
Subject: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [mlaradji](#) on Fri, 18 Feb 2022 22:05:59 GMT

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

Hi all,

I am currently struggling to find a way to link individual demographics to the hepatitis C blood sample results in the 2008 survey. In the specific dataset that contains the hepatitis C results, EGOD5ADT.dta, only the cluster number, household number, and line number are provided, none of which are usable immediately to determine demographics of those that tested positive or negative (other than their cluster GPS coordinates). Does anyone know of a way to find a possible link here to individual IDs?

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [Bridgette-DHS](#) on Tue, 22 Feb 2022 13:25:48 GMT

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

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

The cluster, household, and line number are all you need to identify individuals in any of the data files. For example, in the PR file they are given by hv001, hv002, and hvidx. In the IR file they are given by v001, v002, and v003. You can do a merge (there are actually alternatives) to attach the blood sample results to the right individuals.

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [mlaradji](#) on Tue, 12 Apr 2022 07:12:25 GMT

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

Tom Pullum,

Thank you for responding. While the cluster number, household number, and line number are all present in the PR and OD dta files for 2008, the OB dta file (which contains the hepatitis C blood test results), does not contain cluster numbers and is unable to match with the PR or OD dta files due to an inability to match uniquely. Is there an alternative variable that should be used for this purpose?

I am also slightly confused on variation between h and w prefixes for variables, such as household or line (wnumber vs hnumber, wline vs hline, etc.) Could you explain this briefly?

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [admin](#) on Fri, 15 Apr 2022 19:58:57 GMT

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

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

I am looking at EGOB5AFL.dta and see obclust, obnumb, and obline, which are the cluster id, household id, and line number, respectively. All three are coded for all cases. Have you tried merging with these codes? Please let us know if you still have difficulty.

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [mlaradji](#) on Thu, 05 May 2022 04:37:11 GMT

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

Tom Pullum,

Thank you for responding again, I have explained the issue further below. I also miswrote the issue in my earlier reply but I have corrected it below.

Identifying variables in each dataset:

EGOB5AFL.DTA : household number (obnumb), line number (obline), cluster number (obclust)

EGOD5AFL.DTA : household number (wnumber and hnumber), line number for women only (wline), and no cluster number

EGPR5AFL.DTA : household number (hv002), line number (hvidx), and cluster number (hv001)

I am attempting to merge all 3 through first merging the OB dataset with the the PR dataset along the household number, line number, cluster number, then merging this new dataset ideally along the same variables but there isn't a cluster number for the OD dataset and there is also only a womens line number for OD dataset. I have redownloaded all the datasets above to double check that it isn't just my version that is missing the variable, could you help with this?

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing
Posted by [mlaradji](#) on Thu, 05 May 2022 14:54:31 GMT

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

Update: I was able to figure it out, the variable hpsu is the cluster number in the OD dataset and the wline variable in the OD dataset, despite being labeled as the women's line number, contains both male and female observations. Through using both of these I was able to make the final dataset I wanted with 12,008 observations.

Subject: Re: Individual IDs for 2008 Egyptian Hepatitis C Blood Sample Testing

Posted by [Janet-DHS](#) on Thu, 05 May 2022 16:39:35 GMT

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

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

In EGOD5AFL.dta, hpsu=hv001; hnumber=hv002; wline=hvidx

In EGOB5AFL.dta, obclust=hv001; obnumb=hv002; obline=hvidx.

I think the problem is just that you did not realize that the PSU (primary sampling unit) and cluster are one and the same.
