
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)
