The role of states in changing disparities in the age-specific risk of first birth

Abstract

Unconditional age specific rates are insufficient to study disparities in the timing of life course events when prior life course experiences have varied between the categories under study. In the study of fertility, conditional fertility rates require denominators defined by parity and estimating parity-specific populations for demographic subgroups for subnational regions is difficult. The main objective of this paper is to analyze between-state variation in changing disparities in the timing of first birth. In order to do so, we develop and assess a method to estimate parity-specific birth rates for subgroups by state. Specifically, we estimate age specific risks of first birth (ASFR1) by state and year for demographic subgroups using the American Community Survey. Preliminary findings indicate that, while national-level comparisons show a convergence in ASFR1s between women with varying levels of advantage, this convergence is happening more rapidly in some states and very slowly in others.

Rapid declines in teen fertility have been widely heralded by policymakers and scholars, but a concurrent emerging attention on repeat teen births identifies a conundrum: unconditional age specific rates are insufficient to study disparities in the timing of life course events when prior life course experiences may have varied among the categories under study. However, constricting conditional fertility rates requires both numerators (births) and denominators (populations) by parity. While we have parity indicators on birth certificates in the US, we have no administrative source for population estimates by parity, particularly not by parity, state, age, sex, and other demographic indicators such as race/ethnicity, poverty, or education. This lack of denominators hinders our ability to study disparities in fertility experience in a manner which accounts for prior life course events.

The main objective of this paper is to analyze how states vary over time in between-group differences in the age-specific risk of first birth. States have recently received increased attention for their variation in health and mortality (Chetty et al., 2016; Montez, Zajacova, & Hayward, 2016), and states vary in unconditional age patterns of fertility (Kost & Henshaw, 2014; Martin, Hamilton, & Osterman, 2018) and rates of mistimed and unwanted births (Finer & Zolna, 2014). States also vary in their generosity toward publicly-subsidized family planning (Hasstedt, Sonfield, & Gold, 2017) and the levels of economic opportunity they may afford their residents (Chetty & Hendren, 2018). Thus, for less-advantaged residents, both the means to control fertility and the benefits of delaying fertility vary across states. However, most studies of between-group differences in fertility, including race/ethnic and educational differences, focus on the US as a whole, rather than differences between groups within states. Given the high level of variation across states and the uneven distribution of groups across states, attention to between-group fertility differences within states might illuminate the divers of between-group differences at the national level. Such an examination could also illuminate theories of demographic change like the proposed second demographic transition (Lesthaeghe, 2010) or help formulate new theories of contemporary demographic change in the United States (Zaidi & Morgan, 2017). Practically, group-specific estimates of fertility by state and parity-specific estimates of populations at the state level could potentially inform population projections at multiple levels of geographic aggregation, which rely on fertility estimates.

Because the majority of births are first and second births, and because the share of all births that are first births is increasing, Type I first birth rates drive overall fertility trends (Morgan, 1996). The age pattern of risk of first birth was a subject of substantial demographic inquiry in the 1980s (Chen & Morgan, 1991), and some interest in this risk has continued as the United States illustrated an idiosyncratic bimodal pattern in this rate in the 1990s (Sullivan, 2005). Sullivan found that this bimodality was driven by inequality, particularly by differences in first birth rates by educational attainment, but she found that these differences were attenuating by the early 2000s. More recently, Latin American nations have continued to exhibit bimodal patterns of first birth rates (E. de Lima, Zeman, Castro, Nathan, & Sobotka, 2016; E. E. C. Lima, Zeman, Sobotka, Nathan, & Castro, 2018) and continued differences between highly educated Blacks and Whites in the US have been noted (Nitsche & Bruckner, 2018).

This paper employs published National Center for Health Statistics annual state-specific age-specific fertility rates (ASFRs) from 2007-2016 and estimates annual and pooled-year ASFRs and risks of first birth (ASFR1s) for the US as a whole and for states by demographic subgroups for the years 2005-2016. This abstract focuses on differences by poverty, but the full paper will apply the same method to examine between-state variation in differences in age specific risk of first birth by educational attainment, race/ethnicity, rural/urban residence, and union status.

Methods

In order to examine state-level disparities in the risk of first birth, we develop and assess a method to estimate unconditional and parity-specific birth rates for subgroups by state using population representative survey data from the ACS. Specifically, we estimate age specific fertility rates and age-specific risks of first birth (ASFR1) by state and year for demographic subgroups. We identify women who have had a birth in the past 12 months based on the ACS question, asked of all women 15-44, “Has this person given birth in the past 12 months”. Unconditional age specific fertility rates (ASFRs) are estimated in order to compare with published statistics from NCHS.

In order to estimate conditional age specific risk of first birth (ASFR1), we must identify all first births to women and an estimate of women at risk of a first birth among reproductive age women in the ACS. We identify women's parity using coresident children and we decrement the parity-specific denominator to account for children who have moved away from their mothers in adolescence. Briefly, we identify all coresident children for every reproductive-age woman based on each child's and woman's relationship to the ACS respondent. We consider a woman potentially nulliparous if she has no coresident children.

To account for women whose children have moved out of the house, we estimate the rates at which children move out at ages 15 and beyond and we recursively apply these rates to reduce the potentially nulliparous count to an estimate of the number of women who are actually nulliparous at each age or age group. We use this estimate as our ASFR1 denominator. Our ASFR1 numerator is based on the number were identified as having given birth above and who we consider potentially nulliparous because they have no coresident children. We use cohort progression to recursively decrease this number to account for women who had a birth and are observed to have no coresident children because all of their children have moved out. We use sample size and population estimates to estimate standard errors for all our estimated rates following published guidelines for obtaining standard errors for derived estimates in the ACS from the Census Bureau (Gardner, 2010).

A more complete description of our process for estimating parity and population at risk of first birth is available in Appendix A.

Results

Performance of ASFR estimates from the ACS

In order to assess the validity of our basic approach, we compare our ACS-based unconditional estimates to statistics published by NCHS for each state and year 2007-2016. Our estimates have variable performance, with the ACS sample size driving the conformation of our estimates to the published rate. Examining the deviations between the published rates and our estimates by age-group (Figure 1), our estimates over-estimate risk of first birth at 15-19 years and 20-24 years. The deviation is not as systematic for 25-29 years and 30-34 years, and it reverses in the older age groups. At both 35-39 years and 40-44 years we appear to slightly over-estimate the risk, even at high sample sizes. Within each age group, deviations are highly variable. Unsurprisingly, the variability increases as fertility rates in an age group increase. The most variable deviation is for the age group 30-34 years, with a mean deviation of -11.9 (19.95 SD). The least variable is 15-19 years, with a mean of 9.5 (8.3 SD).

