
Subject: Re: Sibling mortality in maternal IR files

Posted by [Bridgette-DHS](#) on Thu, 10 Mar 2022 12:35:23 GMT

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

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

If you enter "tab midx_01,m" in the IR file, there are 11,280 cases with the value 1 and 12,104 cases with a dot. When you see this pattern for one of the modules, it means that there was subsampling. Half of the households would have been selected for the this module. The subsampling was probably done to reduce data collection costs. If you look at hv117 and hv118 in the PR file, you will also see that men were also subsampled, with a smaller fraction. Usually when there is subsampling you can find a variable in the PR or IR file with labels such as "selected for sibling module" and "selected for men's survey". Unfortunately, I can't find such variables in the files for this survey. They may be in there but I don't see them. In many of the older surveys the subsampling indicator was dropped during data processing.

Subsampling is not left up to the interviewers. There would have been some selection, almost certainly of households, using a random procedure, during the household listing before the cluster was entered by the interviewer teams. The 12,104 consists mainly of women whose household was not selected, but also includes some women who had no eligible siblings. That's probably the only reason why the module was NA for more than half of the women.