Subject: Re: A little question about pooling data
Posted by Reduced-For(u)m on Fri, 06 Apr 2018 22:47:49 GMT
View Forum Message <> Reply to Message

There are several methods discussed in detail on this forum, but the answer is basically "it depends on what you want to do".

Several options include:

1. Not doing any weighting at all, and then sacrificing the idea that your estmate corresponds to any particular "population"

2. Using the DHS weights given, in which case total influence of one survey is equal to its sample size (as above) but within each country the estimate is population-representative...also not great.

3. Re-normalizing the weights within each country (essentially dividing each weightby the sum of weights within that country, so that all the weights for each country/survey add up to 1) giving each survey both a nationally representative estimate and the overall estiamte the interpretation of weighting each survey/country as equally important (so unit of observation is sort of like the country).

4. Taking (3) but then multiplying those weights by some reference population (say, number of households in a country) so that the sum of weights in one survey adds up to the number of households in that country - this is probably the closest you can get to a "population representative" weighting, if the population you want is the people in all of the countries in your sample.

Choice of these optinos depends on what you want to do. Most published work prior to very recently that merges multiple DHS rounds probably did not pay enough attention to the problem, so I can't promise following the previous literature is the way to go (I don't know those papers you mentioned). I tend to lean towards (1) or (3) these days (partly because 4 means Nigeria is basically everything if you do Africa...or India is everything if you do all DHS countries). But there is no clearly 100% right way...at least until you state what you are trying to estimate in terms of the population you want your numbers to represent.

This is all about getting population-level parameters right. For causal effects estimation, there are a whole set of other arguments that apply, but they all basically relate to 1-4 in terms of what can be done, they just differ on why you might prefer one to the other (for instance, if you a prior believe the causal effect is constant across everyone, you actually don't need to weight at all).

Hope this helps.