Subject: Re: About pooled analyses
Posted by Janet-DHS on Wed, 30 Oct 2024 19:41:10 GMT
View Forum Message <> Reply to Message

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

Yes, there have been many posts related to pooling. Here are my own recommendations (as the former director of research for DHS and the author of several methodological reports). I will put them in terms of child health, your topic, using KR files.

First, it CAN be helpful to combine several KR files into one file, that you could call a megafile. You would do this by appending the files. You would have to check each file BEFORE appending them, to make sure that the variables you want to use are coded the same way in each survey. In many surveys in the same country, for example, the codes for region (v024) are not consistent. The codes for where the child was taken for treatment are usually different. Some surveys have variables that are not in earlier surveys--for example, check whether h31a, h31b, h31c are included. It helps to include a code to identify the survey--v000 is NOT a unique identifier of the survey. All of this is data preparation or data processing, not analysis.

Second, you can analyze the combined file in terms of changes, differences, and trends, using the weights as they appear in the surveys. For example, you can test whether there was a statistically significant change between successive surveys in the proportion of children with fever and cough in the past two weeks who were taken for treatment. You can do complex multivariate analyses of changes, differences, and trends.

Third, some users want to analyze the combined file to obtained pooled estimates. They might think it is helpful to estimate the mean proportion (from the previous paragraph) for all surveys in East Africa since 2000. Then they get into questions about weighting and survey effects that have appeared several times on the forum. My recommendation is that you do NOT do this kind of pooling of surveys. There have been a few DHS reports that pooled in the context of a rare outcome, where individual surveys did not have enough cases. But even then I am uncomfortable with pooling because there is not a well-defined reference population and it is not clear what the weights should be.

To summarize, for your question, I think it could be very helpful to combine the KR files to look at trends and possibly relate change between surveys (or lack of change) to health policies and interventions. But I would not recommend that you pool in the sense of trying to calculate overall pooled means