

Opting for Families: Recent Trends in the Fertility of Highly Educated Women

Data Appendix for Vital Statistics Birth Data

There are a number of issues that arise when using the education variables in the Vital Statistics Birth Data. For all analyses, we focus on women who are college graduates or more. College graduates are defined as having at least 16 years of completed school or whose highest degree is bachelor's degree or higher.

Missing Schooling Data for States in Various Years

Data on schooling are missing for a number of states in a variety of years. The states and years with missing data are:

- 1976-77: Years of Schooling is missing for California (5), Texas (44), Washington State (48), New Mexico (32), Arkansas (4), and Idaho (13)
- 1978: Years of Schooling is missing for Texas (44), Washington State (48), and New Mexico (32)
- 1979: Years of Schooling is missing for California (5), Texas (44), Washington State (48), and New Mexico (32)
- 1980-87: Years of schooling is missing for California (5), Texas (44), and Washington State (48).
- 1988: Years of schooling is missing for California (5), Texas (44), Washington state (48), and New York state (33) other than New York City.
- 1989-90: Years of schooling is missing for Washington state (48), and New York state (33) other than New York City.
- 1991: Years of schooling is missing for Washington State (48).

Note that all states report education from 1992 onward.

To correct for missing data for Arkansas, Idaho, and New Mexico, for the years 1976-79, we use data for 1980-87, to calculate the ratio of Arkansas, Idaho, and New Mexico (or just New Mexico) to the 48 states (and the District of Columbia) for which data are available in 1980-87 (i.e. all states other than California, Texas, and Washington State). These ratios are used to adjust births to account for Arkansas, Idaho, and New Mexico for 1976-77 and for New Mexico for 1978-79.

To correct for missing data for New York State other than New York City from 1988-1990, we use data for 1985-87 to estimate the ratio of births in New York State relative to New York City. This ratio is used to adjust the birth data, for 1988-90 to reflect all of New York State.

At this point, the data only exclude California (5), Texas (44), and Washington State (48) for 1976-77 and 1979-88; Texas and Washington State in 1978; and Washington State in 1989-91. To adjust the data for these years, we use data for 1992-2002, calculating the ratio of births in each missing state to data for all states. We use these ratios to adjust data to be nationally representative in all years.

In all cases the adjustments for missing education data were performed for individual years of age and live birth order.

The second approach to the missing data would be to calculate estimates only for the states for which data is available in a given set of years. We have also explored these estimates, and they are similar to those reported.

Missing Individual Information on Schooling for States Reporting Schooling

Even when schooling is reported for a given state, there are missing values for schooling, especially in the early years.¹ We calculate the proportion of women between 23 and 44 with missing data on schooling by year and state and adjust births in each year and state upward by $1/(1-\text{Share Missing})$ based the share missing schooling in that state in that year. Implicitly, this procedure assumes that probability that data on schooling is missing does not vary with schooling, age, and live birth order conditional state and year. The data for 2005 and 2006 do not contain geographic identifiers. In these years, national data (which exclude possessions and territories) on the share of women with missing data on schooling are used to adjust the data.

Conversion from Degrees Obtained to Years of Completed Schooling

Starting in 2003, some states used new birth certificates on which educated is reported according to degrees obtained as opposed to completed schooling as in prior years. The degree-based classification was converted back to one based on years of completed schooling by weighting births by women with a college degree or higher on the new certificates by 1.007599 based on Jaeger [1997]. Population data in the March Current Population Survey, which switch to a degree-based classification in 1992, are adjusted in the same way.

Missing Information on Parity / Live Birth Order

Live-birth order and parity (coded as 1, 2, 3, and 4+) refer to the total number of live births the mother has had up to a point in time, with fetal deaths excluded. When data on live birth order is missing, we assign a live birth order in proportion to the share of births of that live birth order among all births for which live birth order is reported.

Estimating Distributions of Births by Live Birth Order

To calculate the distribution of the number of children that women have from the birth records, one needs to do a bit of manipulating. We start from age 23 and take the rate of women having a 1st, 2nd, 3rd, 4th and further child at age 23 as the birth probabilities at age 23. Consider women born in 1953, who turned 23 in 1976. We calculate the share of

¹ The codebook states, “No attempt has been made to convert years of school completed in foreign school systems, ungraded school systems, and so forth, to equivalent grades in the American school system. Such entries are included in the category ‘not stated.’”

women having a birth at each parity in 1976 using the Vital Statistics data merged to population data from the CPS. To get the distribution of the number of children these women had by age 24, we first calculate the birth rate at each parity in 1977 in the same way as we did for 1976. To get the share of women with exactly 1 child, we then add the share of women who had a 1st child in 1977 to the share of women who had exactly 1 child in 1976 and then subtract the share of women having a 2nd child in 1977.

