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

Posted by rkinoshita on Fri, 27 Sep 2013 12:42:43 GMT

View Forum Message <> Reply to Message

Dear colleagues,

I am a master's student in Epi and currently finishing my master's thesis.

I am using DHS data from Nicaragua looking at the relationship between partner violence and newborn health outcomes, primarily neonatal, infant and under5 mortality. I am using a dataset in which women's and children's datasets were merged. When I merged two datasets, they created duplicated women's records because I am doing birth-based analysis (each birth needs to have their mother's record).

I've run into trouble with a birth-based analysis because the data for mothers who have more than one birth is duplicated, ending up over-representation of women who had more than one child in the analysis of regression. We could limit the analysis to one birth per woman but in this case with a mortality outcome, I prefer to keep all the births data to get adequate power.

My supervisors at university suggested to use vce option of regression in STATA. other suggested options are random effect modeling and use of a hierarchical model. For none of the methods, I am familiar with, and my deadline for submitting thesis is coming soon.

could you please let me know what would be the easiest and most appropriate method in this case, to minimize the over-representation of women with multiple births in my analysis? I am concerned because parity could be associated with other confounder or effect modifiers that I am looking at, and need to control in the final model.

thanks in advance.
Rinko