
Subject: Unique identifiers/duplicate entries in DHS Pakistan Data
Posted by [nqayyum](#) on Tue, 17 Feb 2015 03:43:15 GMT

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

Hi,

I am using the DHS Pakistan child recode data to make a panel dataset using STATA software. In the data however, there seem to be duplicate observations. I used the household, cluster and line number as the unique identifiers but even then there are some duplicate observations.

Is it possible that the same child from a household in a cluster may be interviewed twice at the same time and the information is included twice in the dataset as well?

I am having trouble figuring out why there will be duplicate entries for individuals in the dataset.

Thanks,

Subject: Re: Unique identifiers/duplicate entries in DHS Pakistan Data
Posted by [Bridgette-DHS](#) on Wed, 18 Feb 2015 12:55:34 GMT

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

Following is a response from Nouredine Abderrahim, Senior DHS Specialist:

You need to use the cluster number, the household number, the woman's line number, and the child's line number in the birth history (BIDX) to identify the child. This will ensure that you do not get duplicate cases as reported in your post.

Please keep in mind that some children do not live in the household, and for that reason, you can't use the line number in the household (B16).
