Sex in the (Non) City:
Teen Childbearing in Rural America

By Alison Stewart Ng and Kelleen Kaye

Technical Appendix

Our recent descriptive analyses have shown that the teen birth rate is one-third higher in rural areas compared to the rest of the country. To better understand what factors account for this disparity, we present first-of-its-kind analysis that tackles three questions:

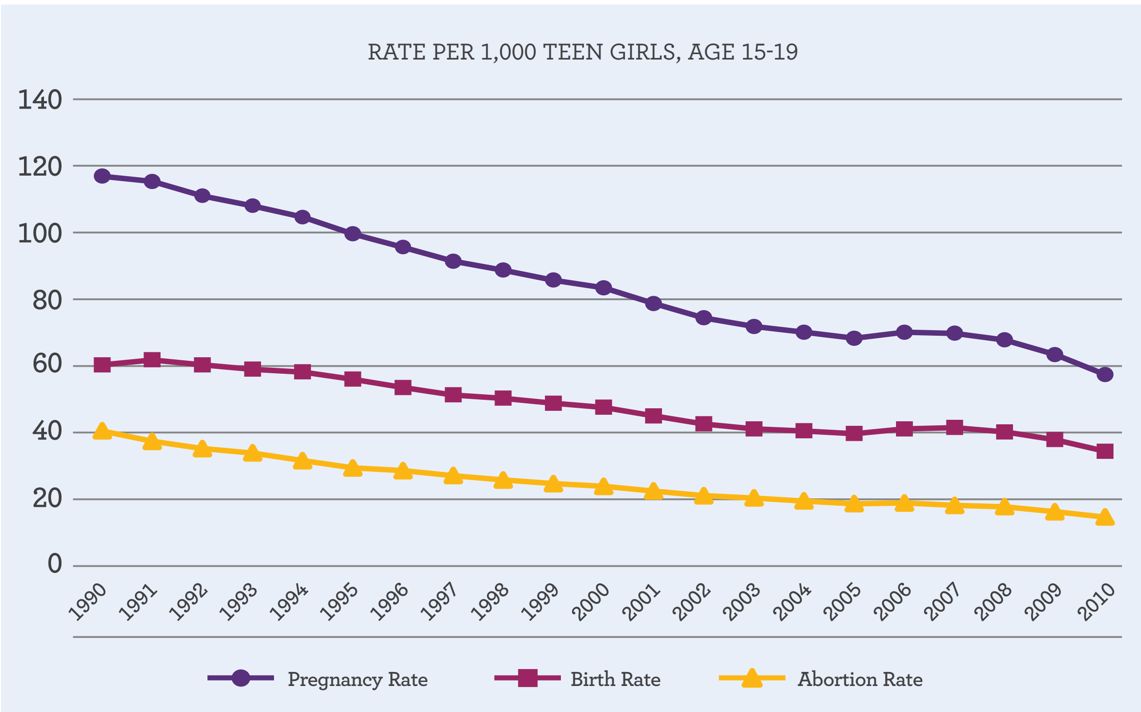
- First, what factors are significant predictors of teen childbearing in all counties—rural and metropolitan—across the United States?
- Second, which of these risk factors are more prevalent in rural counties compared to metropolitan counties?
- Third, putting one and two together, how much does each risk factor contribute to the difference between rural and metropolitan teen birth rates?

Our analytic approach is briefly summarized in the main report, and described in more detail here.

DEPENDENT VARIABLE

Ideally, our analysis would focus on the county-level teen pregnancy rate, as this is the outcome on which prevention efforts focus; however this cannot be reliably and consistently measured across all counties. Therefore, our dependent variable is defined as the *teen birth rate,* measured for every county and county-equivalent (N=3141) as of 2010. We believe patterns in teen birth rates closely reflect patterns in teen pregnancy rates, given that across the counties for which we had both measures, the two were nearly 100% correlated and this remained the case even after controlling for factors such as region, race/ethnicity, and rural status.[a] Furthermore, when we look at national trends over time, we see that rates of teen births, teen pregnancies, and teen abortions have all been falling in lock step.[1]

Figure A1. Trends in Teen Pregnancy, Birth, and Abortion Rates, 1990–2010



U.S. Teenage Pregnancies, Births and Abortions, 2010: National and State Trends by Age, Race and Ethnicity, Guttmacher Institute (Kost & Henshaw, 2014).

We recognize the possibility that in particular counties, a higher teen birth rate could reflect a greater share of pregnancies that result in a birth rather than a higher teen pregnancy rate; however, we control for this to the extent possible. Specifically, we include the state-level teen abortion ratio as a control variable, as well as a set of state fixed effects that account for any state-specific influences not already measured (including policies related to abortion). In addition, for a subset of counties for which it could be measured, we also estimated a model that further controlled for county-level attitudes toward the acceptability of abortion. Our main conclusions were robust to these added controls, suggesting that our results primarily reflect the chances that a teen gets pregnant in the first place, rather than whether a pregnant teen has a birth or an abortion.

[a] County-level abortion data are not available from a single, national source; rather, we pulled them from various state health departments. As such, reporting standards for these data vary, however we believe they are sufficiently accurate to support the basic point that patterns in teen childbearing largely mirror patterns in teen pregnancy.

DATA

The county-level teen birth rate is based on restricted-use data we obtained from the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). Independent variables are also measured primarily at the county level and draw on data from a wide variety of sources including NCHS, the Census Bureau, the Guttmacher Institute, the Department of Transportation, the Health Resources and Services Administration, and the Substance Abuse and Mental Health Services Agency, among others. These data sources are listed at the end of the Technical Appendix. Many additional data sources were explored but ultimately not included in the model, as noted later in this Technical Appendix.

Due to changes in county definitions in recent years, not every data source used the same list of counties. Our dataset was based on the list of counties from vital statistics data in 2010. Sometimes, definitions changed when a county was split into smaller counties. When vital statistics data had the larger county, we aggregated data from the smaller counties using a weighted average based on population, and assigned this average to the larger county in our dataset. When vital statistics data had the smaller counties, we assigned the data from the larger county to each of the smaller counties. However, this affected only 12 counties—less than half of one percent of the counties in our dataset. Data for the City and Borough of Wrangell and the Petersberg Census Area in Alaska were often not available, so we combined these county-equivalents into the former Wrangell-Petersberg Census Area for data completeness. In addition, we removed Kalawao County, Hawaii from our analysis because there were no girls age 15 to 19 living in that county in 2010.

