

---

Subject: Need Clarification: DV weight  
Posted by [Jasminc](#) on Fri, 22 Feb 2019 17:25:42 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

We are trying to find prevalence of IPV in Couples Data (CR dataset) in India from 2005 and 2015 dataset.

We used D005 (domestic violence weight) and divided by 1000000. When we ran the analyses, the number of weighted observations was SMALLER than the number of observations (unweighted).

In order to generate nationally represented estimate for IPV, what do we do for D005? Do we still divide it by 1 million or smaller number?

Thank you,  
Jasmin

---

---

Subject: Re: Need Clarification: DV weight  
Posted by [Bridgette-DHS](#) on Thu, 07 Mar 2019 17:26:03 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

There is not a problem. You should divide by 1 million, as you did. You should use the DV weight, as you did. The weighted and unweighted numbers for subsamples will always differ, and sometimes by a surprisingly large amount--for example 10% or more. This happens regardless of which weight you use.

---

Subject: Re: Need Clarification: DV weight  
Posted by [Jasminc](#) on Fri, 08 Mar 2019 16:50:53 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you, Bridgette, for facilitating the discussion.

Thank you, Tom, for your response.  
We have a follow-up question to clarify the statistical methodology and analyses.

We did what you suggested, and we got the weighted observations.  
For CR (couple) dataset for 2015-16 cycle, we had a total sample of 63,696. Among those, we wanted to see who among women in couple had experienced intimate partner violence (IPV). Therefore, we had a missing response observation of 16,182.

Here are our results:

The weighted observations of IPV among women in couple (DV weight divided by 1 million) is 42,419.

Unweighted observations of IPV among women in couple is 47,514.

We would like to review these numbers with you.

First, as we pointed out earlier, the weighted observations was smaller than the unweighted observations. If the weighted observation (42,419) was to be represented of the number of couples nationally, then it seems way too small. What do you think?

Second, we came across Dr. Stan Becker's literature on the sampling weights for couple data analyses. I saw that you mentioned in other threads regarding Becker's article. You have stated that Becker's methodology was, "although theoretically superior, the effect of using these alternatives is small." After reviewing Becker's literature and your other responses in DHS threads, we would like inquire about the detailed weighting procedures for DHS India Couple dataset.

For India DHS 2015-16 survey, the report states that "601,509 households were interviewed...interviews were completed with 699,686 women... interviews were completed with 112,122 men." We would like to know what is the weighting procedure in order to generate the appropriate weighted results that is represented to the country. For example, if 32.5% among women in couple dataset ever reported IPV (based on our computation and weighted DV by million), then this should represent the national women in couple's experience of IPV. Going back to our weighted observations, 42,419 seems like an undercount of nationally weighted observations compared to the prevalence.

Hope this follow-up question was clearer.

Thank you so much for all your help. I cannot emphasize enough how much your guidance and DHS team support have been tremendously helpful in our research endeavor.

Sincerely,  
Jasmin and the team.

---

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.