Diabetes and amputation incidence in Fyn county, Denmark

Benjamin Schnack Rasmussen, Knud Yderstræde, Henning Bech-Nielsen,
OUH
September 2015
http://BendixCarstensen.com/SDC/Ampu
Version 5.4

Compiled Thursday 10th September, 2015, 00:40
from: /home/bendix/sdc/extn/OUH/BSR/FynAmp

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

Contents

Chapter 1

Introduction to models

1.0 Conjuration

The purpose of a statistical report is to describe statistical considerations and analyses as precisely as possible. The report is intended to be a support for the client in the writing of scientific papers, as well as a documentation for the author (and other statisticians) of the entire statistical analysis.

It is not the intention that the client understand all the technical details of the report, but it is important that the client sees to it that the author's description of the problems is correct, and that the proposed solutions address the research questions adequately.

This report contains some technical aspects that are necessary in order to document the content. They are not essential for the understanding of the concepts conveyed, but they are necessary for the documentation of the computational aspects of practical analyses. While this paper is centered on the use of R-code there in nothing that precludes use of other types of software (except available programming time).

1.1 Amputation rates

If we want to assess the change in amputation rates among diabetes patients over calendar time, we must first define the rates of amputation.

There are different kinds of amputation, and the same person my undergo different types of amputations. To capture this we must make some assumptions:

- Amputations are accurately recorded w.r.t. date and type
- Types of amputation can be ordered by "severity"

1.2 Two simple models

1.2.1 Maximal amputation model

Suppose for the sake of the argument that amputations come in 3 different grades: $A_1 < A_2 < A_3$, and that they are defined as the *maximal* degree of amputation. This means that a patients that has an A_2 amputation (above knee, left, say) and subsequently an A_1 amputation (foot, right, say) does *not* change status at the second amputation because it was "only" an A_1, and the patient was already an A_2 patient. So this approach does not necessarily count all amputation events, it only counts *increases*.

An illustration of this is as follows:

```
> tmat <- matrix( NA, 4,4)
> tmat[col(tmat)>row(tmat)] <- 1
> rownames(tmat) <- colnames(tmat) <- c("DM","A1","A2","A3")
> tmat
   DM A1 A2 A3
DM NA  1  1  1
A1 NA NA  1  1
A2 NA NA NA  1
A3 NA NA NA NA
> library(Epi)
> boxes( tmat, boxpos=list(x=c(14,47,85,77),y=c(52,83,50,9)),
+              col.arr=rep(c("red","black"),each=3),
+              hmult=1.7, wmult=1.5 )
```

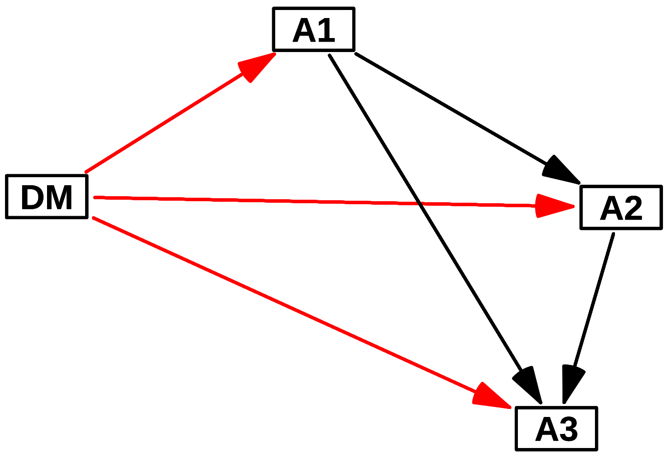

Figure 1.1: *Model for maximal amputation status.*

In the model illustrated in figure 1.1, it will be of interest to see if the three red arrows (first amputations of different severity) change over time in the same way. Likewise the three black arrows would be analyzed in the same way. Further details in the "Models" section 1.4.