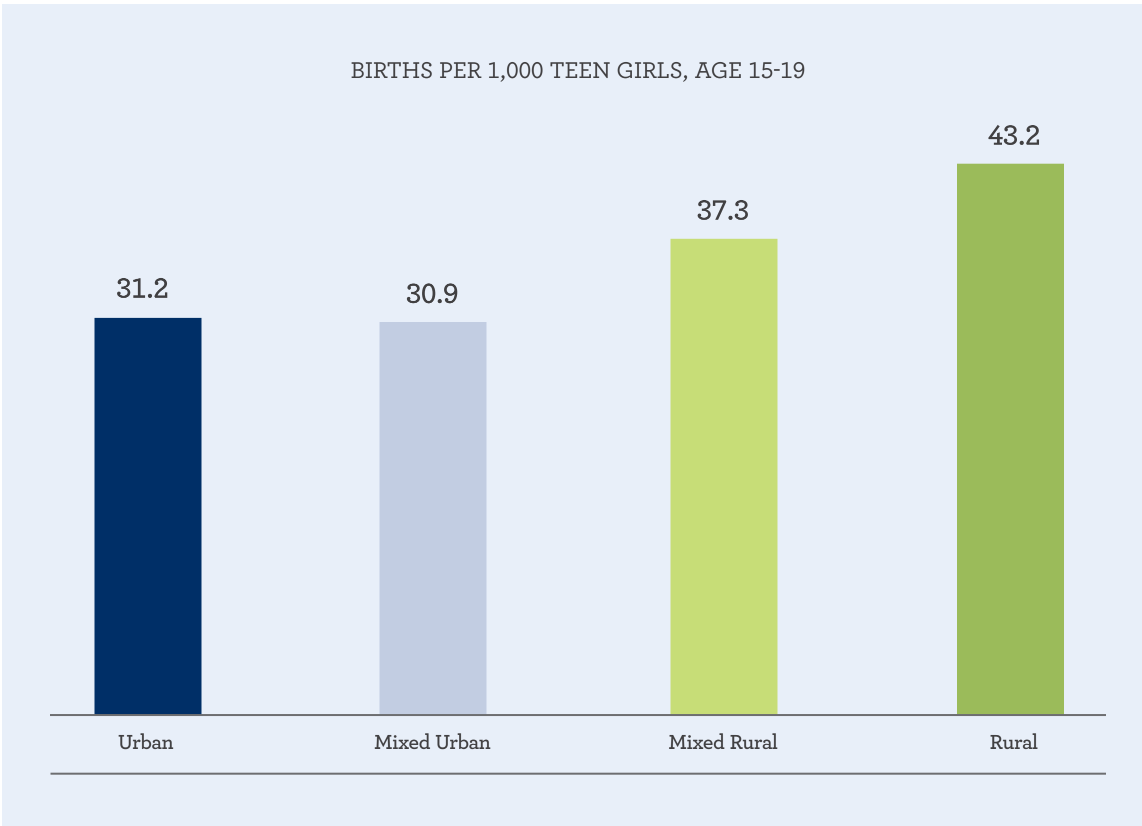
On a few occasions, some of the data for our explanatory variables were only available on a sub-county level, such as a census tract. We aggregated these data to the county level using a weighted average based on the population in each sub-county area. Whenever a sub-county boundary overlapped multiple counties, the value of the county with the most similar racial/ethnic distribution was used. This process was only applied to a few counties on a handful of measures, so the impact on the model is minimal.

RURAL DEFINITION

We considered several different definitions of rural areas. The simplest definition uses population density at the census-tract or census block level, as calculated by the U.S. Census Bureau. Data for our covariates were frequently not available at a sub-county level, so the census definition was not a feasible choice for our model. We considered other commonly used county-based definitions, including the Urban Influence Codes (UIC) and the Rural-Urban Continuum Codes (RUCC).[2] Neither of these definitions differentiate between central cities of large metropolitan areas and their large fringe counties (which are roughly comparable to suburban areas).[3] Because health outcomes are often different in metropolitan centers than in suburban areas, this was an important distinction for our research. Much of the literature recommends the use of RUCAs, or Rural Urban Commuting Areas, in this situation.[4] However, we were unable to use this definition because it is also defined at the sub-county level. As the NCHS Urban-Rural Categorization Scheme is measured at the county level and differentiates between metropolitan centers and suburban areas, we decided this definition was the best fit for our analysis.

Of course, defining rural areas at the county level has its weaknesses. There are rural census-tracts that are located in metropolitan areas, raising the possibility that our results may mischaracterize the disparity between rural and metropolitan teen birth rates. To help assess whether this is true, we also ran an analysis based on an alternative categorization, which classifies each county by the share of the population that is urban or rural, based on population density at the sub-county level.[5] Applying this to our data, we observed the same trend as in our county-based analysis—the teen birth rate increases as urbanization decreases.

Figure A2. Teen Birth Rates According to Urban-Rural Density Codes, 2010



2010 All-County Natality File (Restricted Use). National Center for Health Statistics; Urban-Rural Density Codes (Isserman, 2005).

Teen birth rates within each rural/metropolitan category published in our main report will differ slightly from those we published earlier[6] because the NCHS Urban-Rural Categorization has been updated since then to reflect metropolitan boundaries as of 2010, and is the basis of findings presented in our current report.

MEASURING TRENDS IN TEEN BIRTH RATES BY URBANIZATION

Whether a county is considered rural or metropolitan can change over time, both because the population is shifting, and because the methodology for categorizing counties has been revised periodically. When measuring trends in birth rates by urbanization, there are a number of ways to address these shifts. In the case of rural trends, for example, one could compare all rural counties at the beginning of the period to all rural counties at the end, knowing that the list of counties would be slightly different in each year. Alternatively, one could compare all rural counties at the beginning of the period to those same counties at the end of the period, knowing that a few counties would no longer be rural. Finally, one could include only those counties that were rural both at the beginning of the period and the end, knowing that some other counties would be excluded. We choose the first option, however all approaches yielded similar results, in large part because the roster of counties within each urbanization category changed relatively little over the period (for example 9% changed categorization from the 2006 scheme to the 2013 scheme).[b]

MEASURING SEXUAL ACTIVITY AND CONTRACEPTIVE USE

Results from this section come from our analysis of the National Survey of Family Growth (NSFG), released by NCHS. This in-person survey is nationally representative of the population age 15 to 44, and collects information on family formation, marriage, pregnancy, contraceptive use, and sexual activity, among other topics. Our analysis relies on the female respondent file for 2006-2010, and includes observations for girls age 15 to 19.

MODEL CHOICE FOR MULTIVARIATE ANALYSIS

Ordinary Least Squares (OLS) is often the model of choice because the results are easily interpreted; however it is not a good fit for our data, particularly given that the county-level teen birth rate cannot equal less than zero and given that the teen birth rate increasingly takes on the properties of a count variable rather than a continuous variable as the county size diminishes.[7] Furthermore, the error term is heteroskedastic, with birth rates being less precisely estimated in smaller counties. These last two issues tend to be exacerbated when looking at relatively rare events (like teen births), and are particularly salient given that our question of interest focuses on rural counties.[8] Neither natural log transformation nor OLS with the option for robust standard errors adequately addressed these concerns.[9]

We next considered using a basic Poisson model, which is more appropriate for count data. In this case, our dependent variable was defined as the number of teen births in each county, while including the female teen population in each county as the offset variable, thus still reflecting the teen birth rate in essence. However, this specification violates the core assumption of the Poisson model—that the mean and the variance should be equal.

