

---

Subject: Stata do file for U5 mortality analysis  
Posted by [namsgfr](#) on Mon, 08 Apr 2013 14:59:32 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear DHS,

I am requesting anyone who has done U5 mortality analysis in stata using complete birth history tables to share the do file. I would like to get these skills. Is this Analysis treated like typical survival analysis with stsplot?

Thanks

Geoffrey

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [Reduced-For\(u\)m](#) on Sat, 13 Apr 2013 05:44:37 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I might have some code that I could adapt, or at least present in some way that gets the gist of the program, but it would take some work (they are imbedded in much longer .do files). What kind of U5 mortality rates were you hoping to compute? There are a few ways and they all have slightly different interpretations, and I don't use the standard DHS method (not because mine is better, it was just better suited to the task I had at hand).

The DHS U5 mortality rates come from (I think) estimating age-period-specific mortality probabilities, and then computing the probability that a child doesn't die before age 5. The assumption here is that the age-mortality profile is not shifting over time (so the 36 month old kids born 3 years before the survey are a good representation of mortality chances kids born today will face in 3 years). The complicated part about this involves cohort-timing and age-timing stuff, where they use cohort level survival ratios but some cohorts end up in different age groups and you have to like use a fraction of that cohort. I'm still digesting it, but the DHS manual covers it here on page 90-94: [http://www.measuredhs.com/pubs/pdf/DHSG1/Guide\\_to\\_DHS\\_Statistics\\_29Oct2012\\_DHSG1.pdf](http://www.measuredhs.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf)

I haven't done that. I've done a cohort-level estimation, which they describe as option 2 on page 91. You just compute the fraction of kids born in any year who died before they turned some age. But for U5 it can be a problem (I'm mostly concerned with U1), because you can only get rates starting 5 years before the survey - a kid has to live to age 5 until you know if s/he will survive to age 5 for sure. This method is, for me, more parsimonious, but it won't help if you want a u5 mortality rate estimate for 2010.

Of course, you can always just take the total number of births in the birth history and divide by total number of deaths (option 1 page 91), and that will estimate a parameter that can be compared across surveys, but it is not a U5 mortality rate exactly, since a lot of the kids in the survey haven't lived 5 years yet, so we don't know if they will survive or not later on.



So...which method were you trying to do, or what are you hoping to use the rate for?

---

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [namsgfr](#) on Thu, 18 Apr 2013 07:42:17 GMT

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

---

Thanks very much for your detailed response.

The first method of computing age-period specific mortality probabilities is perhaps what i want to apply, however i just didn't understand how it all adds up. In the end i want to be able to compute NMR, IMR,CMR & U5MR from birth history survey data.

In addition, when you use the age period specific mortality probabilities to compute which ever you want, how then do you compute the CIs and tests of significance if needed?

I'll read up the DHS manual further, wish DHS was conducting an analysis workshop where some of would endeavor to attend and ask all these questions, I believe though this forum is a step in the right direction

---

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [Reduced-For\(u\)m](#) on Mon, 22 Apr 2013 00:10:30 GMT

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

---

Hey there. Sorry it has taken so long to get back to you. I don't have Stata code to do the synthetic cohort method, but I think maybe we can work through a little bit.

First, though...when you ask about CIs and tests...what are you trying to compare? If you want a confidence interval around the mortality rate, we might be able to come up with something, but I haven't ever actually seen that done for this synthetic method. But I think maybe we can get something if we make some assumptions. Let's try this:

First, we need some ingredients, and since I don't have the code, I'll just list them and you can grab the info from the recode files. When you're all done, maybe you can post the code here so we have it. Anyway...

Birthdate (year and month)

Survey date (year and month) - make these into numbers using the time commands in stata: `gen bdate = ym(byear bmonth)`

Age at measurement (compute from birthdate and measure date, to get "ages" for dead children (ugh, I hate writing "dead children"))

Alive? (0/1 indicator)

Age at Death - make this something like 99 if the child is still living, computed maybe from `deathdate - birthdate`

X - any covariates you want to have around for cutting up the sample

sampling weight, strata, cluster (there is code on the DHS FAQs for this under "using data files")



<http://www.measuredhs.com/faq.cfm>)

\*Also, do the "svyset" thing, as described.

