

---

Subject: Challenges and limitations of using big DHS data sets

Posted by [sko16](#) on Sun, 24 Sep 2017 21:30:20 GMT

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

---

Dear,

I am carrying the study primarily based on data from Ethiopian DHS 2000, 2005, 2011 and 2016 EDHS on women file recode. The unit of the study are women in reproductive age (15 - 49). I am planning to see further analysis on women's' maternal health care utilization at different levels (Individual Vs Community Level). Would you give me a clue on which the community Cluster level variables that exist in these Ethiopian DHS? Also, I need your professional explanation on Challenges and limitations of using big DHS data sets. Thank you so much in advance for sharing my concern.

---

---

Subject: Re: Challenges and limitations of using big DHS data sets

Posted by [Bridgette-DHS](#) on Tue, 26 Sep 2017 14:53:22 GMT

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

---

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

If you are using a good statistical package, the size of the files will not be a problem. The main difficulty will be learning how to use the package.

There are no cluster-level variables in the data files. You have to construct them as the means and percentages of individual-level variables for the women in each cluster. (The cluster id code is v001.) If you are using Stata, you can do this with some of the "egen" commands. I hope other forum users can offer specific suggestions.

---

---

Subject: Re: Challenges and limitations of using big DHS data sets

Posted by [boyle014](#) on Fri, 29 Sep 2017 21:48:47 GMT

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

---

Dear sko16,

If you use IPUMS-DHS (at [www.idhsdata.org](http://www.idhsdata.org)), you won't have any trouble with file size. IPUMS-DHS is original DHS data harmonized and integrated. It allows you to select just the Ethiopian samples (under SELECT SAMPLES) and just the variables of interest to you, and download them in a single data file. For most people, this file will be much smaller than even one IR file.

You will find the variables you need under TOPICS -> WOMEN & INFANT HEALTH.

In terms of community level variables, IPUMS-DHS makes it relatively easy to link the Ethiopian DHS with Ethiopian census data, which is available through IPUMS-International ([www.international.ipums.org](http://www.international.ipums.org)).

Here's how the linking works. Both IPUMS-DHS and IPUMS-International have a variable called DHS\_IPUMSI\_ET, which you can find under TOPICS -> GEOGRAPHY. Create a data extract in both projects and make sure they both include this variable. Then collapse the census data that interests you by DHS\_IPUMSI\_ET and merge the collapsed file with your IPUMS-DHS file.

This process will aggregate variables from the census at the level of DHS regions. Over the next year, we will be adding finer-grained environmental and social context variables directly to IPUMS-DHS using the GPS location of clusters.

Your DHS ID and password will work for IPUMS-DHS. You'll have to register for IPUMS-International, but approval is typically very straightforward.

Note that we're currently in the process of adding the Ethiopia 2016 DHS to IPUMS-DHS. We plan to release it by the end of the year.

Hope this helps. Feel free to post again if you have any questions.

Liz Boyle, University of Minnesota