

---

Subject: Merging children in IR with children in PR  
Posted by [Maiwenn Meyer](#) on Mon, 11 Mar 2024 07:55:01 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I am trying to merge the children in the IR files with their long term outcomes in PR files. I have 2 files, one in which I appended all the African IR datasets that I have reshaped to have 1 entry per birth and not 1 entry per woman, and 1 in which I appended all the African PR datasets.

In the IR file, I removed all the births for which the line number of the child in PR is equal to 0 or missing. I have several issues :

1)Using IR : the variables v000, v001, v002, v003, b16, iyr and hhid do not uniquely identify the entries. I have 56,026 duplicates among which 55,940 are in SL5 (23,872) and SN6 (32,068). Could you help me understand why ? For the moment I removed those entries, they represent 2.12% of my dataset.

2)Using PR : I have 9,154 duplicates in terms of all variables. After removing those, I still have 74,032 duplicates in terms of hv000, hv001, hv002, hvidx, hv007, hv112, hhid, all coming from SN6. In the 5 separate files from which SN6 individuals come from ( SNPR6IFL, SNPR6DFL, SNPR6RFL, SNPR7HFL, SNPR7IFL) those variables uniquely identify the entries. Could you help me understand why it is not the case in the appended file ? For the moment I removed those entries, they represent 0.87% of my dataset.

3)When i try merging the 2 files using the following variables I do not recover all the births I should :

```
-hhid which I recovered in IR using the following code : gen hhid = subset(cased,1,12)
-v000 (hv000 in PR)
-v001 (hv001 in PR)
-v002 (hv002 in PR)
v003 (hv112 in PR)
-b16 (hvidx in PR)
-iy (hv007 in PR)
-hhid
```

In the PR file, the line number of the mother (hv112) is not always available depending on the phase. So I separated the dataset into 2 datasets 1 with the entries for which hv112 is missing that I merged without hv112, and 1 in which v112 is not missing that I merged with hv112. Those two databases are exclusive, no individual is in both. Then I appended them and I recovered 2,138,529 births out of the 2,581,109 births that can be merged (with a b16 not missing and different from 0) in IR. Why can't I recover more births ?

Thank you for your help !

---