Subject: Combining PR and KR to get child-specific malaria test results
Posted by coffeehelps on Thu, 25 Aug 2016 15:08:47 GMT
View Forum Message <> Reply to Message

I am trying to merge KR and PR files for various samples, for instance for Angola MIS 2006-2007. I am primarily interest in adding the result of the malaria test from the PR file (variable hml33) to the KR file.

My concern is how to attach the malaria test result to the correct child in the KR file. My intuition is that there should be a unique 1:1 match for children in KR and PR so that the test result is properly assigned. That means that there should be no duplicate for a specific child in either file.

I understand that some children in KR may not appear in PR, and dropped those below. But there are still duplicates in the KR file, as seen below.

Any advice is much appreciated.

. * Start with kids recode KR and merge to person recode PR
.
. use AOKR51FL.dta, clear

. gen hv001=v001

. gen hv002=v002

. gen hvidx=b16
(140 missing values generated)

.
. * Exclude children who have died or live elsewhere
. drop if inlist(b16, 0, .)
(195 observations deleted)

.
. * Would expect to have unique observations - but there are duplicates
. duplicates list hv001 hv002 hvidx

Duplicates in terms of hv001 hv002 hvidx

```
 +---------------------------------------+
 | group:  obs:  hv001   hv002   hvidx |
 |-------------------------------------|
 |    1   776     58      53       7 |
 |    1   777     58      53       7 |
 |    2  1171     96      69       5 |
 |    2  1172     96      69       5 |
```

```
  +-------------------------------------+
.
. * 1:1 merge fails
. merge 1:1 hv001 hv002 hvidx using AOPR51FL
variables hv001 hv002 hvidx do not uniquely identify observations in the master data
r(459);
```

Clean code

* Start with kids recode KR and merge to person recode PR

```
use AOKR51FL.dta, clear
gen hv001=v001
gen hv002=v002
gen hvidx=b16
```

* Exclude children who have died or live elsewhere
```
drop if inlist(b16, 0, .)
```

* Would expect to have unique observations - but there are duplicates
```
duplicates list hv001 hv002 hvidx
```

* 1:1 merge fails
```
merge 1:1 hv001 hv002 hvidx using AOPR51FL
```

Subject: Re: Combining PR and KR to get child-specific malaria test results
Posted by Trevor-DHS on Mon, 29 Aug 2016 16:53:27 GMT
View Forum Message <> Reply to Message

Here is a response from Senior DHS Stata Specialist Tom Pullum:

You appear to have encountered two errors in the KR data. I will report them to our DP staff. Here I will go into quite a bit of detail because this kind of detective work may be useful.

There are two children in this Angola KR file who have duplicate values of v001, v002, and b16. I found the duplicates with these lines:
```
use AOKR51FL.dta, clear
gen n=1
collapse (sum) n, by(v001 v002 b16)
drop if n==1
list, table clean
```

You can thus identify v001 and v002 and b16 for these cases and go back to the KR file to get the

mother's line number, v003:
use AOKR51FL.dta, clear
list v001 v002 v003 b16 if v001==58 & v002==53, table clean
list v001 v002 v003 b16 if v001==96 & v002==69, table clean

Giving this:

```
      v001  v002  v003  b16
870.   58    53    4    7
871.   58    53    6    7


      v001  v002  v003  b16
1327.  96    69    3    5
1328.  96    69    7    5
```

Apparently, in these two households, the same child is listed twice, each time with a different mother. To figure out how this happened, my next step was to list the two households in the PR file, as follows:

use AOPR51FL.dta, clear
. list hv001 hv002 hvidx hv101 hv104 hv105 if hv001==58 & hv002==53, table clean

```
       hv001  hv002  hvidx   hv101    hv104  hv105
6704.   58     53     1     head     male     59
6705.   58     53     2    wife or  female    52
6706.   58     53     3   grandchi   male     22
6707.   58     53     4   other re  female    17
6708.   58     53     5   other re  female     0
6709.   58     53     6   grandchi  female    18
6710.   58     53     7   other re   male      0
```

. list hv001 hv002 hvidx hv101 hv104 hv105 if hv001==96 & hv002==69, table clean

```
        hv001  hv002  hvidx   hv101    hv104  hv105
10662.   96     69     1     head     male     38
10663.   96     69     2    son/daug  female   15
10664.   96     69     3    son/daug  female   17
10665.   96     69     4    brother/   male    18
10666.   96     69     5    grandchi  female    1
10667.   96     69     6    other re   male    23
10668.   96     69     7    other re  female   19
10669.   96     69     8    brother/  female   31
```

I believe that in the first household, the first child with b16=7 should actually have b16=5. In the second household, I think there is simply a duplication of the child on line 5.

To confirm that in the second household there is a duplication of the same child, I do this:
use AOKR51FL.dta, clear
keep if v001==96 & v002==69 & b16==5
list, table clean

I won't copy the output, but all the b variables are duplicated, not just b16--so, yes, one of these cases is a duplicat and should be removed. However, which one? Who is the mother? Is the mother the woman on line 3 or the woman on line 7? In many surveys, I could get this from hv112, line number of the mother, but that's omitted from this survey, so I have to go to the IR file.