In the end, we chose the negative binomial model, a variant of the Poisson distribution that includes the alpha parameter. This parameter allows for the amount of dispersion to vary

[b] Specifically, we used the 1990, 2006, and 2013 NCHS scheme to define rural/metropolitan status of counties in 1990, 2000, and 2010, respectively. See Ingram & Franco (2014).

based on the size of the exposure group—in our case, the size of the female teen population. The use of a negative binomial specification to model county-level rates, particularly rates of relatively rare events, is well documented in the literature. One particularly helpful article discussing the merits of this approach is Osgood (2000).[10]

Because our estimates are based on a nonlinear regression format, we did not calculate a standard R-squared as a measure of fit. Instead, we used Nagelkerke's pseudo R-squared. This measures the reduction in unexplained variance due to the explanatory variables and is adjusted to fall between 0 and 1. This means it is a roughly comparable analog of R-squared.

COVARIATES

We considered a number of measures that reflect risk and protective factors discussed in the literature and among experts in the field. We focused in particular on availability of health services, economic and educational opportunity, transit availability, recreational options, prevalence of substance abuse, and a few other community-level contextual factors, while also controlling for the county's rural/metropolitan status, its racial/ethnic composition, the state-level teen abortion ratio, and state-level fixed effects to account for any unmeasured factors unique to each state. Covariates are measured at the county level unless otherwise noted and are described more fully below.

Access to clinical services. We include five measures to reflect the availability of health services: doctors per 10,000 in the population, publicly funded clinics offering contraception within a 15 mile radius, designation as a primary care Health Professional Shortage Area (HPSA), the HPSA score for those counties that did receive this designation, and percentage of the population without health insurance.

Our measure of doctors includes physicians who work in an office setting and excludes those not actively involved in patient care (e.g. researchers and administrators) as well as federal physicians who may not serve the general public (e.g. those in the military). It is based on data from the American Medical Association master file, as reported in the Area Health Resource file (AHRF), a county-level health indicator database published by the Health Resources and Services Administration (HRSA). Our measure of clinics includes both public and private clinics receiving public funding as identified in the Guttmacher Institute's Census of Publicly Funded Clinics providing contraceptive services. These include, for example, the range of community clinics, private family planning clinics, and hospital clinics.

Both measures are aggregated to the health service area (HSA), rather than the county area. HSAs combine counties into roughly 800 clusters nationwide, and reflect areas that are relatively self-contained units as measured by hospital care. Health service areas were first defined by NCHS in the early 1990s, and our definition uses updated boundaries published by the Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute in 2008.[11] Our decision to use HSA-based measures is based on our review of the literature, which illustrated a problem with measuring access to medical care at the county level. In many cases, the closest physician or clinic might be located across county lines, so measuring access at the county level might be misleading. Indeed, a recent study examining rural shortages of providers compared county-level measures of availability to actual measures of where patients sought care and found that county-level measures underestimated the availability of medical care, particularly in rural areas.[12] Our HSA approach is consistent with a 1991 study that compared the use of counties and HSAs for measuring rural provider availability, and recommended the use of HSAs as a result.[13] Furthermore, our HSA variables provided a far better fit in our model.

Because our estimates reflect the impact of increasing each risk factor by one unit, it is important to scale each variable such that a one unit increase is sensible. For example, estimating the impact of one additional doctor per person would be unrealistic. Instead, our estimates measure the impact of one more doctor per 10,000 in the population, which is reasonable given that there is an average of 14 doctors per 10,000. Specifically, this variable is calculated as the number of doctors within the HSA, divided by the population in the HSA, times 10,000. Similarly, estimating the impact of one additional clinic per square mile would not be realistic. Instead, our estimates measure the impact of one more clinic per 700 square miles (loosely described as a radius of 15 miles for ease of interpretation), which is reasonable given an average of eight clinics per 700 square miles. Specifically, this variable is calculated as the number of clinics within the HSA divided by the number of square miles in the HSA, times 700. Note that multiplying a predictor by a constant (e.g. whether we measure doctors per person or doctors per 10,000 people) does not affect the statistical significance of our results.

We recognize that total square miles may overstate travel distances if the population and facilities are disproportionally concentrated in dense pockets within the county, irrespective of how large or small the county is, how rural or urban it is, or how populous it is. To help correct for this, the square miles within the HSA is multiplied by an index that equals (1–C), where C is a measure of concentration and is based on the following formula:[14]

$$C = \frac{1}{2}\sum\left|\frac{a_i}{A} - \frac{p_i}{P}\right|$$

where a_i and p_i are the area and population respectively for each block group i within the HSA, and A and P are the total area and population respectively for each HSA summed over all block groups in the HSA. As the population becomes more evenly distributed throughout the HSA, C will approach 0, (1–C) will approach 1, and the adjusted square area in our calculations will approach the actual square area for the HSA. However, as the population becomes more concentrated, C approaches 1, (1–C) approaches 0, and the square area used in our calculations is adjusted downward, reflecting the fact that the relevant service delivery area may be smaller than what is implied by the total HSA square area.

Ideally, one would model clinic access based on Geographic Information System (GIS) data showing the actual location of facilities in relation to population centers. This has been done in studies of more limited geographic scope, e.g. particular states or metropolitan areas.[15] However doing this across all counties in the U.S. is beyond the scope of our work. Given that this approach is not an option for us, we believe that adjusting to account for population concentration is an improvement over simply using facilities per square mile, however we recognize it is imperfect at best.

We also examined the effect of other medical providers, including nurse practitioners and physician assistants. While both of these provider types make up a relatively greater mix of the providers available in rural areas, these were not significant protective factors in our model, and we opted for measuring doctors, recognizing this may reflect the importance of a wider array of providers.

We also investigated whether the type of clinic mattered using clinic data from the Guttmacher Institute—that is, whether the clinic was affiliated with the Health Department, a Federally Qualified Health Center (FQHC), a Planned Parenthood clinic, or another type, and we tested this by Title X funding status. In both cases, we simply didn't have enough power in the data to parse out the effects in this way and these more nuanced indicators were dropped from the model in favor of the more basic indicators described above.

There are numerous other sources of data that we explored pertaining to clinic availability. The Area Health Resource File (AHRF) published by HRSA provides breakouts for specific types of public clinics—distinguishing, for example, between Community Health Centers (CHCs) and Rural Health Clinics (RHCs), while the National Association of City and County Health Officials (NACCHO) publishes an annual profile and dataset describing local health departments and the services they offer. Ultimately, the data and variables we described earlier provided the most robust fit for our model, but we note these additional sources of data can still offer valuable insights to researchers and practitioners concerning clinical services at the local level.

