Modelling multiple timescales using flexible parametric survival models

2023 Stata Biostatistics and Epidemiology Virtual Symposium

Hannah Bower

Karolinska Institutet (hannah.bower@ki.se)
Red Door Analytics (hannah.bower@reddooranalytics.se)

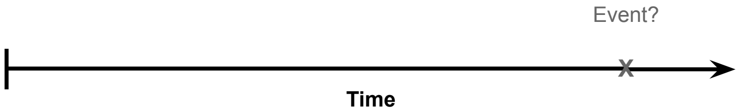
23rd February 2023

About me

- I currently work at Karolinska Institutet (70%), and Red Door Analytics (30%)
- Background in maths and biostatistics
- Defended my thesis at Karolinska Institutet in March 2018
  - *Flexible parametric models for cancer patient survival: loss in expectation of life and further developments*

Timescales

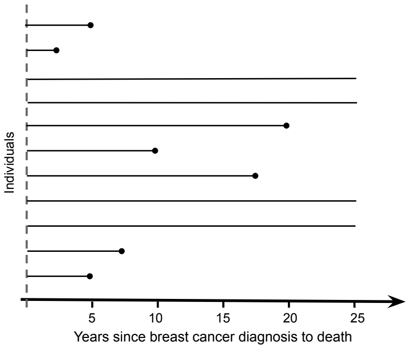
- Defining the timescale(s) of interest is essential in any time-to-event analysis
- Two main components of survival analysis are the event/outcome of interest and time(scale)
- The timescale is defined by the start and stop of follow-up time, and a time origin

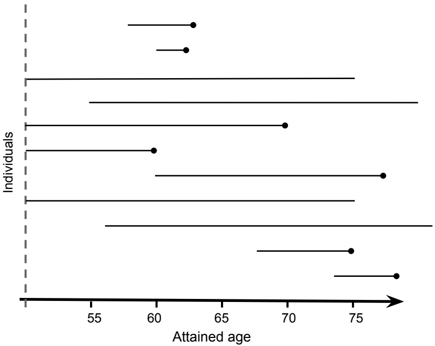


Timescales

- Different timescales could be important for different outcomes
  - Time since a cancer diagnosis to death
  - Attained age for the incidence disease
- Several timescales may also be simultaneously of interest
  - Incidence of breast cancer (attained age, time since childbirth)
  - Mortality rates in the population (calendar year, attained age)
  - Risk of infection after admittance to intensive care unit (time since admittance, calendar time)

Timescales


Individuals
5
10
15
20
25
Years since breast cancer diagnosis to death


Individuals
55
60
65
70
75
Attained age

How to model with multiple timescales?

- Common to model multiple timescales by splitting one one or more timescales[1, 2].
- For example, if we were to use a piece-wise constant exponential (Poisson) model:
  - Split your dataset up according to timescale 1
  - Split your dataset up according to timescale 2
  - Fit a Poisson model to this stacked dataset with
    - Categories for the timescales
    - Could use some smoothed function (e.g. spline)
- This can be computationally intensive.

Example of split data

```
. list id hormon _t0 _t _d in 1/10
```

	id	hormon	_t0	_t	_d
1.	1	0	0	4.9665973	1
2.	2	1	0	5.5251342	1
3.	3	1	0	1.9494031	1
4.	4	1	0	4.9474318	1
5.	5	0	0	2.1136787	1
6.	6	0	0	1.2265907	1
7.	7	1	0	5.9467747	0
8.	8	0	0	5.9166575	0
9.	9	0	0	1.289563	1
10.	10	0	0	5.5141825	0

```
. list id hormon fu _t0 _t _d in 1/10
```

	id	hormon	fu	_t0	_t	_d
1.	1	0	0	0	2	0
2.	1	0	2	2	4.9665973	1
3.	2	1	0	0	2	0
4.	2	1	2	2	5	0
5.	2	1	5	5	5.5251342	1
6.	3	1	0	0	1.9494031	1
7.	4	1	0	0	2	0
8.	4	1	2	2	4.9474318	1
9.	5	0	0	0	2	0
10.	5	0	2	2	2.1136787	1

Flexible parametric models on the hazard scale

- Flexible parametric survival model (FPM) use restricted cubic splines to model the baseline function
- Here the restricted cubic splines model the baseline log hazard function $\ln h(t)$
- With one timescale:

$$\ln(h(t|\boldsymbol{x})) = s(f(t)|\gamma_0) + \boldsymbol{x}'\beta$$

where $s(f(t); \gamma_0)$ represents the spline function, $\boldsymbol{x}$ are covariates, t is time.