\*Note: I'm going to use "global" macros here to store some information, just know these will stay on your computer, and just in case, type "macro drop \_all" at the beginning of your do file...so long as you don't have any macros you are keeping around for good measure.

\*Now, a decision. I think you want to compute just 1 rate from the whole period, interpreted as a "current" mortality rate, yes? Well, then we have to decide whether, for instance, we want to use information for u1 mortality on children born a few years back. I'm going to assume we do. The "deep" assumption, here, is that child mortality hazards are not changing over the survey period. If we need to relax this, let me know.

Now, we are going to compute first the Prob(Death before age 1 month)

\*an indicator for child having died

```
gen u1mort = alive==0 & deathage==1
```

\*Now, we want to know, of the children who were recorded after they lived or would have lived to 1 month of age, what fraction died.

\*And we want to weight everything all right and whatnot

```
svy: reg u1mort if measureage>=1
```

```
global P1b = _b[u1mort]
```

```
global P1se = _se[u1mort]
```

\*this may or may not end up working, but it might be nice to keep this handy

\*Now we are going to do this for all the other ages in one loop

```
forvalues i=2/59 {
```

```
gen u`i'mort = deathage==`i'
```

```
reg u`i'mort if measureage>=`i'
```

```
global P`i'b = _b[u`i'mort]
```

```
global P`i'se = _se[u`i'mort]
```

```
}
```

\*Now we'll compute all the rates you want

\*Under 1 year mortality\*

```
gen u1rate = $P1 * $P2 * ... * $P11
```

\*That should basically allow you to compute any rate you want. Unless I made an obvious arithmetic error in here somewhere, which is possible.

Now, a few disclaimers. There are way more children in one survey who have measureage>1 than have measureage>=59. So your estimates for the lower ages are far more precise. You'll see that in your standard errors. But how to build a CI is a bit hard.



One thing you could do is ignore that and a few other things, and if you assume that all of these estimates are independent draws, you can use a standard formula:  $\text{var}(a*b)$  is described here: <http://www.stata.com/statalist/archive/2005-12/msg00183.html>

But I'm thinking that would get a little bit long. There is probably some better option that the DHS uses when it is computing its synthetic cohorts that I just don't know, that accounts for the fact that the same kids are in all these regressions/means and the sample size issues (although, I think they do this some other way, because maybe they just use each kid once...so that for P1 you are using just kids born 2 months ago, and for P37 you are just using kids born 38 months ago...or something like that).

Real quick, though, you could get a U1 mortality rate and SE (and thus CI) using a slightly different method. For example:

```
keep if measureage>=12
gen u1mort = deathage<12
svy: reg dead
global u1rate = _b[dead]
global u1se = _se[dead]
*and for CI at 95%
global ubound = u1rate+1.96*$u1se
global lbound = u1rate - 1.96*$u1se
```

This method gives an average u1 mortality rate for all children in the survey who were (or would have been) at least 1 year old when the survey happened....so an average rate over the previous 5 years, but excluding the survey year. It's not the synthetic cohort approach, but it gives a nice standard error/CI.

OK. That was basically some ideas more than a real code. I hope it was helpful - I'm not totally sure that it was. I'm happy to iterate on this. It would be nice to have some clean code to post here at the end so that other people can have a template handy. Let me know if I can help anymore. I'll try to get back to you faster.

Anyone else wants to play with this code, re-organize it, or point out why I'm doing something really stupid, that would be appreciated.

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [namsgfr](#) on Mon, 22 Apr 2013 10:10:04 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi,  
Thanks a bunch for these,  
I'll work through them and get back to you soon if I have any issues.  
I'm interested in comparing mortality rates at two different timepoints (4 years apart) to see if it has



significantly changed.

