

A Comment on Jaeger, Joyce, and Kaestner (2018)
Did Reality TV Really Cause a Decline in Teenage Childbearing?

Melissa S. Kearney and Phillip B. Levine¹

July 16, 2018

We maintain the position that our 2015 *American Economic Review* paper “Media Influences on Social Outcomes: The Impact of MTV’s *16 and Pregnant* on Teen Childbearing” (hereafter KL 2015) provides strong empirical evidence that the introduction of the reality TV show “*16 and Pregnant*” led to a sizable reduction in teen birth rates. As we believe it is important to understand the determinants of teen childbearing, we have carefully read the variety of critiques of our work that Jaeger, Joyce, and Kaestner (JJK) have circulated to date. However, a careful consideration of the numerous arguments do not lead us to alter that conclusion.

JJK’s July 2018 NBER working paper (hereafter JJK 2018) follows an IZA working paper they posted in October 2016 (hereafter JJK 2016): <http://ftp.iza.org/dp10317.pdf>.² Shortly thereafter, we posted a lengthy response also as an IZA working paper (hereafter KL 2016): <http://ftp.iza.org/dp10318.pdf>. Since much of the critique is similar, we direct interested readers to our response there and offer only brief remarks here.

To interpret the results reported in JJK 2018, it is important to put them in the context of the full range of analyses that were conducted. We have read four full-length drafts produced by Joyce, Jaeger, and Kaestner on KL 2015. The specific arguments made and results reported have shifted in each subsequent draft. We thus briefly review some of this history here as well.

JJK 2018 Econometric Critiques

1. Parallel Trends by Race/Ethnic Group

In this latest critique, JJK 2018 contend that teen birth rates among black and Hispanic teens were not on parallel trends in television markets, formally called Designated Market Areas or DMAs, with higher/lower baseline MTV ratings in the years before the 2009 introduction of *16 and Pregnant*. As shown in their Figure 2, between around 2007 and 2009, teen birth rates were falling less rapidly for blacks and more rapidly for Hispanics in places with above average MTV ratings. JJK 2018 argue that this demonstrates a violation of the parallel trends assumption necessary for our estimated effect to have a causal interpretation.

We note that trends in the log teen birth rate appear to be parallel among white teens (as shown in their Figure 2, Panel A). When we separate our analysis by race, we find a statistically significant reduction in white teen births. In fact, much of the overall impact on teen births is driven by whites, as reported in Table 2 of KL 2015. A violation of parallel trends does not

¹ University of Maryland and NBER and Wellesley College and NBER, respectively.

² Their NBER working paper will also be published in the *Journal of Business and Economic Statistics*. We were not invited to comment on that submission before it was accepted.

appear to be an issue for white teens and our analysis finds that the introduction of *16 and Pregnant* led to a significant reduction in teen births for them.

JJK 2018 questions whether it is reasonable to observe a statistically significant effect just for whites when blacks and Hispanics were more likely to watch the show, a fact about ratings that we pointed out in our initial paper. Of course, just because the average white woman is less likely to watch the show says nothing about how many “marginal” white women are watching the show. We are therefore not swayed by this criticism. Moreover, the point estimate of the effect for Hispanic teens is actually larger than that estimated for whites, but it is less precise.

JJK 2018 further suggests that the confounding factor driving our results is the differential response of teen births from different racial/ethnic groups to the Great Recession, as measured by the local area unemployment rate. Note that all of the regressions in KL 2015 control for both the time-varying local area unemployment rate and the time-varying racial/ethnic composition of the teen population. Our results are therefore not confounded by any differential change in racial/ethnic composition or economic conditions. What JJK 2018 adds is the interaction of the unemployment rate and the racial/ethnic composition and alternatively the interaction of a linear time trend and the racial/ethnic composition.

