neill 2003; Leonard M. Lopoo and Thomas Delaire 2006), labor market conditions (Rajeev Dehejia and Adriana Llearas-Muney 2004, and Elizabeth Olitmans Ananat, Anna Gassman-Pines and Christina Gibson-Davis 2013), and income inequality (Kearney and Levine 2014b). None of these explanations appear to be definitive.

[2] See http://www.hhs.gov/ash/oah/oah-initiatives/evaluation/grantee-led-evaluation/summary-researchdemonstration.pdf; http://www.hhs.gov/ash/oah/oah-initiatives/evaluation/grantee-led-evaluation/summary-ebps.pdf (last accessed 15 September 2016).

effective intervention to prevent teen pregnancy may explain why a recent study that claimed a reality TV show caused a significant decline in teen birth rates garnered immediate and widespread attention in the national media.[3] The study by Melissa S. Kearney and Philip B. Levine (2015c, henceforth KL) concludes that the MTV reality shows *16 and Pregnant, Teen Mom*, and *Teen Mom 2*, by dramatizing the challenges of pregnancy and childrearing, caused a 4.3 percent drop in teen birth rates between July 2009 and December 2010. This effect is so large that it accounts for a quarter of the total reduction in teen childbearing during this period. KL assert that their estimates represent a causal impact and claim that "…a social media campaign in the guise of a very popular reality TV show… adds a new 'policy mix'" (p. 3598) to current interventions designed to reduce teen pregnancy.

In this paper, we reassess the basic hypothesis and empirical foundations of KL's conclusions. At the heart of KL's paper is an identification strategy that is, at best, problematic. *16 and Pregnant* began broadcasting nationally in June 2009, providing a pre- and post-treatment analysis, but no clear counterfactual.

---

[3] For the initial reaction to the working paper by Kearney and Levine (2014) see:
CNN: http://www.cnn.com/2014/01/13/health/16-pregnant-teens-childbirth/;
The Washington Post: https://www.washingtonpost.com/news/arts-and-entertainment/wp/2014/04/09/how-mtvs-16-and-pregnant-led-to-declining-teen-birth-rates/;
Time: http://time.com/825/does-16-and-pregnant-prevent-or-promote-teen-pregnancy/;
Newsweek: http://www.newsweek.com/why-teen-birth-rate-keeps-dropping-333946;
and *The New York Times*: http://www.nytimes.com/2014/01/13/business/media/mtvs-16-and-pregnant-derided-by-some-may-resonate-as-a-cautionary-tale.html (All last accessed 15 January 2016). For the continuing popularity of *16 and Pregnant* as an explanation for the decline in teen birth rates, see http://www.nytimes.com/2016/07/19/opinion/winning-the-campaign-to-curb-teen-pregnancy.html (Last accessed 24 August 2016).

To overcome this limitation, KL use the audience share watching MTV in the period before *16 and Pregnant* was aired as an instrument for watching *16 and Pregnant*. Both measures of viewership reflect choices, however, and even KL concede that their instrument does not provide a random source of identifying variation in the viewing of *16 and Pregnant*.[4]

KL's research design is essentially a difference-in-differences approach. We show that the assumption of common trends in birth rates in areas with low and high MTV viewership necessary for identifying a causal effect is not met. Placebo analyses using only the pre-*16 and Pregnant* period yield significant results similar to those of KL, suggesting the existence of omitted variables associated with the pre-*16 and Pregnant* trends in both teen birth rates and MTV viewership. We confirm this confounding with non-parametric tests of common trends in teen birth rates by quartiles of MTV viewing in the period prior to *16 and Pregnant* that decisively reject the null of no difference. These pre-period trends in teen birth rates become visually apparent when we replicate and extend KL's "event study" analysis. In addition to examining teen birth rates, KL devote half their empirical analyses to the association between broadcasts of *16 and Pregnant* and Google searches and Twitter tweets for birth control and abortion. KL present these results as evidence of the causal chain that links *16 and Pregnant* to behaviors that lead to lower birth rates. We demonstrate that these social media analyses are so fragile that nothing can be inferred from them.

We conclude that causal claims about the effect of *16 and Pregnant* on teen birth rates are unwarranted. More generally, we show the importance of clearly and robustly establishing that the parallel trends assumption holds in difference-in-difference analyses that use geographic variation for identification.

[4] KL write in footnote 29 (p. 3610): "To be clear, this approach is not designed to identify the effect of assigning a random teenager or young adult to watch *16 and Pregnant*. Individuals select into MTV viewership in all periods."

Our results also underscore the great and greatly underappreciated importance of replication in the social sciences, especially for results that receive widespread and uncritical attention in the national media, potentially leading unproven strategies becoming part of the "policy mix."

I. Identification in Previous Studies of the Impact of Television on Behavior

The previous literature has examined the impact of the introduction of television or specific television content on a variety of outcomes, such as voter turnout (Matthew Gentzkow 2006), tests scores (Gentzkow and Jesse M. Shapiro 2008), women's empowerment (Robert Jensen and Emily Oster 2009); voting behavior (Stefano DellaVigna and Ethan Kaplan 2007); and fertility and divorce (Eliana La Ferrara, Alberto Chong, and Suzanne Duryea 2012). In each of these studies, the authors acknowledge the endogeneity of television viewing and address this problem using plausibly exogenous changes across both space and time in access to television or specific content in a reduced form model and/or in an instrumental variables approach. Indeed, Kearney and Levine (2015b) themselves follow a similar approach in their recent study on the effect of *Sesame Street* broadcasts on long-term educational outcomes. They use the distance to a UHF or VHF television tower and the introduction of *Sesame Street* in different places at different times as sources of plausibly exogenous variation, arguing that that historical FCC decisions regarding the placement of television towers are plausibly unrelated to the characteristics of children who could potentially watch *Sesame Street*.

Unlike these studies, KL do not exploit a plausibly exogenous source of variation in viewership of *16 and Pregnant* to obtain estimates of the show's effect on teen fertility. *16 and Pregnant* started to be broadcast everywhere in the U.S. at a single point in time, providing no clear control group (i.e. where *16 and Pregnant* was not available) to serve as a counterfactual. To address this issue, KL use exposure to *16 and Pregnant*, measured by the Nielsen ratings for the show, and compare areas with greater and smaller viewership. The authors recognize that *16 and Pregnant* viewership is unlikely to be exogenous and, therefore instrument for it using Nielsen ratings for MTV viewership in the period directly preceding the beginning of *16 and Pregnant*. MTV viewership is also unlikely to be exogenous, however, because it is based on socioeconomic and other differences across places. Although KL control for area fixed effects, their instrumental variables approach is vulnerable to time-varying confounders and lacks a strong theoretical basis for its validity.

II. KL's Empirical Framework

KL use an instrumental variables approach to obtain estimates of the effect of *16 and Pregnant* viewership on teen fertility, as described by the following four equations, which we reproduce here:

(1) $$\ln(BR_{jt}) = \beta_0 + \beta_1(Rate16P_j \times Post_t) + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt},$$

(2) $$\ln(BR_{jt}) = \beta_0 + \beta_1\widehat{(Rate16P_j \times Post_t)} + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt},$$

(3)

$$(Rate16P_j \times Post_t) = \beta_0 + \beta_1(MTV0809_j \times Post_t) + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt},$$

and

(4) $$\ln(BR_{jt}) = \beta_0 + \beta_1(MTV0809_j \times Post_t) + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt}.$$

We follow KL's notation, although in actuality the parameters represent different quantities across the different equations, i.e. β_1 in equation (3) is not the same quantity as β_1 in equation (4). In the equations, BR_{jt} is the birth rate by calendar quarter of conception t, in Designated Market Area (DMA) j; $Rate16P_j$ is the average Nielsen rating in DMA j for the shows *16 and Pregnant*, *Teen Mom,* and *Teen Mom 2* (henceforth *16 and Pregnant*) for 12 to 24 year-old viewers for the seven Nielsen sweep months of July 2009, February 2010, July 2010, and November 2010, May 2011, November 2011, and May 2012; $Post_j$ is a dichotomous indicator for the period after the introduction of *16 and Pregnant*; *MTV*0809 is the average Nielsen rating in for 12 through 24 year olds in DMA_j for all MTV shows in the sweeps months of July 2008, November 2008, February 2009, and May 2009, U_{jy} is the annual unemployment rate in DMA j in year y; $\mathbf{X}_{jy}$ is vector that includes the percent population in the DMA that is that non-Hispanic black, and the percent that is Hispanic in calendar year y; θ_t is a set of quarterly dummy variables; and δ_{js} is a full set of DMA × season fixed effects, which implicitly defines DMA fixed effects.

Equation (1) represents the equation of interest that yields estimates of the association between *16 and Pregnant* viewership and teen birth rates. Equation (3) is the first stage regression with ($MTV0809_j \times Post_t$) as an instrument. Equation (2) is the second stage regression and identical to equation (1) except that predicted *16 and Pregnant* viewership derived from equation (2) is used instead of actual *16 and Pregnant* viewership. Equation (4) represents the reduced form effect of MTV viewership on teen birth rates.

KL's study period includes 24 quarters (2005:QI – 2010:QIV) for 205 DMAs for a total (potential) sample of 4,920 observations.[5] Ratings for *16 and Pregnant* are measured during the time slot from 9:00 to 10:00 pm for 12 to 24

[5] The dependent variable is missing in some DMAs in some quarters because there are no teen births. This occurs more frequently in the analyses stratified by age or race/ethnicity.

year-old viewers on Tuesdays in the sweep months.[6] KL average these Tuesday ratings within each month and then average the four months of ratings and assign that value to the six post-*16 and Pregnant* quarters within each DMA. They follow the same procedure for the MTV ratings during the period 2008:QIII to 2009:QII by using the average for the four quarters within each DMA. The measures that KL use for *16 and Pregnant* viewership and MTV viewership therefore vary only in the cross section. Note that because the models also include DMA fixed effects, which absorb the time-invariant determinants of fertility (in equations (1), (2), and (4)) or of *16 and Pregnant viewership* (in equation (3)), the main effect of *16 and Pregnant* cannot be estimated in equations (1), (2), and (4), nor can the main effect of MTV be included in equation (3). Rather, only the interactions between $Rate16P_j$ and $MTV0809_j$ with $Post_t$ can be included on the right hand side of the relevant regressions.

