Subject: Missing data for C-section variable (m17) for births prior to the most recent
Posted by sagedoesdata on Sat, 19 Dec 2020 19:05:19 GMT
View Forum Message <> Reply to Message

Hello DHS Community! I am analyzing Indonesia DHS individual-coded data from years 2002-3, 2007, 2012, and 2017. The program I use is STATA.

I want to analyze C-section data. I am using variable m17. For the last birth, m17_1, this data is very straightforward. However, for previous births, m17_2 through m17_6, there is very significant missing data.

For the 2017 dataset, there are initially 49627 observations, with 34313 observations missing for m17_1. After dropping all observations for individuals with no births (drop if v208 == 0) there are only 43 missing values remaining for m17_1 and 15314 useful values of "yes" or "no". However, for m17_2 (next to last births), there remain 13084 missing values, and only 2273 useful values. Information from other variables bidx_02 (birth number) report that there were 10601 next to last births total in the dataset. Why are there almost 8000 missing values? What do these missing values mean in the context of the C-section data? What should I do to perform a proper analysis?

In addition, in the sample survey listed in Appendix F of the DHS Indonesia 2017 report, it says that cesarean section data (Question 432) was only collected for the last birth and next to last birth. Why does the data then also include m17_3 through m17_6, including data up to five births prior to the most recent? Can I perform a valid analysis on births prior to the last birth?

Subject: Re: Missing data for C-section variable (m17) for births prior to the most recent
Posted by Bridgette-DHS on Tue, 22 Dec 2020 14:44:50 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

This is one of the variables that is given for all children born in the past 5 years. The subscript _1 refers to the most recent birth in the past 5 years; _2 refers to the second most recent birth in the past 5 years, etc. Most women who had a birth in the past 5 years had 1 or 2 and very few had more than 3. In the IR file, m17_3, for example, will be blank or have a dot if the woman only had 0, 1, or 2 births in the past 5 years. These are not MISSING cases. If you want the births to be the units of analysis, you could shift to the KR or BR file. Some other variables are only available for the most recent birth and are blank or a dot for all subscripts other than _1. This issue--the distinction between "missing" and "not applicable"--has come up repeatedly.