Subject: Understanding dataset structure
Posted by popanalyst on Mon, 13 Nov 2017 14:44:07 GMT
View Forum Message <> Reply to Message

Hi, I'm working with the Malawi 2015-2016 DHS data to identify out-of-school adolescent girls and young women aged 15-24 who have an HIV test result. I am a bit confused about all the different datasets and wondering from where I should pull observations. Currently, I have tried to pull from the individual recode and couples recode, and match them to their HIV status. I'm wondering if I should pull from the household roster, or if I have already counted those by using the individual dataset? It looks like I will need to pull current schooling status from the household dataset as well, and I'm wondering how to connect people in the roster to the other datasets, if possible. Thanks!

Subject: Re: Understanding dataset structure
Posted by Liz-DHS on Mon, 13 Nov 2017 15:06:31 GMT
View Forum Message <> Reply to Message

Dear User,
If you are merging datasets, please refer to https://dhsprogram.com/data/Merging-Datasets.cfm, if you want a description of the various datasets and how you can use to analyze, please refer to https://www.dhsprogram.com/data/Dataset-Types.cfm#CP_JUMP_6014. We also recommend The Guide to DHS Statistics https://www.dhsprogram.com/publications/publication-dhsg1-dhs-questi onnaires-and-manuals.cfm and The Standard Recode Manual https://dhsprogram.com/publications/publication-dhsg4-dhs-questionna ires-and-manuals.cfm. We also have various YouTube videos available to assist users.
DHS Dataset Types in 60 Seconds: https://www.youtube.com/watch?v=JGRJZCGiCJw
Introduction to DHS Datasets: https://www.youtube.com/watch?v=fzLNQkkvDeI
Introduction to DHS Data Structure: https://www.youtube.com/watch?v=BmiSDPebmgc
De Jure and De Facto: https://www.youtube.com/watch?v=klClwMgFcCU

After reviewing these resources, if you still have questions, please feel free to post again.
Thank you!

Subject: Re: Understanding dataset structure
Posted by kingx025 on Sun, 22 Apr 2018 02:14:55 GMT
View Forum Message <> Reply to Message

The set of households covered by the household member (PR) files are randomly selected from within primary sampling areas, regardless of whether then include a woman of childbearing age or not. Women of childbearing age within those households (which is most commonly defined as women age 15-49) then receive the long individual women's questionnaire and go into the IR (women's) and couples files. The young women age 15-24 from the household roster are already included in the IR file, so I'm afraid you would be double counting them if you used the same women from the PR file without merging between the file types. Though I don't have personal

experience working with the couples' file, I believe the women in the couples file are a subset of the same women in the IR file.

It may help to think of how the data are collected, in terms of the survey forms. Usually there is just a household survey form (including the roster of household members), a woman's survey form (which also collects information for the children and birth recode files), and maybe a men's file. There is no separate survey for couples, and the women who get the individual women's form are drawn from the larger group of randomly selected households with their household rosters. While it's easy to think of the separate DHS files as having their own separate existence, they are really rearrangements of material collected via the 3 forms of household, women, and men.

We hope to eventually do such merging of IR and PR files within the IPUMS-DHS project, but so far we have only linked across household, women (IR), child (KR), and birth (BR) files. I wonder if you might roughly approximate the education data you need using the data on total years of schooling or highest level of schooling for women in your age group of interest (V106, V133) in the IR files. If you need to link between the IR file and the PR file for the school attendance data, I assume you would use the household id number and the person's line number (V003 in the IR file) and HVIDX (the person's line number in the household).

Good luck!

Miriam King