Finally, our measure of HPSA is defined as any county containing all or part of a primary care shortage area, including those defined based on population group, geographic area, or a single county, using the HRSA Data Warehouse. We included all HPSAs meeting our criteria that have been active since 2010 or earlier. HPSA designated facilities are not included in our measure. For our measure of the percentage of the population without health insurance, we use estimates produced by the Census Bureau within their Small-Area Health Insurance Estimates (SAHIE) program.

Economic and educational opportunities. We included five measures of economic and educational opportunity. Percentage of households that were female headed, percentage of 18-24 year-olds who ever attended college, and percentage of 16-19 year-olds who were idle (neither in college nor working) were taken from the 2005-2009 American Community Survey. We used five-year averages to create more stable estimates at the county level, and we lagged the measures by one year to avoid any potential for reverse causality. For the percentage in poverty we used the Small Area Income and Poverty Estimates (SAIPE) produced by the Census Bureau, and our measure of the overall unemployment rate came from the Local Area Unemployment Statistics (LAUS) series published by the Bureau of Labor Statistics, both lagged one year as well.

We also considered measures of deep poverty, long-term poverty and income inequality (using the Gini index).[16] In the end these measures added little to the explanatory power of the model and we opted for the most straightforward measure of percentage in poverty. In addition to the overall unemployment rate, we also considered the rate for 20-24 year-olds overall and 20-24 year-old men. These too added little to the model's explanatory power and we opted for the overall unemployment rate.

We have heard some practitioners posit that mining communities were unique from other rural communities and at particularly high risk. Rather than singling out mining communities, we include a collection of categorical variables reflecting whatever the dominant economic specialization was for that county (farming, mining, service, manufacturing, and state or federal government). While these data are from 2004, it was the most recent data available as of publication. Only mining was statistically significant, although all variables are retained in the model, with "non-specialized" being the omitted category.

Whether a community is expanding or contracting is another common concern, and to capture this we have included two measures of migration—both the share of the population who moved to the county within the past year and the net migration rate over the past 10 years (that is, number moving in minus number moving out, divided by the 2010 population).

Transit. We considered numerous measures from the National Transit Database, the Journey to Work analyses and County Business Patterns data published by the Census Bureau the Location Affordability Index program run jointly by the Department of Transportation and the Department of Housing and Urban Development, and the Smart Locations Database run by the Environmental Protection Agency. We explored alternative specifications using various combinations of the number of public transit trips per household, the percentage of people in the county using public transit to commute to work, the existence of a public transit system as recorded in either metropolitan or rural transit administrative data, the number of transit employees or establishments in the county, the percentage of households with no vehicles and the average number of vehicles per household. In the end, we found consistent support for the notion that both public and private transportation play a significant role in predicting teen births, and we chose the specific variables that fit the data best.

The three indicators of transportation included in our final model are defined as follows: 1) among households with at least two people, the percentage with less than two cars (taken from the American Community Survey); 2) a 0/1 variable indicating if a fixed-guideway (predominantly rail) transit system exists, and 3) if there is a fixed-guideway system, the percentage of employment located within half a mile of a station (both taken from the EPA's Smart Locations Database). We also considered a term interacting the percentage with less than two cars with the level of urbanization to allow the effect of private transportation to vary between rural and metropolitan areas, however this variable was insignificant and dropped from the model.

Our measures of public transit, focused on fixed-guideway systems, do not reflect the full extent of public transportation available, particularly in rural areas where modes such as bus routes and van service play an increased role.[17] As such, they may not be the best indicators for the purpose of describing transit options in a specific location. However, in the aggregate, public transit in rural areas remains rare even when these alternative modes are considered—with less than 1% of rural residents relying on public modes as a significant form of transportation, and the vast majority of rural counties reporting no public transit system.[18] Therefore, for the broader purposes of assessing the link between transportation options and teen childbearing, we believe our measures are sufficient, especially given that alternative specifications of our model incorporating the limited data we have on these other transit modes yielded largely similar results.

We also note that while our last measure—employment within a half mile of a transit station—does not directly reflect proximity between transit and services, it does offer a proxy of how extensive the public transit system is. It also offers a degree of variation across public transit systems, helping to ensure that our public transit measures don't simply capture the effect of urban centers.

Finally, we also explored variables that combined both public and private transportation into a single index, but were unable to construct a measure that was sensible and informative.

Recreation and risk. Our measure of recreational instruction facilities per 100,000 people draws on County Business Patterns data and includes establishments whose primary function is offering instruction for a wide range of athletic activities to groups or individuals. Overnight and day sports instruction camps are included, while academic institutions are not as their primary activity would be classified as education instead of sports instruction.[19]

Our measures of risky behavior include the percentage of teens who were binge drinking or using marijuana in the past month. These are derived from survey data collected by the Substance Abuse and Mental Health Services Agency (SAMHSA). We note that the SAMHSA variables are based on 10 years of pooled data and represent HRSA service areas that combine some counties, which was the only form in which these indicators were available for public download, so these should be considered general proxies at best.

Other contextual variables. It's reasonable to expect that in communities where young adults marry early—even during their late teens—they will likely start their families early as well and teen birth rates might be higher for this reason. To account for this possibility, our model includes the percentage of teen girls who are married. This measure is lagged one year to account for the potential of reverse causality (especially in the form of "shotgun" marriages).

We also included two measures of religiosity in the county: the rate of total adherents and the rate of adherents within churches that identify as evangelical, both drawing on data from the Association of Religion Data Archives.

Existing literature suggests that it may not be religious adherence or denomination per se that influences individual and family outcomes, but rather the degree of religious participation and the salience of religion in one's life.[20] To help assess this, we ran additional specifications that also included variables reflecting the average frequency of attendance at religious institutions and the average stated importance of religion. These measures draw on data from the Cooperative Congressional Election Study, fielded by Harvard University in partnership with numerous other institutions.

Results based on these additional measures were largely consistent with the literature, though here again we were constrained by serious data limitations. In particular, these additional measures reflect five years of pooled data and, even then, results were either unavailable or based on very small sample sizes for numerous counties. Therefore, we did not include these additional measures in our final model. In the end, we conclude that our analyses are unable to provide much insight into the role of religiosity.

Rural status and other basic controls. A primary motivation for this analysis is our earlier finding that teen birth rates increase as urbanization decreases, in nearly monotonic fashion. Therefore we include a 0/1 variable distinguishing the county as rural or metropolitan, to assess whether this difference remains significant after controlling for a variety of risk and protective factors or instead is explained by these factors. We found that once we included measures for the risk and protective factors of interest, the rural status variable became insignificant. We also ran an alternative specification replacing the rural/metropolitan indicator with a set of categorical variables reflecting the full rural/metropolitan continuum. Our results did not change appreciably, and we opted for the simpler model based on the 0/1 rural/metropolitan variable because these results more easily lend themselves to the decomposition analyses that follow.