1.2.2 Amputation score model

In order to accommodate *all* amputations we could define a *score* for each amputation, and so for patients multiply amputated a *cumulative* score. So if for example the maximal score were 6 we would in principle have a model as in the r.h.s. of figure 1.2.

```
> tmat <- matrix( NA, 7,7)
> tmat[col(tmat)>row(tmat)] <- 1
> rownames(tmat) <- colnames(tmat) <- c("DM",paste(1:6))
> tmat
```

```
   DM  1  2  3  4  5  6
DM NA  1  1  1  1  1  1
1  NA NA  1  1  1  1  1
2  NA NA NA  1  1  1  1
3  NA NA NA NA  1  1  1
4  NA NA NA NA NA  1  1
5  NA NA NA NA NA NA  1
6  NA NA NA NA NA NA NA
> boxes( tmat, boxpos=TRUE, #list(x=c(14,47,85,77),y=c(52,83,50,9)),
+              col.arr=rep(c("red","black"),c(6,15)),
+              hmult=1.7, wmult=1.3 )
```

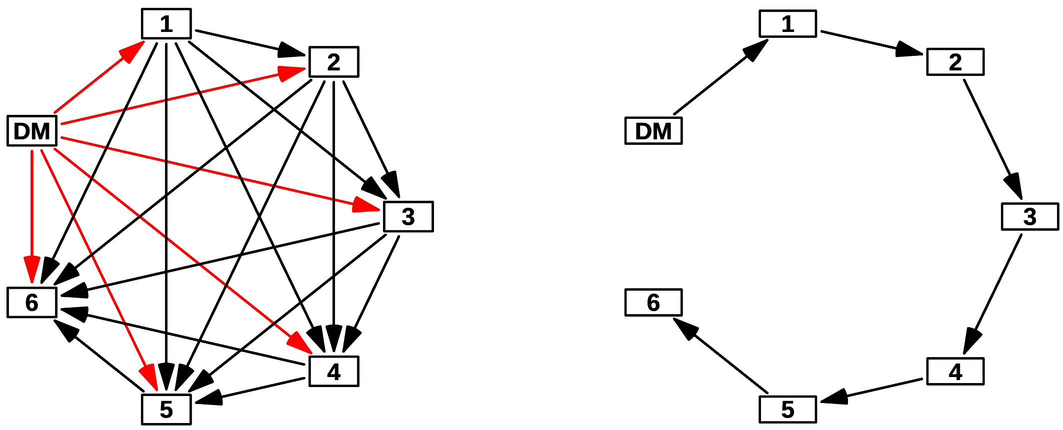

Figure 1.2: *Models for the maximal amputation score.*

Therefore, a simplification would be to look at a model as in figure 1.2, where the focus would be on the patients exceeding a certain degree of amputation.

```
> tmat <- matrix( NA, 7,7)
> tmat[(col(tmat)-row(tmat))==1] <- 1
> rownames(tmat) <- colnames(tmat) <- c("DM",paste(1:6))
> tmat
   DM  1  2  3  4  5  6
DM NA  1 NA NA NA NA NA
1  NA NA  1 NA NA NA NA
2  NA NA NA  1 NA NA NA
3  NA NA NA NA  1 NA NA
4  NA NA NA NA NA  1 NA
5  NA NA NA NA NA NA  1
6  NA NA NA NA NA NA NA
> boxes( tmat, boxpos=TRUE, hmult=1.5, wmult=1.5 )
```

But it would require a bit of data-tampering, because a person moving from score 2 to 5, say, would have to be coded as going from 2 to 3, from 3 to 4 and from 4 to 5, with a minimal (1 day, say) sojourn time in states 3 and 4. Potentially this could also cause problems if some scores were rare, because the amount of risk time in those states would be rather small. The obvious remedy for this would be to pool adjacent states, pooling states with very small amounts of risk time with the subsequent state.

1.3 Data requirements

The requirement of data for analyzing set-ups like these is simply that each diabetes patient be recorded with date of:

- birth
- diabetes diagnosis
- start of follow-up
- end of follow-up
- amputations
- death

It is important to note here that *all* dates of amputations as well as the type (location?) of amputation be recorded. Whether each amputation should give rise to a separate record, or whether the dataset just should have a sufficient number variable to record all dates for the patient with the largest number of amputations is immaterial.

