
Subject: Different findings from the 3 files
Posted by [lka035](#) on Wed, 20 Jul 2016 16:38:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

The Individual recode (IR) contains the child history of all the births born by the women that participated in the survey. I find varying rates for births in the last five years from the different files in the Zambia 2013 DHS data.

The individual recode shows 16,411
FREQUENCIES VARIABLES=V208
/ORDER=ANALYSIS.

Births in last five years
Frequency Percent Valid Percent Cumulative Percent
Valid No births 7058 43.0 43.0 43.0
1 5692 34.7 34.7 77.7
2 3237 19.7 19.7 97.4
3 405 2.5 2.5 99.9
4 19 .1 .1 100.0
Total 16411 100.0 100.0

After restructuring selected variables into cases using the IR file, births in the last 5 years totals to 35,827 (49207-13380)

Births in last five years
Frequency Percent Valid Percent Cumulative Percent
Valid No births 13380 27.2 27.2 27.2
1 19241 39.1 39.1 66.3
2 14245 28.9 28.9 95.2
3 2205 4.5 4.5 99.7
4 136 .3 .3 100.0
Total 49207 100.0 100.0

Household recode file shows 14,043
FREQUENCIES VARIABLES=HC31\$1 HC31\$2 HC31\$3 HC31\$4 HC31\$5 HC31\$6 HC31\$7
HC31\$8
/ORDER=ANALYSIS.

Statistics

	Year of birth	Year of birth	Year of birth	Year of birth	Year of birth	Year of birth	Year of birth
N Valid	9401	3863	645	110	17	5	1
Missing	6519	12057	15275	15810	15903	15915	15919

When I looked at the children recode, births in the last 5 years is 13457

Births in last five years

	Frequency	Percent Valid	Percent Cumulative	Percent
Valid 1	5692	42.3	42.3	42.3
2	6474	48.1	48.1	90.4
3	1215	9.0	9.0	99.4
4	76	.6	.6	100.0
Total	13457	100.0	100.0	

To my understanding the children recode is a subset of the individual recode. Births in the last 5 years should be the same figure in all files.

Kindly requesting if you can shade more light by explaining why the discrepancy in the figures generated from the three different files.

How can I reproduce the children recodes figures from the individual recode file?

Subject: Re: Different findings from the 3 files
Posted by [Trevor-DHS](#) on Fri, 22 Jul 2016 14:21:29 GMT

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

Taking your list in sequence:

1) This is the number of births to women reported in the individual questionnaire.

1 x 5692 = 5692

2 x 3237 = 6474

3 x 405 = 1215

4 x 19 = 76

sums to 13457

2) This looks like you have reformatted the birth history variables into long format and the numbers you show are for all births (42907), not restricted to births in the last 5 years. I don't recommend tabulating v208 after you have reshaped your file as you have now effectively multiplied the number of births in the last 5 years for each women by the total number of births. Instead try the following code:

```
* open data file
```

```
use "ZMIR61FL.DTA",clear
```

```
* tab births in the last 5 years
```

```
tab v208
```

```
* keep only a few variables for this example as reshape can be really slow with a lot of variables  
keep caseid v008 b3* v208
```

```
* rename b3 series to drop 0 on the occurrence for the reshaping to work properly
```

```
rename b3_0* b3_*
```

```
* reshape into records for births
```

```
reshape long b3_, i(caseid) j(idx)
```

```
* drop the cases where there was no nth birth
```

```
drop if b3_ == .
```

* tab to see how many births, limiting to births in the last 5 years (date of interview - date of birth < 60 months)

tab idx if v008-b3<60

* the result is 13457 as in 1)

3) Now you have switched to the household data file and this is a whole different population. The HC series of variables is a series of variables for children under 5 living in the household. This includes all children in the household, irrespective of whether their mother was in the household, or whether the mother was interviewed. This excludes all children who have died. You cannot get a number of births in the past 5 years using the HC series of variables.

4) This is correct and matches the numbers in 1)
