
Subject: India 2019-2021 Missing Data
Posted by [behayes4](#) on Wed, 10 Apr 2024 14:31:12 GMT
[View Forum Message](#) <> [Reply to Message](#)

For many of the demographic variables - education, religion, married, ae, number of children - there is absolutely no missing data in a sample of 724,000+ respondents. Is this correct?

Subject: Re: India 2019-2021 Missing Data
Posted by [Bridgette-DHS](#) on Wed, 10 Apr 2024 19:28:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

These variables do not have any cases with a dot (.), the code for Not Applicable, because they applied to everyone. Most of them do have a few "don't know" cases.

A few variables, such as age and number of children, are not allowed to have "don't know" responses, because that would essentially mean that the case would be thrown out, unusable, wasted.

There are indeed some cases in which a household, woman, or man is selected but is dropped from the sample because they cannot be located or simply refuse to be interviewed. Surprisingly few. All DHS final reports describe the response rate in Chapter 1. In the NFHS-5 report, see section 10 of Chapter 1, page 12.

Subject: Re: India 2019-2021 Missing Data
Posted by [behayes4](#) on Tue, 30 Jul 2024 13:34:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks for these helpful response! How come there were 420,149 women who were selected for the IPV interview but not interviewed? This number is much higher than other nations. I cannot find the detail in the report. Thanks!

Subject: Re: India 2019-2021 Missing Data
Posted by [Bridgette-DHS](#) on Tue, 30 Jul 2024 18:58:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

In the HR file, which has one record per household, there is a variable hv044, "household selected for domestic violence module". It is incorrectly coded "1" for EVERY HOUSEHOLD. This variable

is defective. I believe that hv027, "household selected for male interview" was actually used to select for the DV module as well as the male interview.

You are looking at v044, in the IR file, which has the same label as hv044. It has the following categories:

. label list V044

V044:

- 0 woman not selected
- 1 woman selected and interviewed
- 2 woman selected, but privacy not possible
- 3 woman selected, but not interviewed

Code 3, as you say, has 420,149 cases. I believe these cases arise from the defective coding of hv044. Someone thought that hv044 was correctly labeled and coded, and during data processing they used it to determine eligibility. However, those women are in households that should have been coded 0 on hv044 and NA on v044.

There is no problem here, just a coding error.
