
Subject: Re: Merge wealth index with individual recode - Malawi 2000

Posted by [Liz-DHS](#) on Fri, 23 Sep 2016 00:56:02 GMT

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

Dear User,

A response from one of our experts, Dr. Sarah Staveteig:

Quote:

The household ID (whhid) in the Malawi 2000 wealth index file is a concatenation of the cluster number and the household number. It is similar to the caseid string found in the IR file except that caseid also has the respondent's line number at the end.* To directly merge the IR file with the WI file, you have three options: (1) create a new whhid in the IR file based only on v001 and v002, (2) parse out a numeric cluster number and numeric household number from whhid in the WI file to merge with v001 and v002, or (3) truncate the caseid string in the IR file to match whhid by excluding the respondent's line number.

Option 1 requires that you know the rules for spacing strings within whhid and add a certain number of spaces based on the number of digits in the cluster number. Option 2 involves splitting a string and transforming it into numeric values. Option 3 is the most straightforward, as you only need to know the length of whhid and truncate caseid to that width.

To determine how long whhid is in the WI file, you can click on "variable properties" in the bottom right corner of your Stata screen, or type:

```
describe whhid
```

Quote:

Stata will tell you that it is a 12-character string. Based on that information, you can use the substr command to return that number of characters from caseid in the MWIR41FL file. Namely:

```
gen whhid = substr(caseid,1,12)
```

Quote:

Now, if you merge whhid between IR and WI files they will match.

For clarity, the full code is:

```
use MWIR41FL
gen whhid = substr(caseid,1,12)
merge m:1 whhid using "MWwi42fl"
```

* According to the DHS recode manual, CASEID is typically a concatenation of v001, v002, and v003. To confirm that this is true in Malawi, browse the first few columns of MWIR41FL or open that dataset and type

list caseid v001-v003 in 1/10

Then browse the first few rows of MWwi42fl or open it and type

list whhid in 1/10

You can see that caseid is indeed v001, v002, and v003 concatenated together with extra spaces for alignment and that whhid is similar to caseid except--as it is a household-level dataset--there is no respondent line number.
