

---

Subject: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Yawo](#) on Sun, 10 Mar 2019 21:24:46 GMT

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

---

Hello,

I am getting started on big project investigating stigmatization among those living with HIV in sub-Saharan Africa. Given the relatively low rates of testing in SSA, I needed to pool data across countries to ensure my analyses (HLM) models have enough power.

Consequently, I am pooling data from 33 countries with HIV data: Angola Benin Burkina Faso Burundi Cameroon Chad Congo Congo Democratic Republic Cote d'Ivoire Ethiopia Gabon Gambia Ghana Guinea Kenya Lesotho Liberia Malawi Mali Mozambique Namibia Niger Rwanda Senegal Sierra Leone South Africa Swaziland Tanzania Togo Uganda Zambia Zimbabwe

For some countries, like Kenya, we have HIV data for 2 years: 2003 and 2008, while for others like Zimbabwe, we have data for 2005, 2010 and 2015

So, for each country, I need to:

1. append men and women's data, then merge this with HIV test results for a specific year;
2. repeat the same for the next year,
3. append data of year1, year2, year3;
4. Then finally, pool all these data together to create a dataset for all countries, for all years for which they HIV data.

I have already:

- a) selected a subset of variables I needed to be sure they are consistent across countries, and years; including the various survey-specific variables - stratum, psu, etc
- (b) renamed the male variables, from mv\* to v\*;
- (c) dumped all the datasets (men, women, HIV) into one main datafile, that I will use as my working directory.

Here is the process I have outlined, and I will appreciate some comments and suggestions:

```
/* sort Kenya2003 women's data by key variables */
```

```
use Kenya2003_Individual.data, clear  
sort v001 v002 v003  
save Kenya2003_women1.dta, replace
```

```
/* sort Kenya2003 men's data by key variables - note, variables already renamed from mv* to v* */  
use Kenya2003_male.dta, clear  
sort v001 v002 v003
```

```
save Kenya2003_male1.dta, replace
```

```
/* call Kenya2003 HIV data, rename key variables, then sort */  
use Kenya2003_HIV.dta  
ren hivclust v001  
ren hivnumb v002  
ren hivline v003  
sort v001 v002 v003  
save Kenya2003_HIV1.dta, replace
```

```
/* Append Kenya2003 men to women */  
use Kenya2003_women1.dta, clear  
append using Kenya2003_male1.dta  
save KenyaWomenMen.dta, replace
```

```
/* Merge HIV data into the combined Kenya 2003 men and women file */  
use KenyaWomenMen.dta, clear  
sort v001 v002 v003  
merge merge v001 v002 v003 using Kenya2003_HIV1.dta  
save Kenya2003_HIV_MenWomen.data, replace.
```

```
/* Repeat same steps to create Kenya2008_HIV_MenWomen.data */  
use Kenya2003_HIV_MenWomen.dta  
append using Kenya2008_HIV_MenWomen.dta  
save Kenya2003-2008_MenWomenHIV.dta, replace
```

Cycle the same process through all countries. At the end, append, country datasets to each other.

Questions:

1. Assuming everything is correct, how do I handle the psu, stratum, etc variables needed to create my svyset? Do I do this for each dataset (within a country, within a specific year) or wait until the end, and create a grouping variable to do this?
2. Any other advice to help / facilitate this process?

Thanks very much in advance for any assistance -

best- Yy

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Bridgette-DHS](#) on Thu, 14 Mar 2019 18:35:14 GMT

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

You have to be very cautious about interpreting the results from this kind of pooling. It's likely that the stigma has been declining over time and it less in countries with higher prevalence. The surveys have a wide range in dates. Is it appropriate to include more than one survey from a country? We have had a few relevant reports; here are links to two of them:

<https://www.dhsprogram.com/pubs/pdf/AS35/AS35.pdf> and

<https://www.dhsprogram.com/pubs/pdf/AS40/AS40.pdf>.

Appending the IR and MR files and merging with the AR file, and then appending the merged files from all the surveys is the correct approach, and your Stata code looks good. When you prepare the women's file you need a line "gen sex=2", and when you prepare the men's file you need "gen sex==1" and "rename mv\* v\*". For each survey you need to add a line such as "gen str20 survey\_label="Kenya 2003"" and it is also helpful to have a numerical variable "survey" which is 1, 2, 3, etc. When we do this sort of thing at DHS we loop through the files using sub-programs and local notation. I hesitate to describe that process in detail.

You can alter the weights, clusters, and strata afterwards, in the final merged and appended file. There are many forum postings on the options and strategies for doing this.

