

---

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.