Subject: Merging African countries and discrepancy between births and household file
Posted by nholla on Thu, 06 Aug 2015 18:06:41 GMT
View Forum Message <> Reply to Message

Hello DHS, I had two questions that I could use your help on!

I'm trying to create an overarching analytic file across all sub-Saharan African countries for 3 different years using the birth, household member, and women's files. Is there a recommended approach to doing this/are there any potential roadblocks that I should be aware of? This is what I had planned on doing (I'm using stata):

1. Merge different file types for a particular country for a particular year.
2. Append all pertinent years of interest for a specific country.
3. Append all countries together.

I'm trying to merge each child in the births file to the household member file, using the "child line number in household" variable as a reference. However, I see that there are some children that are "not listed in household." Why are not all of the births not accounted for in the household? I initially thought this might mean the child passed away, but these are recorded missing. In Ghana 2008, for example, this occurs around 3000 times.

Thanks for your help!

Subject: Re: Merging African countries and discrepancy between births and household file
Posted by Bridgette-DHS on Tue, 25 Aug 2015 12:25:55 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

The BR file for Ghana 2008 has one record for each live birth in the birth histories. They are "children" in the sense that everyone in the world is a child, i.e. had parents. If you go to this file and enter "tab b2" you will see that the birth year is as early as 1972. That is, the oldest "child" was 36 years old at the time of the survey. If you enter "tab b16,m" you see that of the 11,888 children in this file, 1413 have died (b16=.) and 2914 are not in the household (b16=0). If you do "tab b2 if b16==0" you get the year of birth of children who are not in the same household as the mother. Some of the children are young, because in much of sub-Saharan Africa there is a lot of fostering, but about half are over age 17.

If you want to merge the PR, IR, and BR files, as you say, then you will not want to--or be able to--include cases in the BR file with b16=. or b16=0. Older BR files do not include b16 at all, and unfortunately if it's missing there is no reliable substitute for it.

I recommend that you reduce to the variables you want before doing the merges.   You will probably want to drop the b variables from the IR file, for example.

Note that the v variables in the BR file are actually from the mother's record. They describe the mother, not the child.

You should add a unique identifier for each survey. Do not rely on v000 or hv000.

Age in the PR file (hv105) will be superseded by age in the IR and BR files. Weight in the PR file (hv008) will be superseded by weight in the IR and BR files.

To match most DHS numbers, reduce the PR file to de facto residents (hv103=1).

If you combine the surveys for a single country, to study trends, you will be able to do several analyses that you cannot (easily) do with separate files. For example, you can easily test for changes over time, graph trends, etc. However, I don't think much is gained by combining countries into a single file, compared with looping through countries in your program. What you will find you are doing with a combined file, I suspect, is mostly analyzing one country at a time, which you could have done with a loop.

What I often do when working with many countries is to process one country at a time, save a collapsed summary file for that country, and append the summary files for the integrated analysis.