[6] KL use ratings for a broad age range from a single time in the later evening (9:00-10:00 p.m.) when a new episode of *16 and Pregnant* was broadcast. Using this as a measure of the total exposure to *16 and Pregnant* for teenagers has some obvious problems. Previous episodes are broadcast at various times on the same day as the new episode, which is also rebroadcast at different times during the week. KL could have used what Nielsen terms "Gross Rating Percentages," which averages the percent of households watching any broadcast of *16 and Pregnant* at any time during the sweep month. This double counts any individual who watches the same show more than once, but is perhaps a better measure of exposure. Nielsen also collects separate ratings for teens aged 12-17 separately from young women 18 to 24 years old. The younger teens are more likely to watch *16 and Pregnant* earlier in the day than women 18 to 24. By aggregating ratings across age groups, KL potentially reduce noise in the data, but at the expense of accurately measuring exposure. KL also average Nielsen ratings from seven sweep months drawn from 2009-2012 and assign this average to conceptions from 2009:III to 2010:QIV, leading to another source of mis-measurement of exposure to *16 and Pregnant*. Teens who conceived in August of 2009, for example, are assigned a rating that pertains to future broadcasts to which they could not have been exposed. Ratings almost tripled in the first year suggesting important variation by quarter of conception.

The identification strategy implied by equations (1) through (4) is a comparison of teen birth rates before and after the introduction of *16 and Pregnant,* stratified by levels MTV viewership in the year before the show began. This is essentially a difference-in-differences approach.[7] The key identifying assumption is therefore parallel trends in teen birth rates by levels of MTV ratings in the years prior to *16 and Pregnant.*

III. Placebo Tests

For estimates from equation (3) to have a causal interpretation, MTV ratings should be unrelated to trends in birth rates conditional on time and DMA fixed effects except through their relationship with *16 and Pregnant* ratings. Any association with MTV ratings would suggest omitted variables bias and be a violation of the exclusion restriction required for the validity of the instrumental variables estimator, as MTV should only affect birth rates through its relationship with *16 and Pregnant*. In Table 1 we present results from a series of reduced form and IV regressions in which we artificially start a placebo "show" in sequential quarters. Each regression includes 24 quarters, the same length of analysis used by KL: 18 prior to the beginning of the placebo show and 6 after, including the quarter that the "show" starts. We estimate the reduced form

[7] Trudeau (2016) also uses a difference-in-differences strategy in her analysis of *16 and Pregnant* and teen sexual behavior. She compares sexual activity of younger teens ages 15-18 to older women ages 19-22 and then birth rates of 15-16 year olds to 19 and 20 years before and after *16 and Pregnant*, i.e. she uses the older women as the comparison group for the younger ones. The strategy is highly questionable because both groups are exposed to *16 and Pregnant* and they differ dramatically in their levels of sexual activity and birth rates. Although Trudeau includes area fixed effects, small variations in trends can yield spurious findings. She finds no effect on birth rates when she contrasts changes between younger and older women, but statistically significant declines among younger teens in a first-difference model interacted with the Nielsen ratings of the show by designated market area (DMA).

equation (4) and IV equation (5) and use MTV viewing from 2008-2009 as the instrument.

Each row of Table 1 presents the estimated coefficients and standard errors from the estimation of the reduced form and instrumental variables models from a different 24 quarter- period. For example, in row (1) the 24-quarter period begins in 1999:Q1 and ends in 2004:QIV, with the placebo show beginning in 2003:QIII. Row 25, which is boxed in the table, reproduces KL's reduced form and instrumental variables results from their Table 1 columns (4) and (3), respectively. The shaded rows include periods that include some of the actual post-*16 and Pregnant* period in the placebo post-"show".[8]

For the reduced form, we find that 10 of the 19 estimated coefficients in rows (1) through (19), which use data exclusively from the pre-*16 and Pregnant* period, are statistically different from zero at conventional levels. That all of the estimates except for row (1) are negative and roughly equal to KL's reduced form result in magnitude strongly suggests that MTV viewing is not a valid instrument and that the parallel trends assumption required for the difference-in-differences strategy does not hold.[9] Because the sample period is fixed at 24 quarters, the

---

[8] To be consistent with KL, we weight all of our analyses of birth rates by the female population aged 15-19 in the DMA in the quarter of observation. We have also estimated all of the results presented here without weights (Jaeger, Joyce, and Kaestner 2016) and have replicated all of the results in KL without using weights (results available from the authors by request). The results without using weights are generally qualitatively similar to those using weights, although the coefficient estimates change somewhat. For example, the unweighted IV estimate in row 25 of Table 1 is -4.504 with standard error of 1.343, implying that *16 and Pregnant* can explain over 50 percent of the decline in teen birth rates in the first 18 months of the show (see KL, p. 3616), an amount that we find implausible. Other parameter estimates (for example, those by race reported by KL in their Table 2) change by even larger amounts when weights are not used, suggesting potential specification errors (Haider, Solon, and Wooldridge 2015).

[9] Adjusting for multiple comparisons in Table 1 using the Carlo E. Bonferroni (1936) correction or the somewhat more powerful Bonferroni-Holm (Sture Holm

first stage relationship between MTV and *16 and Pregnant* ratings is quite stable, with the first-stage coefficient on MTV Ratings ranging from 1.457 in row (5) to 1.524 in row (17); the first-stage coefficient is 1.513 in row (25), which corresponds to KL's analysis period. The IV results therefore reflect those from the reduced form.

The IV estimates in Table 1 are similar to KL's "IV Event Study" (p. 3614) in the sense that we use data on MTV and *16 and Pregnant Ratings* that post-date the period under study. For example, in row (13) our IV estimate suggests that teen birth rates decrease 1.958 percent from a one rating point increase in our placebo *16 and Pregnant* show, which is very similar to the estimate of -1.88 that KL report in their Figure 6 and Appendix Table B1. Our

---

1979) correction leads to rejection of none of the null hypotheses that the coefficients in both the reduced form results and the instrumental variables results are equal to zero, including KL's results in row (25). For the purposes of these tests, we treat the reduced form and instrumental variables results separately. Both the Bonferroni and Bonferroni-Holm multiple comparison procedures fix the so-called *familywise error rate* (the rate of committing *any* Type I error) and tend to have substantially less power than single comparison procedures. By fixing the familywise error rate, the probability of committing a Type II error (not rejecting one of the hypotheses when, in fact, it is not true) can increase quite substantially. An alternative procedure is to focus on the so-called *false discovery rate*, which is less conservative and fixes the share of null hypothesis rejections that are false rather than fixing the probability that *any* null hypothesis rejections are false. We have performed the Benjamini-Hochberg (1995) procedure and do not reject any of the hypotheses implicit in Table 1 when we fix the false discovery rate at 5 or 10 percent. If the false discovery rate is increased to 13.7 percent for the reduced form and 11.9 percent for instrumental variables, we still reject the null hypothesis that the coefficient is zero for 7 of the 19 pre-"show" periods for the reduced form and 8 of the 19 pre-"show" periods for instrumental variables when we consider all 36 multiple comparisons. If we limit our attention only to the tests in which the data are fully from the pre-*16 and Pregnant period*, we reject the null hypothesis of a zero coefficient for 10 of the 19 regressions, with a false discovery rate of 16.9 percent for the reduced form and 8 of 19 with a false discovery rate of 14.7 percent for instrumental variables. These calculations are available from the authors by request.

estimate captures unobserved and time-varying factors within the DMA that are correlated with teen birth rates and *16 and Pregnant* viewership.

KL's decision to restrict their analysis to a period of 24 quarters is arguably limited, given the profound trends in teen births rates over the past 25 years and the centrality of the parallel tends assumption to the validity of their difference-in-differences analysis. Figure 1 shows the natural logarithm of teen births for all teens and separately for non-Hispanic whites, non-Hispanic blacks, and Hispanics from 1990 to 2014. Teen births fall continuously from 1991 to 2005 after which they rise briefly between 2005 and 2007. After 2007, however, birth rates begin a rapid decline that precedes the airing of *16 and Pregnant.*

To capture more of the long decline in birth rates, in Table 2 we perform a similar placebo exercise, but now hold the analysis period fixed and use all of the pre-*16 and Pregnant* quarters from 1999:QI to 2009:QII. The quarter in which the placebo "show" starts sweeps through the period such that there are always at least 18 pre-"show" quarters and at least 6 post-"show" quarters. For example, in row 1 there are 18 quarters from 1999:QI to 2003:QII in the pre-"show" period and the placebo "show" begins in 2003:QIII, giving 24 quarters from 2003:QIII to 2009:QII in the post-"show" period. In row 19, there are 36 quarters in the pre-"show" period from 1999:QI to 2007:QIV and 6 quarters in the post-"show" period from 2008:QI to 2009:QII.[10] We find that all the reduced form and IV estimates are negative and of approximately the same magnitude as the actual KL estimates, which one would not expect if KL's instrument met the exclusion restriction. For the reduced form, the last 5 coefficients in rows (15) through (19) are statistically significantly different from zero at the 10 percent level while for

[10] The relationship between the reduced form and IV estimates changes across different placebo tests because first stage estimates change as the percentage of zeroes (i.e. the pre-"show" period) in MTV and *16 and Pregnant* viewership vary. The first stage coefficient is largest in row (1) at 1.539 and smallest in row (10) at 0.195.

the IV estimates those in rows (1) through (4) and those in rows (15) through (18) are statistically significant at the 10 percent level. Notably, *16 and Pregnant* was not on the air during any of the periods in Table 2 and yet we find consistently negative and often statistically significant estimates of its effect.

