Subject: Zambia 2018 -- duplicates in cluster number and household number
Posted by nchesterman on Fri, 12 Aug 2022 15:47:54 GMT
View Forum Message <> Reply to Message

Hi all,

I am attempting to merge variables from the male and female datasets onto the household dataset. The DHS help site for this (https://dhsprogram.com/data/Merging-Datasets.cfm) indicates that v001 and v002 (cluster and household numbers) should be unique for the datasets. However, I am finding that these two variables do not uniquely identify all observations; there are many duplicates of these two variables in the dataset.

I also checked my variables of interest, and found that the data differs between duplicated cases. So the households are not true duplicates.

Has anyone else encountered this issue for the Zambia dataset, and have suggestions for how to successfully merge on cluster and household numbers with the household dataset?

Thanks,
Nathan

Subject: Re: Zambia 2018 -- duplicates in cluster number and household number
Posted by Janet-DHS on Thu, 01 Sep 2022 20:30:33 GMT
View Forum Message <> Reply to Message

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

In the PR file, individuals are identified by hv001 hv002 hvidx. In the IR file, by v001 v002 v003. In the MR file by mv001 mv002 mv003. For this kind of merge, you just use those three files. NOT the HR file, which perhaps you have been using.

We apologize for the delay in this response and hope it will still be useful.