On top of this, other relevant clinical variables at start of follow-up is required. If clinical variables recorded *during* follow-up is required, separate records for each recording date are required. In that case the “date of start of follow-up” would then be replaced by “date of clinical record”, and the earliest of these would be taken as the start of follow-up.

1.4 Models

With data as outlined above, it is possible to set up statistical models for the transition rates, either separately for each singe rate, or jointly for more rates, if we for example want to see if calendar time of follow-up influences the amputation rates.

It will for example be possible to see how amputation rates change by calendar time, controlled for the composition of the diabetes population with respect to sex, age, duration of diabetes *and* current amputation status.

The model to be used will be Poisson models for suitable time-split data; it is not possible to use a Cox-modeling approach as several time-scales will be involved: age, diabetes duration, calendar time and possibly time since last amputation. The multistate approach does not itself preclude the use of Cox-models, it is the need for more than one time-scale that does.

Chapter 2

The data base

In this chapter we read in the diabetes and amputation data and set up an appropriate analysis database.

```
> options( width=120 )
> library( Epi )
> print( sessionInfo(), l=F )
R version 3.2.1 (2015-06-18)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
[1] Epi_1.1.69

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7     MASS_7.3-42      parallel_3.2.1   survival_2.38-3 etm_0.6-2       splines_3.2.1
[8] lattice_0.20-31
```

We have 4 data files at our disposal, 3 relating to the Danish Diabetes register, and 1 to the amputations, all with person-id in the form of a cpr-number:

2.1 Diabetes Register

The Danish National Diabetes Register, NDR:

```
> ndr <- read.csv2( "./data/t_diabetes.asc", header=TRUE )
> whd <- grep( "D_", names(ndr) )
> for( i in whd ) ndr[,i] <- cal.yr( ndr[,i] )
> names( ndr ) <- substr( tolower( names(ndr) ), 3, 30 )
> with( ndr, addmargins( table(table( cpr ) ) ) )
     1    Sum
497232 497232

> str( ndr )
'data.frame':        497232 obs. of  12 variables:
 $ cpr       : num  1.01e+08 1.01e+08 1.01e+08 1.01e+08 1.01e+08 ...
 $ foddto    :Classes 'cal.yr', 'numeric'  num [1:497232] 1900 2000 2000 1901 2001 ...
 $ sex       : Factor w/ 2 levels "K","M": 1 1 1 1 2 2 2 2 2 1 ...
 $ inkldto   :Classes 'cal.yr', 'numeric'  num [1:497232] 1990 2006 2009 1993 1994 ...
 $ inklaarsag: Factor w/ 6 levels "blod2i5","blod5i1",..: 3 5 5 3 6 5 5 5 5 5 ...
 $ dodsdto   :Classes 'cal.yr', 'numeric'  num [1:497232] 1991 NA NA 1994 NA ...
 $ lpr       :Classes 'cal.yr', 'numeric'  num [1:497232] 1991 2006 2009 NA NA ...
```

```
 $ fodt      :Classes 'cal.yr', 'numeric'  num [1:497232] 1990 NA NA 1993 NA ...
 $ blod2i5   :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA NA NA NA NA NA NA ...
 $ blod5i1   :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA NA ...
 $ ins       :Classes 'cal.yr', 'numeric'  num [1:497232] NA 2006 2009 NA NA ...
 $ oad       :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA 1994 ...
> summary( ndr )
      cpr                 foddto        sex            inkldto          inklaarsag          dodsdto
 Min.   :1.010e+08   Min.   :1889   K:239392   Min.   :1942   blod2i5:   380   Min.   :1971     Min.
 1st Qu.:8.055e+08   1st Qu.:1927   M:257840   1st Qu.:1995   blod5i1:187642   1st Qu.:1998     1st
 Median :1.513e+09   Median :1939              Median :2002   fodt   : 48758   Median :2003     Medi
 Mean   :1.562e+09   Mean   :1940              Mean   :2001   ins    :  6855   Mean   :2003     Mean
 3rd Qu.:2.308e+09   3rd Qu.:1951              3rd Qu.:2008   lpr    :150186   3rd Qu.:2008     3rd
 Max.   :3.113e+09   Max.   :2011              Max.   :2012   oad    :103411   Max.   :2012     Max.
                                                                               NA's   :310870   NA's
      fodt            blod2i5          blod5i1             ins               oad
 Min.   :1990     Min.   :1993     Min.   :1990     Min.   :1994     Min.   :1994
 1st Qu.:1996     1st Qu.:1997     1st Qu.:1996     1st Qu.:1995     1st Qu.:1998
 Median :2002     Median :2002     Median :2002     Median :2002     Median :2004
 Mean   :2001     Mean   :2002     Mean   :2001     Mean   :2002     Mean   :2003
 3rd Qu.:2005     3rd Qu.:2007     3rd Qu.:2007     3rd Qu.:2007     3rd Qu.:2009
 Max.   :2012     Max.   :2012     Max.   :2012     Max.   :2012     Max.   :2012
 NA's   :303518   NA's   :431893   NA's   :206898   NA's   :375954   NA's   :223838
```