We also control for the racial/ethnic composition of the population and the teen state abortion ratio, as well as state fixed effects that capture any additional factors unique to each state that are not already controlled for, including policies related to abortion. These latter variables are included, not because the role of abortion policy is a substantive focus of our work, but because access to abortion can influence how well birth rates proxy for pregnancy rates, as noted above.

Additional variable transformations. Several of our variables have non-linear relationships with the teen birth rate. For example, we found that the percentage living in poverty and the percentage uninsured have a quadratic relationship. This reflects the notion that, for example, the marginal effect of poverty on the teen birth rate attenuates at the highest levels of poverty. In order to be consistent, we tested quadratic specifications with every continuous variable, and included it wherever it was significant and sensible to do so. In addition to poverty, this includes female-headed households and educational attainment, as well as the percentage that are uninsured and idle. The exception is our measure of clinics per square mile, which also enters as a nonlinear term, although in this case we included the original term and the square root.

Comparing effect sizes. Ideally, one would be able to assess which risk factors were more important than others—e.g. whether clinic access was more important than poverty. However this is challenging given the vastly different units of measurement across risk factors. We explored a number of ways that researchers often use to produce effect sizes that are more comparable across covariates, but they were not a good fit for our data.

For example, one approach is to transform all covariates into natural logs—in this case our results would tell us the percentage change in the dependent variable given a percentage change in the independent variable. This approach was not possible with our data given that many of our covariates had a large number of zeros. This is sometimes addressed by replacing each zero with a constant that is close to zero, for example changing all zeros to a value of 0.1, however we found that our results were highly sensitive to the value we chose, and we concluded this was not a reliable transformation of the data. Another approach is to standardize the data—in this case, our results would tell us the percentage change in the dependent variable given a one standard deviation change in the independent variable. However, many of our covariates have large standard deviations, such that a change of one standard deviation would be an unrealistic change to model. A third method is sequential modeling, such that one first estimates an equation with a limited set of covariates, for example those associated with health services, then adds more explanatory variables incrementally, observing how the predictive power of the model increases at each step. However, when the covariates are correlated, as many of ours are, the results from this exercise depend on the order in which the variables are added, and thus are not terribly insightful.

In the end, we simply scaled each covariate such that a one unit change was reasonable relative to its mean value. Although this does not facilitate a comparison of effect sizes across covariates, it does yield what we consider to be credible results.

MULTIVARIATE RESULTS

As noted in the full report, we use a negative binomial specification to model the relationship between the county level teen birth rate and several independent variables, mostly measured at the county level as well. In Table A1, we show the main results of our multivariate analysis, including the Incident Risk Ratio, or IRR, and the *p* value. The IRR is calculated by raising e to the power of each coefficient, and largely corresponds to the column labeled "% Change in TBR for Each Unit Change in Risk Factor" from Table 2 (found in the main report). The exception is that Table 2 presents a combined impact for any factors that were included both linearly and nonlinearly (e.g. poverty and poverty squared), whereas these terms are listed separately in Table A1. In addition, Table A1. presents a more complete list of variables included in the model.

Table A1

Risk Factors	IRR	p value	sig
Economics, Education, and Community Growth			
% of population in poverty	1.035	0.000	***
% of population in poverty, squared	0.999	0.000	***
% of households that are female headed	1.067	0.000	***
% of households that are female headed squared	0.999	0.000	***
% of population unemployed	0.991	0.001	**
% of 18-24 year-olds ever in college	1.018	0.000	***
% of 18-24 year-olds ever in college, squared	1.000	0.000	***
% of 16-19 year-olds idle	1.008	0.000	***
% of 16-19 year-olds idle, squared	1.000	0.008	**
% net migration, 2000-2010	0.997	0.000	***
% of population living in another county last year	0.992	0.001	**
Economic Specialization (=1 if yes)			
Mining	1.059	0.029	*
Federal-State government	0.980	0.206	
Farming	0.982	0.413	
Manufacturing	1.018	0.133	
Service	1.005	0.730	
Nonspecialized (reference)			
Access to Health Services			
% of population uninsured	1.060	0.000	***
% of population uninsured, squared	0.999	0.000	***
Doctors per 10,000 people	0.998	0.024	*
Publicly funded clinics within 15 mile radius	1.000	0.109	
Publicly funded clinics, square root	0.972	0.000	***
Primary care HPSA designation (=1 if yes)	0.927	0.003	**
Primary care HPSA score	1.003	0.100	
Transportation			
% of 2+ person households with fewer than 2 cars	1.013	0.000	***
Any employment near fixed public transit stop (=1 if yes)	0.962	0.049	*
% of employment near fixed public transit stop	0.995	0.005	**
Recreation and Risk			
Recreational instruction facilities per 100,000 people	0.993	0.000	***

Table A1 (Continued)

Risk Factors	IRR	p value	sig
% of teens binge drinking past month	1.003	0.007	**
% of teens using marijuana past month	0.998	0.118	
Other			
% of teen girls married	1.005	0.001	**
% of population religious adherents	1.000	0.001	**
% of population evangelical adherents	1.000	0.000	***
% of population that is foreign born from Latin America	1.002	0.620	
Rural Status (=1 if rural)	0.997	0.820	
Control Variables			
Race/Ethnicity			
% Non-Hispanic Black	0.998	0.003	***
% Hispanic	1.005	0.000	***
% American Indian/Alaska Native	1.003	0.019	***
% Asian/Pacific Islander/Other	0.998	0.356	
% Non-Hispanic White (reference)			
State-level teen abortion ratio	0.998	0.320	

^Also includes state fixed effects, not shown

DIFFERENCES ALONG THE RURAL/METROPOLITAN CONTINUUM

Not every factor associated with higher teen birth rates is unique to rural counties. Some are prevalent in rural and metropolitan counties alike, while others are actually less prevalent in rural counties compared to metropolitan counties. The main body of our report highlights a number of these factors, showing how they vary across the full rural/metropolitan continuum. In addition, Table A2 shows the full range of the risk factors we found to be significant, and how they vary between rural and metropolitan counties.

Additional exploration of these differences by region is beyond the scope of this report, however there are some interesting patterns geographically, and we include additional columns showing rural/metropolitan differences for the four main census regions as an added informational resource.