Also, i have lots of records with missing data i.e. didn't know age/age at death. Which imputation method is ideal here!

Thanks

Geoffrey

---

---

Subject: Re: Stata do file for U5 mortality analysis

Posted by [Reduced-For\(u\)m](#) on Mon, 22 Apr 2013 20:38:34 GMT

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

---

You are just gonna hate me for this:

You want to compare rates four years apart, but using the same survey? I'm actually a bit afraid that that won't work very well. I'm linking to a paper on child age misreporting, and I've been worried for a while in my work. The paper focuses on calculating total fertility rates, but the problem extends to mortality (since the denominator is number of children born). The gist is that there are a "surprising" number of children born about 5 years before a survey, and "surprisingly" few born 4 years before a survey. The author thinks it is so that enumerators can avoid doing some of the long child health/anthropometry sections. It is worse in some surveys than others, so you should check on your survey and see what it says (you could plot total births by year too and see if you get a funny jump).

Here is the paper. Check out Figures 5 and 8 for what can go wrong.

<http://paa2010.princeton.edu/papers/101547>

Can you use two different surveys? They are usually 4-5 years apart, and if the exact 4/5 years doesn't matter, you could use survey means to compare and make things easier.

Sometimes I feel like I just make everything worse. But I'm glad that the demo code was at least a little bit helpful.

I'll keep an eye out here for your response.

---

---

Subject: Re: Stata do file for U5 mortality analysis

Posted by [Reduced-For\(u\)m](#) on Mon, 22 Apr 2013 23:04:51 GMT

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

---

FIX: \$P1, \$P2, etc. are probabilities of dying at some age. To get the mortality rate from that, you need to compute

global P1live = 1-\$P1



Then, the Probability of living is  $1 - (\$P1_{live} * \$P2_{live} * \dots)$  up to  $P_a$  - the highest age under consideration.

What  $\$P1 * \$P2 * \dots$  calculates a fun, nonsensical probability about dying at multiple ages (missing the whole  $\text{Prob}(\text{dying at age 2} \mid \text{died at age 1})$  part...

I should just admit that I was never that good at probability calculations. My bad.

---

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [doc14163](#) on Thu, 27 Aug 2015 11:17:10 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thanks alot for the stata code illustration.

However i am facing a similar trouble.

Firstly i will like to confirm if the "measureage" variable is just the age of all the observations as at the survey date.

Secondly when i impliment the command "svy: reg u1mort...." and try to store the estimate in a global macro i.e `_b[u1mort]` i am told that

`[u1mort]` is not found. I dont know if we just forgot to include a dependent variable in the "svy: reg u1mort...." command

Thanks

---

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [Reduced-For\(u\)m](#) on Thu, 27 Aug 2015 19:11:50 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Yes - the "measure age" or "age at measurement" is just the age a child has (if it is alive) or would have had (if it died already) at the time of the survey.

As for the problem with "`_b[u1mort]`" - it should be there if you included it in your regression. I would look at the results of the regression in Stata directly and see if the `u1mort` is getting dropped from the regressions for some reason. If it is actually being estimated, then you probably have a typo in your code with the "`_b[]`" command. If it is being dropped, there is something wrong with either your identifying of death properly (you could just summarize your "child is dead" variable to check) or with your regression specification (co-linear dummy variables or something maybe?).

---

---

Subject: Re: Stata do file for U5 mortality analysis  
Posted by [doc14163](#) on Fri, 28 Aug 2015 00:00:52 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thanks a million for the clarification

---

---



Subject: Re: Stata do file for U5 mortality analysis  
Posted by [Janet-DHS](#) on Wed, 06 Mar 2024 16:59:08 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you for submitting your question. Could you please provide some more detailed information so we can better advise you?

Can you let us know:

- Which survey you are using (Include country name and year)?
- Which data files you are referring to?
- Which software you are using (Stata, SPSS, R, etc.)?

If you are trying to match a Table in a final report, please also indicate which table and which estimate you are trying to match.