
Subject: Difference in N when HIV data merged
Posted by [behayes4](#) on Wed, 26 Feb 2020 15:09:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you. I'm going through systematically to merge country HIV data with men and women datasets. The biggest issue I've caught so far is Burundi 2016-2017. There are 21,955 people in the HIV dataset. When I match and merge in men and women datasets, I am getting 15,904 who have HIV data. I cannot identify why this drop off happens.

Here is how I am recoding for the initial merge to match HIV data:

```
recode mv001 (else=copy) into HIVCLUST.  
recode mv002 (else=copy) into HIVNUMB.  
recode mv003 (else=copy) into HIVLINE.  
EXECUTE.
```

```
recode v001 (else=copy) into HIVCLUST.  
recode v002 (else=copy) into HIVNUMB.  
recode v003 (else=copy) into HIVLINE.  
EXECUTE.
```

I am then doing a one to many merge (Syntax is country specific but here is an example).

```
DATASET ACTIVATE DataSet40.  
SORT CASES BY HIVCLUST HIVNUMB HIVLINE HIV03 HIV05.  
DATASET ACTIVATE DataSet39.  
SORT CASES BY HIVCLUST HIVNUMB HIVLINE HIV03 HIV05.  
DATASET ACTIVATE DataSet40.  
MATCH FILES /FILE=*  
  /TABLE='DataSet39'  
  /BY HIVCLUST HIVNUMB HIVLINE HIV03 HIV05.  
EXECUTE.
```

I am then simply tracking the Ns in an Excel file sheet to identify where missing data appears. So far the other 4 nations I coded have about a 2-5% dropoff between HIV full dataset and then the Ns reported when the data is merged.

Anyone else have this issue?