Table A2 Variation in Risk Factors Between Rural and Metropolitan Counties, by Region

	Total		Northwest		Midwest		South		West	
	Rural	Metro	Rural	Metro	Rural	Metro	Rural	Metro	Rural	Metro
Economics and Education										
Poverty	18%	14%	14%	11%	15%	12%	22%	16%	16%	14%
Female-headed households	11%	12%	10%	11%	8%	10%	14%	13%	9%	10%
Unemployment	9%	9%	9%	8%	8%	9%	10%	9%	8%	10%
College enrollment	39%	46%	45%	53%	44%	50%	35%	43%	37%	46%
Idleness	10%	8%	8%	6%	7%	6%	12%	10%	9%	8%
Net migration	0%	7%	2%	1%	2%	4%	2%	10%	-3%	10%
From outside of county	7%	7%	6%	5%	6%	6%	7%	8%	8%	7%
Mining	6%	1%	2%	0%	2%	0%	7%	2%	10%	2%
Access to Health Services										
Uninsured	19%	16%	13%	11%	15%	13%	22%	19%	23%	18%
Doctors per 10,000 people	12	17	15	20	11	16	12	16	13	18
Clinics within 15 mile radius	3	15	6	59	2	6	4	9	3	19
HPSA Designation	55%	46%	56%	38%	51%	40%	55%	44%	62%	74%
Transportation										
Households w/fewer than 2 cars	22%	22%	23%	27%	18%	19%	25%	23%	21%	21%
Has fixed public transit stop	0%	15%	3%	48%	0%	8%	0%	8%	1%	33%
% of jobs near public transit stop	0%	1%	0%	7%	0%	0%	0%	0%	0%	2%
Recreation and Risk										
Recreational facilities per 100,000	2	3	4	5	1	3	1	3	4	4
Binge drinking	25%	24%	28%	27%	28%	27%	22%	21%	25%	23%
Other										
Marriage among teen girls	3%	2%	1%	1%	2%	2%	4%	3%	3%	3%
Religious adherents	53%	48%	37%	48%	57%	47%	55%	49%	45%	41%
Evangelical adherents	25%	20%	6%	5%	19%	16%	37%	29%	11%	10%

All factors are defined as described earlier in technical appendix.

DECOMPOSITION ANALYSIS

In a linear regression model, the average value of an outcome of interest conditional on a set of explanatory variables can be calculated as the average value of each explanatory variable (or $\bar{X}$) times its coefficient (or β), summed across all explanatory variables:

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}_1 + \cdots + \beta_n \bar{X}_n \tag{1}$$

Where n equals the number of explanatory variables. By extension, the difference between the average Y for group a and the average Y for group b equals the average X for group a (or $\bar{X}_a$) minus the average X for group b (or $\bar{X}_b$), times its coefficient, summed over all X's:

$$\bar{Y}_a - \bar{Y}_b = \beta_1 * \left(\bar{X}_{a1} - \bar{X}_{b1} \right) + \cdots + \beta_n * \left(\bar{X}_{an} - \bar{X}_{bn} \right) \tag{2}$$

Furthermore, the difference in Y between one group and another that is attributable to how much a particular factor varies between those two groups can be calculated as $\beta_1 * \left(\bar{X}_{a1} - \bar{X}_{b1} \right)$, using X_1 as an example.

While this cannot be replicated exactly within a negative binomial equation, it can be approximated fairly closely. With a negative binomial, the mean of Y conditional on our variables of interest is instead equal to:[21]

$$\frac{1}{N} * \sum_{i=1}^{N} \exp(\beta_1 X_{1i} + \cdots + \beta_n X_{ni}) \tag{3}$$

Where i represents each observation, N equals the total number of observations, and n equals the number of explanatory variables in the model. Evaluating this expression at the mean of each X does not equal the arithmetic mean of Y as in equation (1) but through additional calculations (not shown), it can be demonstrated that it does equal the *geometric* mean of Y conditional on X. Based on this and taking the natural log of both sides, we approximate the contribution of a given risk factor to the rural/metropolitan disparity in teen birth rates as follows, using factor X_1 as an example:

$$\frac{\beta_1 * \left(\bar{X}_{rural,1} - \bar{X}_{metro,1} \right)}{\beta_1 * \left(\bar{X}_{rural,1} - \bar{X}_{metro,1} \right) + \cdots + \beta_n * \left(\bar{X}_{rural,n} - \bar{X}_{metro,n} \right)} \tag{4}$$

As an added check, we also estimated our multivariate model using ordinary least squares, which is linear and enables us to do a more traditional breakdown consistent with equation (2). These alternative results (not shown) were largely consistent with the results we present in the body of our report, lending credibility to our conclusions.

This analysis estimates how much of the difference in teen birth rates between rural and metropolitan counties is attributable to, say, the fact that the average poverty rate is higher in rural counties compared to metropolitan, while assuming that each additional percentage point in the poverty rate has the same impact everywhere—that is, for each risk factor, we allow $\bar{X}$ to vary between rural and metropolitan counties but we hold β constant. Decomposition analysis can be used to take account of rural/metro differences in both $\bar{X}$ and β for each risk factor,[22] however these calculations are difficult in a negative binomial framework. Nonetheless, it is possible to assess the role of both $\bar{X}$s and βs for the set of risk factors *taken as a whole*. Using a technique that extends this decomposition methodology to non-linear equations[23] we determined that the vast majority of the rural/metropolitan disparity in teen birth rates is due to differences in the prevalence of risk factors (that is, $\bar{X}$s) rather than differences in their impact (that is, the βs) —between 89% and 98% depending on how certain parameters are specified. These results suggest that our reliance on a single set of βs in the decomposition results we present is reasonable.

VARIABLE SOURCES

The specific data sources we use to construct our variables are listed below:

- National Center for Health Statistics
  - » *Births:* National Center for Health Statistics. (2013). 2010 All-County Natality File (Restricted Use) [Data file]. Available from http://www.naphsis.org/programs/vital-statistics-data-research-request-process; National Center for Health Statistics. (2003). 2000 All-County Natality File (Restricted Use) [Data file]. Available from http://www.naphsis.org/programs/vital-statistics-data-research-request-process; National Center for Health Statistics. (1996). 1990 All-County Natality File (Restricted Use) [Data file]. Available from http://www.naphsis.org/programs/vital-statistics-data-research-request-process.
  - » *Urbanization:* Ingram, D.D., & Franco, S.J. (2014). 2013 NCHS Urban–Rural Classification Scheme for Counties. Vital and Health Statistics, *2*(166), 1-81. Retrieved from website http://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf; Ingram, D.D., & Franco, S.J. (2012). NCHS Urban–Rural Classification Scheme for Counties. Vital and Health Statistics, *2*(154), 1-72. Retrieved from website http://www.cdc.gov/nchs/data/series/sr_02/sr02_154.pdf.
  - » *Population, Race/Ethnicity:* The National Center for Health Statistics (NCHS). (2011). *April 1, 2010 Bridged-Race Population Estimates* [Data file]. Available from http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm; The National Center for Health Statistics. (2003). April 1, 2000 Bridged-Race Population Estimates [Data file]. Available from http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm; Centers for Disease Control and Prevention (CDC). (2006). *Modified Age, Race, and Sex Data From The 1990 Census* [Data file]. Available from http://wonder.cdc.gov/.
  - » *Sexual Activity, Contraceptive Use:* The National Center for Health Statistics. (2014). The National Survey of Family Growth, 2006-2010 Female Respondent File [Data file]. Available from http://www.cdc.gov/nchs/nsfg.htm.
- Economic Research Service
  - » *Net migration, Economic Specialization, Square Mileage:* U.S. Department of Agriculture, Economic Research Service. (2011). The Atlas of Rural and Small Town America [Data file]. Available from http://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america.aspx.

