
Subject: Re: Weighting between DHS and non DHS surveys
Posted by [Bridgette-DHS](#) on Thu, 24 Jan 2019 13:44:46 GMT
[View Forum Message](#) <> [Reply to Message](#)

Please see the following response from Senior DHS Stata Specialist, Tom Pullum:

Benjamin--I think the word you are looking for is "append". That's when two files are combined into one file by, basically, stacking them. If one file has 10,000 cases and the other has 5,000 cases, then the combined file has 15,000 cases. This can be done when working with two DHS surveys, say, because both surveys will have almost exactly the same variable names and categories of variables. For example, when appending two IR files, v013 will be age in five-year intervals in both surveys. But if you have a DHS survey and a non-DHS survey, with different variable names and potentially different categories of variables, then appending will only be useful if you rename and recode the variables in one survey to match the other survey. That's what you would have to do, if the second survey was a MICS survey, for example, because MICS uses different variable names.

Perhaps you want to do this so you can pool the two surveys and get an estimate of something that draws from both surveys. I personally would avoid doing this (but there's no law against it...). The two surveys probably refer to different dates, and the reference time for any estimates will be blurred. From a statistical perspective, a pooled estimate should weight the surveys in proportion to their size. For example, if E1 and E2 are the estimates from the two surveys, and the surveys have N1 and N2 cases, respectively, then the pooled estimate would be $E = (N1 * E1 + N2 * E2) / (N1 + N2)$. But if E1 and E2 are different, especially if they are statistically significantly different, then I'd prefer just to report E1 and E2 and not calculate the pooled estimate E.
