

---

Subject: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Manasi\\_C](#) on Fri, 03 Feb 2023 22:47:36 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I had a couple of questions about working with the responses to "wife beating" justification questions (Question 936 in NFHS 4, women's questionnaire):

1. Missing values in NFHS 4 and 5 women's datafiles: the responses to these variables for example variable v744a have ~85% values missing whereas there were just 27 missing values in the NFHS 3 file. I wanted to understand whether these questions were not asked to all women despite the sample size increase or is there some other specific reason? Any guidance on whether we can consider them as random or not would be very helpful.

2. District level Aggregation: I want to study the responses to these questions conditional on some district level covariates given that NFHS 4 and 5 have district identifiers. However, I wanted to check if I would need to construct new weights to do so or would the individual women's weight be appropriate? In past queries, you've recommended dropping some districts that were split when comparing NFHS 4 and 5. Does this imply any adjustment to the weights afterwards?

Thanks in advance for your help.

---

Subject: Re: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Bridgette-DHS](#) on Mon, 06 Feb 2023 18:15:35 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS staff members, Tom Pullum and Fred Arnold:

Response to Part 1. NFHS-5 included a module of questions on domestic violence that was administered to women age 18-49 in the subsample of households selected for the state module. A similar module was also included in NFHS-4, although the NFHS-4 module was administered to women age 15-49, not 18-49. In both NFHS-4 and NFHS-5, households that were not selected for the state module skipped Section 11 on domestic violence, as well as Sections 8-10. In accordance with the World Health Organization's guidelines on the ethical collection of information on domestic violence, only one eligible woman per household was randomly selected for the module in both NFHS-4 and NFHS-5. In both NFHS-4 and NFHS-5, the domestic violence module was not administered to eligible women if privacy could not be obtained after at least three attempts during the survey.

Response to Part 2. Are you thinking of constructing district-level variables based on individual-level responses? For example, you could construct a measure of education using E=0

if a woman has no formal schooling, and E=1 if she any. (The boundary between 0 and 1 could be specified in different ways.) Then the district-level mean of E would be the proportion of women who have E=1. If you do that, you could use either d005 or v005 as the weight. For an individual-level analysis of DV variables you would definitely use d005 for a weight. However, for a contextual variable you could use v005, assuming that all women, not just those who respond to the DV module, provide the context. This is your choice--and there would be almost no difference between the two means. You would not have to calculate any new weights.

---

---

Subject: Re: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Pankhuri](#) on Fri, 13 Sep 2024 07:05:45 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,  
I am working with NFHS 5 for my masters dissertation. I am working on women's empowerment and the use of contraceptives. The missing values for questions on wife beating, land ownership are above 80%. I am confused about how I should treat these missing values since I can't drop these variables as they are extremely relevant to my question. My professor asked me not to impute values and dropping all the missing values would result in significant loss of data. Is there any treatment for this?

---

---

Subject: Re: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Bridgette-DHS](#) on Fri, 13 Sep 2024 16:13:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

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

In this survey, 1/6 of households were selected for the interview of men. There is a variable in the HR and PR files called hv027, "household selected for male interview". In households with hv027=1, all males were interviewed. Otherwise, no males were interviewed.

In the IR and KR files, there is another variable called ssmod, "household selected for the state module". This is exactly the same as hv027. It would have been helpful if the same variable name and variable label could have been used, but that was not done.

The variables you are using are NA, or not applicable, for the women with ssmod=0. You will see this if you enter "tab v744a smod, m".

So what you are seeing is due to subsampling. The "missing" cases are actually "not applicable" cases because the women who lived in households not selected for the male interview were not asked the questions about wife beating. Your analysis must be limited to the subsample,

approximately 1/6 of the total sample, in which the relevant questions were asked.

---

Subject: Re: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Pankhuri](#) on Fri, 13 Sep 2024 18:14:20 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi,  
Thankyou for your reply. I looked at the pattern of missing values and realized that it was a structural thing. Only few women who were selected for domestic violence section were asked about women empowerment questions. Thus, I reduced my sample to currently married women who gave the interview for domestic violence section i.e the 1/6 households(v044=1, v502=1 - because a few of my variables were only asked to currently married women), could you please tell me how I should adjust the sample weights for this. Should i use the d005 (weight for domestic violence section since v044 =1 implies women who were selected and interviewed for domestic violence section) as my weight or v005(sample weight for women) as my weight and should I use the all-woman factor for weighting since I want estimates for all women and not just currently married women. If yes, then which all woman factor should I be using and how? I am pretty confused with the weights thing since I am considering women from the domestic violence sample, but the questions I am interested in are not codes as v744, 743 etc and not d-coded (part of domestic violence record).

Best,  
Pankhuri

---

Subject: Re: Questions "wife beating": Missing values and district level info: NFHS 4 and 5 v/s NFHS 3

Posted by [Bridgette-DHS](#) on Fri, 20 Sep 2024 10:46:46 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

All of the India surveys are all-women surveys. For such surveys you do not need to use the all-women factors. You can see this by entering the Stata command "summarize aw\*". The means of the factors are 100 and the standard deviations are 0, meaning that they are coded 100 for all women. 100 is a scale factor; when the aw factors ARE used, the values are divided by 100.

All-women factors are relevant for surveys limited to EVER married women; no DHS surveys are limited to CURRENTLY married women.

The convention is to use d005 if your analysis includes variables from the DV module or is limited to women who were included in the module. Otherwise you use v005. You do not need to

construct new weights.