Flexible parametric models on the hazard scale

- Note that it is more common to fit FPM on the log cumulative hazard scale
- To maximise the likelihood when modelling on the log hazard scale we have to numerically integrate
- FPMs are flexible, easy to include time-dependent effects, software allows for nice predictions [3].

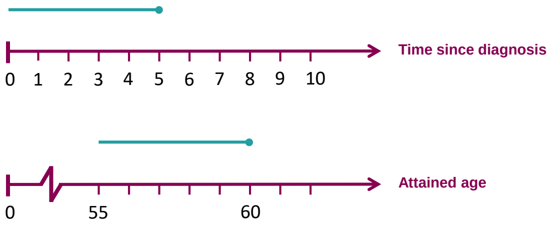
Multiple timescales, again

To incorporate multiple timescales we utilise the fact that timescales increase with the same unit [4, 5, 6]

- One timescale is a function of the other
- Have to consider where the origin of each timescale is

Multiple timescales, again

To illustrate, say an individual is diagnosed with a disease at age 55 and has follow-up for 5 years:



Multiple timescales, again

Then we can extend this idea to the following:

- t_{diag} = time since diagnosis of a disease
- t_{age} = attained age
- a = age at diagnosis (constant offset)

Then, we can write

$$t_{age} = t_{diag} + a$$

Flexible parametric model with two timescales becomes:

$$\ln(h(t_{diag}|a, \boldsymbol{x})) = \underbrace{s_1(f(t_{diag})|\gamma_1)}_{\text{Time since diagnosis}} + \underbrace{s_2(f(t_{diag} + a)|\gamma_2)}_{\text{Attained age}} + \boldsymbol{x}'\beta$$

Why use multiple timescales?

- In most situations, using one primary timescale should be OK
  - Survival estimates from models with one timescale may be biased when the hazard rate of the event of interest is actually a function of two timescales [7]
- Part of the research question
- If it is of interest, there is now a user-friendly Stata command `stmt` [8]
  - Note that these models can also be fitted using `stmerlin`, a more generalised command (survival analysis using `merlin` [9])

Modelling multiple timescales in Stata using stmt

- `stmt` is a Stata command which fits multiple timescales using FPMs on the log hazard scale [8]
- First timescale is specified using the `stset` command
- Second is specified in `stmt` options

Title

stmt — Modelling multiple timescales using flexible parametric survival models on the log hazard scale

Syntax

stmt [*varlist*] [*if*] [*in*] [, time1(*sub-options*) time2(*sub-options*) time3(*sub-options*) *options*]

Modelling multiple timescales in Stata using `stmt`

- Covariates, and interactions with the timescale can be specified
- Numerical integration performed via Gauss-Legendre quadrature
- Requires `rcsgen` and `stpm2` to be installed
- Analytic derivatives for the score and Hessian are included to increase speed and accuracy when maximising the likelihood

`stmt` syntax

`stmt` syntax

`stmt` *varlist* [*if*] [*in*], `time1(`*`sub-options`*`)`
[`time2(`*`sub-options`*`)` `time3(`*`sub-options`*`)` `timeint(`*`int_list`*`)`
`timeintknots(`*`int_list`*`)` `timeintbknots(`*`int_list`*`)`
noconstant nodes(#) noorthog nohr verbose from(*matrix*) inith(*varname*)
maximise_options]

`stmt` syntax

Time-scale specific sub-options include:

- The offset between the two timescales (`start()` option)
- Number and position of knots of restricted cubic splines
- Interactions with the timescale, i.e. HRs which change over the timescale
- Subgroup-specific timescale

stmt postestimation command

It is also possible to predict from the fitted model using the `predict` command

- The linear predictor and hazard function
- Can specify which values of each timescale to predict over

predict syntax

predict *newvar* [*if*] [*in*], { <u>h</u>azard |xb}
time1var(*varname*) time2var(*varname*) time3var(*varname*)
[at(*varname* # [*varname* # ...]) ci nodes per zeros level(#)]

