

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

These are good questions. I will give recommendations but there is some room for different approaches.

I would definitely combine the surveys into a single file, but when you do this, you need an identifier for the survey. I would probably just construct `survey=1` or `2` for 2008 and 2018, respectively. You will need unique identifiers for strata and clusters. I won't take the time to look up the stratum identifiers for these two surveys but let's assume it was `v023` in each of them. You construct the combined stratum identifier with `"egen stratum_ID=group(survey v023)"`. You construct the combined cluster identifier with `"egen cluster_ID=group(survey v001)"`.

In each survey, `v005` has been normalized so that the sum of the weights is the total number of cases (times 1000000), and therefore the mean weight in each survey is 1 (times 1000000). (Stata automatically re-normalizes `pweights` and that gets rid of the factor of 1000000.) When you combine the surveys, the overall mean of `v005` will also be 1000000. You do not need to do anything with `v005`. That is, you do not need to do any re-scaling or re-normalizing.

You then use `svyset` including adjustments for weights, clusters, and strata.

I recommend that you only use the combined file for looking at differences between the two surveys. The aggregate of the two survey (e.g. the CPR for the pooled surveys) is not meaningful. If, say, you were combining surveys from several countries for a pooled analysis then you might want to re-scale the weights to take account of differences in sample sizes and/or population sizes but I would be very cautious with such an analysis, and fortunately that's not what you are talking about. Successive DHS surveys in the same country are well-suited for a difference-in-differences approach.

Good luck.