
Subject: Re: Need Clarification: DV weight
Posted by [Bridgette-DHS](#) on Fri, 08 Mar 2019 18:37:59 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is another response from Senior DHS Stata Specialist, Tom Pullum:

It's good that you are being careful, but I think you are being a little too concerned about the weights. Here are the rules or conventions that DHS follows:

For household data, use hv005;
For women's data and children's data, use v005
For men's data, use mv005;
For couples' data use mv005
For DV data, use dv005;
For HIV data, use hiv05.

This hierarchy is based on a tendency for nonresponse to increase as you move down the list.

The factor of 1 million is included only to move the decimal point to the right. Some weight procedures require an integer weight. Those are the only reasons for the factor of 1 million. You do not actually have to remove that factor when doing statistical models, as with pweight in Stata, because the weights are automatically divided by the total weight, so that the mean pweight becomes 1, i.e. the weighted and unweighted sample sizes are forced to be equal. You can easily confirm this. Run a model (a model that uses pweight) with the weights as they are given, then multiply the weights by ANY number whatever, and then run the model again. You will get exactly the same estimates, test statistics, confidence intervals, etc. To repeat: Stata always re-normalizes the weights to have a mean of 1.

I would avoid the label "missing" as in "a missing response observation of 16,182". These cases are "not applicable". They are women who were not selected for the DV subsample, as indicated by v044.

I repeat what I said elsewhere about Stan Becker's couples weights. They are theoretically superior but will not give different results. We do not calculate those weights and I cannot give you a program to do that.

I don't understand "42,419 seems like an undercount of nationally weighted observations compared to the prevalence". The quality of the estimate does not depend on the size of the population or the prevalence in the population, but on the size of the sample. This is a large sample, and certainly at the national level the estimates will have narrow confidence intervals. At the level of the state, or below, they will not be as good, of course. And I expect that non-sampling errors are potentially more serious than sampling errors, especially for a sensitive topic.

The module was only administered to a fraction of the women in order to keep data collection costs down, but because the overall sample was so large, the number in the subsample is larger than the total sample of women in many other DHS surveys. In order to avoid bias, subsampling

is always done in a random manner.

Let us know if you have other concerns.