Example

- 2982 females diagnosed with breast cancer
- Our outcome of interest is death (due to any cause)
- We follow patients from primary surgery
- Grade of cancer is our exposure of interest (2 or 3)
- Timescales of interest
  1. Time since primary surgery
  2. Attained age

stset

`stset` with time since surgery as timescale 1:

```
 . stset survtime, f(dead==1) scale(12)
Survival-time data settings
         Failure event: dead==1
Observed time interval: (0, survtime]
     Exit on or before: failure
     Time for analysis: time/12
--------------------------------------------------------------------------
      2,982  total observations
          0  exclusions
--------------------------------------------------------------------------
      2,982  observations remaining, representing
      1,272  failures in single-record/single-failure data
 21,270.702  total analysis time at risk and under observation
                                                At risk from t =         0
                                     Earliest observed entry t =         0
                                          Last observed exit t =  19.28268
```

stmt with one timescale

```
. stmt grade, time1(df(5)) nolog
Log likelihood = -3023.3924                            Number of obs = 2,982
```

	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
xb						
grade	1.659792	.1152621	7.30	0.000	1.448582	1.901798
rcs						
__t1_s1	.1130917	.0309638	3.65	0.000	.0524038	.1737795
__t1_s2	.1179876	.0291052	4.05	0.000	.0609425	.1750326
__t1_s3	-.1213544	.0299503	-4.05	0.000	-.1800559	-.062653
__t1_s4	-.0914425	.0299644	-3.05	0.002	-.1501716	-.0327134
__t1_s5	-.026898	.0318642	-0.84	0.399	-.0893507	.0355546
_cons	-4.12488	.1960625	-21.04	0.000	-4.509155	-3.740604

```
Note: Estimates are transformed only in the first equation to hazard ratios.
 Quadrature method: Gauss-Legendre with 30 nodes
```

stmt with two timescale

We use the `time2()` option with the `start()` sub-option to specify our second timescale, attained age

```
. stmt grade, time1(df(5)) time2(df(3) start(agesurgery)) nolog
Log likelihood = -2956.7515                              Number of obs = 2,982
```

	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
xb						
grade	1.609036	.1118154	6.84	0.000	1.404151	1.843816
rcs						
__t1_s1	.078144	.0316819	2.47	0.014	.0160487	.1402394
__t1_s2	.1307119	.0291879	4.48	0.000	.0735047	.187919
__t1_s3	-.1186399	.0300058	-3.95	0.000	-.1774502	-.0598297
__t1_s4	-.0946102	.0300088	-3.15	0.002	-.1534263	-.0357941
__t1_s5	-.030988	.0318674	-0.97	0.331	-.0934469	.0314709
__t2_s1	.2271669	.0260366	8.72	0.000	.1761361	.2781977
__t2_s2	-.2377882	.0251054	-9.47	0.000	-.2869938	-.1885825
__t2_s3	-.0887876	.0266178	-3.34	0.001	-.1409575	-.0366178
_cons	-4.048158	.1961963	-20.63	0.000	-4.432695	-3.66362

```
Note: Estimates are transformed only in the first equation to hazard ratios.
 Quadrature method: Gauss-Legendre with 30 nodes
```

stmt with two timescales

We can add an interaction using the `tvc()` and `dftvc()` options:

```
. stmt grade, time1(df(5) tvc(grade) dftvc(2)) time2(df(3) start(agesurgery)) nolog
Log likelihood = -2953.3856                                Number of obs = 2,982
```