In order to illustrate state-specific trends in the conformation of our estimates to NCHS statistics, we display unconditional ASFR estimates from the ACS with 95% error bars together with NCHS published statistics for a selection of states by year, 2007-2016. Figure 2(a-e) displays this for California, Colorado, Georgia, New York, and Texas. In general, NCHS statistics are within the 95% error bars, even in smaller states. However, the underestimation of our estimates at earlier ages and overestimation of our estimates at older ages is evident in these graphs.

The next version of the paper will similarly compare ASFR1s for each state and year for which they are available.

State ASFR1 estimates

Preliminary findings from our ACS-based estimates of ASFR1 by subgroup indicate that, while national-level comparisons show a convergence in ASFR1s between women with varying levels of advantage, this convergence is happening more rapidly in some states and very slowly in others. In this abstract, results are presented for differences in fertility between poor (below 185% of the Federal Poverty Level - FPL) and non-poor (above 185% FPL). We focus on differences in fertility by poverty in part because our method allows ASFR1s by income to be calculated at the state level for the first time and also because 185% FPL is the cut-off for Medicaid eligibility in many states or Emergency Medicaid eligibility for childbirth in states which have not expanded Medicaid.

In order to illustrate the time-trends in US ASFR1 over the period under study, Figure 3(a-c) displays annual conditional risk of first birth (ASFR1) for all women and by poverty status for 2005-2016. The first panel shows ASFR1 for all US women and demonstrates large decreases in risk of first birth at younger ages and relatively little change in risk of first birth at older ages. The shape of the curve on the left-hand side demonstrates less-rapid rates of increase, consistent with attenuating bimodality over this period. Comparing the panels for poor women (FPL<185%) and non-poor women (FPL>185%), it is clear that the largest declines in the risk of first birth at younger age groups are among poor women. Even at the latest year, 2016, the modal age at first birth is one age-group younger among poor women compared with non-poor women. This decomposition also shows the reason for the relatively flat, almost bimodal shape of age-specific risk of first birth in the earlier years – poor women and non-poor women were exhibiting very distinct patterns at that time. In the intervening years, while they still exhibit distinct patterns, the age pattern of risk of first birth has become more similar for these two groups at the national level.

Turning to differences within states, Figure 4(a-b) displays annual ASFR1 curves by poverty status for 2005-2016 for two large states, California and Texas. Analogous curves for 5 year periods (2005-2010 and 2011-2016) are displayed in Figure 4c for a smaller state, Colorado. These figures illustrate that the ASFR1 curves for poor and non-poor women started out more similar in shape and scale in California than in Texas and nearly converged in California, while remaining distinct in shape, mode, and to a lesser degree scale, in Texas. The case of Colorado illustrates a challenge of this method for smaller states. Smaller sample sizes make estimates of rates such as these unstable for subgroups. Aggregating 5 years, we still see less stability in Colorado's estimated rates compared with California and Texas. However, the figure demonstrates a trend which appears to be somewhere between Texas' non-convergence and California's near-convergence.

Sorting US states by either vote percentages for Obama or Clinton or per capita spending on subsidized family planning, we notice that California appears to be part of a group of states in which poor and non-poor ASFR1 curves nearly converge by the end of the period. By contrast, Texas appears to be part of a group of states where poor and non-poor ASFR1 curves begin with more distinctly different age patterns at the beginning of the period and, while they become less distinct over time, their convergence is less advanced by the end of the period. The full paper will illustrate this pattern by computing summaries of the estimated ASFR1 curves by subgroup (extending beyond poverty to compare groups by educational attainment, race/ethnicity, rural/urban residence, and union status). We plan for these summaries to include parameterizations of the curves so that these comparisons and the trends we summarize may be quantified.

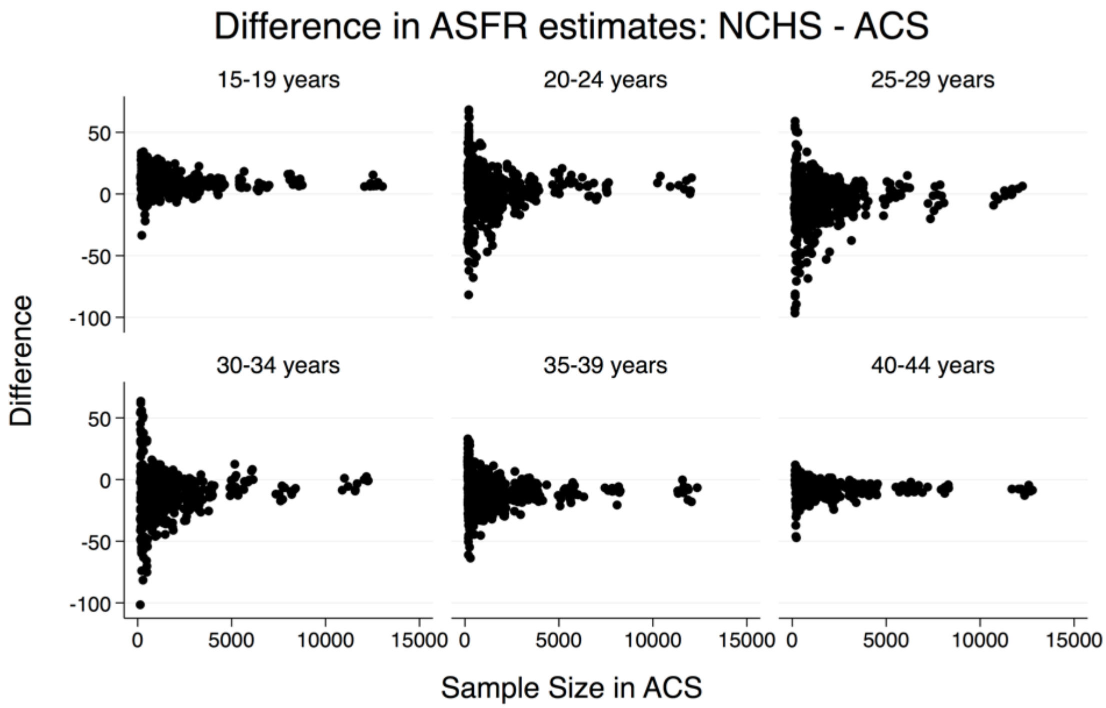

Figure 1. Differences between annual ASFR estimates from ACS and published NCHS ASFRs, 2006-2016

