
Subject: Merging HIV data with Couples recode
Posted by [tedrigecho](#) on Tue, 11 Mar 2014 12:42:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

Helo there,

Is there any one who can help me on how I can merge HIV test result with couple data set?

Best regards,
Tewodros

Subject: Re: Merging HIV data with Couples recode
Posted by [Sarah-DHS](#) on Tue, 11 Mar 2014 14:19:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

Could you tell us which software package you are using to try and merge the datasets (Stata, SPSS, etc?).

Thanks,
Sarah at DHS

Subject: Re: Merging HIV data with Couples recode
Posted by [tedrigecho](#) on Wed, 12 Mar 2014 06:49:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Sarah,

Thank you for your very fast replay. I'm using SPSS. I am analyzing the association between household/couples characteristics and HIV discordance result in Ethiopia 2011 DHS. I have few questions:

1. How can I merge HIV data set with other data sets (couples and Household data sets). How can I make sure whether I did proper merging of these data set? I mean is there any way I can replicate and check my merging process is correct?
2. How can I select men who are in HIV sero-discordance status (I mean men who are married/in union and have different HIV serostatus from their partners)?

Thanks again,
Tewodros

Subject: Re: Merging HIV data with Couples recode
Posted by [Trevor-DHS](#) on Wed, 12 Mar 2014 15:19:41 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Tewodoros

1. I recommend that you start by reading the following page:
<http://www.dhsprogram.com/data/Merging-Datasets.cfm>

Then sort the datasets you are using according to the variables that will be used for merging:
Household (HR): sort cases by HV001 HV002.
Couple (CR): sort cases by V001 V002 V003.
HIV test results: sort cases by HIVCLUST HIVNUMB HIVLINE.

Let's next take the example of merging the household (HR) data with the couple's data. You need the following step:

```
match files
  /file=*
  /table='xxhr60fl_sorted.sav'
  /rename (hv001,hv002=v001,v002)
  /by v001 v002.
execute.
```

where xxhr60fl_sorted.sav is the sorted version of the HR dataset.

Then, let's take the example of merging the HIV test results with the couple's data. you have to do this in two steps - first the women and then the men.

For the women, we will create renamed variables with the women's test results.

```
match files
  /file=*
  /table='xxar60fl_sorted.sav'
  /rename (hivclust,hivnumb,hivline,hiv01,hiv02,hiv03,hiv05=v001,v002,
v003,whiv01,whiv02,whiv03,whiv05)
  /by v001 v002 v003.
execute.
```

where xxar60fl_sorted.sav is the sorted version of the HIV test results dataset.

Now for the men, you first need to re-sort the couples data by the men's line number within the cluster and household:

```
sort cases by v001 v002 v034.
```

And finally merge the HIV test results for the men, creating separate renamed variables for the men's test results.

```
match files
  /file=*
  /table='xxar60fl_sorted.sav'
  /rename (hivclust,hivnumb,hivline,hiv01,hiv02,hiv03,hiv05=v001,v002,
v034,mhiv01,mhiv02,mhiv03,mhiv05)
```

/by v001 v002 v034.
execute.

To test the result, you need to first weight the data. For the couple's data we generally use the men's weight (mv005), however, when tabulating the HIV test results for couples, we use the HIV weight from the men's HIV test result (in this example, that is mhiv05, renamed from hiv05).
compute wgt = mhiv05/1000000.
weight by wgt.

Finally you can crosstabulate whiv03 with mhiv03 and compare the results to the table on couple's HIV status in the DHS reports.

I did this for the Zimbabwe DHS 2010-11 survey, comparing to table 14.12 in the report. See the attached example program.

2. If you follow the example above, then you will be able to select the cases where (whiv03 = 0 and mhiv03 = 1) or (whiv03 = 1 and mhiv03 = 0), i.e. where women's and men's HIV test results differed, but both were tested.