The results in Tables 1 and 2 cast substantial doubt on the validity of KL's empirical strategy. If their approach was valid, and MTV viewership "randomly" assigned DMAs into those that are more and less exposed to *16 and Pregnant*, then the estimated coefficients in Tables 1 and 2 should not (nearly) all be of the same sign and few if any should be statistically significantly differ from zero. Yet many of the coefficients from placebo tests that occurred wholly before *16 and Pregnant* was broadcast are statistically different from zero, and most are of the same sign and magnitude of KL's actual estimates.

IV. Direct Tests of Parallel Trends

We now directly test the parallel trends assumption in the pre-*16 and Pregnant* period. In Figure 2, Panel A, we show the trends in the natural log of teen birth rates by year of conception and quartile of MTV ratings in the year before *16 and Pregnant* was first broadcast. The gradient of birth rates in MTV ratings is clearly positive, with the lowest rates occurring in the DMAs with lowest viewership of MTV and the highest birth rates in DMAs with highest viewership. Log birth rates in the second, third, and fourth quartiles begin to converge after 2006, suggesting that log birth rates in the highest quartiles are declining faster. In Panel B of Figure 2 we show residual log birth rates after removing DMA and period fixed effects. The convergence in birth rates becomes evident as the residual log birth rates cross after 2006 with log birth rates in the first and second quartiles increasing and those in the third and fourth quartiles decreasing. It is also worth noting that, conditional on DMA and period fixed effects, the functional form of the residual log birth rates is complicated even at this highly aggregate level and varies across DMAs.

To test more rigorously whether there were parallel trends in log birth rates in the pre-*16 and Pregnant* across quartiles of MTV watching, we estimate allow for a non-parametric relationship and interact the quarter of conception indicators with indicators for quartiles of MTV or *16 and Pregnant* ratings:

$$\ln(BR_{jt}) = \beta_0 + \sum_{i=1}^{17} \alpha_t Quarter_t + \sum_{i=1}^{17}\sum_{k=2}^{4} \beta_{tk}(Quarter_t \times Quartile_k) + \mathbf{X}_{yt}\gamma + \beta_2 U_{jy} + \theta_t + \delta_j + \epsilon_{jt} \tag{6}$$

$Quarter_t$ is the calendar quarter of conception for those who give birth and $Quartile_k$ is an indicator that groups MTV or *16 and Pregnant* ratings by quartile. The rest of the specification in equation (6) is exactly the same as used by KL in equations (1) through (4) above.

Table 3 shows the results of testing whether the interaction terms between the quarterly fixed effects and quartile indicators for MTV ratings (in the left columns) or *16 and Pregnant* ratings (in the right columns) are jointly equal to zero. We show *F* statistics for testing the hypothesis that the interaction between quartile indicators and period indicators are jointly equal to zero as well as for testing the hypothesis that an indicator for above median viewing interacted with period indicators are jointly equal to zero. Each entry in the table is from a separate regression and the table shows the *F* statistics for the relevant joint hypothesis tests.[11] In row (1) we examine KL's 2005-2009 period. We reject the

[11] Strictly speaking, the test statistics based on the clustered and heteroskedasticity-consistent variance-covariance matrix are valid only asymptotically. As the number of degrees of freedom in the denominator of the *F* statistics goes to infinity, the reported test statistics multiplied by the numerator degrees of freedom will approach the limiting χ^2 distribution. The *p*-values of the

joint hypothesis that that the interaction between MTV quartile indicators and the period effects are equal to the main period effects (F=1.95, p<0.001) and also reject the null hypothesis that trends in birth rates of DMAs with above the median viewership of MTV are the same as those below the median (F=1.64, p=0.057). When we do the same analyses stratifying by quartile of *16 and Pregnant* ratings, we reject the null hypothesis that the interactions with period indicators are jointly equal to zero (F=1.46, p=0.034). In rows (2), (3), and (4) we extend the time period backwards incrementally to 2003, 2001, and 1999, respectively. In each case, we reject the null hypothesis that the quarterly effects are the same across the quartiles of MTV ratings and *16 and Pregnant* ratings. Focusing on the results for being above the median in ratings, there are clearly different trends in all periods when using MTV ratings and in the two longer periods using *16 and Pregnant* ratings.

The results in Table 3 provide evidence against the assumptions required by KL's research design to produce a causal estimate of the effect of *16 and Pregnant* on teen births. Viewership of *16 and Pregnant* is clearly endogenous based on its association with pre-show trends in teen birth rates, conditional on period and DMA fixed effects as well controls for race/ethnicity and unemployment. For KL's identification strategy to be valid, MTV ratings should only be related to birth rates through their relationship with *16 and Pregnant*, which did not exist before 2009. The results indicate that this condition is violated, suggesting that there are systematic differences in trends in birth rates across DMAs that are related to MTV viewership. Given the complicated patterns in trends evident in Figure 2, attempts to control parametrically for differential trends in birth rates across DMAs are likely to be subject to overfitting, *p*-

---

tests shown are more conservative (larger) than from the corresponding asymptotically valid χ^2 statistics, however, and so we show these instead.

hacking, and identification through functional form, providing a less than convincing strategy for making causal inferences.[12]

V. Replicating and Extending KL's Parallel Trend Assessments

KL present an "event study" in their Figure 5 and Appendix Table B1 as a means of testing the parallel trends assumption. The characterization of the exercise as an event study is something of a misnomer. Event-study designs in the applied microeconomics literature estimate the average change in outcomes before and after an event that typically occurs at different calendar times for different units (e.g. Jacobson, LaLonde and Sullivan 1993; Jacobson and Royer 2011; Bailey 2012; Bailey and Goodman-Bacon 2015). *16 and Pregnant* affected all DMAs simultaneously and calendar time and event time are the same in their analysis. KL's choice of using the four quarters prior to the start of *16 and Pregnant* as the reference category is arbitrary, obscures existing trends in the birth rates, and gives the appearance of a discontinuity when, in fact, none exists.

To illustrate these points we replicate Figure 5 in KL using the coefficients from the following regression:

$$\begin{aligned}\ln(BR_{jt}) = \beta_0 + \sum_{i=1}^{14} \alpha_i (MTV0809_j \times PreQ_i) + \\ \sum_{i=19}^{24} \beta_i (MTV0809 \times PostQ_i) + \phi U_{jy} + \theta_t + \delta_{js} + \epsilon_{jt}.\end{aligned} \tag{5}$$

The coefficients α_i show the average difference in log birth rates for varying levels of MTV viewership by quarter of conception in the pre-*16 and Pregnant*

[12] In Jaeger, Joyce, and Kaestner (2016) we presented models with DMA-specific quadratic trends as well as MTV quartile-specific quadratic trends. We also tested models with quartiles of MTV viewership as instruments instead of the more restrictive linear specification used by KL. In these results we obtain a smaller but still statistically significant IV estimate of the impact of *16 and Pregnant* than KL. Given the results in Figure 2, however, DMA quadratic trends are unlikely to be appropriate.

period (quarters 1 to 14, 2005:QI-2008:QII) relative to the reference category (quarters 15 to 18, 2008:QIII-2009:QII). The β_i coefficients show the same contrast but for the six quarters after the introduction of *16 and Pregnant* (quarters 19 to 24: 2009:III-2010:QIV). KL find that they cannot reject the joint null hypothesis that $\alpha_1 = 0, \alpha_2 = 0, \dots, \alpha_{14} = 0$, but they can reject the null hypothesis that $\beta_{19} = 0, \beta_{20} = 0, \dots, \beta_{24} = 0$.

Figure 3, Panel A replicates KL's event study analysis.[13] Like KL, we show least squares fit lines through the coefficients for both the pre-*16 and Pregnant* and post-*16 and Pregnant* periods.[14] The vertical line at 2009:QIII marks the quarter in which *16 and Pregnant* was first broadcast. Unlike KL, we show the actual four quarters of the reference period.[15] When the periods are appropriately spaced, the "discontinuity" at the introduction of *16 and Pregnant* appears to be more part of a trend that began before the show began. The absence of a noticeable discontinuity at the point the show began would undercut what

[13] The note to KL's Figure 5 states: "Estimates reflect coefficients on 2008–2009 MTV ratings interacted with quarter from a regression model controlling for DMA × season fixed effects." We were not able to replicate their figure with that specification. We *were* able to replicate their figure when we also included the DMA-specific unemployment rate in the model (but not the race/ethnicity variables). All of the models in Figures 2 through 5 therefore include the unemployment rate as a regressor to retain comparability to KL's results. Our general points hold, however, if we follow KL's specification as described.

[14] The relevance of these fitted lines is unclear, as they are not derived from the estimated regression and do not represent formal statistical tests of a break in trend. We therefore omit them from our subsequent event study analysis in Figure 4.

[15] KL's plot of the event study coefficients in Figure 5 is misleading. Each tick mark on the horizontal axis represents a quarter of the year and the vertical axis shows the magnitude of coefficients at each quarter. But only a single tick represents the reference period, which comprises four quarters. This exaggerates the sharpness of the falloff in teen births rates in the post-*16 and Pregnant* period by making a full year appear to be only one quarter.

they term the "… strong visual support for the notion of a causal effect" (KL p. 3614). In Panel B of Figure 3 we change the reference period to 2005:QI to 2005:QIV and plot the coefficients from estimation of equation (5), *mutatis mutandis*. There is little apparent discontinuity at 2009:QIII, however, and the 95 percent confidence intervals are noticeably wider than in Panel A. The pattern in Panel B of Figure 3 is also consistent with the unconditional trends in teen birth rates, which began to decline after 2007, roughly two years before *16 and Pregnant* was first broadcast. We cannot, however, reject the null hypothesis that the coefficients in the pre-*16 and Pregnant* period are jointly equal to zero (F=1.46, p=0.128).[16]

This conclusion changes when we incrementally extend the event study analysis temporally in both directions. In Figure 4, Panels A, B, and C present results from estimating equation (5) for the periods 2003:QI-2011:QIV, 2001:QI-2012:QIV, and 1999:QI-2013:IV, respectively.[17] In each panel, like in Panel B of Figure 3, the first four quarters of the period of analysis is the reference period. The downward trend in teen birth rates in the years prior to *16 and Pregnant* and the lack of a clear discontinuity at 2009:QIII is quite apparent in all three panels and we strongly reject the null hypothesis that the pre-*16 and Pregnant* coefficients are jointly equal to zero in all three periods of analysis with F statistics of 2.13 (p=0.003), 2.83 (p<0.001) and 3.41 (p<0.001) in Panels A, B, and C, respectively.[18] Our result that the pre-*16 and Pregnant* relationship

[16] Note that the F-test on pre-*16 and Pregnant* coefficients is invariant to the choice of reference period. KL erroneously report the p-value from this test as 0.21 on p. 3613.

