

---

Subject: Re: pooling countries to run fixed effect

Posted by [Bridgette-DHS](#) on Thu, 17 Jun 2021 14:36:02 GMT

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

---

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

The variables v001 and v021 are exactly the same in virtually all surveys. There are a handful of old surveys in which one of them is missing, in which case you have to use the other. (For example, if v021 is empty, you would have to use v001.) I believe there is one old survey from Egypt in which v001 and v021 differ, and priority should be given to v021. My general rule would be this: use v021 when it is present, and when it is not, use v001. That will cover all surveys. However, I believe v001 is safe for all surveys except that old one in Egypt.... Similarly, in most recent surveys v022 and v023 are identical and are the stratum. Either can be used. However, for some surveys the stratum variable is different. There is a file in our GitHub site that gives the strata for all surveys.

If "stratumid" and "clusterid" are the correct variables in each survey, then you can use "egen group" to construct the combined ID's as "egen clusterid\_all=group(clusterid survey)" and "egen stratumid\_all=group(stratumid survey)". Then construct svyset. These steps have appeared on the forum several times. Adjustments to the weights have been discussed on the forum many times, along with cautions about pooling surveys. Within DHS, we pool surveys when analyzing a variable for which there are very few respondents in a single survey, or when analyzing trends within a single country, or when analyzing differences between surveys or countries.

You do not need to combine the clusters and strata into some kind of new variable, if that's what you were thinking. Svysset and svy will properly nest the clusters within the strata, and should be used for any estimation command regardless of what variables are in the model. The weights, clusters, and strata are characteristics of the cases and are determined by the sample design. They have nothing to do with any specific variables. Hope this is helpful.