File Attachments

1) [Merging example syntax.sps](#), downloaded 756 times

Subject: Re: Merging HIV data with Couples recode
Posted by [tedrigecho](#) on Thu, 13 Mar 2014 06:58:44 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Trevor,

Thanks for your help. Let me try what you showed me though I'm not good with syntax. If there is anything I will come back to you.

Best regards,

Subject: Re: Merging HIV data with Couples recode
Posted by [tedrigecho](#) on Fri, 14 Mar 2014 14:01:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Trevor,

I'm back again. Your suggestions were very helpful. Now I'm able to do the merging process with understanding. However, I have some confusion on the result I'm obtaining after the merging of the data sets. As I explain myself before I'm not good at using syntax, thus I proceed the process simply using different menu buttons/just by clicking. So;

1. To merge the household (HR) data with the couple's data. I sorted using hv001,hv002,hv003 for

HR and v001,v002,v003 for CR. Then after renaming hv001,hv002 = v001,v002 I merged the two data set. Saved with file name HH_CR.sav

2. To merge couple's data with HIV data. First I used HH_CR.sav as couple data to merge it with HIV data set. First I renamed the variables of HIV data sets for women(hivclust,hivnumb,hivline,hiv01,hiv02,hiv03,hiv05=v001, v002, v003,whiv01,whiv02,whiv03,whiv05). Then sort it. Finally merged with HH_CR.sav using v001,v002, v003. Finally saved the file with file name HH_CR_WHIV.save.

3 Then using the same file, I merge/add variables for Men. I mean, in this stage first I renamed variables of HIV data set for men (hivclust,hivnumb,hivline,hiv01,hiv02,hiv03,hiv05=v001,v002, v034,mhiv01,mhiv02,mhiv03,mhiv05). Then sorted both HH_CR_WHIV.save and HIV files using v001,v002,v034. Finally merged using these three variables (v001,v002,v034).

Unfortunately, when I try to do the computation for HIV discordant [(whiv003=1 & mhiv003=0)|(whiv003=0 & mhiv003=1)], HIV concordant positive(whiv003=1 & mhiv003=1), Women HIV positive discordant (whiv003=1 & mhiv003=0) and Men HIV positive discordant (whiv003=0 & mhiv003=1) it is different from EDHS report. So, here are some of my concerns;

1. Did I do something wrong in the sorting and merging steps? If so, what's it? (For instance in my case HIV discordant is 2.6% whereas in DHS report it 1.1%)
2. If not, what could be the possible solution? what shall I do?

3. One more thing, is the unit of analysis will remain the same after merging such data sets? or it will be affected because of merging process?

I hope I will not take your time too much to get your assistance. I so glade to receive your kind assistance and appreciated kindly.

Best regards,

Subject: Re: Merging HIV data with Couples recode
Posted by [Trevor-DHS](#) on Mon, 24 Mar 2014 18:13:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Tewodros,

First, a basic tenant of science is that your results must be reproducible. If you are using a point and click approach then your results are difficult to reproduce. Users are expected to use syntax to produce their analysis and to permit others to be able to reproduce the analysis. For the information you provide me it is impossible for me to tell what you have done exactly. From your description, I am assuming that you have a made a mistake with which file is your base file when merging, but I can't tell from the description. I highly recommend that you switch to using syntax files and use the syntax that I provided as a basis.

In answer to your questions:

1. Yes, you probably made a mistake in the merging steps, but I can't tell what mistake.
2. Learn to use syntax files for your analysis.
3. The unit of analysis will be the unit of analysis from your base dataset. The order in which you use the files in your merging steps is very important and will affect the unit of analysis. As I suspect you have an error in the merging, you probably do not have the right unit of analysis, but I can't tell from the information you provide.

Subject: Re: Merging HIV data with Couples recode
Posted by [user-rhs](#) on Mon, 24 Mar 2014 21:09:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Tewodros,
In SPSS, the output from things you do essentially contain the syntax that is executed in the background by SPSS when you point and click. If you have the output from your merge, you can copy and paste into a Word document or text editor, and the DHS guys might be able to help you better if they can see what you did.