	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
xb						
grade	1.514219	.1233225	5.09	0.000	1.290816	1.776287
rcs						
__t1_s1	.5111851	.2011859	2.54	0.011	.116868	.9055022
__t1_s2	.3295628	.2191507	1.50	0.133	-.0999648	.7590903
__t1_s3	-.098055	.0353463	-2.77	0.006	-.1673326	-.0287775
__t1_s4	-.091684	.0300071	-3.06	0.002	-.1504969	-.0328711
__t1_s5	-.0309256	.0319089	-0.97	0.332	-.093466	.0316147
__t2_s1	.2274068	.0260295	8.74	0.000	.1763899	.2784236
__t2_s2	-.2386316	.0251059	-9.51	0.000	-.2878382	-.189425
__t2_s3	-.0892667	.0266396	-3.35	0.001	-.1414793	-.0370541
__t1_s_grade1	-.1539901	.0712874	-2.16	0.031	-.2937108	-.0142695
__t1_s_grade2	-.068185	.0770798	-0.88	0.376	-.2192586	.0828886
_cons	-3.884596	.2262782	-17.17	0.000	-4.328093	-3.441099

```
Note: Estimates are transformed only in the first equation to hazard ratios.
 Quadrature method: Gauss-Legendre with 30 nodes
```

Predictions from stmt

We define our own timescale variables

- `time1` (time since surgery, 0-15 years)
- `time2` (attained age, 40 to 70 years)

and then predict the hazard using the `predict` command:

```
. predict h2, h time1var(time1) time2var(time2) at(grade 3) ci
```

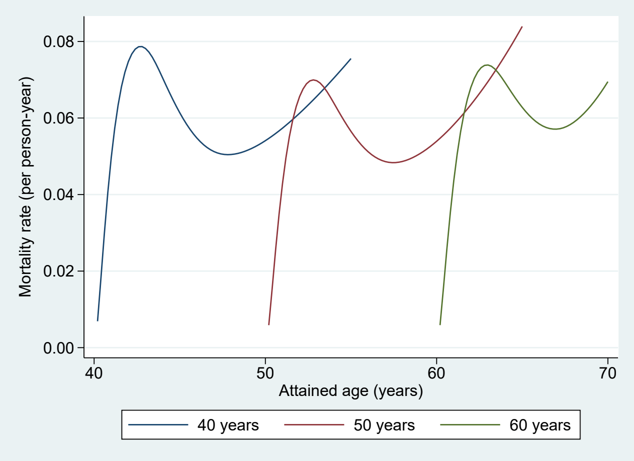
Predictions from stmt

```
. list time1 time2 ageatsurgery h2 in 1/10
```

	time1	time2	ageats~y	h2
1.	.2	40	39.8	.00693705
2.	.2	40.1	39.9	.00691959
3.	.2	40.2	40	.00690231
4.	.2	40.3	40.1	.0068852
5.	.2	40.4	40.2	.00686827
6.	.2	40.5	40.3	.0068515
7.	.2	40.6	40.4	.00683491
8.	.2	40.7	40.5	.00681849
9.	.2	40.8	40.6	.00680223
10.	.2	40.9	40.7	.00678614

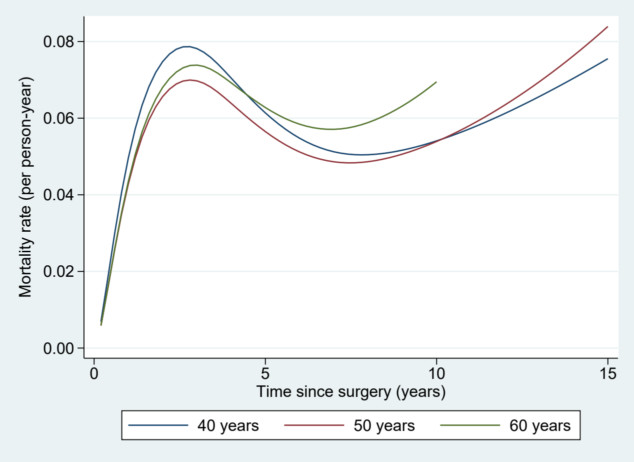
Different plots from these predictions

Fixing age at surgery, we can plot the mortality rate across attained age:



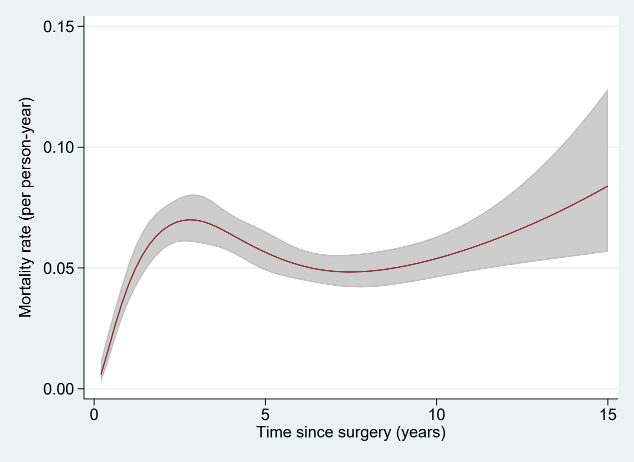
Different plots from these predictions

Fixing age at surgery, we can plot the mortality rate across time since surgery:



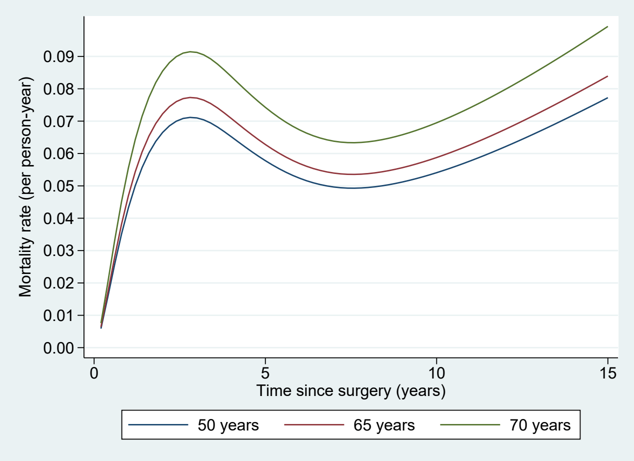
Different plots from these predictions

Can also show confidence intervals (here for mortality rate for 50 year old at surgery):



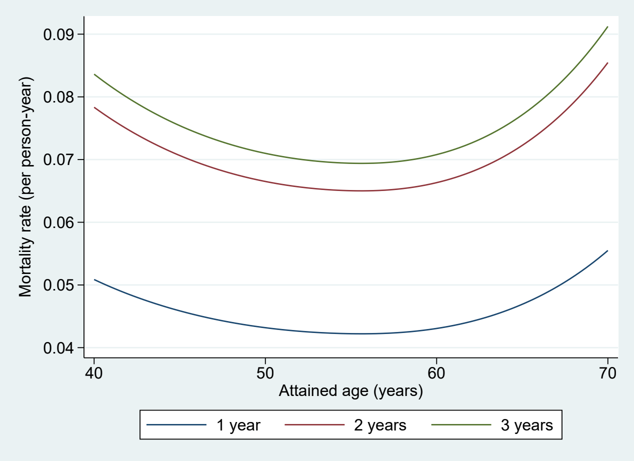
Different plots from these predictions

Now we fix the attained age timescale:



Different plots from these predictions

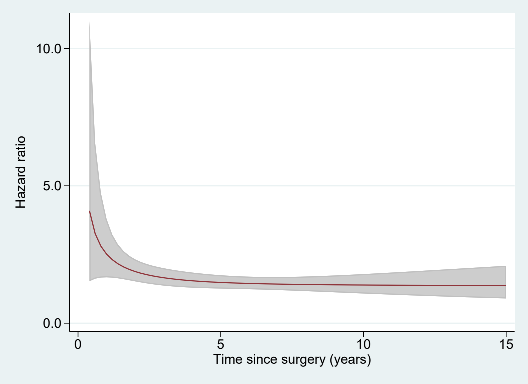
Or we can fix the time since surgery timescale:



Hazard ratio using predictnl

```
 . predictnl lnhr= ///
> ln(predict(h time1var(time1) time2var(time2) at(grade 3))) ///
>- ln(predict(h time1var(time1) time2var(time2) at(grade 2))) ///
>, ci(lnhr_lci lnhr_uci)

. gen hr=exp(lnhr)
. gen hr_lci=exp(lnhr_lci)
. gen hr_uci=exp(lnhr_uci)
```



Other information

- See our Stata Journal paper [8], and paper by Batyrbekova et al. [7]

The Stata Journal (2022)
22, Number 3, pp. 679–701 DOI: 10.1177/1536867X221124552