As described here, we calculated the share of women having a 1st, 2nd, etc. birth in each year. To get shares from the numbers in the Birth Records, we divide by population figures constructed from the CPS. This method introduces noise because of sampling error in the CPS and because of any differences in reporting in the CPS and Birth Data. While the underlying data are large, individual age-education cells, especially for college graduate women, are considerably smaller. We are not aware of more precise data on the age-education distribution of the population.

Our procedure for estimating the distribution of children born for women assumes that the distribution of children born is equal to the birth rates at age 23. A potential problem arises if college-graduate women have children before age 23 but not again at age 23. If such a woman has children later, we will “subtract” her from lower parity, but she was never counted in the first place. Appendix Figure 1 shows the share of 23-year old women with a given number of children calculated from the Census and CPS and from the Vital Statistics Birth Data. While the estimates from the CPS are noisy because of small samples, the estimates from the Vital Statistics Birth Data are much less noisy. Nevertheless, the two series are quite similar in levels. Under the assumption that the Census and CPS accurately reflect the number of children present and that our procedures with the Vital Statistics produce accurate birth numbers, these estimates imply that most college-graduate women do not have children before age 23, and that most of the ones who have children at age 23 are having their first child. Thus, our use of fertility at age 23 to initialize the birth distribution appears reasonable. Data (not reported here) show that the size of the college-graduate population increases substantially from ages 21 and 22 to age 23 and that the age 23 population is considerably more stable over time.

Because more women complete college as they get older, our method involves one important assumption. To get the correct distribution of the number of children for women at age 24, we need to assume that the distribution of the number of children at age 23 is the same for women who have completed college by age 23 and for women complete college at age 24. If the fraction is larger (smaller) for the latter group (those completing at later ages), then our estimate of fertility is downward (upward) biased. This bias increases as the share of women completing college at later ages increases.

This bias can be illustrated with a simple example. Consider women born in 1953. Assume that the population of college graduates in 1976 when they are 23 years old is 1,000, and that it is 1,100 in 1977. So there are 100 women completing college at age 24. Suppose that 5% of the 1,000 women who had completed college at age 23 in 1976, the fraction with exactly 1 child is 5%, and the fraction with exactly 2 children is 2%, and none of them have more than 2 children. Further assume that at age 23 of the 100 women

who will complete college at age 24 in 1977, the fraction with 1 child is 8%, and none of them has more than 1 child (although these shares are not observed in our data). In 1977, we observe that 44 out of the 1,100 college graduates (4%) have a 1st child, and 11 (1%) have a 2nd child (and that there is no other fertility among college graduates).

The correct distribution of parities is:

Share with exactly 1 child = $(5\% * 1,000 + 8\% * 100 + 44 - 11) / 1,100 = [5\% * (1,000/1,100) + 8\% * (100/1,100)] + 4\% - 1\% = 8.27\%$;

Share with exactly 2 children = $(2\% * 1,000 + 0\% * 100 + 11) / 1,100 = [2\% * (1,000/1,100) + 0\% * (100/1,100)] + 1\% = 2.82\%$

With our method, we get:

Estimated share with exactly 1 child = $5\% + 4\% - 1\% = 8\%$ (above the actual value of 8.27%)

Estimated share with exactly 2 children = $2\% + 1\% = 3\%$; (below the actual value of 2.82%).

The first estimate is biased downward because a greater share of the women who completed college between 1976 and 1977 had children than among the women who had already completed college in 1976. Similarly, the later estimate is biased upward because fewer of the women who completed college between 1976 and 1977 had children than among the women who had already completed college in 1976.

Miss-measurement of population will affect our estimates of the distribution of children too. The most obvious effect is to introduce noise in our estimates. More subtly, because the population estimates are on the denominator of the calculations, pure measurement error will bias our estimates of the share of women with each number of children upward and the share with no children (the residual category) downward. This bias does not appear to be too strong insofar as it would imply that estimates of childlessness would decline with age in the Vital Statistics relative to the Census / CPS, whereas our estimates of childlessness from the Vital Statistics tend increase with age relative to the Census / CPS.

Appendix Figure 1. Comparing the Fertility in the Birth Records and CPS at age 23.

