Subject: Appropriate handling of missing values in analysis
Posted by gebretsh@gmail.com on Thu, 08 Dec 2022 13:07:48 GMT
View Forum Message <> Reply to Message

Dear DHS data experts,
I hope you realize how important your help is in correctly analyzing DHS data.
I have personally benefited much from this help forum.
Now, I would like to ask one practical question with regard to the handling of missing cases in DHS.
I have been analyzing the EDHS data with the help of DHS guide to statistics released in 2018. According to this very
important guide, I created variables in such a way that missings were coded to denominators, at least for some variables.
For example, for the variable mass media exposures, the guide says "Missing values are excluded from the numerators, but included in the denominator." For background characteristics of respondents, it says, "Missing values or "Don't know" responses are shown separately in the percent distributions."
Now, my question, is it appropriate to recode the missing cases into denominator as a standalone category of that particular variable or to recode the missing and or dk into a less disadvantageous category (like no education or no occupation) of a variable. For example, for maternal and partner occupation, is it appropriate to recode the missing and or dk into "not working" group or category, or to let the missing/dk be in a separate "missing/dk" category/group?
My aim is to use the variables in a regression model, not just in a descriptive statistic.
The problem that I would face if I recoded the missing/dk in a separate group is that they would have too small sample to be used in a regression model.

Your expertise would help me to tease apart the dilemmas here in my analysis.
Also note that this question applies to all variables namely maternal age, maternal and partner education, maternal and partner occupation, mass media exposure, whether a child is wanted/planning status, skilled antenatal and delivery care. I mention these variables thinking that your advice may differ by variables