Subject: Re: Merging HIV data with Couples recode

Posted by sinaiemail 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 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: