

---

Subject: Reproduction of marriage data in report  
Posted by [JanSchuele](#) on Thu, 28 Aug 2014 22:10:33 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

In the research of my thesis I'm using the Bangladesh DHS data to look at child marriage. Now I thought that I was doing the analysis correctly, but when comparing some basic numbers of mine with the ones in the official reports, they differ.

Let us take for example the recent BDHS 2011 final report, page 48, table 4.1 "Current marital status". Though I was also looking at age 10-14, there we can only take age 15-19. So in summary, 45.7% of the 4,306 women age 15-19 have ever been married. HOW IS THIS NUMBER DERIVED?

I took 2011 household file (for SPSS), which firstly I had to restructure to get the single/ unit of HH-members. Then taking one of the equal variables "Currently, formerly, never married" or "Current marital status" for females shows that 47.83% of 4,815 girls age 15-19 have ever been married (,when the dataset is weighted with household sample weight/1000000) The rate is higher, and has approx. 500 more valid cases!! (All women factor is not necessary here.) I had the same problem for all survey years.

CAN YOU TELL ME WHAT I DID WRONG?

-for the report this household dataset must have been used, because in women/individual dataset there is no data on unmarried women, and the household-member dataset somehow shows even less cases than the report for those two marriage variables: 835 at age 15-19.

-weights: Household weight should be the right one here, because that's what you get when you devide the household population weight by the number of HH members. Even no other weight is available here.

-Did they always do some cleaning of data and thereby remove these cases?

-> I don't see the cause of these differences.

WOULD BE HAPPY ABOUT A SOON ANSWER, BECAUSE MY THESIS SUBMISSION DEADLINE IS APPROACHING.Thank you!

---

---

Subject: Re: Reproduction of marriage data in report  
Posted by [JanSchuele](#) on Wed, 03 Sep 2014 09:01:29 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Through searching the forum, I found others who had the same problem, for example here: [http://userforum.dhsprogram.com/index.php?t=msg&goto=778&S=5dcb679460c61ba540f2f0c0c0f91b2b&srch=reproduce#msg\\_778](http://userforum.dhsprogram.com/index.php?t=msg&goto=778&S=5dcb679460c61ba540f2f0c0c0f91b2b&srch=reproduce#msg_778/) /"Rwanda DHS 2010: Discrepancy in result of Hb measurements"

1. A helpful answer was that you still have to use certain filters. In my case, I had to filter out those who were not "de facto" residents. The de-facto-variable, which helped me to get almost the same results as in the report was "slept last night" (in the household). That way, I (still) only have 4383 instead of the correct 4306. Additionally using the residence-variable brings it below the correct number!

There still is something else to consider, which I don't know!

2. Then besides weight, I read that there is also stratification and clustering (in SPSS done in the menu for complex samples", in STATA: svyset). I'll try that and see whether there is a change.

3. The Guide to DHS statistics (for example p.69) shows that even when both married and unmarried women are looked at as denominator (in Household or household member dataset), the ever-married sample denominators are adjusted by the all women factors. So this one I will also try, but have to get that variable from the women/individual dataset, which is more complicated.

-> Now in the meantime I tested the complex samples command, but in frequencies there has been little difference, maybe for other tests there will be more. Am now with 4382

-> I used all women factor (total), divided by 100, and then multiplied with the HV/V-105 weight. The problem is that these women factors are all 1.0 or bigger, for example 1.3, so my total numbers become bigger, not smaller.

-> When searching for more possible filter variables, I tested V135 about residence or visiting of the ever-married woman:

There are 76 cases, who slept there, but are visitors. 76 is exactly the difference between the report number and my result.

So I guess, they took the slept variable for all women plus in addition the residence variable for ever-married women. I'm trying to implement all my findings now.

...unfortunately that number 76 was for all ages, for ages 15-19 such are only 16, which would bring me to 4367.

I don't know what else to do!?!  
  

---

---

Subject: Re: Reproduction of marriage data in report  
Posted by [Trevor-DHS](#) on Sat, 06 Sep 2014 17:21:03 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Jan,

To reproduce the results in the report, the tabulation is run using the women's recode (IR) file. The following code will reproduce the proportion's ever married by age group:

\* Set up svy parameters

gen wt=v005/1000000

svyset v021 [pw=wt], strata(v023)

\* All women factor

gen aw = awfact/100

\* Check the denominator

tab v013 [iw=aw\*wt]

\* Ever Married

gen evermarr = (v502 == 1 | v502 == 2)

\* Use the ratio of ever married and all women

svy: ratio evermarr/aw, over(v013)

The important piece of information here is that with an ever-married sample you need to inflate the denominator by the all women factor to have all women as the denominator, but not inflate the numerator. The simplest way to do this is to use separate variables for the numerator and denominator and take the ratio of the two as above.

