

Multiple timescales in rate models using Poisson and Cox regression

SDCC

<http://bendixcarstensen.com/>

July 2019

1

Compiled Friday 12th July, 2019, 10:55
from: /home/bendix/sdc/extn/IARC/MultiMorb/r/VVex.tex

Bendix Carstensen Steno Diabetes Center Copenhagen, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bcar0029@regionh.dk b@bxc.dk
<http://BendixCarstensen.com>

Contents

1	Mortality and insulin use in DM patients	1
1.1	Splitting follow-up in intervals	2
2	Main-effects model for mortality	3
2.1	Poisson	3
2.2	Cox	4
3	Marginal effect of time since insulin	4
4	Joint effects of age and duration	5
5	Age\timesduration interaction	8
5.1	Age at insulin start	8
5.2	General interaction	9
6	Estimation and reporting of multiple timescales — practical hints and advice	13
6.1	Context	13
6.2	Methods	13
6.3	Results	13
6.4	Conclusion	13

1 Mortality and insulin use in DM patients

We use the example dataset of mortality from the `Epi` package, and set up the follow-up in a `Lexis` object.

```
> library( Epi )
> library( popEpi )
> library( survival )
> print( sessionInfo(), l=F )
R version 3.6.0 (2019-04-26)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.6 LTS

Matrix products: default
BLAS: /usr/lib/openblas-base/libopenblas.so.0
LAPACK: /usr/lib/lapack/liblapack.so.3.0

attached base packages:
[1] utils      datasets  graphics  grDevices  stats      methods   base

other attached packages:
[1] survival_2.44-1.1 popEpi_0.4.4      Epi_2.37

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.0      lattice_0.20-38  zoo_1.8-4        MASS_7.3-51.1
 [5] grid_3.6.0      plyr_1.8.4       nlme_3.1-139     etm_1.0.4
 [9] data.table_1.12.0 Matrix_1.2-17    splines_3.6.0    tools_3.6.0
[13] cmprsk_2.2-7    numDeriv_2016.8-1 parallel_3.6.0    compiler_3.6.0
[17] mgcv_1.8-28

> data( DMLate )
> dml <- Lexis( entry=list(Per=dodm, Age=dodm-dobth, DMdur=0 ),
+              exit=list(Per=dox),
+              exit.status=factor(!is.na(dodth),labels=c("DM","Dead")),
+              data=DMLate )

NOTE: entry.status has been set to "DM" for all.
NOTE: Dropping 4 rows with duration of follow up < tol
```

The 4 patients dropped are persons dead on the day of diagnosis.

We cut the follow-up at insulin start, and introduce a new timescale, but do not split the (non-precursor) state `Dead`.

```
> system.time(
+ dmi <- cutLexis( dml, cut = dml$doins,
+                 pre = "DM",
+                 new.state = "Ins",
+                 new.scale = "tsI",
+                 split.states = FALSE ) )
  user system elapsed
 0.685  0.012  0.697

> summary( dmi, t=T )

Transitions:
  To
From  DM  Ins  Dead  Records:  Events: Risk time:  Persons:
DM   6157 1694 2048     9899     3742  45885.49     9899
Ins    0 1340  451     1791      451   8387.77     1791
Sum  6157 3034 2499    11690     4193  54273.27     9996

Timescales:
  Per  Age  DMdur  tsI
  ""   ""   ""   "Ins"
```

We see that the time scale `tsI` is defined as time since entry into the state `Ins`. This implies that the value of `tsI` is `NA` for the follow-up before insulin start.

Also we make a graphical overview of the data and transitions:

```
> boxes( dmi, boxpos=TRUE, scale.R=1000, hmult=1.5 )
```

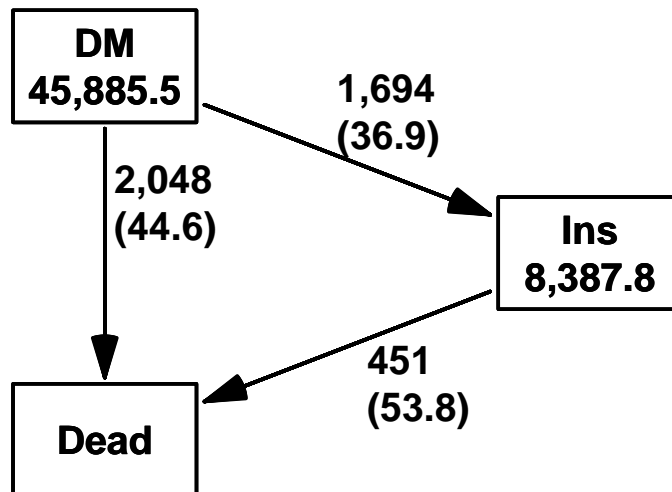


Figure 1: Overview of the states and transitions.

