Remarks on handling life-events as predictors of T1D occurrence

SDC / JeBe
http://bendixcarstensen.com/SDC/JeBe
November 2013
Version 1

Compiled Tuesday 29th November, 2016, 13:54
from: /home/bendix/sdc/coll/jebe/lifeevents.tex

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

Contents

1 Single life-events

We assume that a number of life-events of similar type can occur during the (early) life of a child and we want to assess to what extent these influence the occurrence rate of T1D (incidence rates).

Suppose that life-events of type $X_j, j = 1 \ldots$ occur at ages t_j, and assume for the sake of the argument that the incidence rate of T1D as a function of age, a, is $\lambda(a)$ if no life-events occur.

We can contemplate two conceptually different models for the influence of life-events on T1D occurrence:

- occurrence of X_j at t_j influences incidence rates primarily as a function of $a - t_j$ (of course only for $a > t_j$) — that is as a function of time *since* life-event.
- occurrence of X_j at t_j influences incidence rates primarily as a function of t_j, that is, as a function of age *at* life-event.

In principle there is nothing that requires us to choose one or the other, both can be accommodated.

1.1 Mathematical formulation

The prerequisite for the following is that we have tools to estimate a non-linear dependence of occurrence rates on age and time. The machinery for this is, briefly described, to take the follow-up for each person and split it in small intervals (1 or 3 months, say) and fit a model where incidence rates depend on the time (age, times since life-event, time at life-event) as calculated in each interval. The crucial point is estimation of the effects of these as *smooth* continuous effects of the time scales.

The simplest possible model is the model where incidence rates depends *only* on age (well, separately for each sex), to see what kind of function to expect, see the leftmost curves in figure 3.3 on p. 21 of http://bendixcarstensen.com/SDC/T1APC/t1apc.pdf.

The effect of life-events are described *relative* to these age-specific incidence curves as rate-ratios or hazard ratios (HRs); in the most general form:

$$\lambda(a|X_j \text{ at } t_j) = \lambda(a) \times [\zeta(t_j) \times \xi(a - t_j)]^{I(t_j > a)}$$

where $I(t_j > a)$ is an indicator function which is 0 for $a < t_j$ and 1 otherwise, and

- $\zeta(t_j)$ is the HR associated with life-event occurrence at age t_j — assumed to be the same regardless of age. Thus if X_j occurs at t_j, the assumption is that the age-specific rates if T1D jumps by a factor $\zeta(t_j)$, but otherwise has the same shape by age. Thus if $\zeta(t)$ is uniformly equal to 1, then there is no *general* effect of X_j on the T1D occurrence.
- $\xi(d_j) = \xi(a - t_j)$ is the effect of occurrence of X_j as a function of the time since occurrence of X_j (duration, d_j). If there is no immediate effect, we would expect this to start at 1 (that is for $d_j = 0$) and then possibly increase over time.

However, the model above assumes a particular form of the effect of the time *at* life-event and the time *since* life-event. A more general form would be to have a completely general *interaction*:

$$\lambda(a|X_j \text{ at } t_j) = \lambda(a) \times \kappa(t_j, a - t_j) \tag{1}$$

That is, the occurrence of X_j has an effect that depends in a completely general way on age at occurrence and time since occurrence of the life-event. The reporting of such effects would be as a set of curves showing how the occurrence rates (T1D incidence) depend on age for different ages of life-event.

1.2 Graphical example

Here we generate a graph showing an example of such effects of a *single* life-event and how to report them.

```
> # Generate a plot and read 20 points from mouse-clicks
> plot(seq(0,20,,50),seq(2,11,,50))
> zz <- locator(20)
> rates <- zz$y
> ages <- zz$x
> save( ages, rates, file="bogus.Rda" )
```