[17] *16 and Pregnant* (including *Teen Mom* and *Teen Mom 2*) continued to be a popular shows through 2013.

[18] Kearney and Levine (2016) contend that their "event study" in KL Figure 5 shows a break in trend whereas as their reformulation of our "event study" figure shows a continuation of trend (Kearney and Levine 2016, p.10). We find the

between MTV viewing and log birth rates in the event study analysis is statistically significant is not an artifact of greatly extending the period of analysis. In KL's period, these coefficients are very close to the 10 percent critical value and by extending the pre-*16 and Pregnant* period even by only two years, (in Panel A of Table 3), we reject the null that pre-treatment coefficients are jointly equal to zero.

Overall, we find that there is no evidence to suggest that *16 and Pregnant* altered the existing downward trend in birth rates that began long before its first broadcast. By using only four and one half years of data in the period before *16 and Pregnant*, KL sectioned off a time period in which teen birth rates were at roughly similar levels between 2005 and 2008 due to the rise and fall in teen birth rates over these years. When we extend the period of analysis prior *to 16 and Pregnant* back just two years there is no visual evidence of discontinuity and the identifying assumption that teen birth rates are unrelated to MTV viewership no longer holds.

VI. Social Media and the Causal Chain

KL use Twitter feeds and Google searches to provide evidence of a potential mechanism linking *16 and Pregnant* to lower teen birth rates. If tweets and Google searches about birth control and abortion allow teen girls to acquire new information that leads to increases in prevention or voluntary termination of pregnancies there would be, presumably, fewer teen births. By establishing a

---

distinction puzzling because any significant trend in teen birth rates related to levels of MTV viewership prior to *16 and Pregnant* is a violation of their identifying assumptions. Moreover, the seemingly upward trend in teen birth rates in KL's Figure 5 in the post period is in the wrong direction. The ratings of *16 and Pregnant* tripled between 2009 and 2010 and the number of quarters in which 12-24 years olds were exposed to show was also greater in 2010 than 2009. A break in trend consistent with much greater viewership and longer exposure to the show should have been a faster *decline* and not an upturn.

correlation between *16 and Pregnant* and these online activities, KL suggest that it is reasonable to infer that the program lowers teen birth rates. Although KL's results are somewhat mixed, they conclude that there is sufficient evidence to support a causal link between broadcasts of the *16 and Pregnant*, increases in social media mentions of ways to prevent births (birth control and abortion), and, by extension, teen births.[19]

In Table 4, we present results from time series analyses of national Twitter activity about birth control and abortion related to measures of potential exposure to *16 and Pregnant*. In Panel A we follow KL and present results using the "in season" period, comprising only days in which a new season of *16 and Pregnant* was being broadcast weekly. Columns (1) and (4) of panel A replicate KL's results using Twitter from their Tables 3 and 4 in which they use the log tweet rate about birth control and abortion, respectively, as dependent variables. In part (i) of Panel A we replicate KL's estimate that the tweets about birth control increase 12 percent on the first day an episode is broadcast and 23 percent on the next day relative to the other five days of the week after the broadcast. In columns (1) and (4) of Panel A, part (ii), we replicate KL's results in which they use the log tweet rate about *16 and Pregnant* as a proxy for exposure to the

[19] 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). In their response to Jaeger, Joyce, and Keastner (2016), they write, however, that "[T]he JJK [Jaeger, Joyce, and Kaestner 2016] paper also works to demonstrate the fragility of the social media regressions in the KL paper that illustrate spikes in Google Searches and Tweets containing the terms 'birth control' after the show was introduced. That analysis was secondary and suggestive, the data was necessarily limited, and we would not be surprised if the results are sensitive to various weighting schemes." (Kearney and Levine 2016). Kearney and Levine's (2016) description of the social media analysis as "secondary and suggestive" is surprising, given that fully half of the data analysis in KL is devoted to social media. In Kearney and Levine (2014a), which garnered much attention from the national press, they argued that the analysis of social media and its effect on attitudes was a primary contribution.

show.[20] They also find here a statistically significant relationship between *16 and Pregnant* and Twitter activity about birth control and abortion.

In all of these regressions, KL use the total number of daily tweets as weights. This means that days with increased tweeting will receive more weight. For example, Michael Jackson died on 25 June 2009, the same day that "Amber," the third episode of season 1 of *16 and Pregnant*, aired. This date would get more weight in KL's regressions if Michael Jackson's death caused a spike in Twitter activity. In columns (2) and (5) of Panel A, we re-estimate all four of the regressions in columns (1) and (4) of Panel A without using weights and report standard errors that are calculated using the method of Whitney K. Newey and Kenneth D. West (1987) to allow for two autoregressive lags in the disturbance.[21] Merely by not weighting, we find that the relationship between tweets about birth control is no longer statistically significantly related to *16 and Pregnant* using either the broadcast indicators or tweets.[22]

Two thirds of *16 and Pregnant* episodes were first broadcast on Tuesdays with the other episodes distributed between Sundays, Wednesdays, Thursdays and Fridays. There is, therefore, sufficient variation in the broadcast schedule to control day-of-week effects, which possibly arise because, for example, people

[20] About this specification KL write: "The model has the advantage of being able to account for variation in the popularity of particular episodes of the show. In essence, placing an indicator for new release on the right-hand side captures exposure to an episode of average popularity. Instead, by using a specific measure of popularity of each individual episode taken from the same medium (i.e., tweets about birth control as a function of tweets about *16 and Pregnant*), our model is more powerful" (p. 3621).

[21] We calculated Durbin-Watson statistics for the models in columns (1) and (3) of Panels A and B and in all cases easily rejected the null hypothesis of no serial correlation in the residuals. These results are available from the authors by request. Allowing for additional lags in the Newey-West procedure did not substantively change the estimated standard errors.

may tweet less when they are at work. There may also be seasonality in Twitter activity and we therefore explore whether KL's results are sensitive to controlling for these day of the week and month of year differences. In columns (3) and (6) of Panel A, we repeat the analysis from columns (1) and (4) (including weighting by the number of daily tweets), but add day-of-the week and month-of-year fixed effects. Including these controls causes any association between tweeting about birth control and abortion both measures to disappear.

KL's sample in these results is limited to the 336 days in which *16 and Pregnant* was "in season" out of the 1,455 available days with positive tweets between 1 January 2009 and 31 December 2012 as presented in KL's Figure 7.[23] Days in which *16 and Pregnant* was not "in-season" arguably provide a better baseline for ongoing tweet activity related to contraception and abortion. In Panel B, columns (1), (2), (4) and (5) of Table 6 we show estimates from the exact same specifications as in in Panel A, but use all of the available data. In part (i) of Panel B, the coefficients on the indicator for days with episodes of *16 and Pregnant*, or just after, now have the wrong sign, are of small magnitude, and/or are statistically insignificant. In part (ii), we find a modest association between tweets about *16 and Pregnant* and that for birth control if we use weights, but the rest of the estimates in part (ii) of Panel B do not reflect any association between *16 and Pregnant* and tweets about birth control or abortion.

In columns (3) and (6) of Panel B part (i), we now include an indicator for the day an episode was broadcast along with six lags for the subsequent days along with the day-of-week and month-of-year fixed effects as in Panel B. In addition to the coefficients on the first and second of these seven indicators, we also show their average and its standard error. The results indicate that daily tweets for birth control and abortion are, on average, between 10 and 13 percent

[23] There are 1,461 total days in this period, but KL use log tweet rates and 6 days are lost due to zero tweets.

lower during a season of *16 and Pregnant* than when a new season is not being broadcast. In Panel B(ii), we repeat the exercise from Panel A(ii) but using the full period and find no association between the log tweet rate for *16 and Pregnant* and the tweet rate for birth control and abortion. Rather than bolstering their claim that *16 and Pregnant* caused teen birth rates to fall, their use of social media raises additional doubts about the credibility of their conclusions.[24]

VII. Discussion and Conclusion

The evidence presented by KL is inadequate to conclude anything about the causal effect of *16 and Pregnant* on teen birth rates. Their analysis fundamentally rests on the assumption that areas in which young adults watched less MTV prior to *16 and Pregnant* were no different with regard to trends in teen birth rates than areas in which MTV was more popular. We find that there *were* differences between these areas. This is seen most clearly in the placebo tests in which we found an "effect" of *16 and Pregnant* before it existed, strongly suggesting that the parallel trends assumption necessary for a causal interpretation of KL's results does not hold. The non-parametric tests of homogeneity in birth rate trends by quartile of MTV viewership in the pre-*16 and Pregnant* period and the visual evidence from the event study figures reinforce our conclusion that there were unequal trends in birth rates across areas with high MTV ratings relative to those with lower MTV ratings. KL's social media analysis formed a core part of their argument, but we also find that these results do not support KL's conclusions. By only analyzing days during a season of *16 and Pregnant*, they neglect to exploit an arguably more appropriate baseline of tweet activity. When

[24] KL's last effort to link *16 and Pregnant* with online activity uses within-state variation between the pre-*16 and Pregnant* and post-*16 and Pregnant* periods. This exercise suffers similar sensitivity to selected samples and weighting as their analysis of teen birth rates. Results are available from the authors by request.

we control for day of week and month fixed effects, KL's social media results disappear.

