

---

Subject: Re: Merging data files in Stata

Posted by [Bridgette-DHS](#) on Wed, 28 May 2014 13:21:46 GMT

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

---

Following is another response from Tom Pullum:

With four surveys from the same country, my suggestion of multi-level analysis would not apply.

Your outcome or dependent variable is whether the child died before the first birthday (0=no, 1=yes). Some predictors of interest--birth attendants and place of delivery--are usually available for all children born in the past five years. Others, such as antenatal care, are usually only available for the youngest child born in the past five years. You have to be careful because of the fact that the subgroup of youngest children usually has a better chance of surviving than the other children born in the past five years (the youngest child tends to have a longer preceding birth interval). You also have to take into account that children born less than a year before the survey have not had full exposure to the risk of dying before the first birthday, i.e. are censored.

You should not do just a linear probability model. This is never recommended for a binary outcome, although the fact is that if the fitted probabilities are in the range from .3 to .7, approximately, then a model that is linear in the probabilities will agree pretty well with a model that is linear in the log odds (or logits). Since your probabilities are small, you should do a logit (or probit) model or a log probability model, which is basically equivalent to a hazard model or survival model. I assume you are using a statistical package, and it's easy to do it the right way.

Last year DHS did a Further Analysis report on neonatal mortality in Rwanda using the 2000, 2005, and 2010 surveys: <http://dhsprogram.com/pubs/pdf/FA88/FA88.pdf>. It may help.

---