1.1 Splitting follow-up in intervals

We can split the FU in smaller intervals of say, 2 months (1/6 year) using either `splitLexis` (from `Epi`) or `splitMulti` (from `popEpi`), the latter slightly faster and more convenient:

```
> system.time( dma <- splitLexis( dmi, breaks=seq(0,100,1/6), time.scale='Age' ) )
```

```
  user system elapsed
 3.219  0.532  3.751
```

```
> system.time( dms <- splitMulti( dmi, Age=seq(0,100,1/6) ) )
```

```
  user system elapsed
 2.441  1.241  3.257
```

```
> summary( dma )
```

Transitions:

From	To	DM	Ins	Dead	Records:	Events:	Risk time:	Persons:
DM	DM	281394	1694	2048	285136	3742	45885.49	9899
Ins	Ins	0	51647	451	52098	451	8387.77	1791
Sum		281394	53341	2499	337234	4193	54273.27	9996

```
> summary( dms )
```

Transitions:

From	To	DM	Ins	Dead	Records:	Events:	Risk time:	Persons:
DM	DM	281392	1694	2045	285131	3739	45879.34	9898
Ins	Ins	0	51647	450	52097	450	8386.80	1791
Sum		281392	53341	2495	337228	4189	54266.14	9995

Note that the latter—the result from `splitMulti`—loses a bit of the follow-up; this is because `splitMulti` by default excludes follow-up (risk time and events) outside the range of the specified breaks, so follow-up beyond 100 years of age is excluded.

2 Main-effects model for mortality

We now model mortality by age and time since insulin, as main effects, additive on the log-scale, multiplicative on the rate-scale. Hence, this is often termed a proportional hazards model.

We first set `tsI` to 0 for the Alive state, and then define knots for the modeling:

```
> dms$tsI <- ifelse( is.na(dms$tsI), 0, dms$tsI )
> nk <- 5
> ( a.kn <- c(40, with( subset(dms,lex.Xst=="Dead"),
+                   quantile(Age+lex.dur,(1:nk-0.5)/nk) ) ) )
      10%      30%      50%      70%      90%
40.00000 60.27817 71.30732 77.66461 82.69897 89.80479
> ( i.kn <- c( 0, with( subset(dms,lex.Xst=="Dead" & lex.Cst=="Ins"),
+                   quantile(tsI+lex.dur,1:(nk-1)/nk) ) ) )
      20%      40%      60%      80%
0.0000000 0.3088296 1.1241615 2.5472964 4.9171800
```

2.1 Poisson

A Poisson model approach to mortality by insulin status using the `poisreg` family from the `Epi` package:

```
> pm <- glm( cbind(lex.Xst=="Dead",lex.dur) ~ Ns(Age,knots=a.kn)
+                   + Ns(tsI,knots=i.kn)
+                   + lex.Cst + sex,
+                   family=poisreg, data = dms )
```

There is a wrapper for this exploiting the structure of `Lexis` objects; there are also `from` and `to` arguments, but the default for these is all (de facto) absorbing states as `to` and all immediately preceding states as `from` states, but you will be notified. Although it is called `glm.Lexis` it is *not* as S3 method for `Lexis` objects.

```
> Pm <- glm.Lexis( dms, form = ~ Ns(Age,knots=a.kn)
+                   + Ns(tsI,knots=i.kn)
+                   + lex.Cst + sex )
stats::glm Poisson analysis of Lexis object dms with log link:
Rates for transitions: DM->Dead, Ins->Dead
> c( deviance(Pm), deviance(pm) )
[1] 26794.99 26794.99
> identical( model.matrix(Pm), model.matrix(pm) )
[1] TRUE
```

2.2 Cox

Note that in the Cox-model the age is used as response variable, slightly counter-intuitive. Hence the age part of the linear predictors is not in the model:

```
> cm <- coxph( Surv(Age, Age+lex.dur, lex.Xst=="Dead") ~
+             Ns(tsI, knots=i.kn) + factor(lex.Cst) + sex,
+             data = dms )
```

There is also a Lexis wrapper for Cox models:

```
> Cm <- coxph.Lexis( dms, form= Age ~ Ns(tsI, knots=i.kn) + factor(lex.Cst) + sex )
```

```
model survival::coxph analysis of Lexis object dms:
Rates for transitions DM->Dead, Ins->Dead
```

```
> ci.exp( cm ) / ci.exp( Cm )
```

	exp(Est.)	2.5%	97.5%
Ns(tsI, knots = i.kn)1	1	1	1
Ns(tsI, knots = i.kn)2	1	1	1
Ns(tsI, knots = i.kn)3	1	1	1
Ns(tsI, knots = i.kn)4	1	1	1
factor(lex.Cst)Ins	1	1	1
sexF	1	1	1

```
> round( cbind( ci.exp( Cm, subset=c("tsI", "Ins", "F") ),
+             ci.exp( Cm, subset=c("tsI", "Ins", "F") ) ), 3 )
```

	exp(Est.)	2.5%	97.5%	exp(Est.)	2.5%	97.5%
Ns(tsI, knots = i.kn)1	0.277	0.181	0.426	0.281	0.183	0.432
Ns(tsI, knots = i.kn)2	0.362	0.271	0.483	0.365	0.273	0.488
Ns(tsI, knots = i.kn)3	0.104	0.052	0.210	0.103	0.051	0.210
Ns(tsI, knots = i.kn)4	0.420	0.333	0.529	0.425	0.336	0.536
lex.CstIns	6.189	4.803	7.976	6.187	4.791	7.990
sexF	0.674	0.622	0.731	0.675	0.623	0.731

Thus we see that the Poisson and Cox gives pretty much the same results for the regression parameters. You may argue that Cox requires a smaller dataset, because there is no need to subdivide data in small intervals *before* insulin use. But certainly the time *after* insulin inception need to be if the effect of time since insulin should be modeled properly as here.

