Subject: Re: Malawi - school attendance rates - not matching report
Posted by geoK on Thu, 04 Oct 2018 15:37:52 GMT
View Forum Message <> Reply to Message

Thank you for your kind reply. I have been using the guide to replicate the NAR indicator for secondary school (school attendance ratios page). I still encounter a few issues, as well as my final percentage for Secondary school NAR in Malawi is still around 15% (instead of 17.4%).

I would like to ask a few questions, please, related to the step by step process described in the guide.

1) when it comes to merging the PR file and the BR file quite a few line numbers (b16) are missing - although variable B3 is actually not missing. Therefore that person is either not merged into the new file, or it is but it has all missing values for HV122 and other PR variables. What should i do with them? I can see that some of them are dead (B6).
I suppose, when doing the merging, I should prioritize retaining information from the PR file (hv122), and impute b3 in case it is missing, correct?

2) At page 3 of the School attendance ratios section of the guide (see below), you describe 2 ways to calculate the age at the start of the school year, 1 is by the merging i was talking about at point 1) above, and 1 by imputation. can you please advice whether you usually use one among the 2, or both, and what has been used in Malawi 2015-16 please?

"To calculate the age at the start of the school year a century month code of date of birth (cmcdob) is needed for all persons age 5-24 using one of two approaches is used:

1)   Merge the birth history data for b3 from the BR file onto the PR file using cluster (hv001/v001), household (hv002/v002), and line number in the household roster (hvidx/b16). If the person is listed in the birth history of any women (BR file), the century month code (CMC) of the date of birth of the person (b3) is used as cmcdob.

2)   If the CMC date of birth of the person is not known, a century month code for the date of birth is randomly imputed using a uniform distribution between the possible bounds of the date of birth, calculated as the CMC date of interview minus 12 times the age in years as a maximum and 11 months earlier as a minimum:

max_cmcdob = hv008 - hv105*12

min_cmcdob = max_cmcdob - 11

cmcdob = random(min_cmcdob, max_cmcdob)."

Can you please advice on commands in STATA or SPSS to impute cmcdob = random(min_cmcdob, max_cmcdob).

3) Finally, it would be very useful to know the absolute number of people of secondary school age for this survey please (age 14 to 17), so that i can have a reference to check if i am doing things well. I couldn't find it anywhere in the final report.

thank you!