What may have affected the acceleration in the decline in birth rates that begin in 2008, if not *16 and Pregnant*? The obvious candidate is the Great Recession and its aftermath. Sobotka, Skirbekk and Philipov (2011) show that younger women's fertility is more to recessions than responsive to recessions than older women's. Goldstein, et al. (2013) demonstrate this relationship help in Europe during the Great Recession. Cherlin, et al. (2013) confirm that this relationship also held in the U.S. during that time and show that that poorer women and Hispanic women experienced especially sharp declines at the Great Recession's onset. Although KL control for the overall unemployment rate, it is possible that using a more refined measure of unemployment for teen woman would alter their results.[25] There also may be meaningful interactions between changes in unemployment, longer schooling durations, and changes in more effective contraception.

Are there general lessons to be learned from this exercise beyond demonstrating that causal claims about *16 and Pregnant*'s impact on birth rates are not justified? We believe that there are. Given the importance of the parallel trends assumption in differences-in-differences studies, studies that use geographic variation to identify a causal effect should clearly test the robustness of this assumption over long and varying periods. The parallel trend assumption is not specific to particular time periods and difference-in-difference estimates should be invariant to modest extensions to the periods of analysis. Only when the assumption is met over substantial and varying periods can the resulting estimates

[25] Data from the Current Population Survey indicate that the unemployment rate for 15 to 24 year-old, non-Hispanic whites, non-Hispanic blacks and Hispanics increased by 8.8, 10.5 and 15.8 percentage points, respectively, between 2007 and 2010. These are substantial increases. These calculations are available from the authors by request.

be given a causal interpretation. Perhaps more importantly, our revisiting KL's analysis adds to the growing evidence in economics and other social sciences that replication is important and necessary. Without replication, the myriad problems with KL's analyses would not have come to light and there would be no opportunity to correct the public record about the effect of reality television on teen reproductive activity.[26] Because KL's study attracted extensive media coverage, policymakers may believe that "nudges" like those represented by *16 and Pregnant* are effective when, at least in this case, no causal link has been proven. Getting the answer right, which depends on both revisiting analyses with the original data and replication of the "experiment" in different contexts, should have a higher priority in economics and social sciences journals.

[26] We want to emphasize that such replication exercises require the cooperation of the authors of the original paper. Phil Levine was exemplary in providing us with programs and responding to our many requests while we were trying to reproduce KL's results.

References

Ananat, Elizabeth Oltmans, Anna Gassman-Pines, and Christina Gibson-Davis. 2013. "Community-Wide Job Loss and Teenage Fertility: Evidence From North Carolina." *Demography* 50 (6): 2151–2171.

Bailey, Martha J. 2012. "Reexamining the Impact of Family Planning Programs on US Fertility: Evidence from the War on Poverty and the Early Years of Title X." *American Economic Journal: Applied Economics* 4 (2): 62–97.

Bailey, Martha J., and Andrew Goodman-Bacon. 2015. "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans." *American Economic Review* 105 (3): 1067–1104.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Bonferroni Carlo E. 1936. "Teoria statistica delle classi e calcolo delle probabilità." *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 8: 3–62.

Boonstra, Heather. 2014. "What Is Behind the Declines in Teen Pregnancy Rates?" *Guttmacher Policy Review* 17 (3): 15–21.

Cherlin, Andrew, Erin Cumberworth, S. Philip Morgan, and Christopher Wimer. 2013. "The Effects of the Great Recession on Family Structure and Fertility." *The Annals of the American Academy of Political and Social Science* 650 (1): 214–231.

Dehejia, Rajeev, and Adriana Lleras-Muney. 2004. "Booms, Busts, and Babies' Health." *The Quarterly Journal of Economics* 119 (3): 1091–1130.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122 (3): 1187–1234.

Friedberg, Leora. 1998. "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data." *The American Economic Review* 88 (3): 608–627.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics* 121 (3): 931–972.

Gentzkow, Matthew, and Jesse M. Shapiro. 2008. "Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study." *The Quarterly Journal of Economics* 123 (1): 279–323.

Goldstein, Joshua, Michaela Kreyenfeld, Aiva Jasilioniene, and Deniz Dilan Karaman Örsal. 2013. "Fertility Reactions to the 'Great Recession' in Europe: Recent Evidence from Order-Specific Data." *Demographic Research* 29 (4): 85–104.

Hamilton, Brady E, Martin, Joyce A., Michelle J.K. Osterman, Sally C. Curtin, and T.J. Matthews. 2015. *Births: Final Data for 2014. National Vital Statistics Reports.* Volume 64, Number 12. Hyattsville, MD: National Center for Health Statistics.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *The American Economic Review* 83 (4): 685–709.

Jacobson, Mireille, and Heather Royer. 2011. "Aftershocks: The Impact of Clinic Violence on Abortion Services." *American Economic Journal: Applied Economics* 3 (1): 189–223.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner. 2016. *Does Reality TV Induce Real Effects? On the Questionable Association Between 16 And Pregnant and Teenage Childbearing*, Discussion Paper 10317. IZA. http://ftp.iza.org/dp10317.pdf.

Jensen, Robert, and Emily Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *The Quarterly Journal of Economics* 124 (3): 1057–1094.

Kaestner, Robert, Sanders Korenman, and June O'Neill. 2003. "Has Welfare Reform Changed Teenage Behaviors?" *Journal of Policy Analysis and Management* 22 (2): 225–248.

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

Kearney, Melissa S., and Phillip B. Levine. 2014b. "Income Inequality and Early Nonmarital Childbearing." *Journal of Human Resources* 49 (1): 1–31.

Kearney, Melissa S., and Phillip B. Levine. 2015a. "Investigating Recent Trends in the U.S. Teen Birth Rate." *Journal of Health Economics* 41 (May): 15–29.

Kearney, Melissa S., and Phillip B. Levine. 2015b. *Early Childhood Education by MOOC: Lessons from Sesame Street*. Working Paper 21229. National Bureau of Economic Research. http://www.nber.org/papers/w21229.

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

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

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4 (4): 1–31.

Lindberg, Laura, John Santelli, and Sheila Desai. forthcoming**.** "Understanding the Decline in Adolescent Fertility in the United States, 2007-2012. *Journal of Adolescent Health.*

Lopoo, Leonard M., and Thomas DeLeire. 2006. "Did Welfare Reform Influence the Fertility of Young Teens?" *Journal of Policy Analysis and Management* 25 (2): 275–298.

National Research Council. Panel on Adolescent Pregnancy and Childbearing. 1987. *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*. Edited by Cheryl D. Hayes and Sandra L. Hofferth. 2 vols. Washington, D.C.: National Academy Press.

Newey, Whitney K., and Kenneth D. West. 1987. "Hypothesis Testing with Efficient Method of Moments Estimation." *International Economic Review* 28 (3): 777–787.

Santelli, John S., Laura Duberstein Lindberg, Lawrence B. Finer, and Susheela Singh. 2007. "Explaining Recent Declines in Adolescent Pregnancy in the United States: The Contribution of Abstinence and Improved Contraceptive Use." *American Journal of Public Health* 97 (1): 150–156.

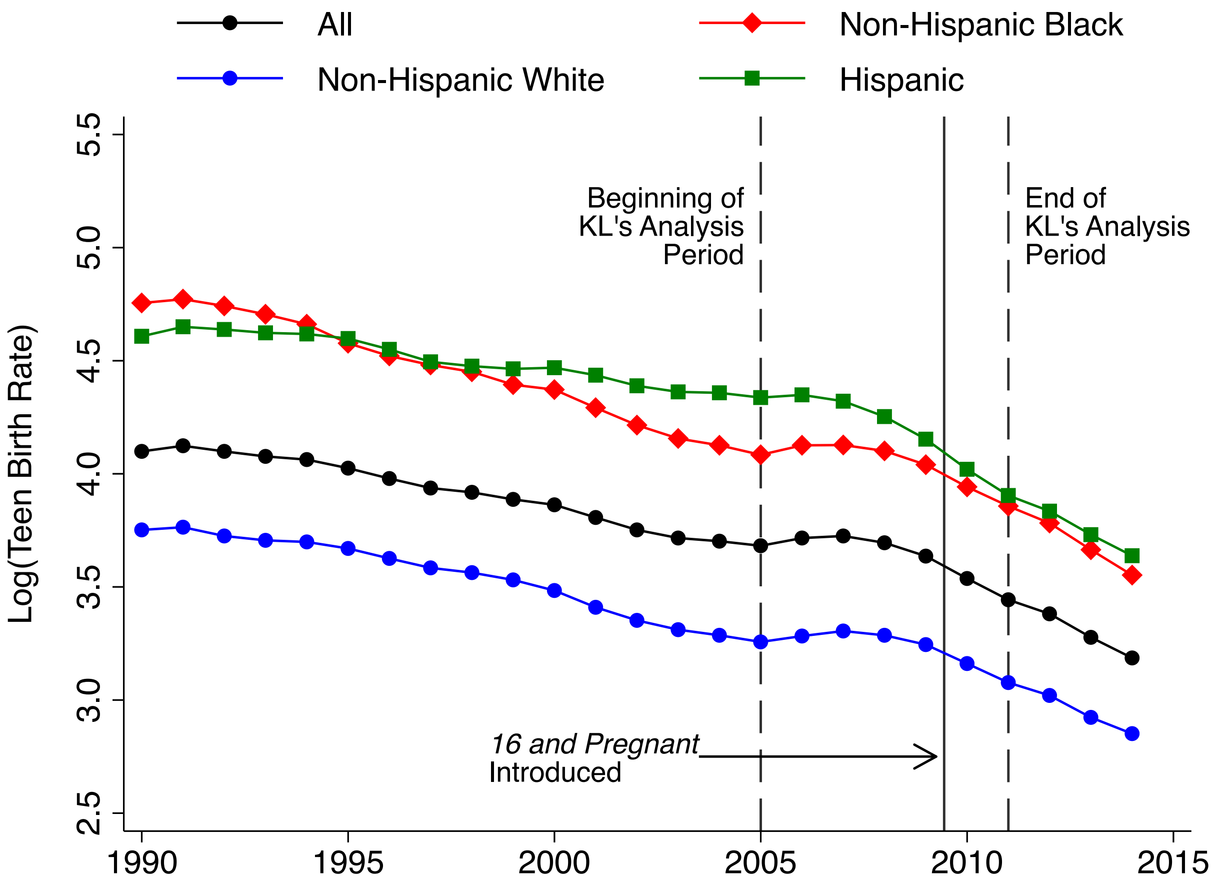
Sobotka, Tomáš, Vegard Skirbekk, and Dimiter Philipov. 2011. "Economic Recession and Fertility in the Developed World." *Population and Development Review* 37 (2): 267–306.