Here is the code that generates the plots, using ad-hoc assumptions about the effects:

```
> library( Epi )
> load( file="bogus.Rda")
> m0 <- lm( log(rates) ~ Ns(ages,knots=seq(2,18,,5)) )
> pa <- seq(0,20,0.1)
> pr <- predict( m0, newdata=data.frame(ages=pa) ) * 7
> a3  <- ifelse( pa> 3, 1.5, 1 )
> a7  <- ifelse( pa> 7, 1.2, 1 )
> a12 <- ifelse( pa>12, 1.1, 1 )
> p3  <- exp( pmax( 0, pa- 3 ) * 0.05 )
> p7  <- exp( pmax( 0, pa- 7 ) * 0.03 )
> p12 <- exp( pmax( 0, pa-12 ) * 0.01 )
> allcrv <-  cbind( p12*pr,    p7*pr,    p3*pr, pr,
+                   a12*pr,    a7*pr,    a3*pr, pr,
+               p12*a12*pr, p7*a7*pr, p3*a3*pr, pr )[,c(4,1:3,8,5:7,12,9:11)]
> clr <- c("red","limegreen","blue","black")[4:1]
> wh <- c(1:31,NA,32:71,NA,72:121,NA,122:201)
> par( mfrow=c(1,2) )
> matplot( pa[wh], allcrv[wh,],
+          type="l", lwd=c(8,3,3,3), col=clr,
+          lty=rep(c("11","42","solid"),each=4), log="y", ylim=c(5,50),
+          xlab="Age at follow-up", ylab="Incidence rate of T1D")
> abline(v=3,col="blue")
> abline(v=7,col="limegreen")
> abline(v=12,col="red")
> matplot( pa[wh], allcrv[wh,]/pr[wh],
+          type="l", lwd=c(8,3,3,3), col=clr,
+          lty=rep(c("11","42","solid"),each=4), log="y", ylim=c(0.5,5),
+          xlab="Age at follow-up", ylab="Incidence rate-ratio of T1D")
> abline(v=3,col="blue")
> abline(v=7,col="limegreen")
> abline(v=12,col="red")
```

The coloured effects in the right panel in figure 1 are the terms that were named κ in formula 1. Note that each point on the coloured curves correspond to a particular combination of t_j (age at life-event) and $a - t_j$ (time since) life-event, so the completely

