

---

Subject: Total number households reported vs v002

Posted by [Glory](#) on Sat, 17 Oct 2020 02:50:42 GMT

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

---

Hi,

I'm trying to combine data sets from 2003 and 2008 kenya DHS. However it appears the reported total number of households(in the DHS reports) do not correspond to the total number in the "v002" variable for both surveys. Also the total number of clusters is far more than total households in both surveys (400 vs 203) specifically in the women file. This doesnt seem right.

Please let me know if there is something I'm missing.

Thanks

---

---

Subject: Re: Total number households reported vs v002

Posted by [Bridgette-DHS](#) on Mon, 19 Oct 2020 14:00:22 GMT

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

---

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

The cluster id (v001) is unique for each cluster. However, the household id (v002) is within clusters. That is, each cluster has a household 1, etc. Similarly, line number (v002) is numbered within households. Perhaps that is the source of the discrepancy.

To find out how many households there are you could enter a line such as "egen hh=group(v001 v002)" and then "summarize hh". The maximum value of hh would be the number of households in the survey.

---

---

Subject: Re: Total number households reported vs v002

Posted by [Glory](#) on Mon, 19 Oct 2020 14:40:49 GMT

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

---

Thank you for the quick response.

I doublechecked the procedure to estimate total numebr of households as advised. The numbers of households in the data sets(individual women file) still do not correspond to the reports for the 3 survey years.

2003 - 6159 vs 8561

2008/09 - 6458 vs 9057

2014 - 24705 vs 36430

Please let me know if there is something I have missed.

Thank you.

---

---

Subject: Re: Total number households reported vs v002  
Posted by [Bridgette-DHS](#) on Mon, 19 Oct 2020 18:12:10 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

To get the number of households in the survey you should use either the HR or PR files. What I said before will give you the number of households in the IR file, i.e. households in which there were eligible and interviewed women. Many households do not have any women who were eligible (hv117=1) and interviewed. That accounts for the difference.

---

---

Subject: Re: Total number households reported vs v002  
Posted by [Glory](#) on Tue, 20 Oct 2020 09:20:05 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

This was helpful, many thanks. I also wanted to clarify if the clusters and households from all surveys are randomly selected from the same master list such that there is a potential temporal correlation within clusters and households since a psu/household may be selected multiple times across surveys.

Thank you.

---

---

Subject: Re: Total number households reported vs v002  
Posted by [Bridgette-DHS](#) on Tue, 20 Oct 2020 13:03:52 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

The clusters are enumeration areas from the latest census, sampled with probability proportional to size. If a census occurs between two surveys, the sampling frame is updated. There is no correspondence between the cluster numbers in a DHS survey and the numbering system for the enumeration areas in the sampling frame, or between the clusters numbers in two successive surveys (except for the continuous surveys in Peru and Senegal). There is no way to know whether a specific enumeration area has appeared in two DHS surveys, but it is believed that the chance of this is extremely small. The chance that the same household would appear in two DHS surveys is much smaller than that. You can safely consider two successive surveys to be

statistically independent.

---

---

Subject: Re: Total number households reported vs v002

Posted by [Glory](#) on Thu, 22 Oct 2020 13:08:25 GMT

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

---

This was helpful. Thank you.

---

---

Subject: Re: Total number households reported vs v002

Posted by [Anonymous](#) on Thu, 18 Mar 2021 14:05:42 GMT

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

---

I doublechecked the procedure to estimate total numebr of households as advised. The numbers of households in the data sets(individual women file) still do not correspond to the reports for the 3 survey years.

---

---

Subject: Re: Total number households reported vs v002

Posted by [Bridgette-DHS](#) on Thu, 18 Mar 2021 14:45:24 GMT

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

---

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

The total number of households in the sample is obtained from hv002 in the PR or HR file. It should be the number of records in the HR file, because households are the cases in that file. Some households do not have any eligible women and therefore will not show up in the IR file.

---