Reviewer's report

Title: Estimating the period mean age at first birth from household surveys

Version: 1
Date: 2 March 2015

Reviewer: Thomas Spoorenberg

Reviewer's report:

This research note investigates two methods of estimating the mean age at first birth from the information collected in the Demographic and Health Surveys (DHS). Because accurate civil registration data are usually not available for the computation of this important indicator in statistically less developed countries, the results presented in this analysis are potentially important.

Major Compulsory Revisions

1. A series of points related to the discussion of the results needs to be addressed. The authors compare M(t) and M*(t)—two estimates of the mean age at first birth. While both indicators aim at measuring similar thing, they differ on a number of points. First, M(t) is computed using DHS micro data sets, whereas M*(t) is computed based on published national reports. Second, M(t) is computed on an age range from 0 to 49 using single-age data, while M*(t) is based on an age range extending from 0 to 40 using five-year age group data. Each estimation method of the mean age at first birth relies on different data and, more importantly for the issue investigated in this research note, different age ranges. The authors seem not to have this point in mind when discussing the results given by each method (second paragraph of the ‘Results’ section). The use of five-year age group data can only return rougher—and possibly overestimated?—estimation of any mean age as it assumes implicitly that the event (here the first birth) is distributed evenly over an age group for which the age at mid-interval of the five-year age group (e.g. for age group 20-24, age 22.5) is classically used in computation. On the contrary, by using single age data one can compute with much more precision a mean age. As the objective of this research note is to investigate which method between M(t) and M*(t) is to be used with the most confidence to estimate mean age at first birth in statistically less developed countries, the authors need to discuss further this point.

2. As detailed in the last paragraph of the ‘Methods’ section, the computation of M*(t) is made on the age group 15-19 to 40-44 (indeed, age 0-4 to 40-44, otherwise Formula (2) would return too low a mean age at first birth). The authors would need to clarify in the text the reasons for excluding the last age group of the reproductive life (45-49). For the sake of clarity, the authors would also need to mention that they use the information on the number of children ever born (CEB) classified by the age of the mothers collected in DHS (and classically published in DHS reports) to compute M*(t). Finally, an annex presenting the formula in discrete format and offering an example of the computation using CEB
data from a given country would be useful. The $M^*(t)$ formula is given in a continuous format in the text, but an illustration using a discrete form would benefit many potential readers unfamiliar with the continuous notation. This would increase the ‘public utility’ of the paper, especially in statistically less developed countries.

3. The authors do not talk also about the fact that surveys imply sampling of households or individuals from a general population. Consequently, for all demographic indicators derived from survey data sets, confidence intervals should be computed, applied and used in comparative analyses. As the authors compare two sets of indicators of mean age at first birth that are computed based on different variables (but from the same survey), the issue of the confidence intervals should be factored into the argumentation.

4. Another point that would deserve better treatment by the authors is to discuss how the known biases inherent to the information collected in full birth history in DHS could eventually affect (or not) the results of each estimation method and the conclusions. As the authors know well, full birth history data collected in DHS suffer of some biases affecting the collection of information referring to (but not only) the 5-year period preceding the survey date (e.g. see Arnold 1990, Schoumaker 2014). The mean age at first birth $M(t)$ used in this analysis is computed based on the births in the three-year period before the survey and could therefore potentially be affected by such bias. While it is true that the birth order should not be related to or affected by such bias (but is indeed more related to a threshold date, and selective intentional and/or unintentional birth omissions or transferences), the authors do not mention how these known biases could eventually affect (or not) their results and conclusions. The authors do also not discuss the implications of their findings in the light of the conclusions of recent studies on the intrinsic bias related to nationally sample surveys (see for example Hertrich and Lardoux’s (2014) discussion related to the study of marriage data). The only quality issue that is touched upon in the current version of the text refers to the fact that the median ages for different cohorts do not overlap (Figure 3). The authors interpret this as the by-product of the sole misreporting of the age at first birth. Yet, a number of alternative factors can potentially contribute to this situation, among which the representativeness and selection of the sample between two or more sample surveys can play an important role (for example, Hull and Hartanto (2009) have shown that DHS samples have missed single women in Indonesia affecting ultimately the estimation of the level of fertility). As the main objective of this research note is on testing and comparing two methods to estimate the mean age a first birth in statistically less developed countries, the text would gain from a (even brief) discussion of these quality issues.

Minor Essential Revisions

1. As written in several instances in the text, the analysis is conducted on DHS data for 62 countries, whereas the abstract cites 64 countries. Please correct appropriately.
2. In their explanation and illustration of the time plotting of the median age at first birth (Paragraph 6 in the ‘Results’ section and Figure 3), a reference to the work of Feeney (1995, 2014) would be indicated. In addition, the authors could give the full computation of the example included in the text: “[I]f women aged 30 to 34 report a median age at first birth of 20 years in a survey conducted during 2010 then this data point is plotted at 1990.0 years (i.e., 12.5 years before 2010.5).” The authors could include in the text that it is assumed that women aged 30 to 34 are on average 32.5 years old. As the median age at first birth is 20 years old, it is therefore estimated that these women have had their first birth 12.5 years before the survey (i.e., age at survey − median age at first birth = 32.5 − 20 = 12.5). The reference date to which the median age at first birth applies is therefore 12.5 years before the survey date (i.e., reference date of survey − time before the survey to which the median age at first birth refers = 2010.5 − 12.5 = 1998.0). This could be added either in the text or in a note.

3. Figure 1: please add a legend identifying the set of markers to the two distinct clusters.

4. Figure 3: Please add a decimal in the scale of the y-axis. I would also add some markers to indicate the different surveys in the mean (period) age at first birth (black series).

Discretionary Revisions

1. In their comparison of M(t) and M*(t) (Figure 1), the authors could comment further (maybe in a note) on how we can better understand the situation when M(t) is higher than M*(t).

References cited in comments


Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests