

---

Subject: Mismatch of number of clusters

Posted by [MrZ](#) on Wed, 09 Jun 2021 16:37:02 GMT

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

---

Dear DHS - Team,

I'm trying to merge GPS data, household and children data files for a number of countries. I've noticed that sometimes there is a mismatch between the number of clusters (as given in v001 and hv001) across different survey files. For example, in the Namibia 2013 data, there are 550 clusters in the household file and 537 clusters in the children file. However, it may well be the case that some clusters do not contain any children, which would explain a slightly smaller number of clusters in the children file.

In addition, the number of locations in the GPS files do match the number of clusters in the survey files relatively seldom, even though the discrepancies are usually rather small. However, in some cases the difference is large, e.g., the Democratic Republic of Congo GPS file from 2013-14 holds 492 locations of clusters while the household file lists 536 clusters.

I am a bit worried that these "missing" clusters could cause a mismatch between GPS coordinates and survey information. Can we be reasonably sure that, e.g., "cluster no. 52" will be the same cluster in all survey and GPS files, even when the number of clusters are not perfectly aligned?

Thank you!

---

---

Subject: Re: Mismatch of number of clusters

Posted by [Bridgette-DHS](#) on Thu, 10 Jun 2021 20:21:39 GMT

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

---

Following is a response from DHS Senior Sampling Specialist, Mahmoud Elkasabi and DHS Geospatial Technologist, Tom Fish:

There can be fewer clusters in some other files than there are in the HR and PR files just because of how the eligible cases are distributed. For example, as you say, in "the Namibia 2013 data, there are 550 clusters in the household file and 537 clusters in the children file". That's because there were 13 clusters with no children under five. This difference does not indicate an error.

For the question in your last paragraph, we are confident that the clusters in the DHS and GPS files are correctly matched. Someone from the GIS team will respond to your question about locations and clusters and the DRC example.

There are no discrepancies between the published GPS file and the recode; the GPS file has 536 clusters listed just like the recode. There are, however, 44 clusters that we could not verify the location of and are at 0, 0 and are marked as missing (MIS). This process does not affect other clusters. We are sure that the 492 clusters with verified locations (GPS and GAZ) are in the correct places. Please note that all clusters have been displaced by a small amount to protect the

privacy of our respondents. For more information on the process, please see Spatial Analysis Report 7 on The DHS Program's Website.

---

---

Subject: Re: Mismatch of number of clusters  
Posted by [MrZ](#) on Fri, 11 Jun 2021 06:07:15 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thanks for the elaborate and helpful answer!

---