Subject: Re: Selecting one child per mother
Posted by Bridgette-DHS on Tue, 29 Jan 2019 13:18:35 GMT

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I assume that you are using the children in the KR file. In that file, the mothers are identified by v001 v002 v003. I suggest the following:

egen mother_id=group(v001 v002 v003)
gen rn=uniform()
sort mother_id rn
egen sequence=seq(), by(mother_id)
keep if sequence==1
drop rn sequence

This procedure sorts the children of each woman in a random order, and then selects the child with the smallest random number. Because of the random step in it, the results will not be replicable, at least not exactly. Every time you run it, you will get a slightly different sample of children. That's inevitable if you do it at random An alternative would be to use all children, with a multi-level adjustment for the similarity of children of the same mother. Another alternative would be to take just the youngest child, for example, the child with bidx=1. However, that will introduce some bias (see https://www.dhsprogram.com/pubs/pdf/MR14/MR14.pdf).