
Subject: Re: Norms variables [[URGENT]]
Posted by [Bridgette-DHS](#) on Fri, 24 Mar 2023 12:38:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

As stated on page 2 of the final report on the NFHS-5, the full sample was designed to be representative at the district level. This means that for variables included in the full sample, such as number of children, the district-level estimates are unbiased and have standard errors that are below some desirable threshold.

For variables in a subsample, such as the DV variables, the estimates continue to be unbiased (because the subsampling is random) but the standard errors are larger (because of the smaller number of cases). In other words, the caution about "representativeness" relates mainly to uncertainty in the estimate rather than potential bias. There is a possibility that you will incorrectly identify some districts as extreme when it's just that they have few cases.

You can proceed and then calculate (using `svyset` and `svy`) the confidence intervals for your district-level estimates. You will see that the confidence intervals are wide, but it's really a judgment call as to whether they are "too" wide. You may be able to group adjacent districts together. Mainly you have to be careful about potentially misdiagnosing specific districts as being very high or very low on your indicators.