It is unclear why the time trend of births among black and Hispanic teens or the differential response of black and Hispanic teen fertility to the local area unemployment rate would necessarily be a driver of white teen birth rates. A plausible explanation for the attenuation of the estimated effect when these additional interaction terms are included is overfitting the data. Race/ethnicity are correlated with pre-period MTV viewership in the pre-period. When JJK includes these variables interacted with a time-trend, they are introducing variables that are highly correlated with our instrument (pre-period MTV ratings interacted with a post-June 2009 indicator). It is thus not surprising that the estimated coefficient on the predicted *16 and Pregnant* ratings variable is attenuated.

Focusing specifically on their analysis of white teen births, JJK 2018 provides a specification where they find no statistically significant effect of predicted *16 and Pregnant* ratings. To do so, they include DMA-specific linear time trends (JJK 2018 page 12 and Table 4). Our entire empirical approach rests on a comparison of trends over a short period of time -- the 24 quarters between 2005-2012 -- across DMAs based on time-invariant baseline MTV viewership rates. The inclusion of DMA-specific time trends leaves little variation left to identify an effect of the show.

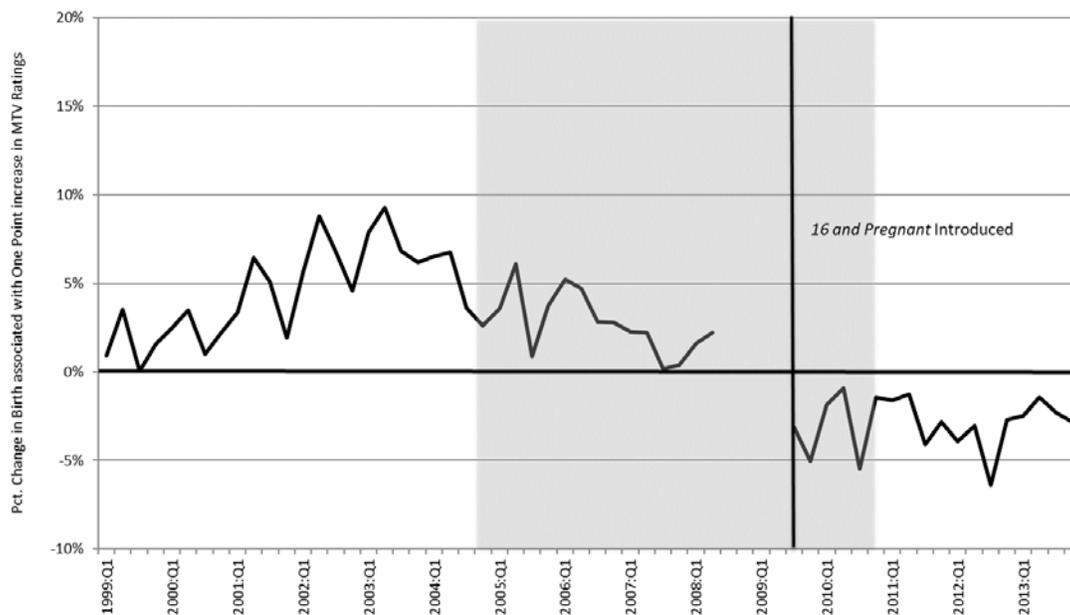
To be clear, our results would be stronger if they were robust to all possible alternative specifications. The fact, though, that one can construct an alternative specification without obvious justification that attenuates the main coefficient of interest should not alter the main conclusion. This is particularly true when a reasonable case can be made that this specification may be over-fitting the data. We leave it to individual readers to weigh the value of the somewhat peculiar specifications in JJK 2018 against all the evidence we present in KL 2015 and KL 2016.

2. Placebo Tests

The second observation that JJK 2018 offers is that if they run a series of overlapping placebo tests during earlier windows of time, some placebo start dates are associated with an estimated reduction in teen birth rates (JJK 2018 Table 5). We have previously responded at length to this line of critique in KL 2016, but we repeat some of that discussion here.

