Subject: Selecting appropriate weights when IR and MR files are pooled
Posted by gebretsh@gmail.com on Thu, 20 Apr 2023 13:56:25 GMT

View Forum Message <> Reply to Message

Dear DHS experts,
As usual, I am happy to thank you for your important contribution to the good utilization of DHS data analysis.
I plan to do a study on the prevalence of and factors affecting tobacco use in Ethiopia using the 2016 Ethiopia DHS (EDHS) among HIV positives and negatives.
Now, besides doing it separately, I want to pool the IR and MR files together, to which the AR file would be merged. In the pooled data (that contains the IR and MR files), there are two weights: the women's (v005) and the men's (mv005).
Now, my question is, which weight should be used to estimate the prevalence of tobacco and do regression on it using the pooled data among HIV positives?

The reason for pooling is to get an idea of the overall prevalence of the problem in the population (among women and men).

Thank you for your help
Regards,

Subject: Re: Selecting appropriate weights when IR and MR files are pooled
Posted by Bridgette-DHS on Fri, 21 Apr 2023 12:27:33 GMT

View Forum Message <> Reply to Message

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

I would approach this by appending the IR to the MR file (or vice versa), with the following modifications. Add in a variable called sex, coded 1 for the cases in the MR file and 2 for the cases in the IR file, the same coding as hv104, and rename the mv* variables to v* (before the append). Then sort by v001 v002 v003. Rename the cluster, household, and line variables in the AR file, sort, and merge with the IR+MR file.

In analyses that include the HIV result, hiv03, you should use the HIV weight, hiv05. Do not use v005 (originally mv005 for men). hiv05 is the recommended weight, adjusted for nonresponse for HIV testing, which can be substantial.

I would hesitate to restrict to HIV positive cases, but would want to compare the distributions of the positives and negatives.

Subject: Re: Selecting appropriate weights when IR and MR files are pooled

Dear Dr. Tom, I am grateful for your much-needed assistance with my question.
As a follow-up, I would like to be sure whether my merge of the AR file to the pooled file is correct.
I did the merge using the one-to-one merging technique (based on the DHS guide). I get the following result:

Result

not matched   3,572
from master   2,595 (_merge==1)
from using   977 (_merge==2)

matched   25,776 (_merge==3)

In the pooled file, there are 28,371 observations (both women and men), and in the AR file, 26753 observations.
Is my merge correct?

My second question is you advised me to use the HIV weight even if my outcome variable is not HIV test result. Now, when I checked the AR file,
I did not find a strata variable. In my regression model, I plan to account for cluster (v001), HIV weight and strata variable. How can I get the strata or is it not necessary to account for stratification at all in my regression analysis? This seems odd to me because in all other cases, all the three design elements are available (v001, v023 and weight variable).

Finally, I try to replicate Table 3.10.2 on men's Tobacco smoking, page 58 in the 2016 EDHS. I get 5.4% after recoding mv463aa. The table produces two conflicting findings: on the left-hand side, it reported 4.3% (for any type of Tobacco), but on the right-hand side (under the frequency of smoking heading), 3.5% and 1.9% are add up to 5.4%, which is exactly I found. Which figure is correct on the prevalence of smoking any tobacco in the table?
Does the 5.4% include smokeless tobacco use?
Thank you again.
Regards,

Subject: Re: Selecting appropriate weights when IR and MR files are pooled

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

Your results looked ok but the only way I could confirm was by doing it myself. I'll paste the Stata

lines below because they could be useful to others. I get the same results.

```
* Specify workspace
cd e:\DHS\DHS_data\scratch

use "...ETAR71FL.DTA", clear
gen cluster=hivclust
gen hh=hivnumb
gen line=hivline
sort cluster hh line
save ARtemp.dta, replace

use "...ETIR71FL.DTA", clear
gen sex=2
save IRtemp.dta, replace

use "...ETMR71FL.DTA", clear
gen sex=1
rename mv* v*
append using IRtemp.dta
gen cluster=v001
gen hh=v002
gen line=v003
sort cluster hh line
merge 1:1 cluster hh line using ARtemp.dta
tab _merge
```