2.1.1 Place of residence

Place of residence for the patients in the NDR is in a special file with one record per known place of residence:

```
> bop <- read.csv2( "./data/t_bopael.asc", header=TRUE )
> ( whd <- grep( "D_", names(bop) ) )
[1] 2 3
> names( bop ) <- substr( tolower( names(bop) ), 3, 30 )
> for( i in whd ) bop[,i] <- cal.yr( bop[,i], format="%Y-%m-%d" )
> str( bop )
'data.frame':        572228 obs. of  5 variables:
 $ cpr   : num  1.41e+09 1.41e+09 1.40e+09 1.41e+09 1.41e+09 ...
 $ start :Classes 'cal.yr', 'numeric'  num [1:572228] 2000 2009 1995 2005 1990 ...
 $ slut  :Classes 'cal.yr', 'numeric'  num [1:572228] 2005 NA 2008 NA 1992 ...
 $ kom   : int  101 101 101 101 101 101 101 101 101 101 ...
 $ region: int  1084 1084 1084 1084 1084 1084 1084 1084 1084 1084 ...
> with( bop, addmargins( table(table( cpr, exclude=NULL ) ) ) )

     0      1      2      3      4      5      6      7      8      9     10     11     12     13
     1 449448  30930   8865   3176   1569    798    406    265    124     96     54     32     22
    17     19     20     21     33     34    Sum
     3      2      3      2      1      1 495827
```

From the latter we see that we are missing place of residence on at least some of the persons in the NDR, since we have only 495,827 different cpr-numbers in the address base, but 497,232 persons in the diabetes register.

We merge the NDR with the residential records and find out which place of residence to use — the id-number of the municipalities on Fyn is 4xx, so these are easy to fish out, and it will give us the diabetes population in the Fyn municipalities, or rather anyone who has ever resided in Fyn. For control and annotation we read the number and name of the municipalities:

```
> ( knam <- read.table( "knam.txt", header=TRUE ) )
```

```
   kno       knam
1  410 Middelfart
2  420     Assens
3  430    Faaborg
4  440 Kerteminde
5  450     Nyborg
6  461     Odense
7  479  Svendborg
8  480    Bogense
9  482  Langeland
10 492  \xc6r\xf8
```

```
> bop <- subset( bop, kom>400 & kom<500 )
> bop$kom <- factor( bop$kom, levels=knam$kno, labels=knam$knam )
> cbind( with( bop, table(kom) ) )
```

```
            [,1]
Middelfart  4137
Assens      5133
Faaborg     5642
Kerteminde  2913
Nyborg      4077
Odense     19241
Svendborg   6020
Bogense     3285
Langeland   2520
\xc6r\xf8    970
```

```
> str( bop )
```

```
'data.frame':        53946 obs. of  5 variables:
 $ cpr   : num  1.81e+09 1.80e+09 1.71e+09 1.80e+09 1.81e+09 ...
 $ start : num  2008 1990 2012 2010 2002 ...
 $ slut  : num  NA 1998 NA NA NA ...
 $ kom   : Factor w/ 10 levels "Middelfart","Assens",..: 5 5 5 5 5 5 5 5 5 5 ...
 $ region: int  1083 1083 1083 1083 1083 1083 1083 1083 1083 1083 ...
```

Now we have records on all diabetes patients who at some point has been resident in Fyn, but we still have multiple records on persons, so we construct for each person the time-span from the earliest entry to Fyn (`start`) to the latest exit. Since we will only be following persons from 1996-01-01 to 2012-01-01, we truncate the Fyn-time to this period.