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Yawo](#) on Fri, 15 Mar 2019 00:37:49 GMT

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

---

Tom and Bridgette: thanks very much for your response. I am still reconsidering my models, and would probably end up focusing on a few years.

I have a few questions re your suggestions: you advised that:

Quote:

1. When you prepare the women's file you need a line "gen sex=2", and when you prepare the men's file you need "gen sex==1"

Is the goal here to ensure that each case in the women's file has a value of 2, and each case in the men's file has a value of 1? And is the command for women, gen sex==2 or gen sex=2?

Quote:

2. For each survey you need to add a line such as "gen str20 survey\_label="Kenya 2003""

I believe the purpose of this command is to generate a string variable labeling each survey ...

Quote:

and it is also helpful to have a numerical variable "survey" which is 1, 2, 3, etc.

And the goal here is to ensure that within each country, we have a variable that indicates whether it is survey 1, survey 2, survey 3.? So, for Kenya, we will have two values, 1 (2003) and 2 (2008). Is that the case?

Thanks very much, .... with much appreciation. Yawo

---

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Bridgette-DHS](#) on Fri, 15 Mar 2019 12:26:10 GMT

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

---

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

You are welcome.

Sorry for the double "=" sign. I should have said "gen sex=2". The reason for putting that variable ("sex") into the appended IR+MR file is that you will want to change the "mv" variables in the MR file to "v" variables, but once you do that, you need a way to distinguish whether the case comes from the IR file or the MR file. The IR and MR files do not contain a variable that is explicitly the sex of the respondent. The coding I suggest is the same is for hv104, with 1 for males and 2 for females. At some point I would also put

```
label define sex 1 "male" 2 "female"  
label values sex sex
```

As for numbering the surveys, you can do that however you want, but the most important thing is also to have a string label such as "survey\_label" because the DHS files do not explicitly include such a label. You cannot rely on v000, for example, to identify the survey. It often happens that two successive surveys have the same value of v000. You can construct the label afterwards using the first two characters of v000, which are the country code, and the year(s) of fieldwork, v007, but I always insert the survey\_label during the construction of the large file, rather than in a later step.

---

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Yawo](#) on Sat, 16 Mar 2019 18:09:15 GMT

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

---

Thanks very much, Bridgette. I appreciate all your assistance.

Will report back if there are any issues.'

Best, Yawo

---

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Yawo](#) on Mon, 18 Mar 2019 14:08:57 GMT

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

---

Bridgette:

I asked this/related questions a bit earlier but got no response - so I am raising it here in the proper context.

for my research, it is important for me to know whether those who are HIV+ are indeed aware of their status (know they are positive). I know the HIV-testing model was tactically confidential and therefore results were not released to participants.

But I am wondering if a combination of two variables - could allow us to make this sort of inference:

- Variable V828 indicates that a sizable number of respondents have had HIV test in the past received the results (and by implication are aware of their status - positive or negative).
- From the HIV biomarker dataset, I call pull out those who are HIV-positive from HIV03.

Will the combination of V828 (HIV status) + HIV03 (positive status) [those who are aware of their status and have tested positive] give me the number who are aware of their HIV+ status? If not, is there any other of getting a sense of those who are HIV+ and know their status?

Thanks , Yawo

---

---

Subject: Re: Merging Women-Men-HIV Data: Different Countries, Different Years  
Posted by [Bridgette-DHS](#) on Fri, 29 Mar 2019 12:50:26 GMT

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

---

Following is a response from Senior DHS Specialist, Joy Fishel:

There are a couple of different ways to estimate the "First 90" percentage of PLHIV who are aware of their status using DHS data.

1. We do have a few surveys in which we directly asked people for their HIV status--Malawi DHS 2010, Uganda AIS 2011, Namibia DHS 2013, and Mozambique AIS/MIS 2015. We also wrote a report assessing the accuracy of self-reporting of HIV status which concludes that it is poor: MR10
2. The UNAIDS "mid-point method" in which you take the average of the percentage of people ever tested for HIV and received the result of their last test and the percentage who were tested for HIV in the past 12 months and received the result of their last test, among people who are HIV-positive according to the survey test result. Note that with this method, you only get

population averages, not a value for each individual. For a detailed definition, see indicator HTS.1 in this guide: <https://www.who.int/hiv/pub/guidelines/strategic-information-guidelines/en/>

UNAIDS has since moved on from this mid-point method. They have concluded that awareness of status cannot be measured accurately from survey data alone and have worked with the HIV Modeling Consortium at Imperial College to develop a statistical model to estimate the first 90. I don't know what public information is yet available about this technique. There may be some info on their website: <http://www.hivmodelling.org/>

Best,  
Joy

---