
Subject: Malawi Micronutrient Survey design issue
Posted by [IsmailZoutat](#) on Thu, 23 Mar 2023 22:01:02 GMT
[View Forum Message](#) <> [Reply to Message](#)

I'm having some issue with the 2016 Malawi micronutrient survey. I'm using the biomarker dataset to investigate iodine status of school aged children and women of reproductive age.

While in the process of setting the survey design I noticed that the the typical variable for stratum was not present (v024). With respect to that I followed protocol and grouped region and urban/rural residence (mtype x mregion) on stata. Following on from that i then set a new variable named "wght" which took the sample weight variable labelled "mweight" and divided it by 1000000.

As the primary sampling unit variable, typically v021 or any variation of it, was not present I used the "mcluster" as a suitable alternative.

I then set the survey design as follows:

```
svyset mcluster [pw=wght], strata(stratum)
```

Now it all worked fine, however when took to assessing school aged children's characteristics i noticed some discrepancies between the values I was calculating vs those reported in the MNS 2017. I've included the discrepancies in the screenshots below.

Now they are very close to each-other but there should be no reason why my output should be any different to those in DHS report. The number of children are the same (n=800). From what I can only assume, either i've done something wrong in setting the survey design or theres something else at play here. Either way I would much appreciate anyone who can help solve this issue.

Just to add the values in the final MNS report screenshot are said to be weighted. So it's not that they are crude unadjusted values.

File Attachments

- 1) [Screenshot 2023-03-23 at 21.56.09.png](#), downloaded 118 times
 - 2) [Screenshot 2023-03-23 at 21.55.20.png](#), downloaded 122 times
-

Subject: Re: Malawi Micronutrient Survey design issue
Posted by [Bridgette-DHS](#) on Fri, 24 Mar 2023 13:06:32 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You need to merge the micronutrient files back with the PR file. For example, the file for women,

MW_WRA.dta, should be merged with MWPR7AFL.dta, matching mcluster mnumber m01 in MW_WRA.dta with hv001 hv002 hvidx in MWPR7AFL.dta. The cluster number can be obtained as mcluster = hv001 = hv021. All three are the same. The weight will be mweight. The stratum is hv023 (not hv024). Similar for the other micronutrient files.

You will find many examples of merges on the forum but let us know if the steps are not clear.

Subject: Re: Malawi Micronutrient Survey design issue
Posted by [IsmailZoutat](#) on Fri, 24 Mar 2023 19:25:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi that seems to work, however when using hv023 as the stratum i find that i get this issue when tabulating school aged children by region:

```
{  
Note: Missing test statistics because of stratum with  
      single sampling unit.  
Note: Missing standard errors because of stratum with single sampling unit.  
}
```

Subject: Re: Malawi Micronutrient Survey design issue
Posted by [IsmailZoutat](#) on Fri, 24 Mar 2023 19:48:13 GMT
[View Forum Message](#) <> [Reply to Message](#)

So I followed the instructions and merged the biomarker for women and SAC datasets with the PR file.

similarly i set the following svy design:

```
svyset hv001 [pw=wght], singleunit(centered) strata(hv023)
```

whereby wght is mweight/1000000

the estimates im getting for proportion of SAC males and females by region is still a little off. i've attached a screenshot of my latest output against the reported proportions within the MNS report.

Furthermore, to deal with strata with a single sampling unit I set the svy deign with singleunit(centred). I assume thats the correct way to go? what am i doing wrong?

Kind regards
ismail

File Attachments

1) [Screenshot 2023-03-24 at 19.47.37.png](#), downloaded 127 times

Subject: Re: Malawi Micronutrient Survey design issue
Posted by [Bridgette-DHS](#) on Mon, 27 Mar 2023 19:07:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

I just looked at the MNS data in an effort to help with your questions. Unfortunately, my conclusions are that (a) the data are dirty and (b) I can't match the published results. Note that DHS had a secondary role in this component of the Malawi 2015-16 survey. It was mainly conducted by CDC (the U.S. Centers for Disease Control and Prevention).

There are four data files in Stata format from the MNS: MW_WRA.dta, MW_MEN.dta, MW_PSC.dta, and MW_SAC.dta. The files allegedly include the MNS results for Women of Reproductive Age, (age 15-49) Men (age 15-49), Pre-School Children (age 0-4) and School Age Children (age 5-14). In each file, the cluster id is mcluster, the household id is mnumber, and the line number is m01. These are supposed to match with hv001, hv002, and hvidx in the PR file. These files also include mweight, m04 (sex), m07 (age) and some other variables that are in the PR file. It is possible that the errors are with line number but I didn't explore that. [You previously asked about the WRA file and I suggested merging with the IR file, but looking at all 4 files together I think the merge should be with the PR file.]

Using Stata lines pasted below, I combined the four types of files into one and then merged with the PR file. I find many errors. For example, some cases appear to be misclassified--they are not in the correct file. hv104 and m04 do not always agree. hv105 and m07 do not always agree.

When I reconstruct the subsample of school age children, I get 800, matching the 800 in the report table that you give. However, this is the unweighted total. I don't match the breakdown the table gives by hv104, hv024, or hv025. I tried renormalizing mweight to match a total of 800 weighted cases, as well as unweighted cases, but I still do not match the breakdown by hv104, hv024, hv025.

In this situation, I recommend that you go through the steps to construct workfiles as shown in the Stata code, even if you do not match the published results. Good luck.

* Program to prepare the MNS data for analysis

```
cd e:\DHS\DHS_data\scratch

use      e:\DHS\DHS_data\MNS\MW_WRA.dta, clear
gen type=1
append using e:\DHS\DHS_data\MNS\MW_MEN.dta
replace type=2 if type==.
append using e:\DHS\DHS_data\MNS\MW_PSC.dta
```

```
replace type=3 if type==.
append using e:\DHS\DHS_data\MNS\MW_SAC.dta
replace type=4 if type==.
```

```
label variable type type_of_case
label define 1 "WRA" 2 "MEN" 3 "PSC" 4 "SAC"
label values type type
```

```
rename mcluster cluster
rename mnumber hh
rename m01 line
sort cluster hh line
save MW_MNS_sorted.dta, replace
```

```
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\MWPR7AFL.DTA", clear
rename hv001 cluster
rename hv002 hh
rename hvidx line
sort cluster hh line
merge cluster hh line using MW_MNS_sorted.dta
```

```
tab _merge
keep if _merge==3
drop _merge
tab hv105 hv104 if type==3 [iweight=mweight/1000000]
tab hv105 hv104 if type==4 [iweight=mweight/1000000]
* We see classification errors across the initial files; revise the types
```

```
gen typer=3 if hv105<= 4
replace typer=4 if hv105>= 5 & hv105<=14
replace typer=1 if hv105>=15 & hv105<=49 & hv104==2
replace typer=2 if hv105>=15 & hv105<=59 & hv104==1
```

```
label define typer 1 "Women 15-49" 2 "Men 15-59" 3 "Children 0-4" 4 "Children 5-14"
label values typer typer
tab type typer,m
```

```
* renormalize mweight
gen mweightr=.
summarize mweight if typer==1
replace mweightr=round(1000000*mweight/r(mean)) if typer==1
summarize mweight if typer==2
replace mweightr=round(1000000*mweight/r(mean)) if typer==2
summarize mweight if typer==3
replace mweightr=round(1000000*mweight/r(mean)) if typer==3
summarize mweight if typer==4
replace mweightr=round(1000000*mweight/r(mean)) if typer==4
```

* Check the distribution for children age 5-14
tab1 hv104 hv024 hv025 if typer==4 [iweight=mweightr/1000000]
