Subject: Re: Pakistan 2014 IPV analysis -- multi-level weights
Posted by Reduced-For(u)m on Sun, 08 Nov 2015 21:40:34 GMT
View Forum Message <> Reply to Message

I'm not sure there is a "right" answer here, although surely some wrong ones. I'll give a couple of thoughts:

1 - a weight is just a person's probability of selection (re-scaled in a DHS-manner, but still basically a probability weight). This does not vary by community, because there is random sampling within a cluster. So everyone in each PSU will have the same weight, and so you can just use the individual weights.

2 - Even adding in information from the Men, if the unit of analysis is still the individual woman you would want the women's weights.

3 - if you are trying to estimate "causal effects" of the community level variables, and you think those relationships are independent of the criteria used for sampling villages (so that, say, all women respond the same to a 1sd change in their "decision making" then you don't need to weight. Weighting buys you population-level mean estimates, but if there are "constant treatment effects" regardless of covariates, you may not need to weight*.

4 - cluster-level measurements are based on too few observations to be meaningful in and of themselves - as you say, there are wildly under-powered. A couple of things you could do: a) by averaging over many clusters, you can still get good estimates of community level variables, but each individual cluster-level point-estimate would be very, very noisy. But they may still mostly "agree" in some sense; b) so if in your hierarchical model you allow each cluster an unconstrained cluster-specific effect (like treating each cluster as a mini-experiment), you could look at those individual point-estimates on a scatter plot (say Beta across some variable you think would affect Beta); c) and then you could start restricting those Betas to have some particular distribution (a random slope model) and see how that changes your overall point estimate as you make your priors on the distribution of Beta more/less informative. I think this makes sense as a kind of model-checking or informal/additional inference procedure. A leave-one-out cross-validation approach might make sense too, depending on how you end up thinking about each of these within-cluster estimates.

I hope something there helped.