
Subject: DHS survey design

Posted by [Elisabetta](#) on Wed, 25 May 2022 11:05:41 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello,

i have a problem with the analysis of weighted data. I'm currently working on my merged dataset KR+PR (my target is children under 5 years and i need variables from both datasets) using the following survey design:

```
kenya2020 %>% mutate(wt = V005/1000000)
```

```
DHSdesign <- svydesign(id = ~V021, strata = ~V023, weights = ~wt, nest = TRUE, data = kenya2020)
```

everything works until i want to generate predictors for malaria prevalence with the variables in PR: microscopy and RDT.

If I use only PR and the corresponding weight, strata and Id : HV005, HV021 and HV023 for generating the predictor I get percentages that match the report: micro : 3%, RDT = 4.5% the problem is that i want to generate weighted frequency tables and run the svyglm function with my independent variables but I want to use the same target population I used for the other predictors (the merged dataset KR+PR) but if I use my merged dataset I run in a error because i have missing data for ID so i can only use the DHS design of KR (V005, V021, V023) to obtain weighted values for MICROSCOPY and RDT indicators. In this case I get values that are lower compared to those I get using the correct survey design (Because the merge takes into account less children (only those present in KR)).

the question is, can I use the above mentioned survey design to assess malaria prevalence or do I need to find a different solution? In that case can you help me understand how ?

Thank you in advance.

Best regards,
Elisabett

Subject: Re: DHS survey design

Posted by [Janet-DHS](#) on Thu, 26 May 2022 20:37:30 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is response from DHS Research & Data Analysis Director, Tom Pullum:

When you merge the children in the PR and KR files, you will find some who are in the KR but NOT the PR file, and some who are in the PR but NOT the KR file. I think that's why you get a discrepancy. You just need to decide which children to include.

The children in the PR file are a representative sample of all children in the household population. The KR file is limited to children who are living with the mother (and the mother is alive). It includes children who have died (but those children have little data, particularly biometric data).
