

---

Subject: Re: Mismatch When Merging KR and PR recode files due to disparities in age variables

Posted by [Bridgette-DHS](#) on Fri, 28 Jun 2024 12:24:21 GMT

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

---

Following is a response from Senior DHS staff member, Tom Pullum:

DHS strongly recommends using weights, as well as the other svy adjustments for clusters and strata. If you include weights, the estimates become unbiased. The other adjustments provide robust standard errors because they take into account the stratified two-stage sample design.

When you pool the four surveys into a single large file, you need to construct a variable to distinguish the surveys, for example `survey=1, 2, 3, 4`. The clusters and strata have to be re-numbered to distinguish all four surveys. For example, you could have `"egen clustered=group(survey v001)"` and something similar for strata. (The stratum code is `v023` in the two most recent Kenya surveys but may be different in the earlier two.) The `svyset` statement would be something like `"svyset clusterid [pweight=v005], strata(stratumid) singleunit(centered)"`. Note that in this statement `v005` is NOT altered. This will be ok for all analyses I can think of EXCEPT for analyses that do not distinguish the surveys. For example, staying with the original weights would be questionable if you tried to calculate the mean of `v201` (children ever born) in the four surveys, because that mean would be biased toward the largest of the four samples, and that's not desirable. But I would say that there is no reason to calculate such a mean. The mean of `v201` (for example) should be calculated within each survey, but not for all surveys pooled, because you need a reference date for each estimate of `v201`.

You can find more discussion of weights in the Guide to DHS Statistics (<https://www.dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>), in the FAQ for the user forum, and within the forum itself, if you search topics and keywords.