3 Marginal effect of time since insulin

When we plot the marginal effect of `tsI` from the two models we get almost exactly the same:

```
> nd <- data.frame( tsI=seq(0,15,,151), lex.Cst="Ins", sex="M" )
> nr <- data.frame( tsI= 2 , lex.Cst="Ins", sex="M" )
> ppr <- ci.exp( pm, list(nd,nr), xvars="Age" )
> cpr <- ci.exp( cm, list(nd,nr) )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( nd$tsI, cbind(ppr,cpr), lty=c(1,2), log="y", plot=T,
+           ylab="Mortality RR", xlab="Time since insulin" )
> abline( v=2, h=1, lty=3 )
```

So the duration effect is the same from the two modeling approaches.

We will also want the RR relative to the non-insulin users — recall these are coded 0 on the `tsI` variable:

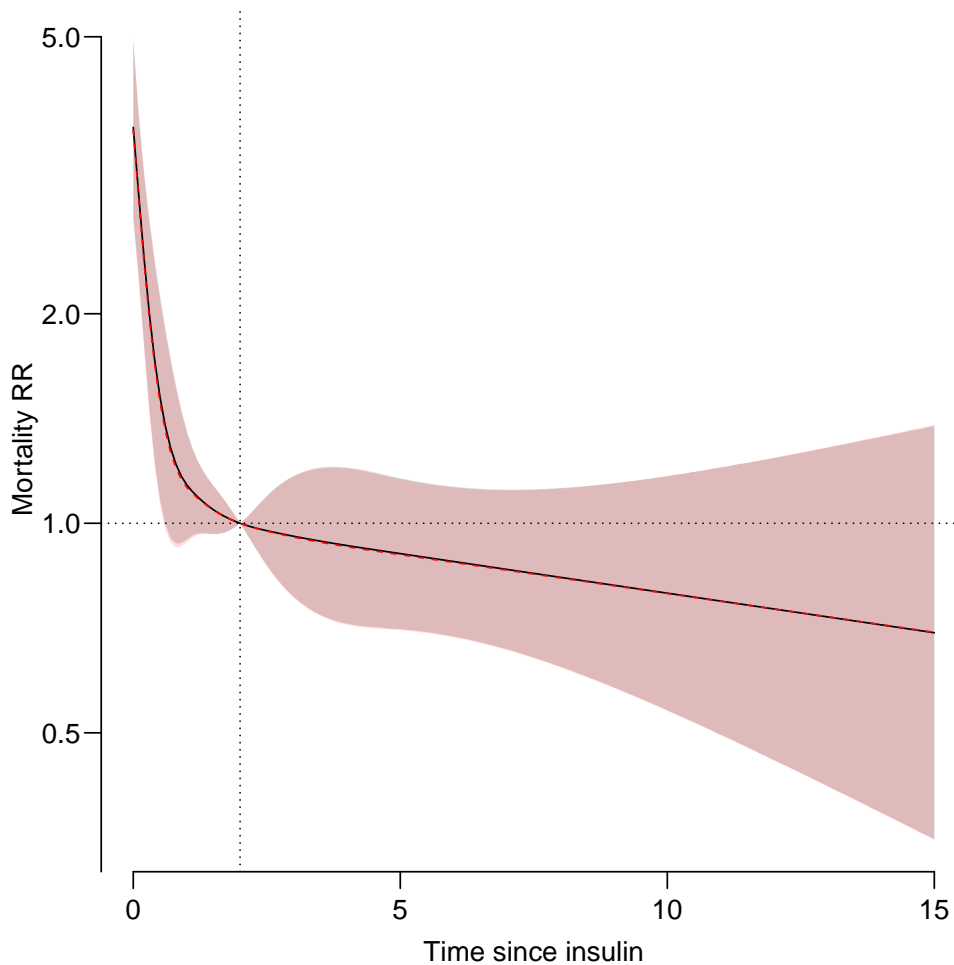


Figure 2: The naked duration effects relative to 2 years of duration, black from Poisson model, red from Cox model. The two sets of estimates are identical.

```
> nd <- data.frame( tsI=seq(0,15,,151), lex.Cst="Ins", sex="M" )
> nr <- data.frame( tsI= 0 , lex.Cst="DM" , sex="M" )
> ppr <- ci.exp( pm, list(nd,nr), xvars="Age" )
> cpr <- ci.exp( cm, list(nd,nr) )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( nd$tsI, cbind(ppr,cpr), lty=c(1,2), log="y", plot=T,
+          ylab="Mortality RR, Ins vs. noIns", xlab="Time since insulin" )
```

This is the effect of increasing duration of insulin use for a fixed age which is a bit artificial, so we would like to see the *joint* effects of age and insulin duration. What we cannot see is how the duration affects mortality *relative* to current age, so in the absence of a baseline mortality we cannot see how mortality changes by age in absolute terms.

4 Joint effects of age and duration

This is however only readily available from the Poisson-model; the age-effect is profiled out of the Cox-model so not available from that without serious complications. But from the Poisson model we can easily get the *joint* effects of age and insulin duration.