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

Trudeau, Jennifer. 2016. "The Role of New Media on Teen Sexual Behaviors and Fertility Outcomes – The *Case of 16 and Pregnant*." *Southern Economic Journal* 82(3): 975-1003.

Wolfers, Justin. 2006. "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." *American Economic Review* 96(5): 1802-1827.

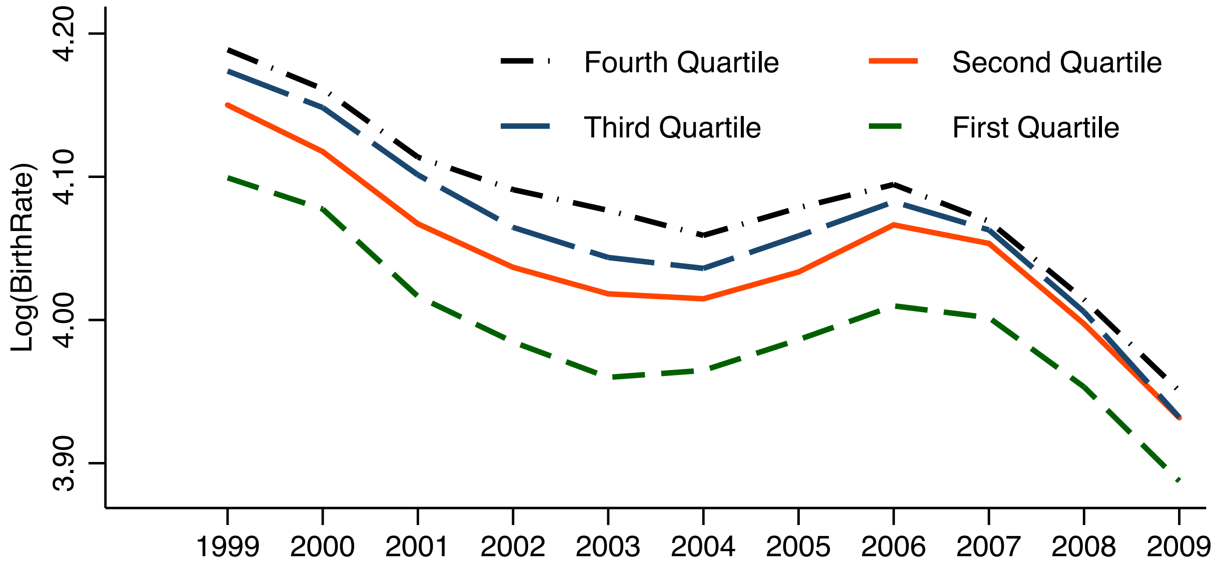
Figure 1
Log(Teen Birth Rate) for Race/Ethnicity Groups



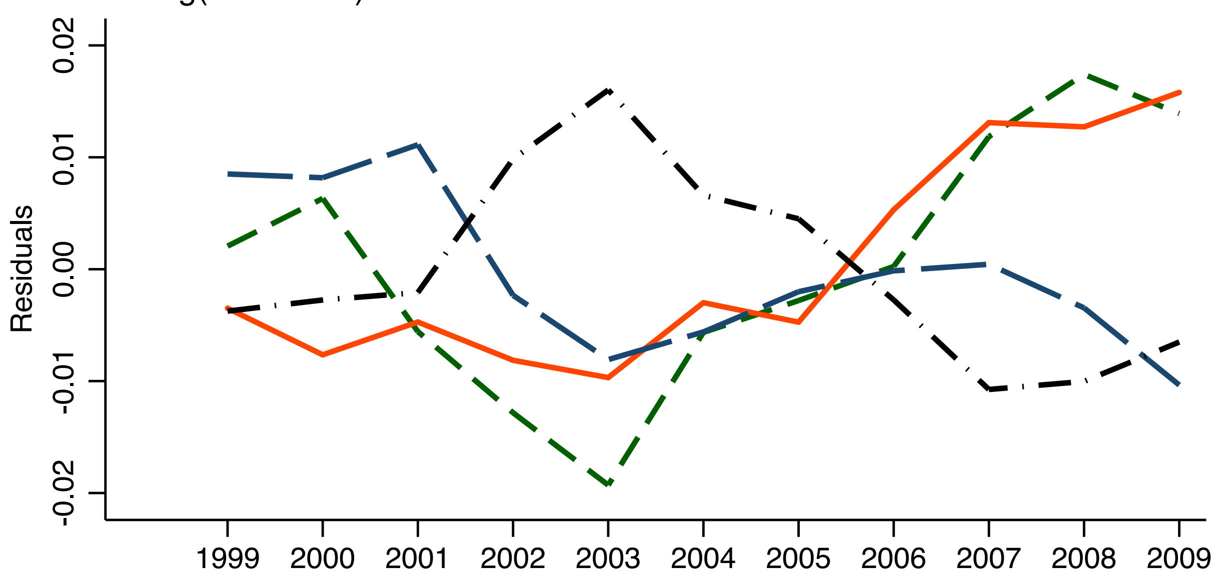
Source: Hamilton, et al. (2015)

Figure 2
Log(Birth Rate) and Residuals by MTV Ratings Quartile

Panel A: Log(Birth Rate)



Panel B: Log(Birth Rate) Residuals

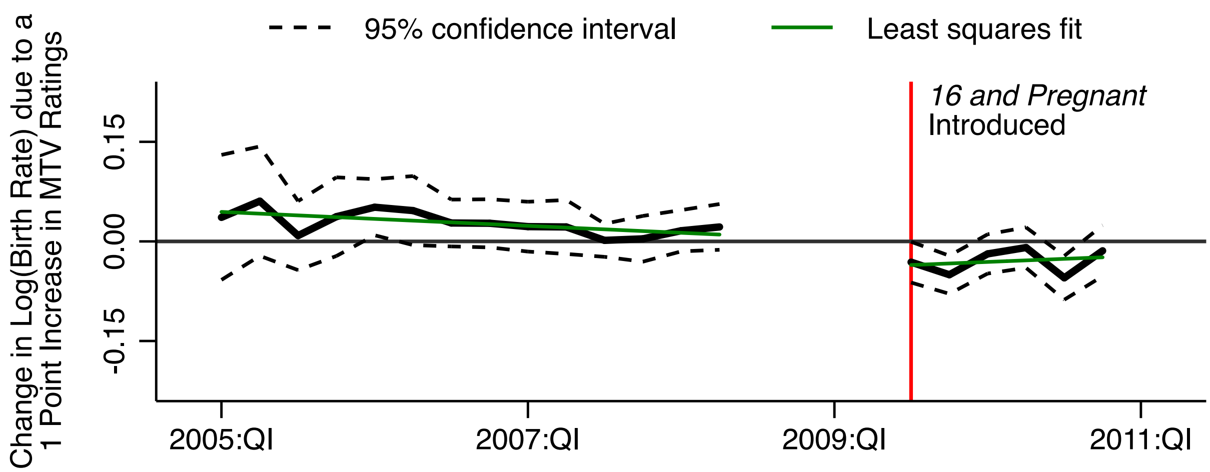


Source: Authors' calculations using Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

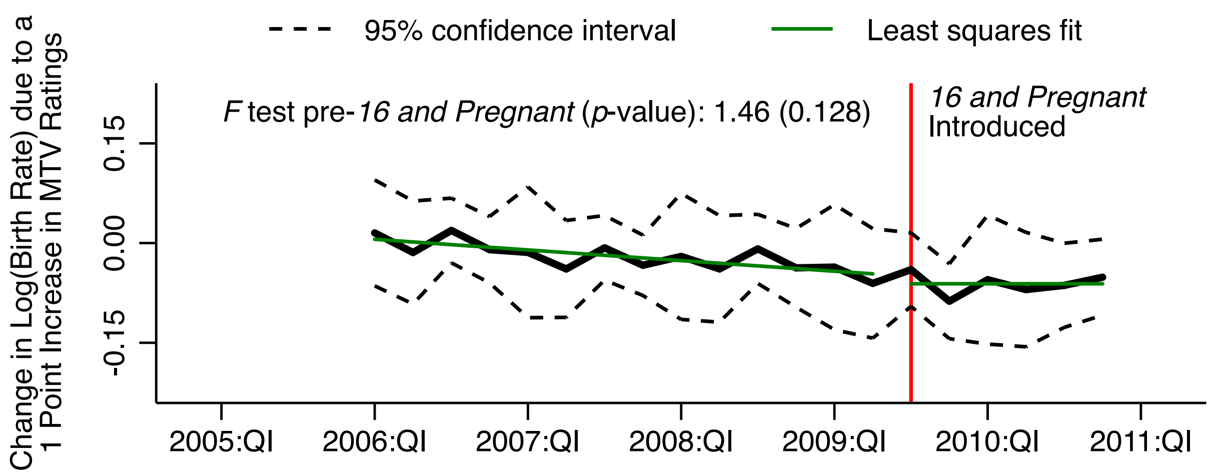
Notes: Panel A shows weighted average annual Log(Birth Rate) for teenagers by quartile of MTV ratings where the weights are the female population aged 15-19 in the DMA. Panel B shows weighted average annual residuals from a weighted regression of quarterly Log(Teen Birth Rate) on DMA fixed effects and time period fixed effects where the weights are the female population aged 15-19 in the DMA.

