

---

Subject: Re: Data analysis using multiple countries different survey years

Posted by [Reduced-For\(u\)m](#) on Fri, 17 Oct 2014 02:17:13 GMT

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

---

I think the answer here will depend a lot on what is the ultimate aim of your analysis. What is your outcome? Are you looking to identify the effect of some particular covariate across countries, or a whole bunch of determinants of some covariate? Matching as you are describing it is usually used to identify some particular causal effect, but apparently you want to match across countries instead of within them (note: GDP will match perfectly within countries) and match on country-year level covariates? Or are you collapsing these surveys down into one observation?

As for your list:

If you think that MM at age 18 is what matters, women should have a value of the appropriate MM at age 18 for their covariate. If you think the MM the year prior to the survey is what matters, you should include that.

Same with GDP - is it the GDP they faced at 18 that matters (even for a 15 year old?) or is it the GDP they faced last year.

All these will answers will depend on your outcome and your theoretical framework. But regardless of what you are trying to estimate, matching a woman in Ghana in 2007 with a woman in Zambia in 2007 (or 2011, or whenever) sounds like an odd approach.

Some things to definitely look out for:

1 - You will have to re-normalize your survey weights in some way or another. See the discussions in the thread.

2 - Many variables are not comparable across countries, an obvious example being household wealth index which is only country and survey round specific.

3 - You will have to adjust your standard errors to deal with the aggregate variables in your regressions, perhaps clustering on something like age-by-country. Clustering on PSUs will not be sufficiently flexible here.

4 - Anything that varies with age will, in this context, vary with time as well. GDP keeps going up for the most part over time. If there is a general secular trend in your outcome variable over age, this will be artifactually correlated with a secular trend over "year" or "time", and thus you can easily end up with a spurious correlation between your GDP measure and your outcome which is actually a function of age-at-measurement.

Those are a few things to definitely look out for, but I don't think anyone here can provide you with better help on your matching/year problem without understanding what you are trying to estimate and why you think matching is appropriate.