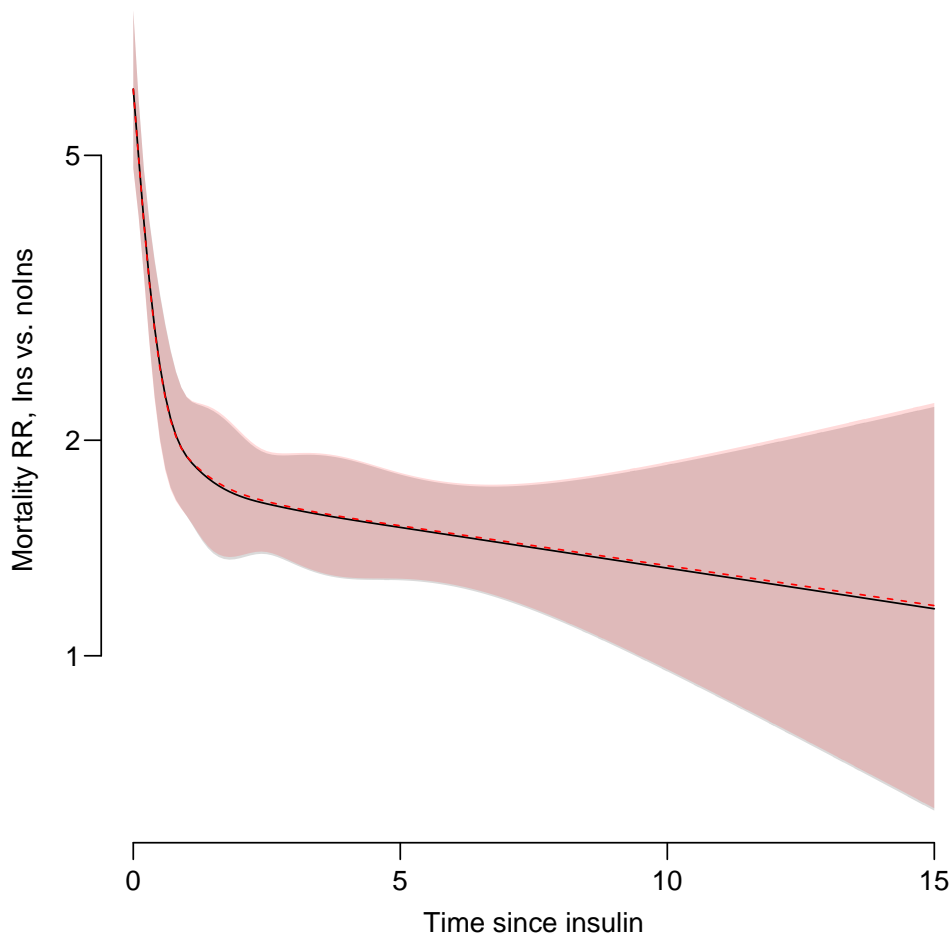


Figure 3: *Insulin duration effect (state Ins) relative to no insulin (state DM), black from Poisson model, red from Cox model. The shape is the same as the previous figure, but the RR is now relative to non-insulin, instead of relative to insulin users at 2 years duration.*

Note here that we define a data frame by age at insulin, `ain` (which not in the model) and time since insulin `tsI` (a variable in the model). From these two we compute the *current* age, `Age` (a variable in the model). Note that in order to get predictions of rates we need to specify all variables in the model, including sex:

```
> ndI <- transform( data.frame( expand.grid(tsI=c(NA,seq(0,15,0.1)),
+                                     ain=seq(40,70,5)),
+                               sex="M",
+                               lex.Cst="Ins"),
+                   Age = ain + tsI )
> ndD <- data.frame( tsI=0,
+                   Age=seq(40,85,0.5),
+                   sex="M",
+                   lex.Cst="DM" )
> head( ndI )
  tsI ain sex lex.Cst Age
1  NA  40  M     Ins   NA
2  0.0 40  M     Ins  40.0
3  0.1 40  M     Ins  40.1
4  0.2 40  M     Ins  40.2
5  0.3 40  M     Ins  40.3
```



```

6 0.4 40 M Ins 40.4
> head( ndD )
  tsI Age sex lex.Cst
1  0 40.0 M    DM
2  0 40.5 M    DM
3  0 41.0 M    DM
4  0 41.5 M    DM
5  0 42.0 M    DM
6  0 42.5 M    DM
> pprI <- ci.pred( Pm, ndI )
> pprD <- ci.pred( Pm, ndD )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( ndI$Age, pprI*1000, lty=1, log="y", plot=TRUE,
+           ylim=c(2,500), ylab="Mortality per 1000 PY", xlab="Age" )
> matshade( ndD$Age, pprD*1000, lty=1, col="red", plot=FALSE )

