
Subject: Household weighted prevalence estimates
Posted by [smugel](#) on Tue, 09 May 2023 18:42:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello everyone,

I am using a number of DHS datasets (HR level) from different countries to estimate the proportion of households using different cookstove fuel types (e.g. prevalence of biomass fuel use). I understand that if I were interested in pooled estimates I would want to either denormalize by scaling each weight to the country's population size (or number of households in this case), or use an equal-country weighting scheme, each with pros and cons. However, rather than pooled estimates, I would like country-specific prevalence estimates for a number of countries in sub-Saharan Africa. I have three questions with respect to using household weights in this case.

1. In reading the forums I came across a note which read to the effect of: weighted counts are not population estimates, but weighted means etc. are population estimates... This was somewhat confusing, so does that mean that a prevalence calculated from the weighted sub-group count numerator and the weighted sample count denominator is a valid representation of the national population prevalence, or not?

2. If I am stratifying by country to generate nationally representative prevalence estimates at the household level, should I use the DHS 'normalized' weights, a 'denormalized' weight scaled to the country population size, or an equal-country weighting scheme? My intuition says that for country-specific prevalence it should not matter because both numerator and denominator weighted counts will be scaled in the same way...

3. If I were to further stratify these prevalence estimates by (a) urban/rural areas, (b) admin-1 levels, and (c) urban/rural areas within admin-1 levels (the two subnational levels for which DHS are also representative), would I use the normalized, denormalized, or equal country weights? Would I need to calculate different weights/strata/PSU designs for each of these levels?

Thank you for consideration of these questions!

Subject: Re: Household weighted prevalence estimates
Posted by [Bridgette-DHS](#) on Wed, 10 May 2023 12:17:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Your intuition is correct. Rescaling the weights with any multiplier will have the same effect on the numerator and on the denominator, Means, proportions, percentages, rates, and ratios are unaffected by re-scaling the weights. This applies to subpopulations too. Means, etc., defined for geographic subdivisions or for wealth quintiles, or any other covariates are unaffected by rescaling. You can just use the weights as they are.

If you were to pool surveys and use the weights in the separate surveys, the overall means, etc.,

would be affected by the relative sizes of the samples. You would then want to re-scale in proportion to the population sizes to get unbiased estimates of the pooled population. But we recommend against such pooling because the surveys are conducted at different times and aggregations are not samples from a well-defined population. Also, the results would be dominated by the largest country. Glad you are not planning to produce pooled estimates.

The observation that "weighted counts are not population estimates" is based on the DHS scaling of the weights, such that the mean of hv005 in the HR file is 1 (multiplied by 1,000,000 to get an integer). Similarly, the mean of v005 in the IR file is 1 and the mean of mv005 in the MR file is 1. If, say, you used the HR file, with hv005 as a frequency weight, to get a count of the number of households that have electricity, the total would be (approximately) the number of households in the sample that have electricity, times 1,000,000. That would NOT be interpretable as an estimate of the number of households with electricity in the country.
