
Subject: Comparability of tobacco prevalence between NFHS 2 & 3

Posted by [sarizwan1986](#) on Tue, 23 Aug 2016 02:57:50 GMT

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

We are using the DHS datasets (of 1998 and 2005) of India for calculating prevalence of tobacco use at the national level and comparing trends over the years. A paper (Bhan N, Srivastava S, Agrawal S, et al. BMJ Open 2012;2:e001348. doi:10.1136/bmjopen-2012-001348) has been published previously on this topic using NFHS 2 and NFHS 3 datasets and we are not sure whether this was the right way to go.

I have a query on using NFHS 2 (1998-99) data for India, with reference to calculation of prevalence of tobacco use at the individual level.

In this dataset the tobacco use variables (sh24, sh26) are available in the file 'IAPR42FL.dta'. This is a household members file. We calculated the prevalence of tobacco use by applying the hv005 sample weight.

1. Our query is whether this prevalence will be representative of the total population at the individual level or only representative of the population at the household level?

This confusion does not exist in NFHS 3 (2005-06), because here the tobacco use variables (v463a to v463g and mv463a to mv463g) are available separately in the female (IAIR52FL) and male (IAMR52FL) files with their own individual weights (v005 and mv005).

2. In this regard will the prevalence calculated from NFHS 2 and NFHS 3 be comparable with each other?

3. To overcome this, in NFHS 2 dataset, can we assign a unique ID to each household and randomly select one individual from each household and generate a dataset for the individual level? IF this is possible how will we calculate the individual weights for this dataset? And will the prevalence calculated from this newly created dataset be representative of the total population of India at the individual level?

Thank you for your time and patience. Hoping for an early reply

Subject: Re: Comparability of tobacco prevalence between NFHS 2 & 3

Posted by [Bridgette-DHS](#) on Tue, 23 Aug 2016 16:06:26 GMT

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

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

In NFHS2, the questions about tobacco use were included in the household survey and were asked about everyone in the household, regardless of sex or age. The information was provided by "the household respondent", an adult member of the household.

In that survey, individual women were only interviewed separately if they were ever-married, and men were not interviewed separately.

In NFHS3, the questions were included in the individual interviews of men and women, who responded about themselves. In that survey, all de facto women who were in the sampled households and in the age range 15-49 were eligible for the women's questionnaire. All de facto men who were in the sampled households and in the age range 15-54 were eligible for the men's questionnaire. "De facto" means the person slept in the household last night and is specified with hv103=1.

You have comparability except for the age ranges and the differences in who was providing the information (the household respondent or the specific person to whom the information applies). You can restrict the IAPR42 file to women age 15-49 and men 15-54 (or restrict all the files to age 15-49) to adjust for the differences in age ranges. There will be a possibility that in the NFHS2 the household respondent was not aware of the smoking status of everyone in the household, or for other reasons misrepresented that status. But there's always the possibility that even when describing their own status, individuals will not be accurate, especially if any stigma is potentially associated with the response. However, all the samples are representative of the population living in households at the time of the respective surveys. Differences between the weights, that is, differences between hv005, on the one hand, and v005 or mv005, on the other hand, will be small and only depend on adjustments to correct for nonresponse. Your suggestion #3 below is definitely not needed.

Subject: Re: Comparability of tobacco prevalence between NFHS 2 & 3
Posted by [sarizwan1986](#) on Wed, 24 Aug 2016 02:41:47 GMT

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

Thanks a lot for the immediate response. It was really eye-opening on several levels. Just to make my understanding clearer, a few statements.

So you are saying that after restricting for ages 15-49 years in both NFHS 2 and NFHS 3, prevalence calculated from the household member file (with weight hv005) and individual men/women file (with weight mv005/v005) are essentially representative of the population at large and therefore comparable to each other, keeping aside the fact that in one was proxy-information and the other was self-information.

Am I right?

Thanks in advance.
A truly great forum this is.

Subject: Re: Comparability of tobacco prevalence between NFHS 2 & 3
Posted by [Bridgette-DHS](#) on Wed, 24 Aug 2016 13:28:07 GMT

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

Another response from Tom Pullum:

If you proceed as you describe--restricting to ages 15-49 and using the weights--then you will be able to make direct comparisons of smoking rates at the dates of the NFHS2 and 3. However, I would not say that these would be estimates of "the population at large". First, the age restriction limits you to the population 15-49, and that should be made clear in your analysis. Second, you are limited to the population living in households. People living in prisons, military bases, religious communities, etc., are not included. That's a small percentage of the total population but such people may well differ from the people living in households, in terms of their tobacco use. You can extend your comparisons to various subpopulations, as well--for example, age groups, men and women, urban and rural, etc.

Subject: Re: Comparability of tobacco prevalence between NFHS 2 & 3

Posted by [sarizwan1986](#) on Wed, 24 Aug 2016 15:25:23 GMT

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

Ok. understood.

Thanks a lot for your time.