Flexible parametric survival analysis with multiple timescales: Estimation and implementation using stmt

Hannah Bower
Clinical Epidemiology Division
Department of Medicine Solna, Karolinska Institutet
Stockholm, Sweden
hannah.bower@ki.se

Therese M.-L. Andersson
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden
therese.m-l.andersson@ki.se

Michael J. Crowther
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden
michael.crowther@ki.se

Paul C. Lambert
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden
and
Biostatistics Research Group
Department of Health Sciences, University of Leicester
Leicester, U.K.
paul.lambert@leicester.ac.uk

Other information

- Red Door Analytics has plans for a course, see `https://reddooranalytics.se/services/training/` for updates
- Others have been working with extending this work:
  - `stmerlin` allows for other models with multiple timescales
  - Extensions to multi-state models to include multiple timescales [10]

Other information

```
 //cox
stmerlin trt,      dist(cox)                                      ///
                 time2(df(2) offset(agec) time noorthog)          ///
                 time3(df(2) offset(yearc) time noorthog)


// royston parmer (FPM on log cumulative hazard)
stmerlin trt,      dist(rp) df(3)  noorthog                       ///
                 time2(df(2) offset(agec) time noorthog)          ///
                 time3(df(2) offset(yearc) time noorthog)

// stmerlin equivalent to stmt
stmerlin trt,      dist(rcs) df(3) noorthog                       ///
                 time2(df(2) offset(agec) time  noorthog)         ///
                time3(df(2) offset(yearc) time noorthog)

//stmt
stmt trt,   time1(df(3)) noorthog                                 ///
         time2(df(2) start(agec) logtoff)                         ///
         time3(df(2) start(yearc) logtoff)                        ///
         nohr


 . timer list
   1:     99.07 /        1 =      99.0690
   2:      0.32 /        1 =       0.3150
   3:      7.07 /        1 =       7.0720
   4:      2.30 /        1 =       2.2970
```

Acknowledgements

Many thanks to my co-authors and colleagues for their work on this:

- Michael J Crowther, Red Door Analytics
- Nurgul Batyrbekova, Karolinska Institutet
- Paul C Lambert, Karolinska Institutet, University of Leicester
- Therese Andersson, Karolinska Institutet

References I

[1] B. Carstensen.
Demography and epidemiology: Practical use of the Lexis diagram in the computer age or: Who needs the Cox model anyway?
Technical report, Department of Biostatistics, University of Copenhagen, 2006.

[2] P. Royston and P. C. Lambert.
Flexible parametric survival analysis in Stata: Beyond the Cox model.
Stata Press, 2011.

[3] P. Royston and M. K. B. Parmar.
Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.
Statistics in Medicine, 21(15):2175–2197, Aug 2002.

[4] Bradley Efron.
The two-way proportional hazards model.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4):899–909, 2002.

[5] D. Danardono.
Multiple Time Scales and Longitudinal Measurements in Event History Analysis.
PhD thesis, UmeåUniversity, 2005.

[6] S. Iacobelli and B. Carstensen.
Multiple time scales in multi-state models.
Statistics in Medicine, 32(30):5315–5327, Dec 2013.

References II

[7] N. Batyrbekova, H. Bower, Dickman P.W., R. Szulkin, P.C. Lambert, and T.M-L. Andersson.
Potential bias introduced by not including multiple time-scales in survival analysis: a simulation study.
Communications in Statistics - Simulation and Computation, 0(0):1–14, 2022.

[8] H. Bower, T.M-L. Andersson, M. J. Crowther, and Lambert P.C.
Flexible parametric survival analysis with multiple timescales: Estimation and implementation using stmt.
Stata Journal, 22(2):679–701, 2022.

[9] M.J. Crowther.
merlin—a unified modeling framework for data analysis and methods development in stata.
The Stata Journal, 20(4):763–784, 2020.

[10] Caroline E. Weibull, Paul C. Lambert, Sandra Eloranta, Therese M. L. Andersson, Paul W. Dickman, and Michael J. Crowther.
A multistate model incorporating estimation of excess hazards and multiple time scales.
Statistics in Medicine, 40(9):2139–2154, 2021.