However, this does not really help you as you are looking for proportions ever married among 10-14 year old girls. For this you should use the household members (PR) file. As you already realized, you can't reproduce exactly the proportion ever married for the 15-19 year old girls given in the Bangladesh report, but you should be very close. Using the code below with the PR dataset should put you in the right direction for what you need. Note the selection in the svy: tab command - it is selecting for girls who slept in the household the previous night (de facto sample). Also note that all women factors are not needed with the household members (PR) file as you already have all women included in the file, whereas the women's recode (IR) file is restricted in Bangladesh to the sample of ever married women.

\* Set up svy paramaters

```
gen wt=hv005/1000000
```

```
svyset hv021 [pw=wt], strata(hv023)
```

\* Age groups

```
gen ageg = int(hv105/5)
```

```
replace ageg = 14 if hv105 >= 70
```

```
replace ageg = 99 if hv105 == 99
```

```
label define ageg 0 "0-4" 1 "5-9" 2 "10-14" 3 "15-19" 4 "20-24" 5 "25-29" ///
```

```
6 "30-34" 7 "35-39" 8 "40-44" 9 "45-49" 10 "50-54" 11 "55-59" 12 "60-64" ///
```

```
13 "65-59" 14 "70+" 99 "Missing"
```

```
label values ageg ageg
```

\* Ever Married

```
gen evermarr = (hv116 == 1 | hv116 == 2)
```

\* Tabulate

```
svy: tab ageg evermarr if (hv103==1 & hv104==2), row count
```

---

Subject: Re: Reproduction of marriage data in report

Posted by [JanSchuele](#) on Sun, 07 Sep 2014 01:19:06 GMT

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

Thank you a lot Trevor! Firstly, Now I can really reproduce the report numbers for age >= 15 in the individual file. Secondly, following your advice on how to work on age <= 15 also worked well, but brought the same numbers, which I already had before (and were close to the report numbers). There are a few more married women in the PR-dataset than in the women/individual dataset, which might be the main cause of different numbers. Though they were eligible, interviews could not always be completed. And although all women factors are calculated at the Household level, even if the absolute numbers in the women dataset are smaller, because of the constant proportions in the all wome factors (ratio all women to ever-married women), at least the percentages (% ever married) should be the same in both procedures.

Is there an explanation for even different marriage rates/percentages? The percentages are often, but not always equal. I found up to several %points difference.

---

---

Subject: Re: Reproduction of marriage data in report  
Posted by [Trevor-DHS](#) on Sun, 07 Sep 2014 22:05:13 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

The difference comes about because not all women are successfully interviewed, as you mentioned, and the non-response rates vary. If the response rate is very high then you will get almost the same result, but if the response rate is a little lower for a group then you may get a bigger difference. You are likely to find the biggest differences in the younger ages groups.

---

---

Subject: Re: Reproduction of marriage data in report  
Posted by [Inez Roosen](#) on Sun, 28 May 2017 12:21:41 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello,

I am trying to reproduce the Bangladesh 2011 report on marital status like the previous post (p. 48 Table 4.1). Using the women's recode file IR, I followed the instructions in the following posts, however I still do not get the numbers in the report. I do get the amount of observations for each subgroup (15-19 has 4306 ) correctly. However, I get the following output, which does not correspond to the DHS Report in Bangladesh:

Linearized  
Over Ratio Std. Err. [95% Conf. Interval]

```
_ratio_1
_subpop_1 .4573656 .0056447 .446279 .4684522
_subpop_2 .8660365 .0011715 .8637355 .8683374
_subpop_3 .9695809 .0004497 .9686977 .9704641
_subpop_4 .9878733 .0000919 .9876928 .9880539
_subpop_5 .9918975 .0000976 .9917057 .9920893
_subpop_6 .9970216 .0001117 .9968023 .9972409
_subpop_7 .9976134 .0001192 .9973792 .9978476
```

As you can see the percentage of married women of 45.7% does correspond to the number Jan posted. What did I do wrong?

I would highly appreciate your help. Thank you very much.

Kind regards,  
Inez

---

Subject: Re: Reproduction of marriage data in report  
Posted by [Trevor-DHS](#) on Mon, 29 May 2017 15:37:12 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

You just need to follow the logic used before which calculated the proportion ever married, and add the following to calculate proportion currently married and proportion never married:

\* Currently married

```
gen currmarr = (v501==1)
```

```
svy: ratio currmarr/aw, over(v013)
```

\* Never married

```
gen nevermarr = aw - evermarr
```

```
svy: ratio nevermarr/aw, over(v013)
```

This will provide you with the estimates in the first couple of columns in table 4.1.

---

---

Subject: Re: Reproduction of marriage data in report  
Posted by [Inez Roosen](#) on Tue, 20 Jun 2017 17:08:36 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear Trevor,

Many thanks for your feedback, this helped a lot.

However, I did notice even when applying the same logic I couldn't reproduce the marriage data as represented in the DHS reports for India (all waves).

Do you know if I need to do something else with the Indian datasets in reproducing the age at first marriage tables?

Thank you in advance!

Best,  
Inez

---