
Subject: Re: Revisiting the topic of weighting data
Posted by [Bridgette-DHS](#) on Tue, 21 Jan 2025 19:14:11 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

I am not enthusiastic about pooling surveys to get some kind of average over a long period of time. However, if you do that, a major issue is that the different surveys will have different sample sizes, and if you don't adjust for that, your results will be most influenced by the largest survey and therefore the conditions at the time of that survey.

Say that n_1, n_2, n_3, n_4 are the four sample sizes and the total is N . Say that the weight variables in the samples are w_1, w_2, w_3, w_4 . You can construct $w_1' = w_1 * N / (4n_1)$, $w_2' = w_2 * N / (4n_2)$, etc. Then, the sum of the weights should be the same in each survey.

You can define the population of interest however you want but it sounds like you want the children under 5 to be the cases, and you would pool the KR files.

Most analysis uses pweights, and they are the weights in svyset. Pweights are automatically normalized in Stata to have a mean of 1 in the separate files and in the pooled file, so it doesn't really matter if you have a factor of 1000000 or something else.
