
Subject: Merge BDSQ7RDT and BDVA7RDT with the household data file

Posted by [farihakabir148](#) on Thu, 01 Feb 2024 06:04:06 GMT

[View Forum Message](#) <> [Reply to Message](#)

I am trying to merge BDSQ7RDT, BDVA7RDT, and household data file, these three files from the 2017-2018 survey of Bangladesh, but I am failing to find a unique identifier. Can you please help me merge these?

Subject: Re: Merge BDSQ7RDT and BDVA7RDT with the household data file

Posted by [Bridgette-DHS](#) on Thu, 01 Feb 2024 16:04:03 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

I recommend that you think of this as merging the SQ and PR data with the VA data. When you use the VA data, your interest is in the children (up to 8) for whom the mother provides information. The lines given below will do the merge, with one record per mother, as in the VA data. Your next step would probably be to reshape this file into a file that has one record per child.

It is not necessary to include "m:1" or 1:m" in the merge command. Let us know if you have questions.

* BD7R: Merge the PR, VA, and SQ files

* Specify workspace

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

* Prepare SQ file for merge

```
use "...BDSQ7RFL.DTA", clear
```

```
gen cluster=coclust
```

```
sort cluster
```

```
save SQtemp.dta, replace
```

* Prepare PR file for merge

```
use "...BDPR7RFL.DTA", clear
```

```
gen cluster=hv001
```

```
gen hh=hv002
```

```
gen line=hvidx
```

```
sort cluster hh line
```

```
save PRtemp.dta, replace
```

* Open the VA file and prepare for merge

* The VA file has up to 8 records per mother, 1 for each reference child,

```
use "...BDVA7RFL.DTA", clear
```

```
gen cluster=qncluster
```

```
gen hh=qnnumber  
gen line=qnmother  
sort cluster hh line  
merge cluster hh line using PRtemp.dta
```

* Reduce to the mothers who are in both the VA and PR file

```
tab _merge  
keep if _merge==3  
drop _merge
```

* Prepare for merge with the SQ file, which has one record per cluster

```
sort cluster  
merge cluster using SQtemp.dta  
tab _merge  
keep if _merge==3  
drop _merge
```

* Save this file, which has one record per mother, and reshape for one record per child

Subject: Re: Merge BDSQ7RDT and BDVA7RDT with the household data file
Posted by [farihakabir148](#) on Thu, 01 Feb 2024 19:30:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much for responding in such a short notice.

I have a few other queries.

1. The last line you mentioned ' Save this file, which has one record per mother, and reshape for one record per child' how do I reshape the data for one record per child?
2. I plan to look at the impact of exposure to salinity on pregnant mothers during their utero period, on hypertension, swelling of feet, and anthropogenic features of children born to these mothers like height stunt.

I did some calculation of conception date and hypertension data and I found that hypertension data is available only for 1960? (I used the household members data file)

can you please tell me what would be the correct way to find the conception date of mothers, how to form hypertension variable?

Subject: Re: Merge BDSQ7RDT and BDVA7RDT with the household data file
Posted by [farihakabir148](#) on Thu, 01 Feb 2024 21:22:02 GMT

I followed through all your steps and yer, the attached screenshot saying "(you are using old merge syntax; see [D] merge for new syntax) variable cluster does not uniquely identify observations in the master data)" keeps showing up.

which one is a unique identifier here? why isn't it working??

File Attachments

1) [Image 2-2-2024 at 3.20?am.jpeg](#), downloaded 311 times

Subject: Re: Merge BDSQ7RDT and BDVA7RDT with the household data file
Posted by [Bridgette-DHS](#) on Fri, 02 Feb 2024 13:07:36 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

Your study is potentially very complex and I can't provide much more help, but I'll make some comments.

The children in the VA file are children who died. You need children who survived, as well as children who died, in order to identify characteristics that are associated with child survival. It could be better to work with the KR file, which has children born in the past 5 years as the cases. The variable b5 in that file tells whether or not the child survived to the date of interview. If the child died, b7 is the child's months of age at death. The KR file includes many characteristics of the child and birth as well as most characteristics of the mother that are in the IR file. If there are some other variables in the PR or IR file that you need, and they are not in the KR file, you can get them with a merge. You can also merge the SQ file with the KR file. The Stata lines I sent earlier would only need simple modifications to do these merges.

If you want to add any variables from the VA file to the children (in the KR file) who died, please identify which variables you think are important and I can provide more detail on the reshaping I referred to.

Where will you find the data on salinity, hypertension, etc., of mothers when they were pregnant? You cannot safely assume that those kinds of biometric characteristics measured at the time of the survey can be extrapolated back to the time of the pregnancy.

DHS-8 surveys have a new variable, p20, which is the estimated duration of a pregnancy. For earlier surveys, including the 2017-18 BD survey, you would have to use 9 months, a constant, as the duration of the pregnancy, and estimate the month of conception as the month of birth minus 9.

Subject: Re: Merge BDSQ7RDT and BDVA7RDT with the household data file
Posted by [Bridgette-DHS](#) on Fri, 02 Feb 2024 13:26:11 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You are describing an annoying (to me) option in the syntax of the "merge" command. Stata wants you to put "m:1" or "1:m" or "m:m" right after the word "merge". (Omit the quotes.) For example, when you merge the SQ file with the PR file, using "cluster", you have many cases in the PR file that will match with a single cluster in the PR file. This requires "m:1" or "1:m". When I only want the cases with _merge=3, the matches, I don't want to think about WHICH of "m:1" or "1:m" I have to use!

The program ran fine for me. I am still using Stata 16. If you are using a later version, you can replace "merge" with "version 16: merge". The merge command changed well before version 16, and you will still get a warning, but version 16 does not FORCE you to specify "m:1", etc. You just have to know the structure of the data and look at the constructed variable "_merge" and keep the cases you want.