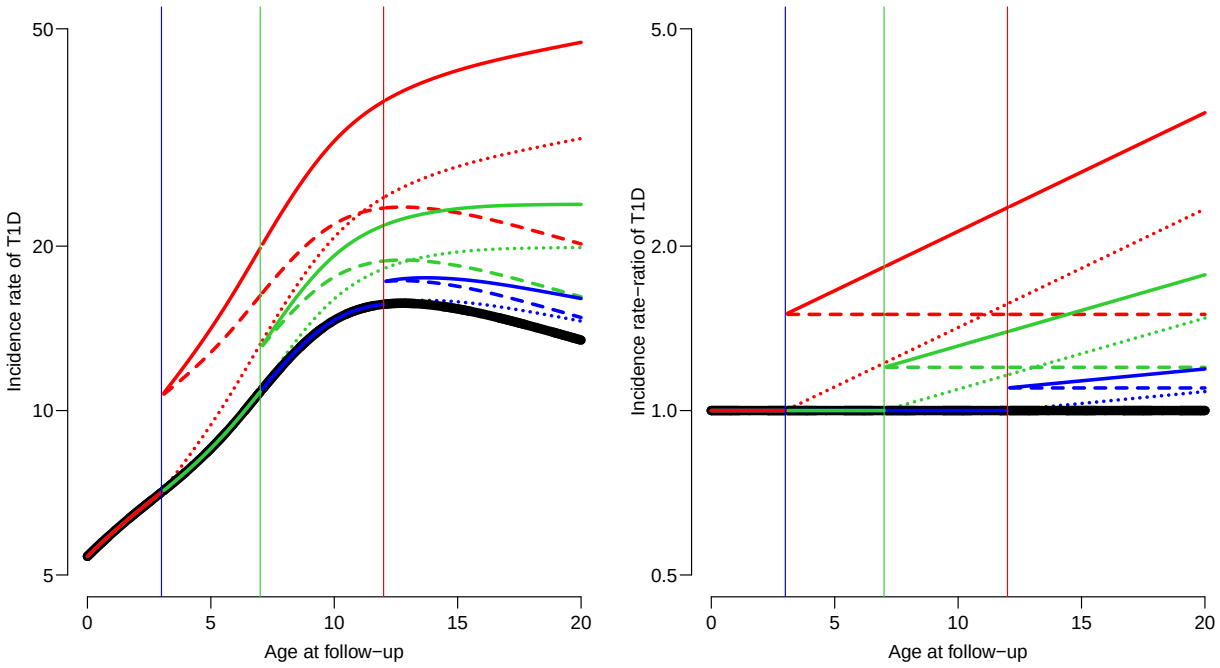

Figure 1: *Example of reporting effects of life-events occurring at ages 3, 7 and 12 years. In the left panel, the thick black line is the incidence rates for children with no life-event. The three coloured lines refer to rates for children seeing a life-event at ages 3, 7 and 12 years respectively.*
The broken *lines are from the model where we assume that the effect of an life-event is immediate and constant after the life-event — this is the type of assumption that is made in a "Cox-model with life-event as time-dependent variable".*
The dotted *lines are from the model where we assume that the effect of an life-event gradually increases from nothing as time since life-event increases. In this example we have just assumed the effect to be log-linear.*
The right hand panel is the corresponding hazard-ratios (incidence rate-ratios) between a persons with a life-event at 3, 7 and 12 years and a person without life-event. The assumption here is that both the immediate and the long-term effects are larger the younger the person is at life-event occurrence. These are of course effects that must be estimated *from the observed data.*

general specification of κ actually allows the coloured lines to have *any* shape and position. In practice we would impose restrictions on the shape and position of the curves to make the problem tractable.

The shape assumption used in figure 1 is that the RRs are linear as a function of time since life-event, but what is not immediately apparent from the figure is exactly how the *slope* of the curves are assumed to depend on t_j — in this case apparently decreasing by increasing t_j. Likewise, we also assumed that the position (essentially the jump at life-event time t_j) was decreasing by t_j in some way.

So in reality we have only assumed that:

$$\kappa(t_j, a - t_j) = f(t_j) + g(t_j)(a - t_j)$$

for some functions f and g, where (as can be seen from the code) $f(3) = 1.5$, $f(7) = 1.2$, $f(12) = 1.1$ (the `as`) and $g(3) = \exp(0.05) = 1.051$, $g(7) = \exp(0.03) = 1.030$, $g(12) = \exp(0.01) = 1.010$ (the `ps`), but in any practical setting we must impose some parametric restrictions on f and g and estimate their shape. But the practical reporting would still be by graphs as shown in figure 1.

2 Several life-events

The curves in figure 1 exemplify one possible way of modeling the effect of *one* particular life-event. The "counting" approach would formally correspond to assuming that all types of life-events had the same effect as function of age at life-event and time since life-event. It would in principle be possible to model effects of different types of life-events separately, assuming an additive effect (on the log-scale).

Note however that this type of assumption is a sort of *independence* type of assumption, where the effect of life-event type X_a is the same, regardless of whether X_b has occurred or not. If we want to include "synergies" it would be possible to assume that the effect of X_a was dependent on, say, the *number* of previous life-events. One simple way of accommodating this would be to define a variable $N(a)$ counting the number of life-events before age a, and expanding the model to:

$$\lambda(a|X_j \text{at} t_j) = \lambda(a) \times \kappa(t_j, a - t_j) \times \exp\Big(\beta\big(N(a) - 1\big)\Big)$$

The reason that we use $N(a) - 1$ is that this makes the last term go away (that is be equal to 1) for the first life-event (where $N(a) = 1$), allowing an interpretation of $\kappa(t_j, a - t_j)$ as the effect of X_j as the *first* life-event, and e^{β} as the factor that attenuates (if $\beta < 0$) or increases (if $\beta > 0$) the effect of X_j per previously seen life-event.

3 Terminological caveat

In this study the *exposures* are *life*-events, whereas the *outcome* of interest is a *disease* event, namely T1D. So there is ample room for misunderstandings, and I suggest that the exposures be always referred to as *life*-events, and that the outcome as T1D occurrence (rates) wherever possible.