

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

Whenever you pool survey results, you have to choose from 3 possible weighting schemes. #1, you just take the unweighted arithmetic mean of the separate prevalences. #2, you use weights which are proportional to the sample sizes for the estimates. I believe STATcompiler includes the n's for most estimates. #3, you use weights that are proportional to the estimated size of the relevant subpopulation (e.g. children age 0-4) at the time of the survey. You can get these from the UN Population Division's WPP (World Population Prospects) 2022 spreadsheets that give annual age distributions for every country.

#1 is simplest but seems intuitively to be a bad idea. For #2, the problem is that the sample sizes are largely arbitrary. It only makes sense for some statistical purposes, such as hypothesis testing. #3 is probably best, although it has the problem that large countries will swamp small countries.

It depends on what you are trying to estimate. If, for example, you are trying to estimate the probability that a child from any randomly selected household in the geographic region of South Asia had diarrhea in the past two weeks, then #3 is definitely the best option, and it will be ok that most South Asian children live in India.

One issue you will face in any region is that DHS does not provide data from every country in the region. Moreover, the coverage is not consistent over time. I don't know how you fill in these holes in the data; all you can do is to list the countries that contribute to each pooled estimate. For the South Asia example, there are 5-year intervals of time during which there was no survey in India. Including India for the intervals when it had a survey, and omitting it when there was no survey, will produce an uninterpretable trend line.

I personally have always avoided this kind of pooling. Apart from the technical issues of weights and missing observations, it masks important differences between the countries in the same region.