Our original paper did not include a formal placebo test, but we reported one in KL 2016 and we discuss it here. The data from our actual analysis in KL 2015 represents the 24 quarters that begin in 2005:Q1 and the intervention occurs in 2009:Q3. For our placebo test we use the same data eliminating those quarters after the intervention and simulate the impact of an intervention occurring in 2008:Q1 with 6 quarters after the intervention, as in our actual analysis. The results of this analysis indicate a statistically insignificant impact of the simulated intervention. The main reduced form estimate from Table 1 in KL is a coefficient (standard error) on 2008-09 MTV ratings of -3.581 (1.512). In our placebo test specification, we get -1.314 (1.696). This placebo analysis provides no indication of non-parallel trends.

Below we reproduce Figure 5 from JJK 2016, which plots data that is helpful to thinking about how to appropriately estimate placebo tests in earlier windows of time. This figure shows that if the sample window is extended back sufficiently far, there appears to be a violation of the parallel trends assumption. Namely, the data appear to follow a quadratic time trend during the 2000 to 2005 window of time. We are unable to estimate placebo models on these earlier years ourselves because currently we do not have access to the restricted birth data from these earlier years (county-level identifiers are available on public release files just for those counties with larger than 100,000 population in those years).



JJK (2016) Figure 5: Reduced Form Event Study, 1999-2013

note: The shaded region is the same used in Kearney and Levine (2015).

We suspect that the placebo tests using these earlier years of data would not fail if quadratic time trends were included in the regression model. Joyce has argued in favor of this approach in two other papers he has written reassessing the work of others. In reference to Donohue and Levitt's work on abortion and crime, Joyce (JHR, 2004) states: "even in models with state and year fixed effects, the relationship between abortion and crime may be biased by differences in within-state growth in cocaine markets over time, a classic problem of omitted variables. A crude solution is to include controls for state-specific linear or quadratic trends." In a current mimeo, Joyce and his coauthor Daniel Dench revisit Hoynes, Miller and Simon's (2015) paper on the EITC and infant health. In this reassessment, they argue for the inclusion of quadratic trends.

JJK 2018 does not include quadratic trends, but JJK 2016 did and the results support our hypothesis. In JJK 2016 Table 5 they report a series of rolling, overlapping placebo tests with the inclusion of quadratic trends. We report an extract of that table here, omitting all sample periods that include some treatment quarters, and ignoring the problem that rolling windows should not be thought of as independent tests. This table reports a negative and statistically significant impact of *16 and Pregnant* in the actual year it was introduced and no evidence of a negative and significant effect in any five-year sample window going back to the year 2000. Only when the placebo test starts in 1999 that it fails. In their current critique they dropped these quadratic trend specifications. The placebo table reported in the current version that omits quadratic trends also drops the 8 placebo tests using data from before 2001 (included in previous versions) that do not yield significant coefficient estimates.

Extract from JJK 2016 Table 5: Placebo Tests of Estimated Reduced Form and Instrumental Variables Impact on Teen Birth Rates Rolling 24 Quarter Periods with Quadratic DMA-Specific Trends