```

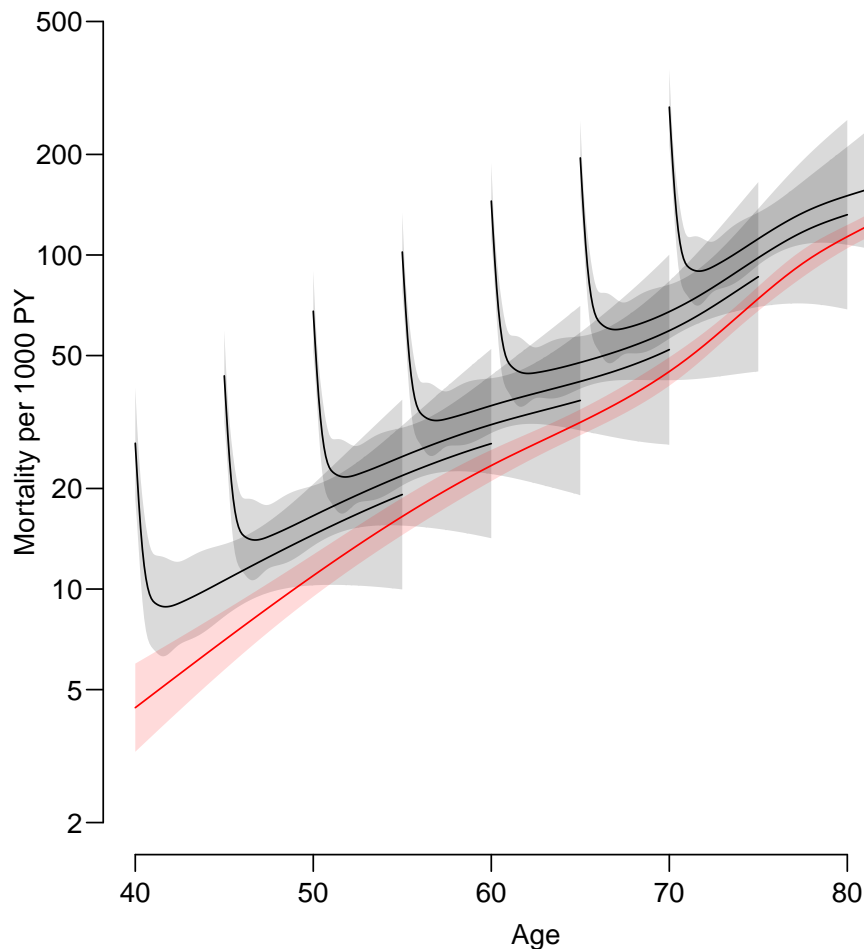


Figure 4: *The joint effect of age and duration of insulin use on mortality. Black curves are for persons starting insulin at ages 40, 45, . . . , and the red curve is for persons not on insulin.*

What can be seen from this plot is that the older age at insulin uptake, the higher the mortality at a given age. So there seem to be a positive effect of age at insulin onset, however only apparently; the model *assumes* that the difference the black curves and the red is the same across age at insulin. The linear effects of current age, Age, and insulin

duration tsI are of course in the model, and by that token also the linear effect of $Age - tsI = age$ at insulin use.

We can also plot the rate-ratios for the different ages at entry:

```
> ndR <- transform( ndI, tsI=0*tsI, lex.Cst=factor("DM"))
> pprR <- ci.exp( Pm, list(ndI,ndR) )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( ndI$Age, pprR, lty=1, log="y", plot=TRUE,
+          ylab="RR between Insulin and no Insulin", xlab="Age" )
> abline( h=1, lty=3 )
```

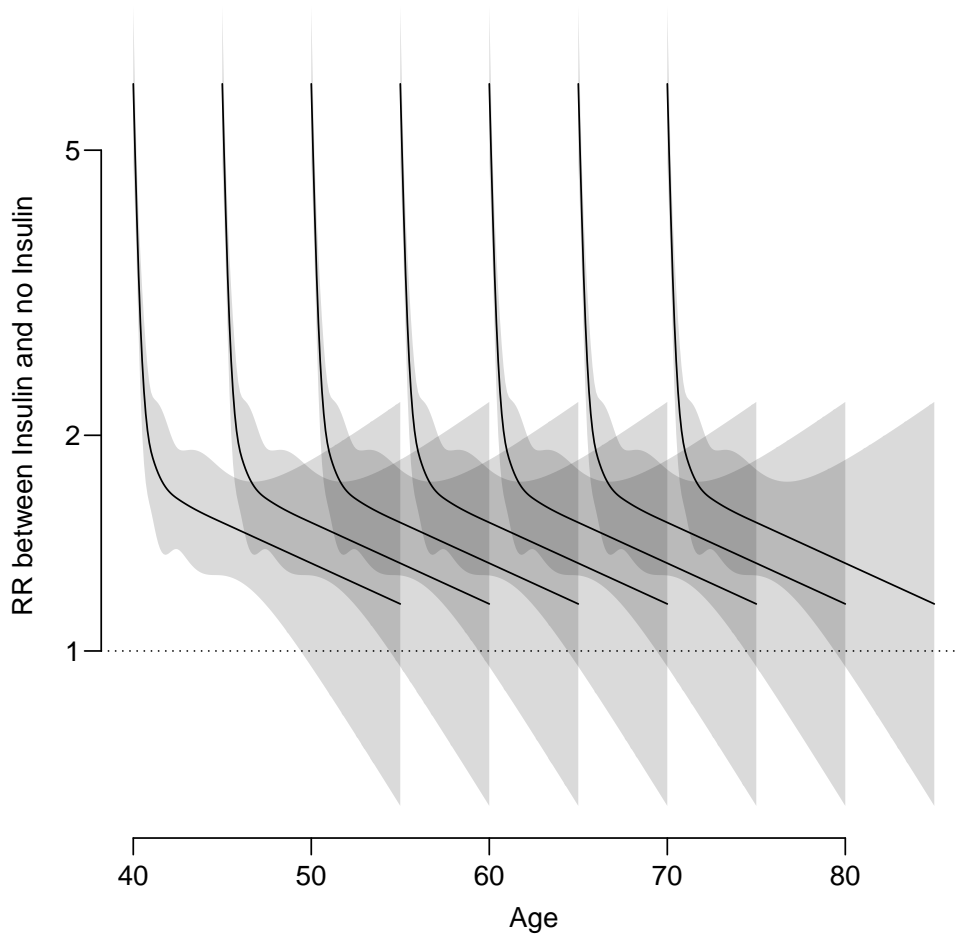


Figure 5: *The effect of duration of insulin use at different ages of follow-up (for ages at insulin start 40, 45, ...). Not surprisingly this is just a lot of identical curves — the model is additive in age and duration terms. Essentially this is just multiple repeats of the curve shown in figure 3.*

