
Subject: Slight difference in HIV numbers in South Africa 2016 dataset (IR, MR)
Posted by [evavanempel](#) on Tue, 01 Dec 2020 21:16:49 GMT

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

Dear DHS team,

I am using the South Africa 2016 men and women dataset, and merged the HIV biomarker dataset into them. I would like to ask for some help as the HIV numbers are slightly different from the SA 2016 report. The numbers are almost the same as in the report but within the age/region/education/wealth groups there sometimes is one (or a few) extra or missing person(s). I have attached my findings of the women dataset (+HIV biomarker) vs. the report for you to see where the numbers differ.

The steps I have taken:

- Rename variables and sort ascending by cluster (V001), number (V002) and line (V003)
- Merge HIV biomarker dataset into women's dataset
- Compute weight variables (V005 and HIV05) and divide by one million
- Turn on weight for HIV (HIV05). SPSS code:
 - o COMPUTE WGT_HIV= HIV05/1000000.
 - o Weight by WGT_HIV.

I am aware that I have to include 'inconclusive (HIV03=9)' in the HIV negative group. However, this does not seem to solve this issue.

I also tried these same steps with another dataset (Zambia 2018 men, women and HIV biomarker dataset) and there are always one or a few persons extra or missing within the age/region/education/wealth groups.

There seems to be a step that I am missing, would you be able to help me with this problem?

Thank you in advance!

File Attachments

1) [DHS forum - attachment.docx](#), downloaded 27 times

Subject: Re: Slight difference in HIV numbers in South Africa 2016 dataset (IR, MR)
Posted by [Bridgette-DHS](#) on Mon, 07 Dec 2020 14:16:16 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I was able to get a match by merging with the PR file and then selecting the women for whom hv117=1 (for men you would select on hv118=1). These are the women (and men) who were eligible for the individual interviews, whether or not they were actually interviewed individually.

I tried this because the biomarkers are part of the household survey and the covariates in the

table are part of the household survey. However, your approach--merging with the IR file (and MR file) was completely reasonable, and the differences are very small.

```
set more off
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ZAAR71FL.DTA", clear

rename hivclust hv001
rename hivnumb hv002
rename hivline hvidx

sort hv001 hv002 hvidx
save e:\DHS\DHS_data\scratch\ZAARtemp.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ZAPR71FL.DTA", clear

keep hv001 hv002 hvidx hv024 hv104 hv105 hv117 hv118
sort hv001 hv002 hvidx
merge hv001 hv002 hvidx using e:\DHS\DHS_data\scratch\ZAARtemp.dta

tab _merge
keep if _merge==3
drop _merge

tab hv024 [iweight=hiv05/1000000] if hv117==1
```

File Attachments

1) [hv024.png](#), downloaded 70 times