Figure 3
Reduced Form Event Study: 2005-2010

Panel A: Reference period is 2008:QIII-2009:QII (Replicating Fig. 5 from KL)



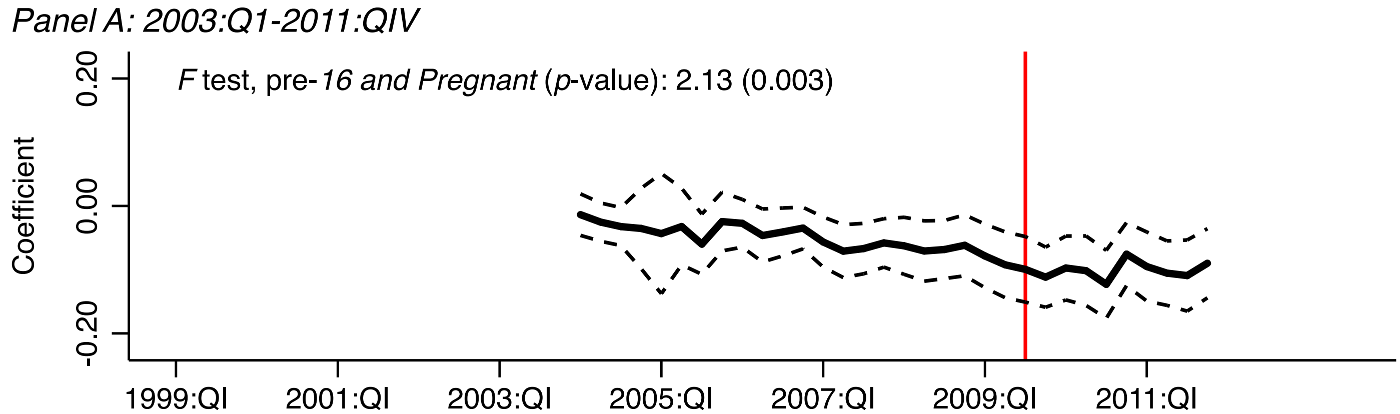
Panel B: Reference Period is 2005:QI-2005:QIV

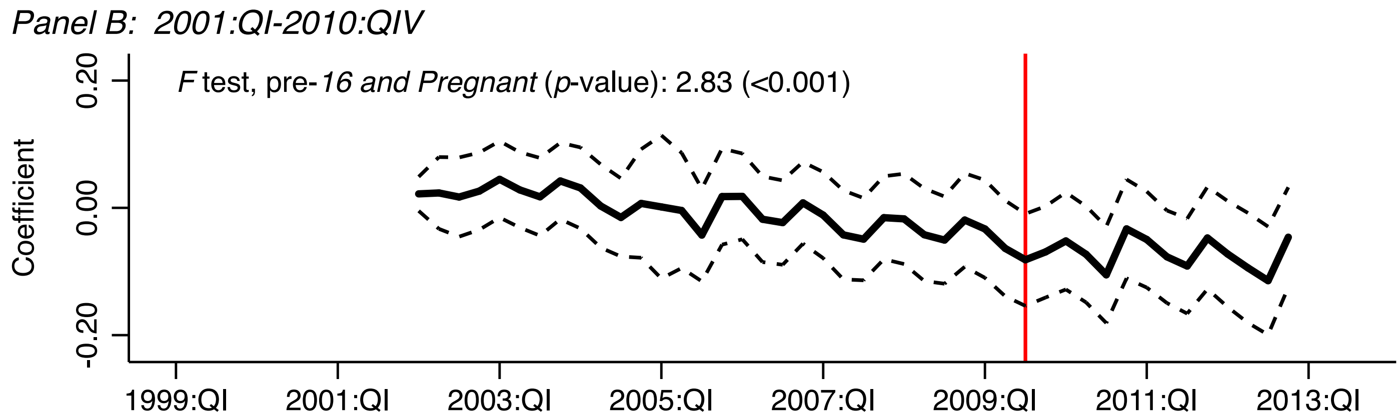


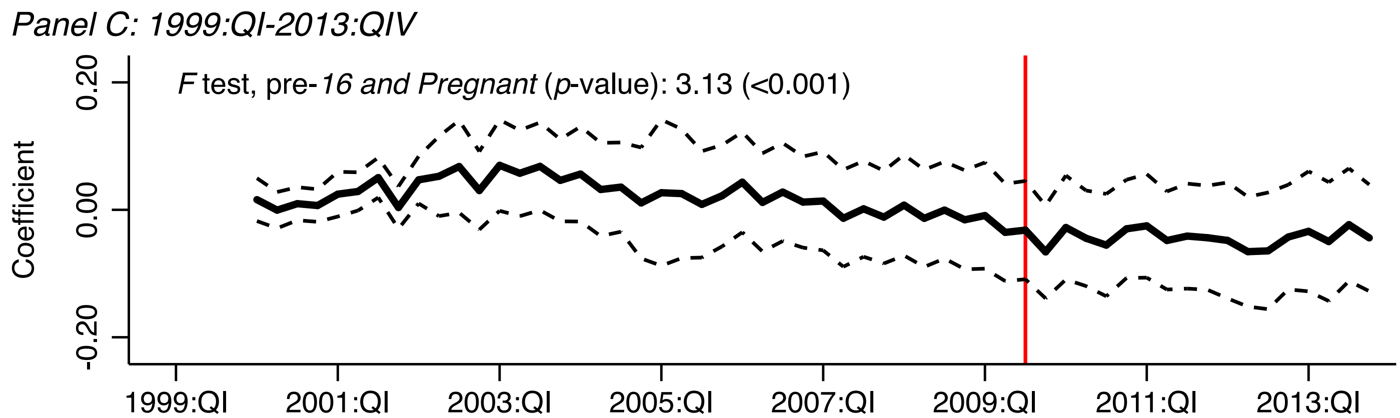
Source: Authors' calculations using Bureau of Labor Statistics Local Area Unemployment Statistics, Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: These figures plot the coefficients of average MTV Ratings in 2008-2009 interacted with period dummy variables. The regressions also include the unemployment rate and a full set of period dummy variables and DMA × season interactions. Confidence intervals are calculated by clustering by DMA. Regressions are weighted by the female population aged 15-19 in the DMA. Panel A uses 2008:QIII through 2009:QII as the reference period. Panel B uses 2005:QI through 2005:QIV as the reference period. Least squares lines in both panels are fitted separately for the pre- and post-*16 and Pregnant* periods. The *F* statistic is for the test of joint significance of the coefficients in pre-*16 and Pregnant* period. Panel A replicates Kearney and Levine's (2015b) Figure 5.

Figure 4
Reduced Form Event Studies







Source: Authors' calculations using Bureau of Labor Statistics Local Area Unemployment Statistics, Bureau of the Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: These figures plot the coefficients of average MTV Ratings in 2008-2009 interacted with period dummy variables. The regressions also include the unemployment rate and a full set of period dummy variables and DMA × season interactions. Confidence intervals, indicated by the dotted lines, are calculated by clustering by DMA. Regressions are weighted by the female population aged 15-19 in the DMA. In each regression the first four quarters are the reference period. The vertical lines indicate the introduction of *16 and Pregnant*.

Table 1

Placebo Tests of Estimated Reduced Form and Instrumental Variables Impact on Teen Birth Rates

Rolling 24 Quarter Periods

	Dates			Reduced Form		Instrumental Variables	
Row	Begin	"Show" Start	End	Coefficient	Std. Err.	Coefficient	Std. Err.
(1)	1999:QI	2003:QIII	2004:QIV	0.397	2.061	0.268	1.262
(2)	1999:QII	2003:QIV	2005:QI	-0.683	2.298	-0.463	1.403
(3)	1999:QIII	2004:QI	2005:QII	-1.604	2.588	-1.091	1.554
(4)	1999:QIV	2004:QII	2005:QIII	-2.692	2.813	-1.839	1.661
(5)	2000:QI	2004:QIII	2005:QIV	-2.681	3.042	-1.840	1.806
(6)	2000:QII	2004:QIV	2006:QI	-2.621	2.865	-1.789	1.684
(7)	2000:QIII	2005:QI	2006:QII	-2.903	2.303	-1.977	1.353
(8)	2000:QIV	2005:QII	2006:QIII	-2.675 *	1.570	-1.808 *	0.954
(9)	2001:QI	2005:QIII	2006:QIV	-2.437 *	1.324	-1.639 *	0.846
(10)	2001:QII	2005:QIV	2007:QI	-1.994	1.299	-1.330	0.837
(11)	2001:QIII	2006:QI	2007:QII	-2.554 *	1.441	-1.693 *	0.948
(12)	2001:QIV	2006:QII	2007:QIII	-2.929 **	1.473	-1.936 **	0.977
(13)	2002:QI	2006:QIII	2007:QIV	-2.968 *	1.521	-1.958 **	0.992
(14)	2002:QII	2006:QIV	2008:QI	-3.144 **	1.569	-2.068 **	1.028
(15)	2002:QIII	2007:QI	2008:QII	-3.341 **	1.508	-2.194 **	1.014
(16)	2002:QIV	2007:QII	2008:QIII	-3.079 **	1.458	-2.021 **	0.966
(17)	2003:QI	2007:QIII	2008:QIV	-2.756 *	1.476	-1.809 *	0.967
(18)	2003:QII	2007:QIV	2009:QI	-2.543 *	1.485	-1.673 *	0.969
(19)	2003:QIII	2008:QI	2009:QII	-2.377	1.558	-1.566	1.002
(20)	2003:QIV	2008:QII	2009:QIII	-2.442	1.580	-1.610	1.006
(21)	2004:QI	2008:QIII	2009:QIV	-2.978 *	1.604	-1.964 *	1.029
(22)	2004:QII	2008:QIV	2010:QI	-3.320 **	1.663	-2.192 **	1.074
(23)	2004:QIII	2009:QI	2010:QII	-3.589 **	1.639	-2.372 **	1.040
(24)	2004:QIV	2009:QII	2010:QIII	-3.898 **	1.611	-2.577 **	1.024
(25)	2005:QI	2009:QIII	2010:QIV	-3.581 **	1.517	-2.368 **	0.942

Source: Authors' calculations using Bureau of Labor Statistics Local Area Unemployment Statistics, Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: Entries in the table are a) in the reduced form, the estimated coefficient on MTV Ratings in 2008:QIII-2009:QII interacted with a dummy variable for being in the "post" period and b) for instrumental variables, the estimated coefficient on *16 and Pregnant* Ratings interacted with a dummy variable for being in the "post" period where the instrument is the regressor of interest from the reduced form regressions. Standard errors, clustered by DMA, are shown in parentheses. All regressions are weighted by the female population aged 15-19 in the DMA at the time of the observation. In each regression there are 18 pre-"show" quarers and 6 post-"show" quarters. All regressions also include the unemployment rate, the percent of the population that is non-Hispanic black and the percent of the population that is Hispanic, 24 quarter fixed effects as well DMA × season fixed effects as regressors. The period analyzed by Kearney and Levine's (2015b) is boxed. In the periods that are lightly shaded, the placebo post period partially includes the actual post-*16 and Pregnant* period. Sample size in rows (1) through (17), and (19) through (21 is 4,918, the sample size in row (18) is 4,917, and the sample size in rows (22) through (25) is 4,919. *** indicates significant at the 1 percent level, ** indicates significant at the 5 percent level, * indicates significant at the 10 percent level.