5 Age × duration interaction

5.1 Age at insulin start

In order to check the consistency of this additivity assumption across the spectrum of age at insulin inception, we can fit an interaction model using a non-linear effect of age at

insulin use (for convenience we use a simple translation of the knots used for Age):

```
> im <- glm.Lexis( dms, formula = ~ Ns(Age      ,knots=a.kn)
+                + Ns(      tsI,knots=i.kn)
+                + Ns(Age-tsI,knots=a.kn-4):lex.Cst
+                + lex.Cst + sex )
stats::glm Poisson analysis of Lexis object dms with log link:
Rates for transitions: DM->Dead, Ins->Dead
```

Note that we are here introducing an extra age-term in the DM group, because the term `Age-tsI` is equal to `Age` in this group (if not on insulin, time since insulin `tsI` is scored 0).

```
> pprIi <- ci.pred( im, ndI )
> pprDi <- ci.pred( im, ndD )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( ndI$Age, cbind(pprI,pprIi)*1000, plot=T,
+          log="y", lwd=c(2,3,3),
+          col=c("black","blue"), alpha=c(0.1,0.1),
+          ylim=c(2,500), ylab="Mortality rate per 1000 PY", xlab="Age" )
> matshade( ndD$Age, cbind(pprD,pprDi)*1000, lty=c(3,1), col="red", plot=FALSE )
```

Here we see that the interaction effect is such that in the older ages the length of insulin use has an increasing effect on mortality.

With the interaction model it would be illustrative to show the RR as a function of age at insulin and age at follow-up:

```
> pprRi <- ci.exp( im, ctr.mat=list(ndI,ndR) )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( ndI$Age, pprRi, lty=1, log="y", plot=TRUE,
+          col="blue", ylab="RR Ins vs. noIns", xlab="Age" )
> abline( h=1, lty=3 )
```

5.2 General interaction

As a further illustration we may want to explore a different kind of interaction, not defined from the duration — here we simplify the interaction by not using the second-last knot in the interaction terms. Note again that the prediction code is the same:

```
> xm <- glm.Lexis( dms, formula = ~ Ns(Age,knots=a.kn)
+                + Ns(tsI,knots=i.kn)
+                + Ns(Age,knots=a.kn[-4]):Ns(tsI,knots=i.kn[-4])
+                + lex.Cst + sex )
stats::glm Poisson analysis of Lexis object dms with log link:
Rates for transitions: DM->Dead, Ins->Dead
> pprx <- ci.pred( xm, ndI )
> pprb <- ci.pred( xm, ndD )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( ndI$Age, cbind( pprIi, pprx )*1000,
+          lty=1, log="y", plot=T, lwd=2,
+          col=c("blue","forestgreen"), alpha=c(0.1,0.2),
+          ylim=c(2,500), ylab="Mortality per 1000 PY", xlab="Age" )
> matshade( ndD$Age, cbind( pprDi, pprb )*1000, plot=F, lwd=2, lty="21", lend="butt",
+          col=c("blue","forestgreen"), alpha=c(0.1,0.2) )
```

Finally we can compare the two interaction models with the age-duration model, we see that there are indeed interaction between age and duration of insulin, but it is not immediately clear which one produces the better fit. From the figure it is pretty evident that the two types of interaction cannot be distinguished.