Row	Dates			Instrumental Variables	
	Begin	"Show" Start	End	Coefficient	Std. Err.
(1)	1999:QI	2003:QIII	2004:QIV	-0.207	1.168
(2)	1999:QII	2003:QIV	2005:QI	-0.788	1.460
(3)	1999:QIII	2004:QI	2005:QII	-1.787*	1.082
(4)	1999:QIV	2004:QII	2005:QIII	-1.906**	0.942
(5)	2000:QI	2004:QIII	2005:QIV	-1.576	1.889
(6)	2000:QII	2004:QIV	2006:QI	-1.036	2.084
(7)	2000:QIII	2005:QI	2006:QII	-0.020	1.383
(8)	2000:QIV	2005:QII	2006:QIII	0.700	0.705
(9)	2001:QI	2005:QIII	2006:QIV	0.404	1.177
(10)	2001:QII	2005:QIV	2007:QI	2.196**	1.006
(11)	2001:QIII	2006:QI	2007:QII	1.386	0.859
(12)	2001:QIV	2006:QII	2007:QIII	0.782	1.150
(13)	2002:QI	2006:QIII	2007:QIV	0.700	1.017
(14)	2002:QII	2006:QIV	2008:QI	-0.241	0.994
(15)	2002:QIII	2007:QI	2008:QII	-1.018	0.876
(16)	2002:QIV	2007:QII	2008:QIII	-0.724	0.574
(17)	2003:QI	2007:QIII	2008:QIV	-0.091	0.664
(18)	2003:QII	2007:QIV	2009:QI	0.375	0.539
(25)	2005:QI	2009:QIII	2010:QIV	-1.591**	0.756

Again, our results would be stronger if one could not construct any placebo test that failed. But the fact that JJK 2018 can find a time period and specification that leads to a placebo test with a significant “finding” does not necessarily mean that there is a confounding issue.

History of JJK Reassessment Efforts

JJK have been reexamining the results of KL since that paper was first released as an NBER working paper in January 2014. The first full-length draft we received from them was in September 2016 (that draft was never published) and it was motivated as a replication exercise. We note that JJK were able to replicate every result in our paper.

Having replicated our results, the initial critique paper then challenged the validity of our instrument by claiming that since the instrumental variable takes on the value of zero in the pre-*16 and Pregnant* period and then jumps to the 2008 DMA-specific level of MTV ratings in the post-period, it is mechanically correlated with the treatment variable.³ Of course, DMA-specific fixed effects remove this mechanical correlation. This critique was dropped in subsequent drafts.

Their initial draft also argued that our instrument was invalid because they could show that alternative instruments produced similar-sized estimates. But any variable that is unrelated to the residual in teen births, but is related to *16 and Pregnant* ratings is a potentially suitable instrument. These specifications, if anything, strengthened our results. These results were dropped in subsequent drafts.

The authors also examined the sensitivity of our results to estimation with and without sample weights. They found larger, but imprecise, estimates in unweighted models. The imprecision is what one would expect when smaller markets are treated the same as markets much larger in size. The results were not significantly different. Again, these findings support the robustness of our results. These results were dropped in subsequent drafts.

JJK 2016 subsequently introduced models that include DMA-specific linear and quadratic trends, as we highlighted earlier. JJK 2016 Table 4 shows that our *16 and Pregnant* result is largely robust to the inclusion of state-specific linear or quadratic trends. Their write-up of that table in JJK 2016 focused on the specification/time period combination that led to an attenuation of the estimated effect. But as we pointed out in KL 2016, the weight of the evidence reported in Table 4 actually shows how robust the finding is. These results were dropped in subsequent drafts.

For the purposes of completeness, we also note that previous versions of JJK’s reassessment, including JJK 2016, criticize our analysis of Google Trends and Twitter data for reasons that are unrelated to parallel trends. We did not address those criticisms in our IZA response because we viewed them as peripheral to the main issues at hand, but we will provide responses to them upon request. We stand by the analyses we present of that internet data in KL 2015. Without getting into the technical details of those econometric models, we note here simply that even a casual look through the Twitter data reveals numerous tweets explicitly saying some version of

³ See the bottom of page 7 in JJK 2018 for a very brief discussion of this issue with no follow-up.

“watching *16 and Pregnant* makes me want to take birth control.” Those data give the clear impression that this show affected viewers’ attitudes.

Conclusion

We are certainly open to revising our assessment of whether the MTV show *16 and Pregnant* led to a reduction in teen births, as our 2015 paper reports, based on the findings of future research. But, our reading of the multiple critiques put forward by JJK, including JJK 2018, do not lead us to alter the conclusions we put forward in KL 2015.