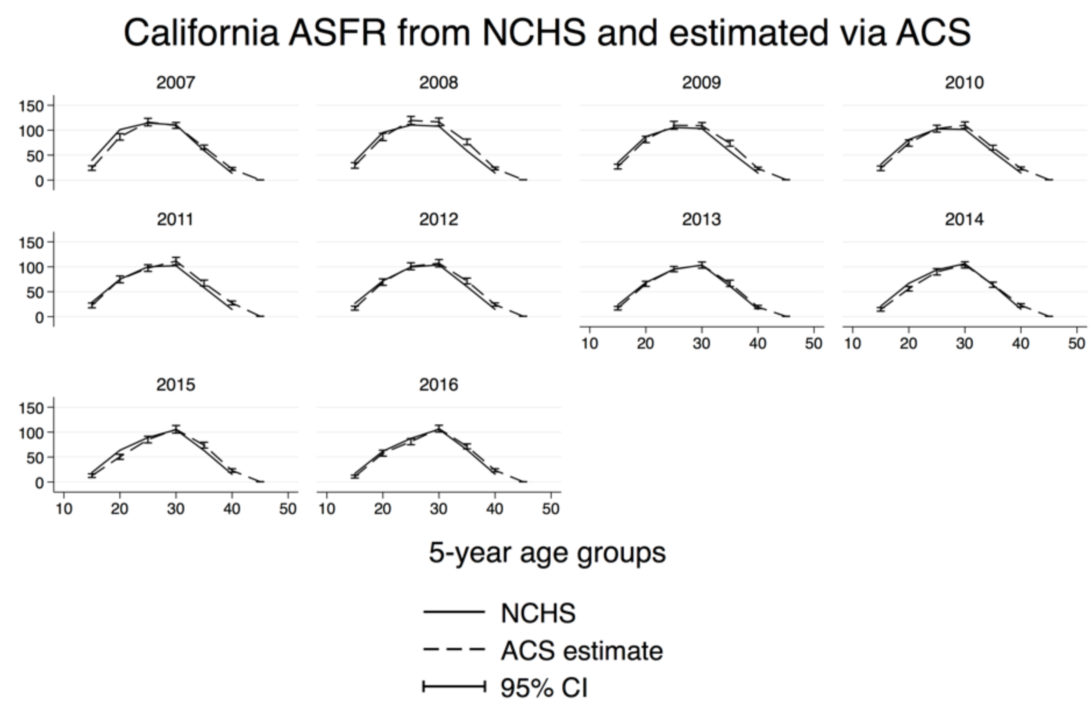

Figure 2a. Comparison of annual California ASFRs, 2007-2016

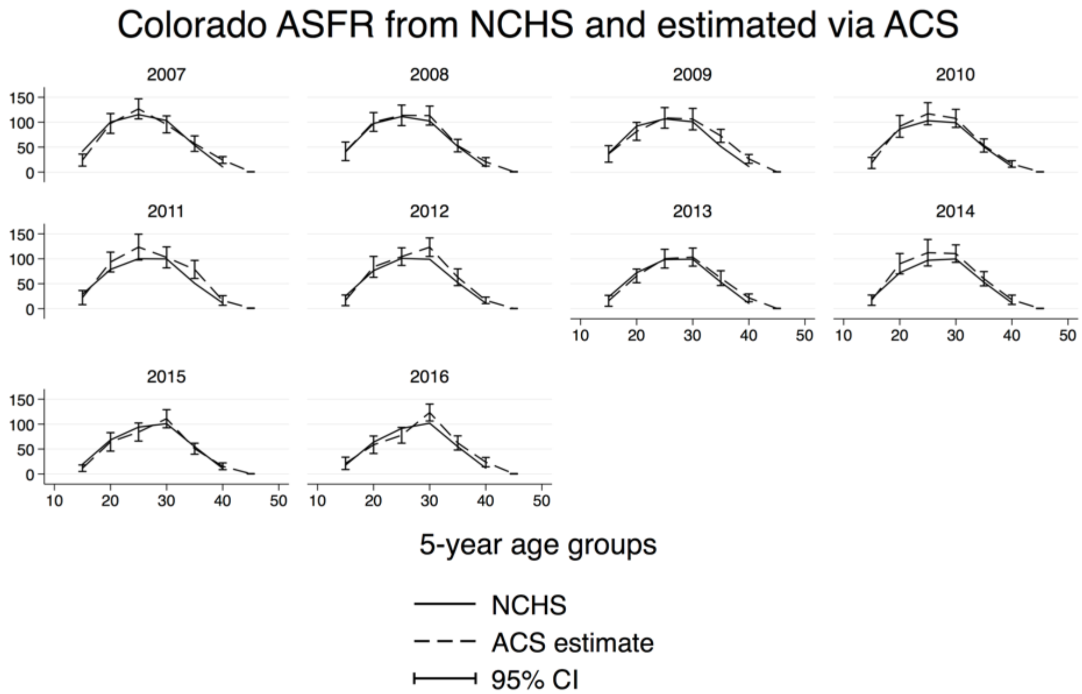

Figure 2b. Comparison of annual Colorado ASFRs, 2007-2016

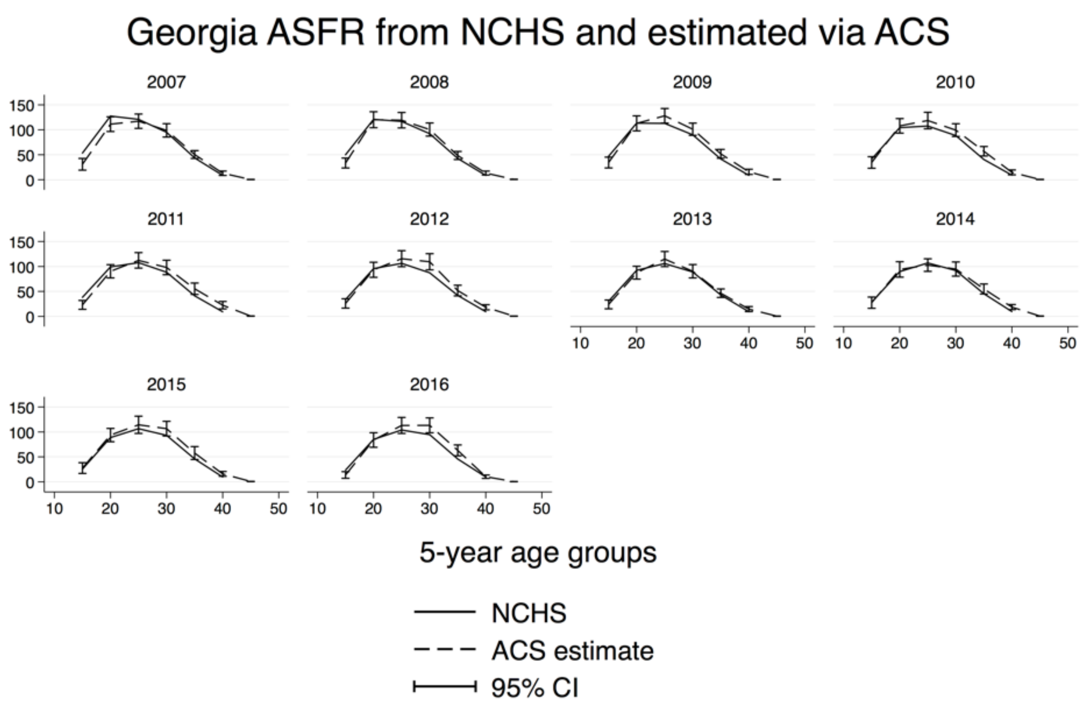

Figure 2c. Comparison of annual Georgia ASFRs, 2007-2016

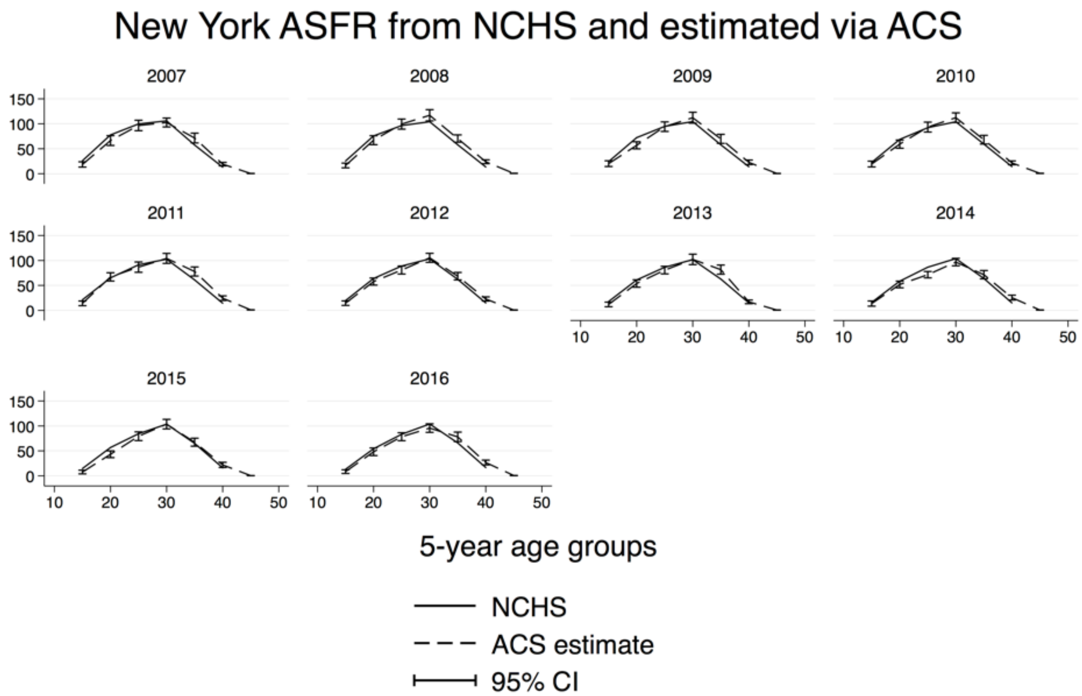

Figure 2d. Comparison of annual New York ASFRs, 2007-2016

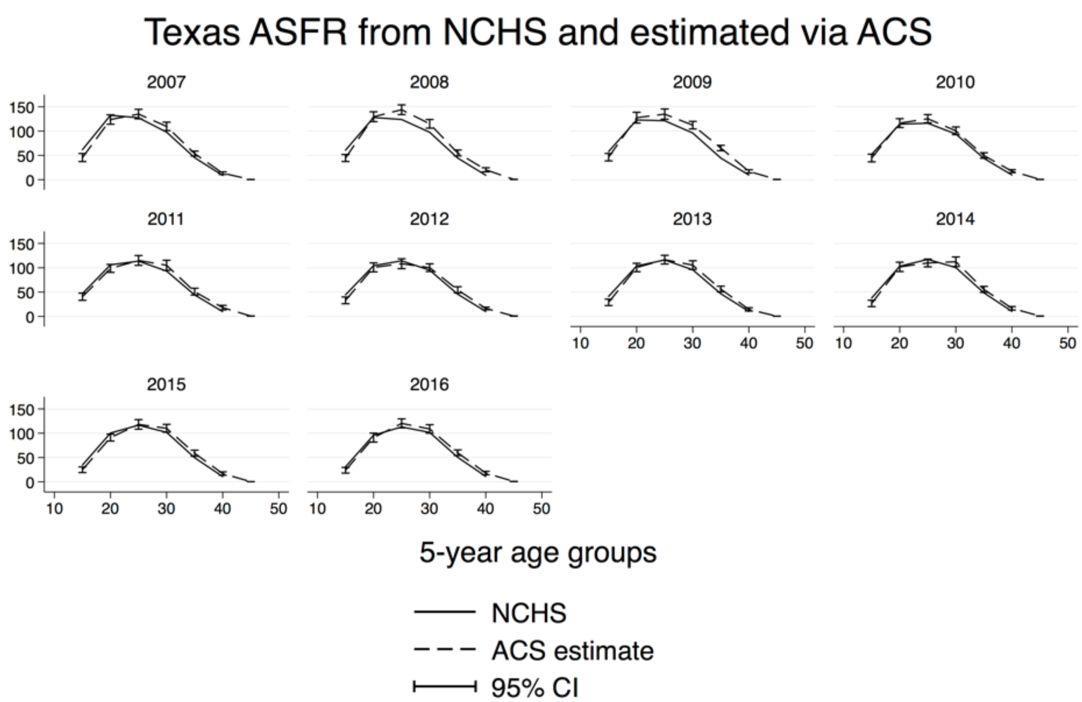

Figure 2e. Comparison of annual Texas ASFRs, 2007-2016

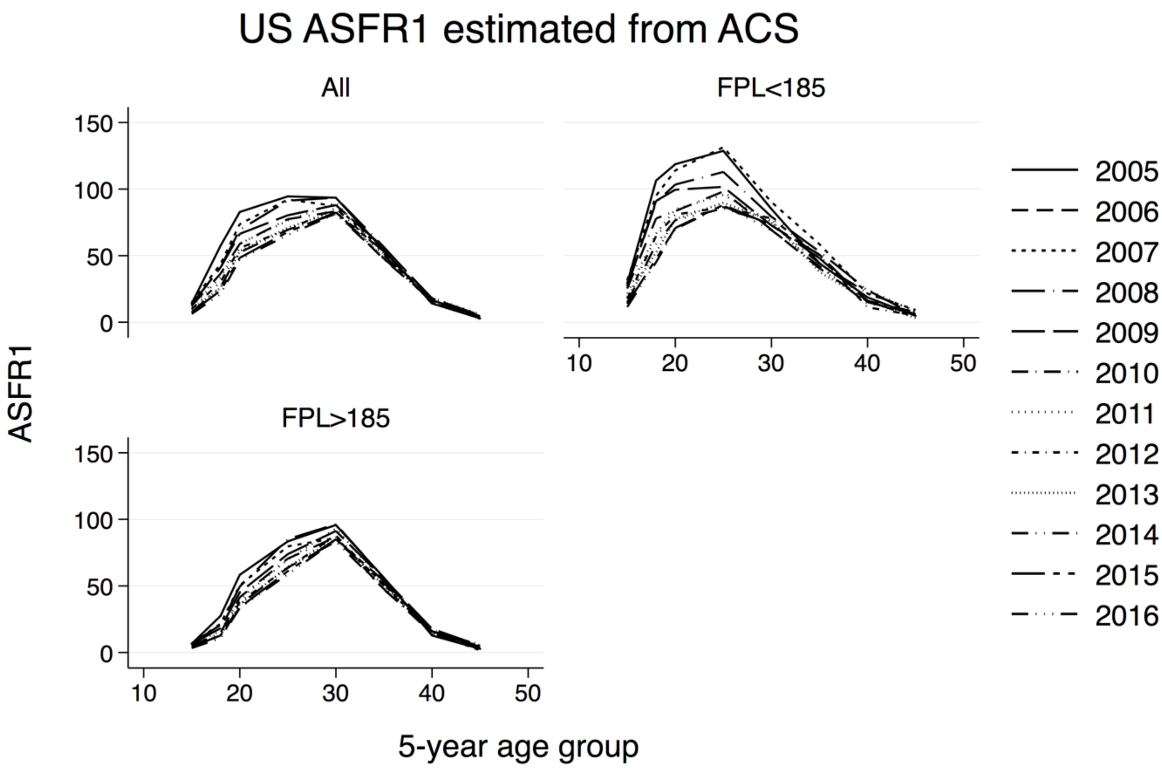

Figure 3. Estimated annual US ASFR1 by poverty status, 2005-2016

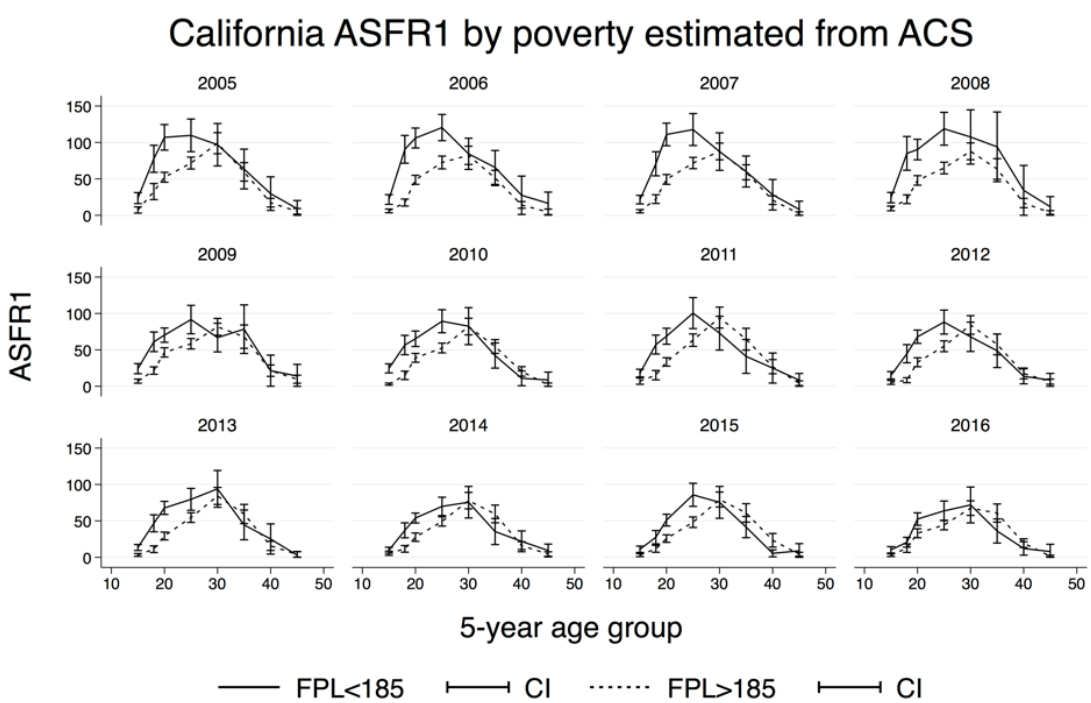

Figure 4a. Estimated annual California ASFR1 by poverty status, 2005-2016

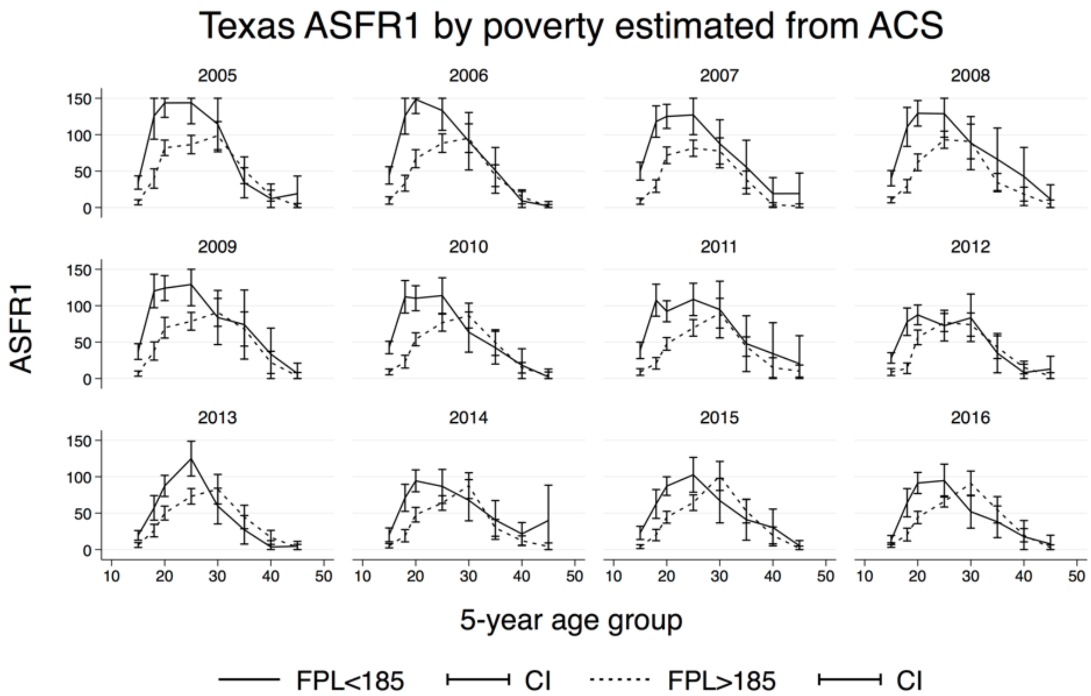

Figure 3b. Estimated annual Texas ASFR1 by poverty status, 2005-2016

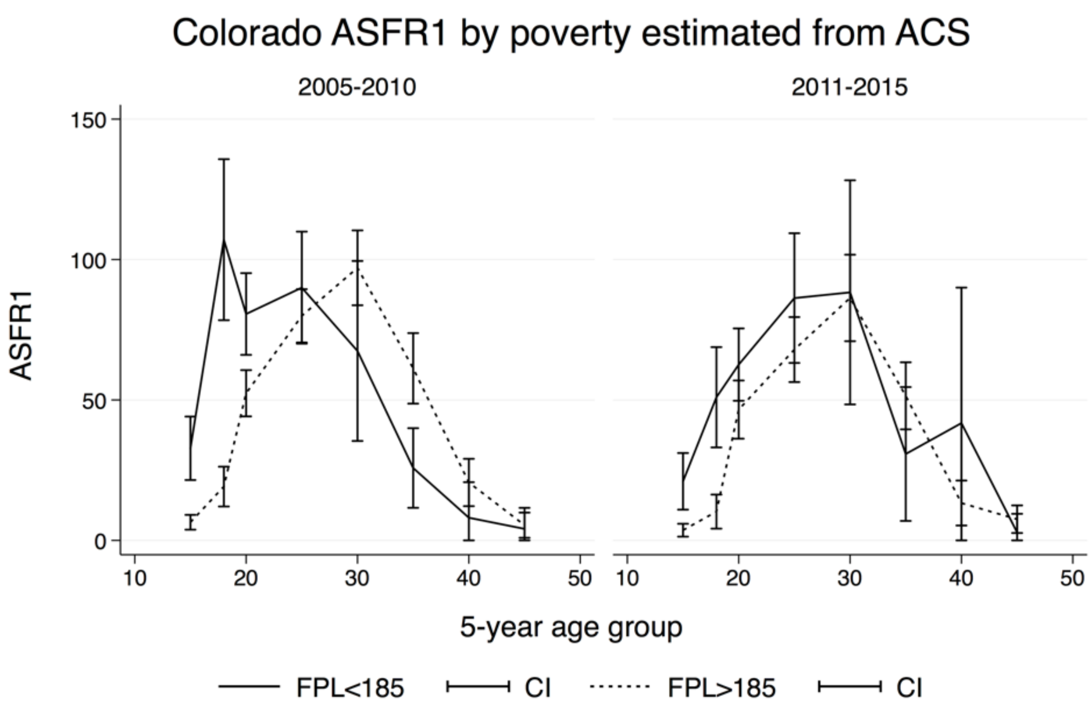

Figure 3c. Estimated annual Colorado ASFR1 by poverty status, 2005-2010 and 2011-2016

