
Subject: Trouble understanding variable nomenclature eg. HML32_01,HML32_02,...
Posted by [SonjaMM](#) on Tue, 07 Jun 2016 00:05:04 GMT

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

I've been looking at the datasets for Mozambique from 2011 and there is something curious that I don't exactly understand, which is the numbering of some variables.

For example variable HML32 = is the final result of a malaria test from a blood smear test. This I understand but the problem is that in the dataset the variable exists as HML32_01, HML32_02, HML32_03...all the way up to HML_24. So there are 24 "versions" of the HML32 variable.

I've tried looking for the meaning of this, I've checked the recode manual, but the recode manual only states the variable as HML32, it doesn't explain the "_xx" part of the variable. At first I thought that maybe I could just combine all the variables and make one single HML32 variable, but I am not sure if I can even do that?

I get the feeling that there is a simple explanation for this, but I would appreciate any help with this.

Thank you.

Subject: Re: Trouble understanding variable nomenclature eg.
HML32_01,HML32_02,...

Posted by [Liz-DHS](#) on Mon, 27 Jun 2016 21:04:00 GMT

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

Dear User,

Here is a response from malaria expert, Cameron Taylor who also noticed you had posted other messages to the forum. Her answer hopefully will help you with all the posts you posted so far.

Quote:

Dear SonjaMM,

Thanks for your question, hopefully this will help better clarify.

It looks like you are in the household file (MZHR62FL). In the household file each row contains information about only one household. So the variables (represented by columns) contains data for each household. Most households have multiple people so we can't just have one variable that contains all the malaria data about everyone in that household. So that is why you are seeing multiple HML32 variables. Keep in mind that for most surveys we are only testing children between the ages of 6-59 months old. That is why you will see a lot of blank data for the HML32_01-HML32_24 variables because most household members were not tested unless they were between the ages of 6-59 months.

For example:

HML32_01 is malaria blood smear test data for person number 1 in the household roster

HML32_02 is malaria blood smear test data for person number 2 in the household roster

etc, up until person number 24 in the household roster.

Depending on your research question the HR file might not be the file that you would like to use for your analysis. Most of the time users who are interested in malaria parasitemia data will use the PR file (in the case of Mozambique MZPR62FL) because your unit of analysis is an individual not a household. This file has all the household members listed as rows in the dataset with one variable (HML32) containing their individual malaria test results. If you are interested in calculating parasitemia via microscopy in children 6-59 months you can use the Stata code below. If you are using another statistical software you can update the coding accordingly.

*Parasitemia (via microscopy) in children 6-59 months

```
gen micmalpos=0
```

```
replace micmalpos=(hml32==1)
```

```
lab var micmalpos "Parasitemia (via microscopy) in children 6-59 months"
```

*Creating weight variable

```
gen wgt=hv005/1000000
```

*Survey setting the data

```
svyset [pw=wgt], psu(hv021) strata(hv024)
```

* Proportion of children 6-59 months old with malaria infection (microscopy)

```
svy: tab micmalpos if hc1>=6 & hc1<=59 & hv103==1 & hml33==0 , ci
```

Let us know if you have any other questions