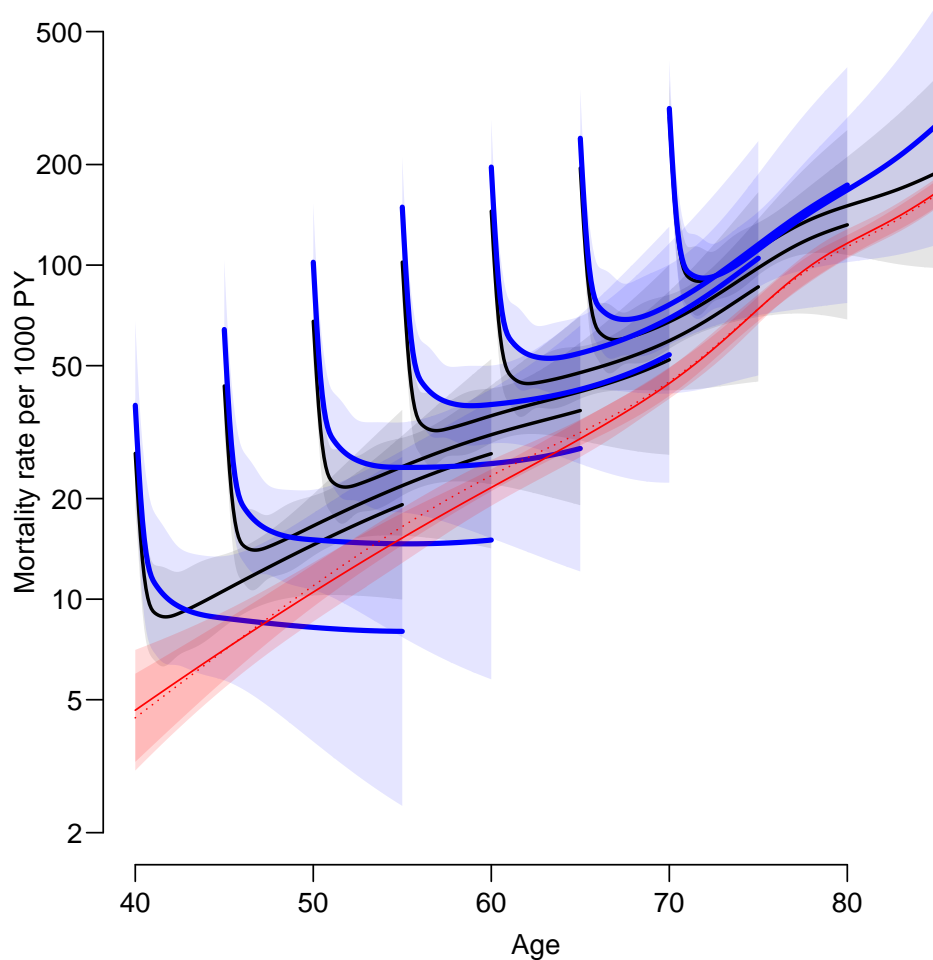


Figure 6: *Age at insulin as interaction between age and duration (blue), the black curves are from the main effects model. The red curves are for the non insulin users, the dotted curve is from the main effects model, the solid curve from the interaction model. We see that the attenuation of the Ins/noIns RR is more pronounced in the younger ages.*

Note that these curves are the combination of two timescales (as well as their difference), and therefore, if we consider one of them the baseline time scale, these quantities will not be available from a Cox model.

```
> anova( xm, Pm, im, test="Chisq")
```