Subject: Re: Selecting appropriate weights when IR and MR files are pooled
Posted by gebretsh@gmail.com on Mon, 24 Apr 2023 14:12:39 GMT
View Forum Message <> Reply to Message

Thank you so much.
Finally, on this topic, I would like to get your help on the following two points:
Lat time you advised me to use the HIV weight even if my outcome variable was not an HIV test result. Now, when I checked the AR file,
I did not find a strata variable. In my regression model, I plan to account for cluster (v001), HIV weight, and strata variables. How can I get the strata, or is it not necessary to account for stratification at all in my regression analysis? This seems odd to me because in all other cases, all three design elements are available (v001, v023, and weight variable).

The second question is: I try to replicate Table 3.10.2 on men's tobacco smoking, page 58 in the 2016 EDHS. I get 5.4% after recoding mv463aa. The table produces two conflicting findings: on the left-hand side, it reported 4.3% (for any type of tobacco), but on the right-hand side (under the frequency of smoking heading), 3.5% and 1.9% add up to 5.4%, which is exactly what I found. Which figure is correct on the prevalence of smoking any tobacco in the table? I want to know the

prevalence of tobacco use (both tobacco that can be smoked and tobacco that is smokeless). Does the 5.4% include smokeless tobacco use?

Subject: Re: Selecting appropriate weights when IR and MR files are pooled
Posted by Bridgette-DHS on Tue, 25 Apr 2023 12:09:00 GMT
View Forum Message <> Reply to Message

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

For your first question--when you merge the AR variables onto the IR+MR file, the clusters and strata will be the same as in the IR and MR files. Only the weight is adjusted. Because of the line "rename mv* v*", the clusters are v001 (or v021) and the strata are v023. The weight is the only thing that changes in svyset--it is not v005 or mv005, but is hiv05.

Second, tobacco use is separated into smoking (mv463aa) and non-smoking (hv463ab). This is an example of a multiple options question--there are many in DHS data. If you enter "tab mv463aa mv463ab" you will see that some men use both types. Yes, 5.4% includes both types.

Subject: Re: Selecting appropriate weights when IR and MR files are pooled
Posted by gebretsh@gmail.com on Wed, 26 Apr 2023 04:59:10 GMT
View Forum Message <> Reply to Message

Dear Dr Tom,
Thank you so much.
Regards,

Subject: Re: Selecting appropriate weights when IR and MR files are pooled
Posted by gebretsh@gmail.com on Wed, 26 Apr 2023 12:54:22 GMT
View Forum Message <> Reply to Message

Dear Dr, Thank you again for your continual support.
I do have some confusion about the variables.
In the MR file, mv463ab is a variable that says "frequency of past smokes or uses of other types of tobacco." So, it measures past behaviors.
However, mv463aa is about current smoking status. One is current, and the other is past, but my interest is in current tobacco use (smoked and smokeless). So, even when I combine the mv463aa and mv463ab assuming that both reflect current measures, the prevalence becomes 6.23% not 5.4%. I am confused here.

Since my aim is to estimate any tobacco use (both smoked and smokeless), I want to combine variables that measure current smoking and current smokeless tobacco use. For the current smoking, I can use mv463aa, but I cannot find the variables that measure current use of

---

smokeless tobacco.

Thank you again.

Regards,

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

My strategy for resolving this kind of issue is to look at the wording of the original questions. I went to the Final Report on the Ethiopia 2016 survey. The men's questionnaire begins on page 469. The questions about tobacco use are 811-815. The variable labels may not be clear, but it appears to me that ALL of these questions are about current use. I believe you can establish a link between each of the questions and each of the variables, and that should clarify the time period they refer to. Hope this resolves the issue.