Table 2

Placebo Tests of Reduced Form and Instrumental Variables Estimates of Impact on Teen Birth Rates Fixed for the Entire Pre-*16 and Pregnant* Period

	Placebo	Reduced Form		Instrumental Variables	
Row	"Show" Start	Coefficient	Std. Err.	Coefficient	Std. Err.
(1)	2003:QIII	-1.343	2.680	-2.067 **	0.985
(2)	2003:QIV	-1.731	2.621	-2.106 **	0.981
(3)	2004:QI	-2.056	2.575	-2.001 **	1.002
(4)	2004:QII	-2.337	2.491	-1.846 *	0.987
(5)	2004:QIII	-2.516	2.421	-1.495	0.995
(6)	2004:QIV	-2.622	2.387	-1.359	1.030
(7)	2005:QI	-2.644	2.300	-1.212	1.053
(8)	2005:QII	-2.666	2.191	-0.833	1.006
(9)	2005:QIII	-2.818	2.187	-0.698	0.988
(10)	2005:QIV	-2.685	2.199	-0.524	0.968
(11)	2006:QI	-2.873	2.227	-0.791	0.943
(12)	2006:QII	-3.160	2.219	-0.877	0.859
(13)	2006:QIII	-3.258	2.189	-0.908	0.785
(14)	2006:QIV	-3.453	2.181	-0.942	0.720
(15)	2007:QI	-3.682 *	2.168	-1.179 *	0.678
(16)	2007:QII	-3.752 *	2.138	-1.340 **	0.663
(17)	2007:QIII	-3.629 *	2.113	-1.708 ***	0.636
(18)	2007:QIV	-3.597 *	2.131	-1.909 ***	0.740
(19)	2008:QI	-3.627 *	2.161	-2.052	1.328

Source: Authors' calculations using Bureau of Labor Statistics Local Area Unemployment Statistics, Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: Entries in the table are a) for the reduced form, the estimated coefficient on MTV Ratings in 2008:QIII-2009:QII interacted with a dummy variable for being in the "post" period and b) for instrumental variables, the estimated coefficient on *16 and Pregnant* Ratings interacted with a dummy variable for being in the "post" period where the instrument is the regressor of interest from the reduced form regressions. Standard errors, clustered by DMA, are shown in parentheses. All regressions are weighted by the female population aged 15-19 in the DMA at the time of the observation. Each regression uses the period 1999:QI to 2009:QII. The only difference in each regression is the quarter in which the placebo "show" begins. All regressions also include the unemployment rate, the percent of the population that is non-Hispanic black and the percent of the population that is Hispanic, 24 quarter fixed effects as well DMA × season fixed effects as regressors. The first stage relationship between MTV ratings and *16 and Pregnant* Ratings ranges from 0.273 (row 13) to 1.539 (row 1). Sample size for all regressions is 8,607. *** indicates significant at the 1 percent level, ** indicates significant at the 5 percent level, * indicates significant at the 10 percent level.

Table 3

Differential Trends in Teen Birth Rates during the Pre-*16 and Pregnant* Period
by Quartiles of MTV Ratings and *16 and Pregnant* Ratings

			Quartiles of			
			MTV Ratings		*16 and Pregnant* Ratings	
Period	Quarters	Obs.	*F*: Quartile × Quarter = 0	*F*: Above Median × Quarter = 0	*F*: Quartile × Quarter = 0	*F*: Above Median × Quarter = 0
(1) 2005:QI-2009:QII	14	3,689	1.953 ***	1.640 *	1.464 **	0.971
			[<0.001]	[0.057]	[0.034]	[0.493]
(2) 2003:QI-2009:QII	26	5,327	2.954 ***	1.844 **	2.041 ***	0.933
			[<0.001]	[0.011]	[<0.001]	[0.559]
(3) 2001:QI-2009:QII	34	6,967	4.317 ***	1.852 ***	3.680 ***	1.574 **
			[<0.001]	[0.005]	[<0.001]	[0.031]
(4) 1999:QI-2009:QII	42	8,607	6.665 ***	2.45 ***	5.355 ***	1.625 **
			[<0.001]	[<0.001]	[<0.001]	[0.015]

Source: Authors' calculations using Bureau of Labor Statistics Local Area Unemployment Statistics, Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: Entries in the table are *F* statistics. *p*-values in brackets. Each entry is from a separate regression. Variance-covariance matrices of estimates are calculated by clustering by DMA. Regressions are weighted by the female population aged 15 to 19 in each DMA. *F*-statistics are from tests that the coefficients on the interactions between quarterly indicators and indicators for the relevant quartiles of either *16 and Pregnant* ratings or MTV ratings are jointly equal to zero. All models also include DMA fixed effects, the percent of black non-Hispanics , the percent of Hispanics in the DMA, and the unemployment rate as regressors. The results in row (1) use the same pre-*16 and Pregnant* period as Kearney and Levine (2015b). *** indicates significant at the 1 percent level, ** indicates significant at the 5 percent level, * indicates significant at the 10 percent level.

Table 4

Estimated Impact of *16 and Pregnant* on Tweets about Birth Control and Abortion

	Birth Control			Abortion		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: National Tweet Rate Using Only "In Season" Daily Data between 1 January 2009 and 31 December 2012						
(i) Impact of 16 and Pregnant Broadcasts (336 days)						
Day new *16 and Pregnant* episode released	0.120 ***	0.065	0.015	0.142 ***	0.176 ***	0.027
	(0.046)	(0.052)	(0.064)	(0.046)	(0.061)	(0.051)
Day after new *16 and Pregnant* episode released	0.229 ***	0.157	0.035	0.212 ***	0.180 ***	0.024
	(0.047)	(0.052)	(0.055)	(0.046)	(0.061)	(0.063)
Replicates Kearney and Levine (2015b) results	Table 3, Column 4	no	no	Table 3, Column 5	no	no
(ii) Impact of ln(tweet rate) of 16 and Pregnant (336 days)						
ln(tweet rate), *16 and Pregnant*	0.077 ***	0.029	0.004	0.064 **	0.007	0.019
	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.030)
Replicates Kearney and Levine (2015b) results	Table 4, Panel A, Column 4	no	no	Table 4, Panel A, Column 5	no	no
Panel B: National Tweet Rate Using All Available Daily Data between 1 January 2009 and 31 December 2012						
(i) Impact of 16 and Pregnant Broadcasts (1,455 days)						
Day new *16 and Pregnant* episode released	-0.045	-0.045	-0.171 ***	-0.015	0.052	-0.112 **
	(0.044)	(0.039)	(0.051)	(0.064)	(0.043)	(0.055)
Day after new *16 and Pregnant* episode released	0.046	0.041	-0.124 **	0.037	0.046	-0.092
	(0.055)	(0.049)	(0.052)	(0.063)	(0.052)	(0.056)
Sum of Day 1 through Day 7 indicators "in season"			-0.126 ***			-0.108 ***
			(0.031)			(0.028)
(ii) Impact of ln(tweet rate) of 16 and Pregnant (1,322 days)						
ln(tweet rate), *16 and Pregnant*	0.035 **	0.011	0.013	-0.020	-0.009	-0.016
	(0.017)	(0.012)	(0.016)	(0.016)	(0.019)	(0.017)
All Panels: Weighting and Covariates						
Includes day of week and month indicators	no	no	yes	no	no	yes
Weighted by number of tweets	yes	no	yes	yes	no	yes

Source: Authors' calculations using Twitter, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: Dependent variable is ln(tweet rate). Panel A, columns (1) and (4) use heteroskedasticity-consistent standard errors (in parentheses) to be consistent with Kearney and Levine (2015b). All other regressions use Newey-West (1987) standard errors (in parentheses) with two autoregressive lags, which were not sensitive to including additional lags. Panel A uses national time series that includes only the day that a new episode of 16 and Pregnant was broadcast and the 6 subsequent days between 1 January 2009 and 31 December 2012. Panel B uses the same specifications as in Panel A but the sample includes all tweets between 1 January 2009 and 31 December 2012. Sample sizes in Panel B(ii) are different from those in Panel B(i) because days with no tweets about 16 and Pregnant are dropped. Estimates in columns (1), (3) and (4), and (6) are weighted by the number of daily tweets. Estimates in columns (2) and (5) use the same sample and specification as columns (1) and (4), respectively, but regressions are not weighted. Regressions in Panels A(i) and B(i) include 7 indicators for the day that a new episode of *16 and Pregnant* was broadcast and the subsequent 6 days. Following Kearney and Levine (2015b), we show the first two of these estimated coefficients. In Panel B(ii), columns (3) and (6) we also show the average of all 7 of these indicators. Data used are the same as in Figure 7 of Kearney and Levine (2015b), but at a daily frequency. *** indicates significant at the 1 percent level, ** indicates significant at the 5 percent level, * indicates significant at the 10 percent level.