

---

Subject: Re: Strange Issues w/ Data Formatting from DHS  
Posted by [Trevor-DHS](#) on Tue, 11 Jun 2024 13:57:14 GMT

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

---

Hi

I don't think I can help much with the issue of pooling data and whether to use weights for your analysis. The thread you provided is a good summary of the issues. The note on denormalization provides information on how to create a weight for pooled data. Remember to create a stratum variable that separates the strata across countries. As for whether you need weights for some types of multivariate analysis, that is a debate that I can't give you a simple answer to - there are people on both sides of the debate. For me, I would tend to use weights unless I had a really good reason not to, but that is my view.

In terms of your code below, one strange thing stood out to me. You have the following line:  
`generate current_age = hv007 - v010`

This appears to be calculating `current_age` from the year of the household interview (not the individual interview) minus the year of birth from the individual interview. This is a very inaccurate method of calculating age (it will be wrong about half of the time on average), especially as there is a current age variable in the dataset (`v012`)! I think you should be using `v012`.

Also, you are calculating indicators for married by age 18 and married by age 15. The married by age 15 indicator is incorrectly including 15. This should use `v511 < 15`, not `<=`.

---