Works Cited

Chen, R., & Morgan, S. P. (1991). Recent Trends in the Timing of First Births in the United States. *Demography*, *28*(4), 513. https://doi.org/10.2307/2061420

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, *133*(3), 1163–1228.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., … Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *Jama*, *315*(16), 1750–1766.

Finer, L. B., & Zolna, M. R. (2014). Shifts in intended and unintended pregnancies in the United States, 2001-2008. *American Journal of Public Health*, *104 Suppl 1*, S43-8. https://doi.org/10.2105/AJPH.2013.301416

Gardner, E. (2010). AMERICAN COMMUNITY SURVEY User Guide ACS Publication No.

Hasstedt, K., Sonfield, A., & Gold, R. B. (2017). Public Funding for Family Planning and Abortion Services, FY 1980–2015. *New York: Guttmacher Institute*.

Kost, K., & Henshaw, S. (2014). US teenage pregnancies, births and abortions, 2010: National and state trends by age, race and ethnicity. *New York: Guttmacher Institute*.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review*, *36*(2), 211–251.

Lima, E. de, Zeman, K., Castro, R., Nathan, M., & Sobotka, T. (2016). Twin peaks: The emergence of bimodal fertility profiles in Latin America. Presented at the European Population Conference, Mainz.

Lima, E. E. C., Zeman, K., Sobotka, T., Nathan, M., & Castro, R. (2018). The Emergence of Bimodal Fertility Profiles in Latin America: The Emergence of Bimodal Fertility Profiles in Latin America. *Population and Development Review*. https://doi.org/10.1111/padr.12157

Martin, J. A., Hamilton, B. E., & Osterman, M. J. (2018). Births in the United States, 2017.

Montez, J. K., Zajacova, A., & Hayward, M. D. (2016). Explaining inequalities in women's mortality between US States. *SSM-Population Health*, *2*, 561–571.

Morgan, S. P. (1996). Characteristic features of modern American fertility. *Population and Development Review*, *22*, 19–63.

Nitsche, N., & Bruckner, H. (2018). HIGH AND HIGHER: FERTILITY OF BLACK AND WHITE WOMEN WITH COLLEGE AND POSTGRADUATE EDUCATION IN THE UNITED STATES.

Sullivan, R. (2005). The Age Pattern of First-Birth Rates Among U.S. Women: The Bimodal 1990s. *Demography*, *42*(2), 259–273. https://doi.org/10.1353/dem.2005.0018

Zaidi, B., & Morgan, S. P. (2017). The Second Demographic Transition Theory: A Review and Appraisal. *Annual Review of Sociology*, *43*, 473–492.

Appendix A: Estimation of parity for reproductive-age women in ACS

In order to estimate the parity of reproductive age women in the ACS, we begin by identifying the mother of each child in each household. Then we estimate women's estimated parity as the sum of their coresident children. In this paper we are only identifying nulliparous women, so we only differentiate between parity 0 and parity 1 or greater.