So implicitly we disregard periods where persons have been outside of Fyn, so we reduce data to a dataset with one record per person, with `eFyn` as the date of entry either to Fyn or to NDR, and `xFyn` as the exit date from Fyn, date of death or end of study, and we plot the exit-dates versus the entry dates before and after we cut them down to the period of interest:

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> xx <- transform( bop, eFyn = ave( start, cpr, FUN = function(x) min(x,na.rm=TRUE) ),
+                       xFyn = ave( slut , cpr, FUN = function(x) max(x,na.rm=TRUE) ) )
> with( xx, plot( eFyn, xFyn, pch=".", xlim=c(1990,2012), ylim=c(1990,2012) ) )
> xx$eFyn <- pmax( xx$eFyn, 1996, na.rm=TRUE )
> xx$xFyn <- ifelse( xx$xFyn==-Inf, 2012, xx$xFyn )
> with( xx, plot( eFyn, xFyn, pch=".", xlim=c(1990,2012), ylim=c(1990,2012) ) )
> x1 <- aggregate( xx[,c("eFyn","xFyn")], xx[,"cpr",drop=F], function(x) x[1] )
> cbind( dim(xx), dim(x1) )
```

```
      [,1]  [,2]
[1,] 53946 49497
[2,]     7     3
```

```
> ndr.fyn <- merge( ndr, x1, all.y=TRUE )
> cbind( dim(ndr), dim( ndr.fyn ) )
```

```
       [,1]  [,2]
[1,] 497232 49497
[2,]     12    14
```

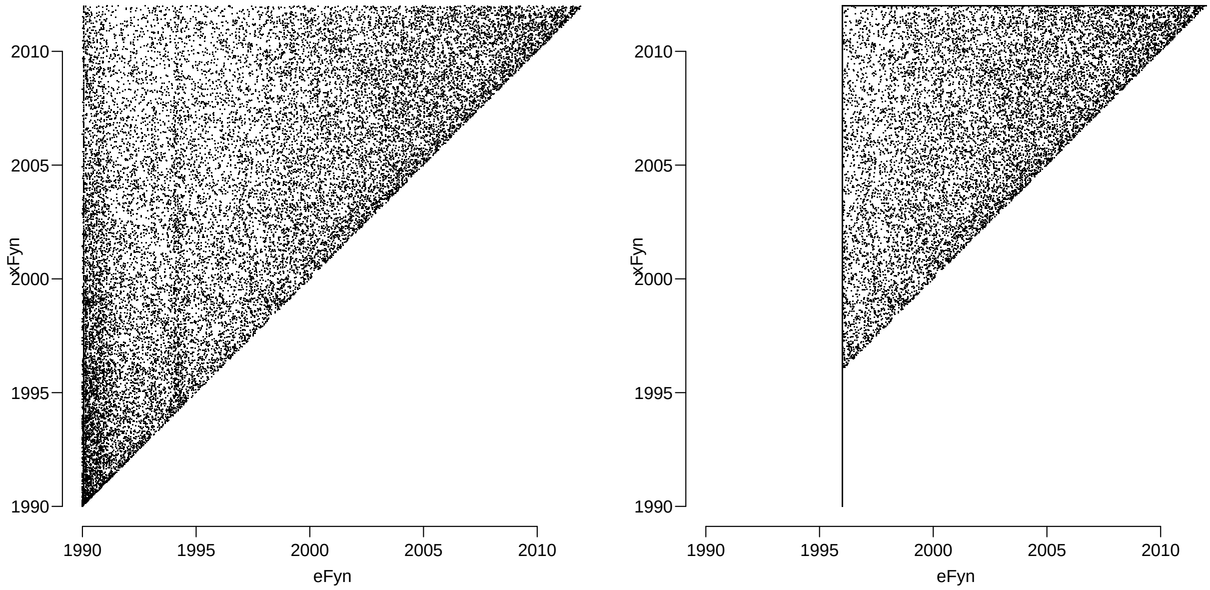

Figure 2.1: *Entry and exit dates for diabetes patients resident in Fyn, before and after truncation to the period 1996-01-01 to 2012-01-01.*

Now we have a file with diabetes patients in Fyn, with dates `eFyn` and `xFyn`, indicating the start and end of their sojourn in Fyn. Patients in the NDR who only resided in Fyn *before* their diagnosis of diabetes should not be included, though:

