
Subject: Re: Contraceptive prevalence rate

Posted by [Bridgette-DHS](#) on Tue, 17 Dec 2024 12:20:55 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS staff member, Tom Pullum:

Whenever I see a substantial number of cases with a dot (for Not Applicable or NA) on a key variable, I enter "lookfor select" to see if there is a selection variable for a subsample. One of the variables that comes up in this survey if you do that is "seligbm":

seligbm byte %8.0g SELIGBM hh selected for biomarker and long/short woman qre

which has the following distribution:

hh selected for biomarker and long/short woman qre	Freq.	Percent	Cum.
-----+-----			
household selected for biomarker and fu	10,053	33.42	33.42
household selected for full woman quest	9,934	33.03	66.45
household selected for short woman ques	10,091	33.55	100.00
-----+-----			
Total	30,078	100.00	

It appears that there was indeed subsampling in this survey. 10,091 women were in households selected for a "short" women's questionnaire that omitted questions about contraceptive use. That is, women with seligbm=3 are NA on v313 (and many other variables). Those women must be ignored for the calculation of the mCPR. This subsampling may affect other variables in your analysis.

This sort of thing is one of the hazards of using DHS data and the GitHub programs. Someone could say that analysts should read about the sampling design before starting to use the data but we (you and I!) usually just plunge in and then find these exceptional features the hard way. The good thing is that you tried to calibrate your estimate against the report and found a discrepancy, and it led to the evidence of subsampling. That's good practice. Subsampling is one of the main reasons why users cannot match reports.