- The Association of Religion Data Archives (ARDA)
  - » *Religious Adherence, Evangelical:* The Association of Religion Data Archives (ARDA). (2012). U.S. Religion Census: Religious Congregations and Membership Study, 2010 (County File) [Data file]. Available from http://www.thearda.com/Archive/Files/Descriptions/RCMSCY10.asp.
- The SEER program at the National Cancer Institute
  - » *Health Service Areas:* Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute. (2008). Health Service Areas (HSA) [Data file]. Available from http://seer.cancer.gov/seerstat/variables/countyattribs/hsa.html.
- U.S. Census Bureau
  - » Education, Female-headed households, Idle, Married teens: U.S. Census Bureau. (2014). 2005-2009 American Community Survey 5-Year Estimates [Data file]. Available from http://factfinder2.census.gov.
  - » Foreign Born, Lived outside of county, Vehicles: U.S. Census Bureau. (2014). 2007-2011 American Community Survey 5-Year Estimates [Data file]. Available from http://factfinder2.census.gov.
  - » *Unemployment:* Bureau of Labor Statistics (BLS). (2010). Local Area Unemployment Statistics (LAUS): Labor force data by county, 2009 annual averages [Data file]. Available from http://www.bls.gov/lau/.
  - » *Poverty:* U.S. Census Bureau. (2010). Small Area Income and Poverty Estimates (SAIPE): State and County Estimates for 2009 [Data file]. Available from http://www.census.gov/did/www/saipe/index.html.
  - » *Uninsured:* U.S. Census Bureau. (2012). Small Area Health Insurance Estimates (SAHIE): State and County Estimates for 2009 [Data file]. Available from http://www.census.gov/did/www/sahie/.
  - » *Region:* U.S. Census Bureau. (2013). Census regions and divisions of the United States. Retrieved from http://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.
  - » *Recreational Instruction Facilities:* U.S. Census Bureau. (2013). *County Business Patterns (CBP)* [Data file]. Available from http://www.census.gov/econ/cbp/.
- Substance Abuse and Mental Health Services Administration (SAMHSA)
  - » *Binge Drinking, Marijuana Use:* Substance Abuse and Mental Health Data Archive (SAMHDA). (2013). National Survey on Drug Use and Health (NSDUH): 10-Year Substate R-DAS (2002 to 2011) [Data file]. Available from https://www.icpsr.umich.edu/icpsrweb/content/SAMHDA/index.html.
- Guttmacher Institute
  - » *Clinics:* Frost, J.J., Frohwith, L., & Zolna, M.R. (2013). Contraceptive Needs and Services, 2010: Detailed County Tables [Data file] New York: Guttmacher Institute. Available from http://www.guttmacher.org/pubs/win/index.html.
  - » *State teen abortion ratio:* Kost, K., & Henshaw, S. (2014). *U.S. Teenage Pregnancies, Births and Abortions, 2010: National and State Trends by Age, Race and Ethnicity*. New York: Guttmacher Institute. Retrieved from http://www.guttmacher.org/pubs/USTPtrends10.pdf.

- Health Resources and Services Administration:
  - » *Doctors:* Health Resources and Services Administration (HRSA). (2013). 2011-2012 Area Health Resources File (AHRF) Access System [Data file]. Available from http://ahrf.hrsa.gov/.
  - » *HPSA:* Health Resources and Services Administration (HRSA). (2014). HRSA Data Warehouse [Data file]. Available from http://datawarehouse.hrsa.gov/.
- U.S. Environmental Protection Agency (EPA)
  - » *Employment near transit:* U.S. Environmental Protection Agency (EPA). (2014). Smart Location Database (SLD) [Data file]. Available from http://www.epa.gov/smartgrowth/webinars/2014-0515-smart-location-database.html.
- State Public Health Departments
  - » *Abortions, Pregnancies where available via linked state public health department websites:* U.S. Department of Health and Human Services (HHS). (2015). State Departments of Public Health. Retrieved from website http://www.foodsafety.gov/about/state/.

ADDITIONAL RESOURCES ON RURAL AMERICA

The following is a partial list of the resources we consulted during the course of our research that may be of interest to the reader. Please note that it is not intended to be a comprehensive list of all of resources available on rural America, and that the links below are valid as of the date of publication.

Additional County-Level Data:

- U.S. Census Bureau, American Community Survey http://factfinder.census.gov/
- Guttmacher Institute, Contraceptive Needs and Services, 2010 County-Level Tablemaker: http://www.guttmacher.org/pubs/win/counties/
- U.S. Department of Housing and Urban Development (HUD), Location Affordability Index: http://www.locationaffordability.info/lai.aspx
- Federal Transit Administration, National Transit Database (NTD): http://www.ntdprogram.gov/ntdprogram/
- Health Resources and Services Administration (HRSA), HRSA data warehouse (HDW): http://datawarehouse.hrsa.gov/
- Commuting (Journey to Work) Data, U.S. Census Bureau: http://www.census.gov/hhes/commuting/
- Cooperative Congressional Election Study, Harvard University: http://projects.iq.harvard.edu/cces/data
- National Profile of Local Health Departments (LHDs), National Association of County and City Health Officials: http://profile-iq.naccho.org/

Resources for Rural Communities:

- Rural Assistance Center: http://www.raconline.org/
- Health Resources and Services Administration (HRSA), Office of Rural Health Policy: http://www.hrsa.gov/ruralhealth/
- National Association of City and County Health Officials (NACCHO): http://www.naccho.org/
- National Association of Community Health Centers: http://www.nachc.com/
- Health Resources and Services Administration (HRSA), About Health Centers: http://bphc.hrsa.gov/about/index.html
- National Association of Rural Health Clinics: http://narhc.org/
- National Rural Health Association: http://www.ruralhealthweb.org/
- HPSA Designation Criteria and Guidelines: http://bhpr.hrsa.gov/shortage/hpsas/
- HPSA State Primary Care Offices: http://bhpr.hrsa.gov/shortage/hpsas/primarycareoffices.html
- Links to each state public health department: http://www.foodsafety.gov/about/state/
- Centers for Medicare and Medicaid Services Rural Health Clinics Center http://www.cms.gov/Center/Provider-Type/Rural-Health-Clinics-Center.html
- Health Resources and Services Administration (HRSA), About Health Centers: http://bphc.hrsa.gov/about/index.html

Rural Health Research Resources:

- Rural Health Gateway: http://www.ruralhealthresearch.org/
- Current and former Rural Health Research Centers (RHRCs): http://www.ruralhealthresearch.org/centers/
- A Compendium of Research and Policy Analysis Studies of Rural Health Research and Policy Analysis Centers: http://www.raconline.org/pdf/research_compendium.pdf
- Carsey School of Public Policy at the University of New Hampshire [Formerly the Carsey Insitute]: http://carsey.unh.edu/
- Southwest Health Research Center at Texas A&M University, Rural Health 2020 Project: http://sph.tamhsc.edu/srhrc/rhp2020.html
- Economic Research Service: Rural Economy and population: http://www.ers.usda.gov/topics/rural-economy-population.aspx
- U.S. Department of Agriculture (USDA) Rural Development: http://www.rd.usda.gov/
- The Henry J. Kaiser Family Foundation: http://kff.org/

Sources

1 Kost, K., & Henshaw, S. (2014). *U.S. Teenage Pregnancies, Births and Abortions, 2010: National and State Trends by Age, Race and Ethnicity*. New York: Guttmacher Institute. Retrieved from http://www.guttmacher.org/pubs/USTPtrends10.pdf.

