MAIN ANALYSIS OF THRIO

```
/* Data Set Description for Main THRio Analysis--thriomaindata.dta (version 12 Stata)
*
This data set has multiple lines per patient; it was split up to handle the time-varying
intervention covariate.

*  numunid:   coded number of clinic, which was the unit of randomization
*  tb     :    0: patient censored on dtfim1 without TB;   1: TB diagnosed on dtfim1
*  interv :    0: control phase during (dtini1, dtfim1) interval; 1: intervention phase
*  dtini1 :   number of weeks since Sept 1, 2005, that the individual patient entered the risk set
*  dtfim1 :   number of weeks since Sept 1, 2005, that the individual patient exited the risk set
*  fakeid :   artificial variable tracking a patient's records


* Stata code to run the main model

use "c:\......\thriomaindata.dta"

stset dtfim1, f(tb) id(fakeid) enter(time dtini1)
 stcox interv, shared(numunid) forceshared


* RESULTS (using Stata v.13)

. stset dtfim1, f(tb) id(fakeid) enter(time dtini1)

                id:  fakeid
     failure event:  tb != 0 & tb < .
obs. time interval:  (dtfim1[_n-1], dtfim1]
 enter on or after:  time dtini1
 exit on or before:  failure

------------------------------------------------------------------------------
    20574  total observations
        0  exclusions
------------------------------------------------------------------------------
    20574  observations remaining, representing
    12816  subjects
      475  failures in single-failure-per-subject data
  2085407  total analysis time at risk and under observation
                                        at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =  208.5714

.  stcox interv, shared(numunid) forceshared
Warning: option shared() is used in the presence of delayed entries or gaps.  The
         results are consistent only under the assumption that the frailty
         distribution is independent of the covariates and the truncation points.
         This is a restrictive assumption, and you should evaluate if it is
         reasonable for your data before you proceed with estimation.

         failure _d:  tb
   analysis time _t:  dtfim1
  enter on or after:  time dtini1
                id:  fakeid

Fitting comparison Cox model:

Estimating frailty variance:

Iteration 0:   log profile likelihood = -4370.7617
Iteration 1:   log profile likelihood = -4370.0091
Iteration 2:   log profile likelihood = -4369.9984
Iteration 3:   log profile likelihood = -4369.9976
Iteration 4:   log profile likelihood = -4369.9976
```

```
Fitting final Cox model:

Iteration 0:   log likelihood = -4382.1117
Iteration 1:   log likelihood = -4370.0927
Iteration 2:   log likelihood = -4369.9976
Iteration 3:   log likelihood = -4369.9976
Refining estimates:
Iteration 0:   log likelihood = -4369.9976

Cox regression --
         Breslow method for ties                Number of obs      =     20574
         Gamma shared frailty                   Number of groups   =        29
Group variable: numunid

No. of subjects =        12816                  Obs per group: min =        80
No. of failures =          475                                 avg =  709.4483
Time at risk    =  2085406.857                                 max =      2273

                                                Wald chi2(1)       =      1.43
Log likelihood  =   -4369.9976                  Prob > chi2        =    0.2324

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      interv |   .8661999   .1041852    -1.19   0.232     .6842848    1.096477
-------------+----------------------------------------------------------------
       theta |   .0486331   .0311977
------------------------------------------------------------------------------
Likelihood-ratio test of theta=0: chibar2(01) =     6.18 Prob>=chibar2 = 0.006

Note: standard errors of hazard ratios are conditional on theta.
Warning: observations within subject belong to different frailty groups.
.
end of do-file
```

Note: results differ slightly from the 0.87 (0.69, 1.10) from the published paper (below); those results were from running the equivalent model in R.

Durovni, B., Saraceni, V., Moulton, L.H., Pacheco, A.G., Cavalcante, S.C., King, B.S., Cohn, S., Efron, A., Chaisson, R.E., Golub, J.E. Effect of Improved Tuberculosis Screening and Isoniazid Preventive Therapy on Incidence of Tuberculosis and Death in Patients with HIV in Clinics in Rio de Janeiro, Brazil: a Stepped Wedge, Cluster-randomised Trial. Lancet Infectious Diseases, 2013. Oct;13(10):852-8.

*/

First lines of data:

	numunid	tb	interv	dtini1	dtfim1	fakeid
1	19	0	0	0	21.14285714	1895
2	19	0	1	21.14285714	208.5714286	1895
3	19	0	0	0	9.142857143	1897
4	19	0	1	9.142857143	208.5714286	1897
5	7	0	0	0	71.57142857	1899
6	7	0	1	71.57142857	208.5714286	1899
7	19	0	0	0	13.14285714	1901
8	19	0	1	13.14285714	208.5714286	1901
9	19	0	0	0	208.5714286	1903
10	19	0	0	0	2.857142857	1905
11	19	0	1	2.857142857	208.5714286	1905
12	19	0	0	0	32.28571429	1909
13	19	0	1	32.28571429	208.5714286	1909