
Subject: Re: Which data file should I use for nutrition outcomes of children under 5 years of age

Posted by [Reduced-For\(u\)m](#) on Mon, 13 May 2013 21:26:56 GMT

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

Hi,

Glad I could help a little. Estimating cross-sectional determinants of child HAZ or time-invariant ones is certainly easier than trying to estimate cohort-based determinants (what I meant by time-variant), but it still requires, in my opinion, a bit more care than some people give it.

In particular, I worry most that people don't sufficiently worry about the distribution of child age-at-measurement across their explanatory variables of interest. I think we say "this is age adjusted height, so age shouldn't be a big predictor", but if you collapse HAZ by age-in-months, and graph it out, you'll realize how important age-at-measurement actually is in DHS countries...because HAZ is a cumulative measure of health/nutrition up until age-at-measurement, older kids have had a lot more time to "lose" HAZ relative to well-nourished children in the reference group.

Just a couple of things to keep in mind: 1) if estimating time-invariant factors (say, rural born or maternal age at birth), make sure that the distributions of child age are similar across X (so, if X is "rural born", overlay a histogram or kernel-density plot of ages for rural and urban born children, and see if they match). 2) if you are using "time semi-variant" things like, say, Asset Quintile, you might have a more pronounced problem in that older parents tend to have both more assets and older children (this could bias your estimates of asset effect downward). 3) if you are using "cohort" variables, such as "drought exposure in-utero", you have to be super-duper careful, because some drought year where lots of kids are exposed will be correlated with some age-at-measurement, and thus induce a spurious HAZ-drought association that is driven by a drought/age-at-measurement association.

The gist is that most people include a linear control for age-in-months, and then write "age is a strong predictor of HAZ", which is true, but almost misses the point. Age is THE best predictor of HAZ in a lot of countries, but it is decidedly non-linear, and the model misspecification error (because it is specified erroneously as linear) is often times correlated with age in such a way that any covariates just accidentally associated with child age will pick up the misspecification error and attribute it to the covariate.

I find that in things like estimating effects of maternal age this affects coefficient estimates just a little bit. In things like in-utero/birth-year economic/health environment (cohort stuff), this affects estimates a whole lot. In between...I don't know, depends on the situation.

So... If you feel like it, once you get your list of determinants down, estimate it a few ways, by specifying age as linear, quadratic, a spline with nodes at each age-in-years, and dummy variables for each age in months, and then post the coefficient estimates on a few of the key determinants of interest for each specification. We can see what kind of difference it makes to your estimates.

Sorry...I'm almost done with a paper on this, and so I talk a lot about it.

Best,
j