RHS

Subject: Re: Merging HIV data with Couples recode
Posted by [sinaimail](#) on Fri, 06 Jun 2014 17:04:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Trevor-

I am also merging couples with HIV data and am having some problems. I am using SAS however, so I will review what I did in that code language. Any suggestions as to where I messed up please let me know.

First I created new variables in both datasets to merge with from the original unique identifiers:

```
data keyhiv.a;  
set keyhiv.kear51fl;  
hivclust_merge=hivclust;  
hivline_merge=hivline;  
hivnumb_merge=hivnumb;  
hivline_merge1=hivline;  
hivline_merge2=hivline;  
run;
```

This one is unique for women

```
data keyhiv.w (keep=hivclust_merge hivnumb_merge hivline_merge1 whiv03 whiv05);  
set keyhiv.a;  
whiv03=hiv03;  
whiv05=hiv05;
```

```
run;
```

And this one unique for men

```
data keyhiv.m (keep=hivclust_merge hivnumb_merge hivline_merge2 mhiv03 mhiv05);  
set keyhiv.a;  
mhiv03=hiv03;  
mhiv05=hiv05;  
run;
```

Now creating the same variables in the couples dataset:

```
data keycpl.a;  
set keycpl.kecr52fl;  
hivclust_merge=v001;  
hivline_merge=v003;  
hivnumb_merge=v002;  
hivline_merge1=v003;  
hivline_merge2=v034;  
run;
```

Sorting

```
proc sort data= keycpl.a;  
by hivclust_merge hivline_merge1 hivnumb_merge;  
run;  
proc sort data= keyhiv.w;  
by hivclust_merge hivline_merge1 hivnumb_merge;  
run;
```

Merging the womens data

```
data keymcpl.keymcpl1;  
merge keyhiv.w (in=x) keycpl.a (in=y);  
by hivclust_merge hivline_merge1 hivnumb_merge;  
if x and y;  
run;
```

Sorting for the mens hiv data:

```
proc sort data=keymcpl.keymcpl1;  
by hivclust_merge hivline_merge2 hivnumb_merge;  
run;
```

```
proc sort data= keyhiv.m;  
by hivclust_merge hivline_merge2 hivnumb_merge;  
run;
```

And merging one more time with mens:

```
data keymcpl.keymcplwm;  
merge keyhiv.m (in=x) keymcpl.keymcpl1 (in=y);  
by hivclust_merge hivline_merge2 hivnumb_merge;  
if x and y;
```

run;

The unweighted final number of couples that I get with HIV data is 1228. Which I understand to be too high- I should only have 1188 before weighting. I don't understand where the extra 40 rows of data (couples) come from in my dataset. So I think the problem is before this step.

Now weighting to do the check against the final report data:

```
data keymcpl.keymcplwm2;  
set keymcpl.keymcplwm;  
wgt = mhiv05/1000000;  
run;
```

```
proc freq data=keymcpl.keymcplwm2;  
tables whiv03* mhiv03;  
weight wgt;  
run;
```

The numbers I get from this are 1294 total couples, 90.98% concordant neg, 2.75% m+ w-, 3.2% m- w+, 3.06% concordant pos. I am very close to the numbers in the final report, but I think it has to do with those extra 40 I have before weighting! Did they exclude any of these couples in the final report? Please help. THANK YOU SO MUCH for your reply.

Colleen:

Subject: Re: Merging HIV data with Couples recode
Posted by [sinaimail](#) on Mon, 09 Jun 2014 15:32:24 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Trevor/anyone else familiar with DHS-

I am also merging couples with HIV data and am having some problems. I am using SAS however, so I will review what I did in that code language. Any suggestions as to where I made a mistake please let me know. I have slightly different (more) rows of couples in my merged dataset than appear to be present in the final DHS report for Kenya 2008-2009.