The mystery child has bidx=1, so I only need to look for the youngest child in the birth histories (child 01) for these two women.
use AOIR51FL.dta, clear
keep if v001==96 & v002==69 & (v003==3 | v003==7)
keep v001 v002 v003 b*_01
list, table clean
These two women have identical entries for the youngest child, and both match with the mystery child. Probably the child should be removed from the birth history of one of the women in the IR file. Based on the order of the household listing, and the relation to head code (hv101), I believe that the child's mother is the woman (age 17) on line 3, NOT the woman (age 19) on line 7.

Therefore I propose that while doing the merge of the KR and PR files, in the KR file you change b16 to 5 for the first child in household 58/53 and you drop the second child in household 96/69:
set more off

use e:\DHS\DHS_data\KR_files\AOKR51FL.dta, clear
keep v001 v002 v003 b*

tab b16,m
drop if b16==0 | b16==.

replace b16=5 if v001==58 & v002==53 & v003==4 & b16==7
drop if v001==96 & v002==69 & v003==7 & b16==5

rename v001 hv001
rename v002 hv002
gen hvidx=b16
sort hv001 hv002 hvidx

save e:\DHS\DHS_data\scratch\AO51temp.dta, replace

use e:\DHS\DHS_data\PR_files\AOPR51FL.dta, clear
keep hvidx hv001 hv002 hv105
sort hv001 hv002 hvidx
merge with the temp file.....
If you do this, you will not get an error message from Stata. Note that you should not permanently alter your AO51 files. If you do, you will regret it later....

Subject: Re: Combining PR and KR to get child-specific malaria test results
Posted by coffeehelps on Mon, 29 Aug 2016 17:46:13 GMT

Thanks very much Tom for this detective work and the detailed description!

There are several other datasets that seem to have similar problems with non-unique observations. I paste my overview below in case that's useful for your CIS DHS. (The variable "mergeid" is my creation and you can ignore it.)

Could you kindly post here or send me a message if/when the source files might be corrected? I may not be able to implement your detailed detective work for all these files and prefer to abstain from making corrections on the fly. For now I'll just drop these duplicates and will update the input data later.

Thanks much!

** List of duplicates
** "mergeid" is not an original DHS variable, please ignore

.. AO2006 - have 4 duplicates

```
+-----------------------------------------------------+
| v000   v001   v002   v003   b16   bidx      mergeid |
|-----------------------------------------------------|
|  AO5    58     53     4     7     1   AO5 58 53 7 |
|  AO5    58     53     6     7     1   AO5 58 53 7 |
|  AO5    96     69     3     5     1   AO5 96 69 5 |
|  AO5    96     69     7     5     1   AO5 96 69 5 |
+-----------------------------------------------------+
```

.. AO2011 - have 2 duplicates

```
+-----------------------------------------------------+
| v000   v001   v002   v003   b16   bidx      mergeid |
|-----------------------------------------------------|
|  AO5    226    32     1     6     1   AO5 226 32 6 |
|  AO5    226    32     2     6     1   AO5 226 32 6 |
+-----------------------------------------------------+
```

.. BJ2012 - have 2 duplicates

```
+-----------------------------------------------------+
| v000   v001   v002   v003   b16   bidx      mergeid |
|-----------------------------------------------------|
|  BJ6    469    11     2     7     1   BJ6 469 11 7 |
|  BJ6    469    11     3     7     1   BJ6 469 11 7 |
+-----------------------------------------------------+
```

.. LS2009 - have 2 duplicates

```
+-------------------------------------------------------+
| v000  v001  v002  v003  b16  bidx      mergeid |
|-------------------------------------------------------|
| LS5   192   177    4   10    1  LS5 192 177 10 |
| LS5   192   177    6   10    1  LS5 192 177 10 |
+-------------------------------------------------------+
```

.. MW2004 - have 2 duplicates

```
+-------------------------------------------------------+
| v000  v001  v002  v003  b16  bidx      mergeid |
|-------------------------------------------------------|
| MW4   499    3    1    3    1  MW4 499 3 3 |
| MW4   499    3    1    3    2  MW4 499 3 3 |
+-------------------------------------------------------+
```

Subject: Re: Combining PR and KR to get child-specific malaria test results
Posted by coffeehelps on Wed, 02 Aug 2017 14:22:37 GMT
View Forum Message <> Reply to Message

Hi DHS colleagues,

I wanted to check on the below thread. As you'll see, there are potential non-unique observations in several KR files. I wondered if you are planning to update those data files?

Thanks!

Subject: Re: Combining PR and KR to get child-specific malaria test results
Posted by Trevor-DHS on Wed, 02 Aug 2017 17:24:03 GMT
View Forum Message <> Reply to Message

Thanks for reporting these cases. We will review these cases and determine a course of action for each one. This may take a little while to complete.
Regards. Trevor