2 Ricketts, T.C., Johnson-Webb, K.D., & Taylor, P. (1998). *Definitions of rural: A handbook for health policy makers and researchers*. Chapel Hill, NC: Cecil G. Sheps Center for Health Services Research at the University of North Carolina at Chapel Hill.

3 Ingram, D.D., & Franco, S.J. (2014). 2013 NCHS Urban–Rural Classification Scheme for Counties. *Vital and Health Statistics, 2*(166), 1-81. Retrieved from website http://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf.

4 Hart, L.G., Larson, E.H., & Lishner, D.M. (2005). Rural Definitions for Health Policy and Research. *Americal Journal of Public Health, 95*(7), 1149-1155.

5 Isserman, A.M. (2005). In the national interest: Defining rural and urban correctly in research and public policy. *International Regional Science Review, 28*(4), 465-499.

6 Ng, A.S., & Kaye, K. (2013). *Science Says 47: Teen childbearing in rural America*. Washington, DC: The National Campaign to Prevent Teen and Unplanned Pregnancy. Retrieved from https://thenationalcampaign.org/resource/science-says-47.

7 O'Hara, R.B., & Kotze, D.J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution, 1*(2), 118-122.

8 Osgood, D.W. (2000). Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology, 16*(1), 21-43.

9 O'Hara & Kotze, 2010.

10 Osgood, 2000; Berk, R., & MacDonald, J.M. (2008). Overdispersion and poisson regression. *Journal of Quantitative Criminology, 24*(3), 269-284; Lord, D., Levine, N., & Park, B.J. (2013). Chapter 16: Poisson Regression Modeling. In *Crime Stat IV: A spatial statistics program for the analysis of crime incident locations*. Washington, DC: The National Institute of Justice. Retrieved from http://www.nij.gov/topics/technology/maps/Documents/crimestat-files/CrimeStat%20IV%20Chapter%2016.pdf.

11 Makuc, D.M., Haglund, B., Ingram, D.D., Kleinman, J.C., & Feldman, J.J. (1991). Health service areas for the United States. *Vital Health Statistics,* (112), 1-102; Author tabulations based on data in Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute. (2008). *Health Service Areas (HSA)* [Data file]. Available from http://seer.cancer.gov/seerstat/variables/countyattribs/hsa.html.

12 Rosenthal, M.B., Zaslavsky, A., & Newhouse, J.P. (2005). The geographic distribution of physicians revisited. *Health Services Research, 40*(6 Pt 1), 1931-1952.

13 Makuc, D.M., Haglund, B., Ingram, D.D., Kleinman, J.C., & Feldman, J.J. (1991). The use of health service areas for measuring provider availability. *Journal of Rural Health, 7* (Suppl. 4), 347-356.

14 Stephan, G.E. (1977). An Empirically Standardized Measure of Population Concentration. *Geographical Analysis, 9*(3), 292-296.

15 Yerramilli, S., & Fonseca, D.G. (2014). Assessing geographical inaccessibility to health care: Using GIS network based methods. *Public Health Research, 4*(5), 145-149; Mao, L., & Nekorchuk, D. (2013). Measuring spatial accessibility to healthcare for populations with multiple transportation modes. *Health & Place, 24*, 115-122.

16 Bee, A. (2012). *Household income inequality within U.S. counties: 2006-2010.* (American Community Survey Briefs ACSBR/10-18). Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration. Retrieved from http://www.census.gov/prod/2012pubs/acsbr10-18.pdf; Farris, F.A. (2010). The Gini index and measures of inequality. *The American Mathematical Monthly, 117*(10), 851-864.

17 American Public Transportation Association (APTA). (2012). *Rural communities: Expanding horizons.* Retrieved from http://www.apta.com/resources/reportsandpublications/Documents/Rural-Communities-APTA-White-Paper.pdf.

18 Brown, D.M. (2004). *Public transportation on the move in America.* Retrieved from the U.S. Department of Agriculture website http://pubs.nal.usda.gov/sites/pubs.nal.usda.gov/files/publictrans_0.htm.

19 U.S. Census Bureau. (2014). *Industry statistics portal: Business data from the U.S. census bureau.* Retrieved from http://www.census.gov/econ/isp/index.php.

20 Lammers, C., Ireland, M., Resnick, M., & Blum, R. (2000). Influences on adolescents' decision to postpone onset of sexual intercourse: A survival analysis of virginity among youths aged 13 to 18 years. *Journal of Adolescent Health, 26*(1), 42-48; Brewster, K.L., Cooksey, E.C., Guilkey, D.K., & Rindfuss, R.R. (1998). The changing impact of religion on the sexual and contraceptive behavior of adolescent women in the United States. *Journal of Marriage and Family, 60*(2), 493-504.

21 Bauer, T.K., & Sinning, M. (2006). *An extension of the Blinder-Oaxaca decomposition to non-linear models.* (RWI Discussion Papers, No. 49). Essen Germany: Leibniz Information Centre for Economics. Retrieved from http://hdl.handle.net/10419/18600.

22 Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal, 8*(4), 453-479.

23 Sinning, M., Hahn, M., & Bauer, T.K. (2008). The Blinder–Oaxaca decomposition for nonlinear regression models. *The Stata Journal, 8*(4), 480-492.

The National Campaign
to Prevent Teen and Unplanned Pregnancy

Mission

Our mission is to improve the lives and future prospects of children and families and, in particular, to help ensure that children are born into stable, two-parent families who are committed to and ready for the demanding task of raising the next generation.

Our strategy is to prevent teen pregnancy and unplanned pregnancy, especially among single, young adults. We support a combination of responsible behavior by both men and women and responsible policies in both the public and private sectors.

When we are successful, child and family wellbeing will improve. There will be less poverty, more opportunities for young men and women to complete their education or achieve other life goals, fewer abortions, and a stronger nation.

www.TheNationalCampaign.org

www.Bedsider.org

www.StayTeen.org