Identifying children's mothers

As the ACS only asks about an individual's relation to the primary respondent (referred to in the rest of this section as *the primary*), we often have to make a guess as to who the mother of a child is based on the relations of both the child and the potential mother to the primary respondent. We begin by crossing all reproductive-aged women (*RAW*s) in the ACS with any persons in her household who are of an appropriate age to be a potential mother-child match. (We will refer to these as "the child" in this section, even though many of them will be adults.) Note that, while we only have fertility data in the ACS for women up to age 50, we keep track of women up to age 51 here for the purposes of estimating the fertility rate for women aged 50 who will have their children move out in the next year. After this crossing, we assign a "mother code" (*mom_code*) to each pair as follows.

First, we assign the pair a *mom_code* of **1** for any clear or explicit mother-child relationship. This means either that one of them *is* the primary respondent and the relationship is explicit, or that both of their relationships to the primary respondent are clear enough to conclude that this is the mother of this child. One example of the latter would be that the potential mother is explicitly the mother of the primary, and the child in question is the primary's sibling. This code is also assigned when the *RAW* in question claims fertility in the past year, and the child in question is zero years old. This will sometimes have the effect of "overcounting" infants (for instance, in the case where there are two co-resident *RAW*s each living with an infant of their own, both women will be counted as the mother of both children), but as we're only concerned here with (a) recent fertility and (b) the existence of previous fertility (and not with the multiplicity of such), this will not be a problem in our analysis.

We assign a *mom_code* of **2** for relationships that are less explicit but still are very likely to be a mother-child relationship. These include situations where the child in question is the *RAW*'s spouse's child, or the *RAW* is the mother-in-law of the child's spouse.

In the case that the child is the grandchild of the *RAW*'s parent, we assign a *mom_code* of **3**. We will approach this code slightly differently than that of the other cases.

We assign a *mom_code* of **4** to pairs who could feasibly be in a mother-child relationship, where both are related to the primary. In the later step where we make a guess as to a child's mother, we'll consider these familial relationships to be more likely a mother-child relationship than non-familial ones. A *mom_code* of **5** is used for pairs where the *RAW* might be the mother of the child based on the relationships to the primary, but neither has a strong familial connection to the primary. Note that we will not utilize either of these codes for children 18 and over.

We assign a *mom_code* of **0** to all other pairs of relationships.

Next we count, for each child, both the number of “potential mothers” a child has in their household of each *mom_code*, the total number of potential mothers (i.e. *RAW*s with which they have a non-zero *mom_code*), and the number of potential mothers they have who have even been married. This will be occasionally used as a “tie-breaker” in determining motherhood.

Also for the purpose of determining the most likely mother of a child in a situation where there are multiple potential mothers, we calculate the average age difference between mothers and their children by considering all of the “clear” mother-child relationships, i.e. all pairs with a *mom_code* of **1** combined with all of the pairs with a *mom_code* of **2** where the child only has one potential mother in the household. We separately calculate the average difference between pairs with *mom_code*=**3** and the child has but a single potential mother in the house.

At this point we are ready to determine the likely mother of a child within their household (if there are multiple potential mothers) using the following rules:

1. If the child has a *mom_code* of **1** with someone, then we estimate that this is their mother and stop. (Each of the following steps stops the process if we “find a mother”.
2. If the child has no *mom*_codes of 1 but does have a *mom_code* of **2** with someone, then we do the following:
    i. If the child has a single potential mother with this *mom_code*, then we estimate that this is their mother. Otherwise:
    ii. If the child has a single potential mother with this *mom_code* who has ever been married, then we estimate this to be the mother. Otherwise:
    iii. We choose as the mother of the child the woman who has the closest-to-average age-difference with the child, as per the tie-breaker calculated above.
3. If there are no **1**s or **2**s but there is a **3**, we repeat the process for 2 but only considering the women with a *mom_code* of **3**. At this point, we stop if we have yet to find a mother for the child and they are 18 years old or older.
4. If there are no **1**s, **2**s, or **3**s, but there is a **4** and the child is under 18, then we repeat the process for 2 for *mom_code*=**4.**
5. If there are no **1**s, **2**s, **3**s, or **4**s but there is a **5** and the child is under 18, then we repeat the process for 2 with *mom_code=***5.**

Children Leaving

Our reliance on coresident children has many limitations, but cohort progression allows us to adjust our observed children with no coresident children (who are potentially nulliparous) to reduce the size of this population to reflect population-level fertility in this cohort at younger ages.

We estimate here the probability that a woman is parous but has no children living with her now at a given age. This is difficult without longitudinal data, so we estimate it indirectly.

Something we *can* estimate, however, is the probability that a woman with co-resident children will have those children move out in the next year. We do this using the following method:

1. First, for each PUMS year Y and for each age X, we can estimate the proportion of reproductive-aged women from that PUMS year who live with an X-year-old child of theirs (after applying the mother-guessing algorithm above.) Call this proportion **p(X,Y)**.
2. Then for each pair of years **Y** and **Y+1** we consider the quantity:

L(X,Y) = (p(X,Y) – p(X+1,Y+1))/(p(X,Y)). We use this as an approximation of the proportion of women living with an **X**-year-old child of theirs in year **Y**, who's **X**-year old child will leave home in the next year (here we use the PUMS person-weights.)

3. Now for each age **X** we take the averages of **L(X,2005), L(X,2006)**... **L(X,2015)** and use this to estimate the probability that an **X**-year-old living with their mother will move out of her house in the next year. Call this probability of leaving **LEAVE(X).**
4. So now, for each woman in the PUMS living with children we can estimate the probability that *all* of her children will leave the house in the next year: if she has children aged $\mathbf{X_1}$, $\mathbf{X_2}$… $\mathbf{X_k}$, then the probability that all of these children will leave (assuming independence) is the product of $\mathbf{LEAVE(X_1)…LEAVE(X_k)}$.

We will use this later in the main recursion.

The Recursion

Before we get to the recursion itself, let's list a few assumptions that we make:

1. Children don't leave their mother's house before the age of 15.
2. Women don't have children before the age of 15. As a consequence of 1. and 2. we have that women under the age of 30 have not had children who have moved out of the house.
3. The fertility rate for women of a given age who have had children previously and had those children move out is the same as that for women of that age whose children will all be moving out in the next year.
4. The population of reproductive-aged women is relatively stable from year-to-year, including fertility rates.

Towards the goal of estimating age-specific risk of first birth (ASFR1), we start by using our mother-guessing algorithm to map which women in the ACS live with children of their own. We apply assumption 4 and combine the populations from 2005-2016 as if they were a single population. Using the results from our mother-guessing algorithm and the note on assumption 2 above, for women aged 15-29 we estimate that the women who have had their first birth that year are exactly the women who live with an infant (0-year-old) of theirs and no other children.