```
> ndr.fyn <- subset( ndr.fyn, xFyn > inkldto )
> dim( ndr.fyn )
[1] 49452    14
```

Thus `ndr.fyn` contains the diabetes patients that have had their diagnosis prior to a sojourn time in Fyn.

2.1.2 Vital status of DM patients

Vital status of persons in NDR is also in a separate file:

```
> vit <- read.csv2( "./data/t_person.asc", header=TRUE, na.strings="." )
> whd <- grep( "D_", names(vit) )
> for( i in whd ) vit[,i] <- cal.yr( vit[,i] )
> names( vit ) <- gsub( "_", ".", substr( tolower( names(vit) ), 3, 30 ) )
> str( vit )
'data.frame':        497232 obs. of  3 variables:
 $ cpr             : num  1.01e+08 1.01e+08 1.01e+08 1.01e+08 1.01e+08 ...
 $ status          : int  90 1 1 90 1 1 90 90 90 90 ...
 $ status.hen.start:Classes 'cal.yr', 'numeric'  num [1:497232] 1991 NA NA 1994 NA ...
> summary( vit )
      cpr                status       status.hen.start
 Min.   :1.010e+08   Min.   : 1.00   Min.   :1969
 1st Qu.:8.055e+08   1st Qu.: 1.00   1st Qu.:1998
 Median :1.513e+09   Median : 1.00   Median :2004
 Mean   :1.562e+09   Mean   :38.16   Mean   :2004
 3rd Qu.:2.308e+09   3rd Qu.:90.00   3rd Qu.:2009
 Max.   :3.113e+09   Max.   :90.00   Max.   :2013
                                     NA's   :289046
```

```
> with( vit, table( status, !is.na(status.hen.start) ) )

status   FALSE    TRUE
     1  288588       0
     3     269       0
     5     189       0
     20      0      58
     30      0       2
     50      0       2
     60      0     130
     70      0     166
     80      0    4004
     90      0  203824
```

The coding of `status` is shown in table 2.1 below.

This means that anyone with a values of `status`>10 should have follow-up terminated at the date `status.hen.start`:

```
> vit <- transform( vit, dox = ifelse( status>10,
+                                      status.hen.start,
+                                      NA ) )
> ndr.fyn <- merge( ndr.fyn, vit[,c("cpr","dox")], all.x=TRUE )
```

Note that for some persons the date of exit, `dox`, exists even if date of death is not defined. Since follow-up is till the end of 2012, we can define `dox` as the date of exit for all persons in the NDR. We also define a date of entry into the study for the diabetes patients as the date from which they are under observation in Fyn:

```
> ndr.fyn <- transform( ndr.fyn, doe = pmax( 1996, eFyn, inkldto,      na.rm=TRUE ),
+                                dox = pmin( 2012, xFyn, dodsdto, dox, na.rm=TRUE ) )
> names( ndr.fyn )
 [1] "cpr"        "foddto"     "sex"        "inkldto"    "inklaarsag" "dodsdto"    "lpr"        "fod
 [9] "blod2i5"    "blod5i1"    "ins"        "oad"        "eFyn"       "xFyn"       "dox"        "doe

