
Subject: Re: Dealing with merged data from different countries
Posted by [Bridgette-DHS](#) on Wed, 07 Aug 2024 12:30:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS staff member, Tom Pullum:

You have done an impressive amount of preparation for this study.

You appended the data from India, Pakistan, and Bangladesh into a single file, but I think mostly you will be analyzing the three countries separately. The main reason I can see for appending the countries is that you can do statistical tests of whether there are differences. I would recommend against making pooled estimates for the three countries combined. India would be such a large part of the total that the pooled estimates would basically be the estimates for India.

You do need to construct unique ID codes for clusters and strata. It would be sufficient to use the following Stata lines: "egen clusterID=group(v000 v001)" and "egen stratumID=group(v000 v023)" and then use clusterID and stratumID appropriately in the svyset command. You do not need to change the weights. You would only need to change the weights if you were going to make pooled estimates, and I have advised against that. For making tests, for example a test of the null hypothesis that the mean difference between v201 and ideal ideal number of children is the same in the three countries, you need to put clusterID and stratumID into svyset, but you would use the survey weights unchanged.

When comparing actual and ideal number of children for women age 45-49, you may want to take child mortality into account. The data files include the number of living sons and the number of living daughters. You could use the sum of those two numbers, rather than v201, which includes children who died. Just a suggestion.