For women aged 30 and above it is possible from our assumptions that she has had children previously who have moved out of the house. Hence, the population of women living with only an infant of theirs and no other children is split into two parts: those who have had previous children and those who have not. It is our goal here to estimate which portion of this population is which; we will use recursion to do this, using a transition diagram (below). But first, for women aged 30+ we estimate **CLFR(Y)** ("children leaving fertility rate"), the fertility rate for women aged **Y** who's children are about to leave the home. We do the following:

1. For each woman such that all of her children (except for possibly any infants) are of age to potentially leave the house in the next year (i.e. all of her non-infant children are 15+) we calculate the probability that all of her non-infant children will leave using the formula described in the section **Children Leaving.** We multiply this probability by her PUMS weight to obtain a new weight for her estimating how many people she represents who (a) have similar fertility and (b) will have their children move out.
2. We use this modified weight to estimate **CLFR(Y)**, the fertility rate for women who's children are about to leave the home. As per assumption 3, this will be the same as the rate for women who have had their children all leave the house already.

Now, we're ready to set up the recursion. For this, we use the transition diagram (Figure A1 on the following page).

Assume that Y is at least 29, and that we have already calculated the portions of the population "on the left" of women who (a) are nulliparous and (b) who have had children, but have no coresident children at the moment (these are the boxes labeled A and B). We can calculate directly from the data the proportions of the population labeled C and D, but we need to refine this and estimate which portion of these are w, x, y, and z. (Then we will be able to estimate the first-birth rate for women of age Y+1.) Using assumption 4 above we can obtain the following equations:

(1) w+x=C

(2) y+z=D

(3) w+y=A

We'd like to be able to also use the equation "x+z=B", but there are more possible contributors to x and z, namely the portions of the population whose children will leave in the next year. So instead we use our estimation **CLBR(Y+1)** of the birth rate for (Y+1)-year-old women who's previous children have left the home. Hence, the portion of (z+x) that is z we estimate to be **CLBR(Y+1).** This gives us the fourth equation:

(4) z = (**CLBR(Y+1)**) * (z+x)

Together, these four equations give us a system of four linear equations in four variables w, x, y, and z. Solving this for y, we get y = EXPRESSION. In theory this should be no greater than either A or D but as we are looking a population samples we set y to be the minimum of EXPRESSION, A and D. Similarly, this is a portion of a population so it must not be negative. In the rare case that this is negative, we set it to be equal to zero (this will sometimes happen with a small sample size.) Once we have solved for y we use equation (3) to get **w=A-y.**

As we have estimated both w and y, we can use these to estimate the portion of women who have just had their first birth, and w becomes A for the next step of the recursion.

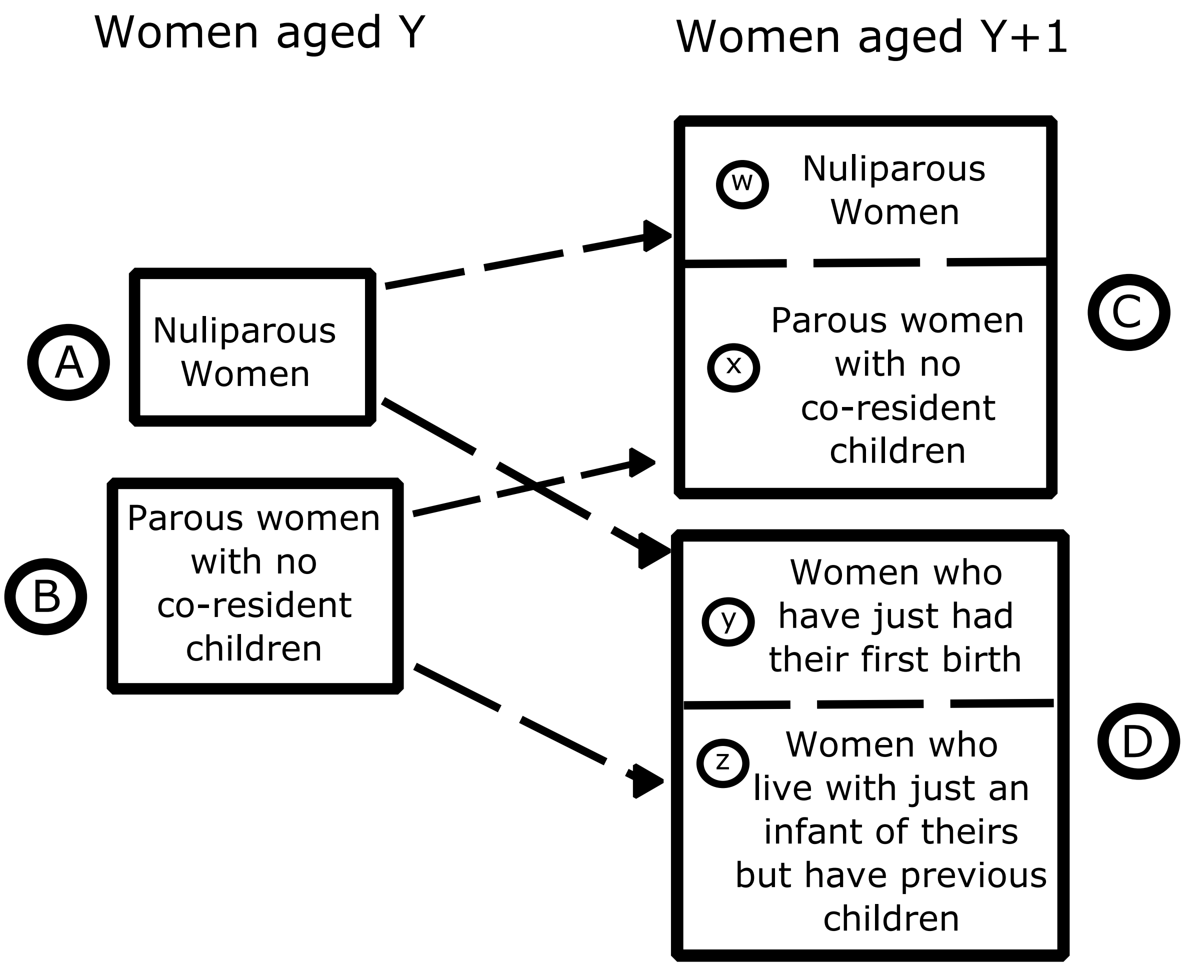

Figure A1. Illustration of recursion