Analysis of Deviance Table

```
Model 1: cbind(trt(Lx$lex.Cst, Lx$lex.Xst) %in% trnam, Lx$lex.dur) ~ Ns(Age,
  knots = a.kn) + Ns(tsI, knots = i.kn) + Ns(Age, knots = a.kn[-4]):Ns(tsI,
  knots = i.kn[-4]) + lex.Cst + sex
```

```
Model 2: cbind(trt(Lx$lex.Cst, Lx$lex.Xst) %in% trnam, Lx$lex.dur) ~ Ns(Age,
  knots = a.kn) + Ns(tsI, knots = i.kn) + lex.Cst + sex
```

```
Model 3: cbind(trt(Lx$lex.Cst, Lx$lex.Xst) %in% trnam, Lx$lex.dur) ~ Ns(Age,
  knots = a.kn) + Ns(tsI, knots = i.kn) + Ns(Age - tsI, knots = a.kn -
  4):lex.Cst + lex.Cst + sex
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	337204	26775			
2	337216	26795	-12	-20.062	0.06592
3	337207	26778	9	17.081	0.04746

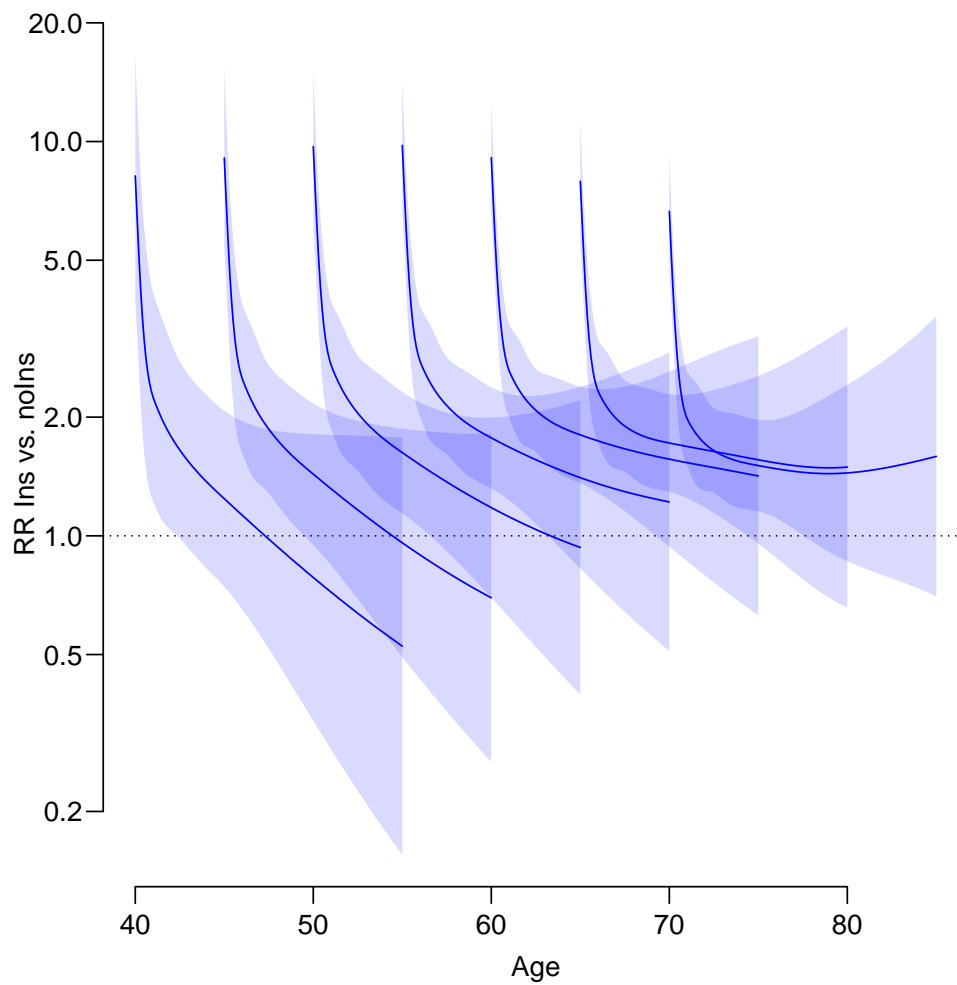


Figure 7: *The effect of duration of insulin use at different ages of follow-up, for different ages at insulin start). Note that this is the only thing you could derive from a Cox-model, the absolute rates as shown in figure 6 are not available. So we from a Cox model we will not know how mortality changes jointly by age and duration.*

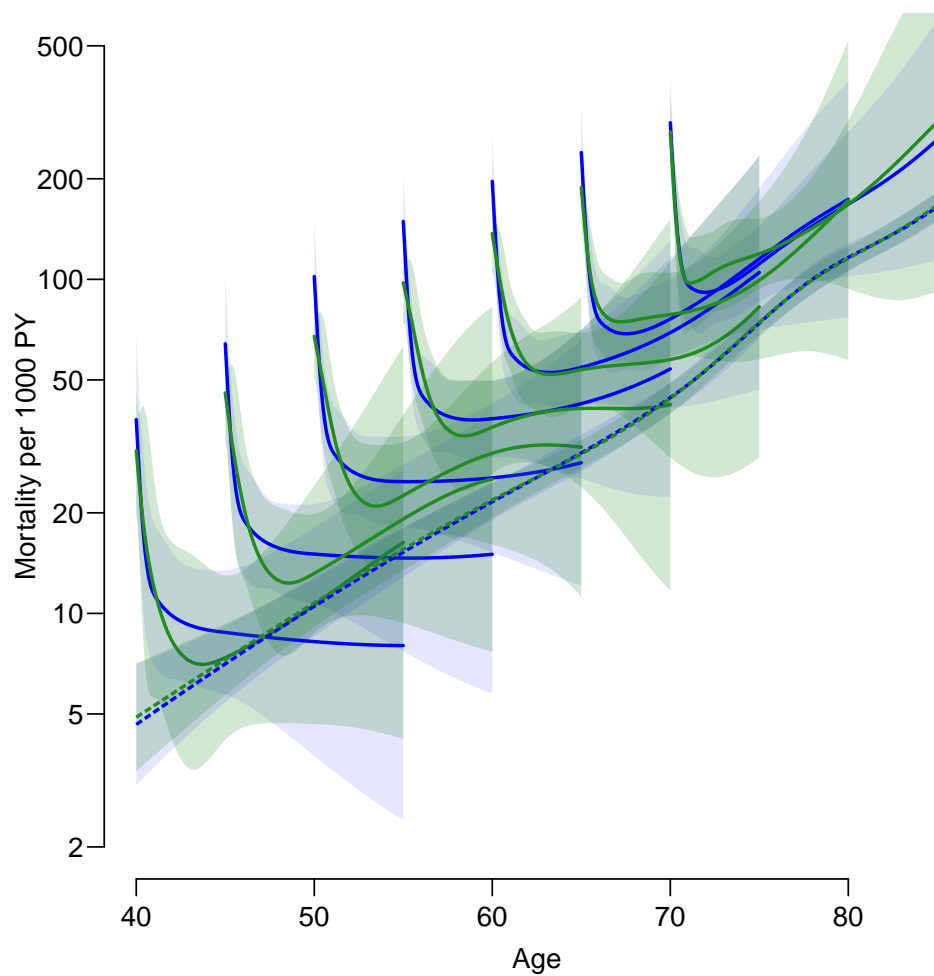


Figure 8: *General spline-by-spline interaction between age and duration (red), and the interaction through age at entry (green). The broken line are mortality among patients not on insulin.*

6 Estimation and reporting of multiple timescales — practical hints and advice

Bendix Carstensen, Senior Statistician, Steno Diabetes Center Copenhagen, Denmark, bendix.carstensen@regionh.dk

Vivian Viallon, Senior Statistician, International Agency for Research on Cancer, Lyon, France, ViallonV@iarc.fr

6.1 Context

When modeling transition rates between health states, be that in clinical or epidemiological studies, it seems to be a widespread practice to use only one timescale for description of rates. While this represents a huge advantage over primitive Markov models with constant transition intensities, it is hardly ever a biologically sensible approach in general.

For example would mortality typically depend not only on current age but likely also on current disease duration (possibly for several diseases), so practical models for transition rates must necessarily include effects of several time scales. And preferably defined from a clinical and/or epidemiological perspective and not from simple technical deliberations as clock forward or clock reset.

6.2 Methods

This can indeed be accommodated using a Cox-model, the practical consequence being a subdivision of follow-up over several records to allow timescales other than the primary, typically by some parametric functions (splines or fractional polynomials, typically). On the other hand, once the data has been subdivided, it is more logical to use a Poisson model to model all timescales the same way.

This approach further allows to report the effect of several timescales simultaneously — age and duration of diabetes normally advance at the same pace, for example.

6.3 Results

We illustrate the practical procedures needed to model several time scales (and the difference between these), and show how the joint effects of more than one time scale can be reported either longitudinally or cross-sectionally.

We will illustrate points with data from an international multi-morbidity study as well as a Danish mortality study of diabetes patients. We shall demonstrate some of the Lexis tools from the Epi package in R that facilitates this type of analyses, both aimed at the realm of rates but also the realm of probability and sojourn times.

6.4 Conclusion

With the advent of modern computing speed the practical problems associated with analysis of the very large data sets the emerging from this approach, a proper continuous modeling of time-effects on has become feasible. The major challenge is reporting several time-scales simultaneously (as the do advance simultaneously in real life). This invariably requires graphical reporting.