

---

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

---