> save( ndr.fyn, file="./data/ndr.fyn.Rda" )
```

Thus we have a register of all diabetes patients with their residence in Fyn from `doe` to `dox`.

Table 2.1: *Coding of the status variable* `status`*, referring to the status as of the date* `status.hen.start`*.*

01	aktiv, bopæl i dansk folkeregister
03	aktiv, speciel vejkode (9900-9999) i dansk folkeregister
05	aktiv, bopæl i grønlandsk folkeregister
07	aktiv, speciel vejkode (9900 - 9999) i grønlandsk folkeregister
20	inaktiv, uden bopæl i dansk/grønlandsk folkeregister men tildelt personnummer af skattehensyn (kommunekoderne 0010, 0011, 0012 og 0019)
30	inaktiv, annulleret personnummer
50	inaktiv, slettet personnummer ved dobbeltnummer
60	inaktiv, ændret personnummer ved ændring af fødselsdato og køn
70	inaktiv, forsvundet
80	inaktiv, udrejst
90	inaktiv, død

2.2 Amputation records

Amputation records from Fyn contains amputations from all persons resident in Fyn at the time of amputation.

We need to derive both sex and the date of birth for the amputees from the CPR-number; that is done according to the rules found in https://cpr.dk/media/167692/personnummeret%20i%20cpr.pdf, which are more pedagogically laid out in http://da.wikipedia.org/wiki/CPR-nummer. The R-implementation of this is here:

```
> check.cpr <-
+ function( x, fishy=FALSE )
+ {
+ # Checks if a CPR-number supplied is too weird, and returns the
+ # the number as a 10-digit character (including the leading 0)
+ # fishy = TRUE set invalid numbers equal to NA
+ if( is.numeric(x) )
+   {
+   if( any( wh <- (x<0101000000 | x>=3112999999), na.rm=TRUE ) )
+     warning( "\nSome fishy CPR-nos at positions:\n",
+              paste( which(wh), collapse=" " ), "\n namely:\n",
+              paste( x[wh], collapse="\n" ) )
+   xx <- formatC( x, format="f", width=10, digits=0, flag="0" )
+   }
+ else
+   {
+   if( any( wh <- (nchar(as.character(x))<10), na.rm=TRUE ) )
+     warning( "\nSome fishy CPR-nos at positions:\n",
+              paste( which(wh), collapse=" " ), "\n namely:\n",
+              paste( x[wh], collapse="\n" ) )
+   xx <- x
+   }
+ if( fishy ) xx[wh] <- NA
+ xx
+ }
> ###
> cpr2date <-
+ function( x, fishy=FALSE )
+ {
+ # Returns the birthdate from the CPR-number.
+ x <- check.cpr( x, fishy=fishy )
+ sixdg <-            substr(as.character(x),1,6)
+ seven <- as.numeric(substr(as.character(x),7,7))
+   yr  <- as.numeric(substr(sixdg,5,6))
+ eight <- paste( substr(sixdg,1,4),
+                 19 + ( (seven>3           & yr<37) |
+                        (seven>4 & seven<9 & yr<58) )
+                    - ( (seven>4 & seven<9 & yr>57) ),
+                 substr(sixdg,5,6), sep="" )
+ as.Date( eight, format="%d%m%Y" )
+ }
> ###
> cpr2sex <-
+ function( x, labels=c("M","F"), fishy=FALSE )
+ {
+ x <- check.cpr( x, fishy=fishy )
+ factor( 2 - (as.numeric(substr(x,10,10)) %% 2),
+         levels=1:2, labels=labels )
+ }
```

We then read the amputation data and tease out the sex and date of birth from the cpr-no.s:

