

---

Subject: Re: accounting clustering effects of women's data when using baby-based analysis

Posted by [Reduced-For\(u\)m](#) on Fri, 27 Sep 2013 18:51:10 GMT

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

---

R,

I think there are two things worth considering.

1 - Standard errors/confidence intervals: In my field, we would use a "clustered" standard errors approach here (one of the "vce" options). To deal with just the multiple-observations-from-same-woman problem, you would want to cluster on the household ID number. But, since the survey is done in a cluster-randomized framework, you would probably want to cluster at a higher (bigger) level to subsume the survey effects, probably PSU (primary sampling unit - its a variable in your data"). If you use the DHS recommended standard errors (see the FAQs on the measureDHS site), that will account for survey clustering too, but I've had trouble figuring out if that is a random effects method or a "clustering" correction (non-parametric V/C matrix). So I would just do it by hand: `reg Y X [pweight=weight], cluster(PSU)` . Note that by clustering on PSU, you are subsuming the woman, so you take care of two problems at once.

2 - point estimates: since you have the same woman in the data multiple times, you will have to adjust your weights to account for this. The weights from the "birth recode" instead of the "woman recode" should take care of this (anyone disagree? I haven't used those weights). You could compare weights and maybe get some idea if this is right.

Anyway, I would cluster at the PSU level for your standard errors and make sure you weight the data to account for the multiple women. That is standard in my field (applied microeconomics), and is increasingly common in other fields. Other people like the hierarchical model, but point-estimate-wise you should get very similar results either way, and the clustering method should be slightly more conservative and easier to implement and defend.