

---

Subject: Sample size Kenya DHS 2014

Posted by [Glory](#) on Thu, 20 Jul 2017 02:23:50 GMT

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

---

Dear All,

I am running some exploratory analysis on the Kenya 2014 DHS. However, more than half (16,790) of the values in g102 (ever circumcised) is missing which seems strange. This missing category in addition to the Yes (4,377) and No(9,912) binary levels brings the denominator to 31,079. The same size (31,079) was observed across the key socio-demographic variables (v013,v025,v024,v190) compared to 14,289 with no missing values in "g102"-ever circumcised . I realized this is way more than the actual survey size of 14,625 reported in the DHS 2014 report. Even after svysetting in stata, this value(31,079) does not change. I have reloaded the dataset and looked up previous threads for similar comments before posting this.

Kindly assist.

---

Subject: Re: Sample size Kenya DHS 2014

Posted by [Bridgette-DHS](#) on Thu, 20 Jul 2017 15:25:25 GMT

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

---

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

In this survey, about half of the households were randomly selected for more thorough data collection. The selection is indicated by hv027 in the PR file, "household selected for male interview". hv027 has the same value for everyone in the same household. The FGC questions were only asked in the households with hv027=1.

I merged the IR and PR files in order to copy hv027 onto the women's files, and then cross-tabulated g102 and hv027, showing the "missing" cases. These are actually "Not Applicable" cases. Below I will paste the Stata lines and the table.

```
use e:\DHS\DHS_data\IR_files\KEIR70FL.dta, clear
keep v001 v002 v003 g102
sort v001 v002 v003
save e:\DHS\scratch\KEtemp.dta, replace
```

```
use e:\DHS\DHS_data\PR_files\KEPR70FL.dta, clear
keep hv001 hv002 hvidx hv027
rename hv001 v001
rename hv002 v002
rename hvidx v003
sort v001 v002 v003
merge v001 v002 v003 using e:\DHS\scratch\KEtemp.dta
```

tab \_merge  
keep if \_merge==3

---

### File Attachments

1) [tab g102 hv027.jpg](#), downloaded 1036 times

---

---

Subject: Re: Sample size Kenya DHS 2014

Posted by [Glory](#) on Thu, 20 Jul 2017 23:58:37 GMT

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

---

Thank you very much for the feedback.

However, I would appreciate a little more clarification as regards the following:

1. The rationale for selecting "hv027" as a basis-a variable that identifies households selected for male interview, while the variable of interest "g102"-female genital cutting only applies to women in households and not men.
2. Kindly provide a justification for the merge procedure as well, as I expect the IR data file for individual women to suffice to estimate the percent (number circumcised/total women interviewed) of FGM in women, just like any other DHS survey.

I was just wondering if one needs to always merge the IR and PR files to derive the estimates of any indicator from the 2014 Kenya DHS? Also, If I understood this explanation correctly, I can disregard the Kenya 2014 DHS report of 14,625 total sample size.

Many thanks

---

---

Subject: Re: Sample size Kenya DHS 2014

Posted by [Bridgette-DHS](#) on Fri, 21 Jul 2017 15:47:35 GMT

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

---

Another response from Senior DHS Stata Specialist, Tom Pullum:

The only reason why I merged the PR and IR files was to show you that the cases with "." are "Not Applicable" or "NA" cases, not "missing" cases, as in "refused" or "flagged". The FGC questions were only asked of a subsample.

The FGC subsample was the same as the households randomly selected for the male interviewer

simply for easier execution of the survey. Costs were reduced by subsampling for FGC and subsampling men and it was decided to subsample households for both purposes in the same way. There is no reason other than that. Your calculations with g102 and other FGC questions should just use the IR file and the cases that are not missing. The sample size for these variables is the number of cases that are not NA.

---

---

Subject: Re: Sample size Kenya DHS 2014  
Posted by [Glory](#) on Sun, 23 Jul 2017 21:54:02 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

This is helpful. Thank you

---