First I created new variables in both datasets to merge with from the original unique identifiers:

```
data keyhiv.a;  
set keyhiv.kear51fl;  
hivclust_merge=hivclust;  
hivline_merge=hivline;  
hivnumb_merge=hivnumb;  
hivline_merge1=hivline;  
hivline_merge2=hivline;  
run;
```

This one is unique for women

```
data keyhiv.w (keep=hivclust_merge hivnumb_merge hivline_merge1 whiv03 whiv05);
set keyhiv.a;
whiv03=hiv03;
whiv05=hiv05;
run;
```

And this one unique for men

```
data keyhiv.m (keep=hivclust_merge hivnumb_merge hivline_merge2 mhiv03 mhiv05);
set keyhiv.a;
mhiv03=hiv03;
mhiv05=hiv05;
run;
```

Now creating the same variables in the couples dataset:

```
data keycpl.a;
set keycpl.kecr52fl;
hivclust_merge=v001;
hivline_merge=v003;
hivnumb_merge=v002;
hivline_merge1=v003;
hivline_merge2=v034;
run;
```

Sorting

```
proc sort data= keycpl.a;
by hivclust_merge hivline_merge1 hivnumb_merge;
run;
proc sort data= keyhiv.w;
by hivclust_merge hivline_merge1 hivnumb_merge;
run;
```

Merging the womens data

```
data keymcp1.keymcp1;
merge keyhiv.w (in=x) keycpl.a (in=y);
by hivclust_merge hivline_merge1 hivnumb_merge;
if x and y;
run;
```

Sorting for the mens hiv data:

```
proc sort data=keymcp1.keymcp1;
by hivclust_merge hivline_merge2 hivnumb_merge;
run;
```

```
proc sort data= keyhiv.m;
by hivclust_merge hivline_merge2 hivnumb_merge;
run;
```



```
And merging one more time with mens:
data keymcpl.keymcplwm;
merge keyhiv.m (in=x) keymcpl.keymcpl1 (in=y);
by hivclust_merge hivline_merge2 hivnumb_merge;
if x and y;
run;
```

The unweighted final number of couples that I get with HIV data is 1228. Which I understand to be too high- I should only have 1188 before weighting. I dont understand where the extra 40 rows of data (couples) come from in my dataset. So I think the problem is before this step.

Now weighting to do the check against the final report data:

```
data keymcpl.keymcplwm2;
set keymcpl.keymcplwm;
wgt = mhiv05/1000000;
run;

proc freq data=keymcpl.keymcplwm2;
tables whiv03* mhiv03;
weight wgt;
run;
```

The numbers I get from this are 1294 total couples, 90.98% concordant neg, 2.75% m+ w-, 3.2% m- w+, 3.06% concordant pos. I am very close to the numbers in the final report, but I think it has to do with those extra 40 I have before weighting! Did they exclude any of these coupels in the final report? Please help. THANK YOU SO MUCH for your reply.

Colleen

Subject: Re: Merging HIV data with Couples recode
Posted by [Bridgette-DHS](#) on Thu, 12 Jun 2014 15:22:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Senior Sampling Expert, Ruilin Ren:

Thanks for your email. Your SAS code is correct and the number of cases (both weighted and unweighted) you got for the couples are correct too. However, the tabulation of table 14.13 in the Kenya report 2008-09 was not done in the standard fashion. The problem is with the treatment of polygamous couples (ie where husbands report they have more than one wife living in the same household). In table 14.13 in the 2008-9 Kenya report, in the case of polygamous men, only the couple including the man with his first wife was included in the tabulation. The standard approach for this table is to include all tested couples, such that a polygamous man would be included in the table for each one of his wives. Please see the SAS code below where the output matches exactly the table 14.13.

Hope this helps.

File Attachments

1) [capture1.jpg](#), downloaded 1666 times
