
Subject: Data discrepancy-the Philippines 2022
Posted by [w@dhs](#) on Sun, 23 Jun 2024 21:09:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

Thank you so much for the data and codes that is made available. It is really helpful. As I tried to observe HIV knowledge data, I looked at the overall aged 15-24 data. In my data set, the total respondents are 27,821 which is the same as the Philippines DHS report. However, the young people data are different. In my dataset, whether I used v012 or v013, I've got 10,388 for aged 15-24. In the Philippines report, I see 10,208.

Do you have any idea why there is a discrepancy.

Thank you so much.

Sincerely
w

Subject: Re: Data discrepancy-the Philippines 2022
Posted by [Janet-DHS](#) on Tue, 25 Jun 2024 14:16:32 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

In the IR file there are 27,821 women, as you say. The weights for the IR file are constructed such that the mean weight is exactly 1 (that's the mean of v005 when it is divided by 1,000,000). Because of that, the weighted and unweighted sample sizes, for all women, are equal.

However, for subpopulations, the weighted and unweighted totals will not be equal. In particular, that's the case for the women age 15-24. If you enter "tab v013 if v013<=2" you will get an unweighted total of 10,388, as you say. But if you enter "tab v013 [iweight=v005/1000000] if v013<=2" you will see that the weighted total is 10,207.501, which rounds (barely!) to 10,208. The frequencies given in the report, like this one, are weighted.
