
Subject: Unique identifier in pooled DHS (countries and years)

Posted by [marion](#) on Tue, 10 Dec 2013 13:40:45 GMT

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

Dear all,

I pooled the birth history datasets for all phases and all countries. My total sample consists of roughly 6,000,000 observations. I created a unique identifier by combining v000+v001+v002+v003+bidx, as it is recommended in the DHS manuals. However this combination does not uniquely identify each line (=each child). There are more than 1,500,000 observations which are not uniquely identified by this combination of variable. My guess is that is problem could be connected to the presents of different waves within the phases (e.g. for Uganda phase 5 there are the datasets UGBR5HDT and UGBR52DT). Do I have to assign a coding for the different waves within the phases? I did not read anything about this requirement in the manuals so I am not sure how to proceed best. Do you have any advises?

Many thanks and warm regards,
Marion
