Hi,

Thank you very much Tom for providing this code and explanation, it is extremely helpful! I am still uncertain as to whether the variables (cluster, strata, survey) need to have unique identifiers (cluster1, cluster2, etc.) for each survey included in the pooled (appended) data set. So whether it should be egen cluster_pooled= group(survey1 survey 2 cluster 1 cluster 2).

And then I am doing the weighting method you describe (scalar TOTWT=1000000 quietly summarize hiv05 scalar T=r(sum)
gen hiv05r=hiv05*TOTWT/T) for each survey before I append them. Is it correct that I can use the variable name hiv05r for each survey or does this need a unique identifier as well?

I know these are stupid questions but I am just having a hard time conceptualizing this process and it's difficult to check whether I am doing it correctly in stata.

Also, I asked in another thread and have not yet heard back but I am doing this with the couples data merged with the HIV data for multiple countries. I am assuming that I should be using all of the women's variables for this weighting process since the men were chosen on the basis of the women. Is this correct?

If someone could help with these lingering questions I would greatly appreciate it!

Cheers!

Dani