```
> amp <- read.csv2( "./data/Amputationsdata version2.csv", header=TRUE )
> names( amp ) <- gsub( "_", ".", tolower( names(amp) ) )
> names(amp)[whd <- c(8,13)]
[1] "amputations.dato" "dod"
> names(amp)[8] <- "doa"
> for( i in whd ) amp[,i] <- cal.yr( amp[,i], format="%d-%m-%Y" )
> amp$cpr <- as.numeric( as.character(amp$cpr) )
> amp$dob <- cal.yr( cpr2date( amp$cpr ) )
> amp$sex <- cpr2sex( amp$cpr )
> table( is.na(amp$cpr) )
FALSE  TRUE
 3964     3
> subset( amp, is.na(cpr) )
     cpr antal   stat.amb alder ampu.kode       amputation grad      doa year  sex diabetes.type.dia
485   NA     1 Stationaer    41    KNGQ19 Knae og underben    2 1996.133 1996 <NA>
3006  NA     1 Stationaer    50    KNHQ14      Ankel og fod    1 2008.166 2008 <NA>
3524  NA     1 Stationaer    28    KNHQ07      Ankel og fod    1 2010.386 2010 <NA>
                hoejeste.grad diabetes.patient id.cpr                                   ampu.txt dob
485  Cpr\xb4s hoejeste grad  Ikke Diabetiker    306                      Amputation paa underben  NA
3006 Cpr\xb4s hoejeste grad  Ikke Diabetiker   1913                   Transmetatarsal amputation  NA
3524 Cpr\xb4s hoejeste grad  Ikke Diabetiker   2228 Eksartikulation af taa i interfalangealled  NA
> amp <- subset( amp, !is.na(cpr) )
> str( amp )
'data.frame':        3964 obs. of  18 variables:
 $ cpr                 : num  1.01e+08 1.01e+08 1.01e+08 1.01e+08 1.01e+08 ...
 $ antal               : int  1 1 1 1 1 1 1 1 1 1 ...
 $ stat.amb            : Factor w/ 2 levels "Ambulant","Stationaer": 2 2 2 2 2 2 2 2 2 2 ...
 $ alder               : int  75 71 71 61 63 63 63 60 62 73 ...
 $ ampu.kode           : Factor w/ 58 levels "B 81031","B 81040",..: 14 14 14 43 19 19 14 14 41 17 .
 $ amputation          : Factor w/ 3 levels "Ankel og fod",..: 2 2 2 1 3 3 2 2 1 3 ...
 $ grad                : int  3 3 3 1 2 2 3 3 1 2 ...
 $ doa                 : num  2001 2006 2006 1998 2000 ...
 $ year                : int  2001 2006 2006 1997 1999 1999 1999 1998 2011 1999 ...
 $ sex                 : Factor w/ 2 levels "M","F": 2 1 1 1 1 1 1 2 2 2 ...
 $ diabetes.type.dialog: Factor w/ 6 levels "","Anden","Auto Oprettet",..: 1 1 1 1 1 1 1 1 5 1 ...
 $ status              : Factor w/ 2 levels "Doed","Ikke doed": 1 1 1 1 1 1 1 1 2 1 ...
 $ dod                 : num  2001 2006 2006 2000 2000 ...
 $ hoejeste.grad       : Factor w/ 2 levels "","Cpr\xb4s hoejeste grad": 2 2 1 1 1 1 2 2 2 2 ...
 $ diabetes.patient    : Factor w/ 2 levels "Diabetiker","Ikke Diabetiker": 2 2 2 1 1 1 1 2 1 2 ...
 $ id.cpr              : int  1 2 2 3 3 3 3 4 5 6 ...
 $ ampu.txt            : Factor w/ 16 levels "","Amputation i ankelled a.m. Syme",..: 3 3 3 1 4 4 3
 $ dob                 : num  1926 1935 1935 1936 1936 ...
```

We now have a database of amputations on identifiable persons, assumed resident in Fyn at amputation.

In order to check the sanity of the amputation data we make a histogram of the date of birth, dates of amputation and the ages at amputation:

```
> par( mfrow=c(2,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( amp, hist( dob, col="blue", border="blue", breaks=1895+0:120, main="",
+       xlab="Date of birth", ylim=c(0,150) ) )
> with( amp, hist( doa-dob, col="blue", border="blue", breaks=0:120, main="",
+       xlab="Age at amputation", ylim=c(0,150) ) )
> abline( v=15, col="red" )
> with( amp, hist( doa, col="blue", border="blue", breaks=1985:2015, main="",
+       xlab="Date of amputation", ylim=c(0,400) ) )
> abline( v=c(1996,2012), col="red" )
> with( amp, hist( dod, col="blue", border="blue", breaks=1985:2015, main="",
+       xlab="Date of death", ylim=c(0,400) ) )
> abline( v=c(1996,2012), col="red" )
```

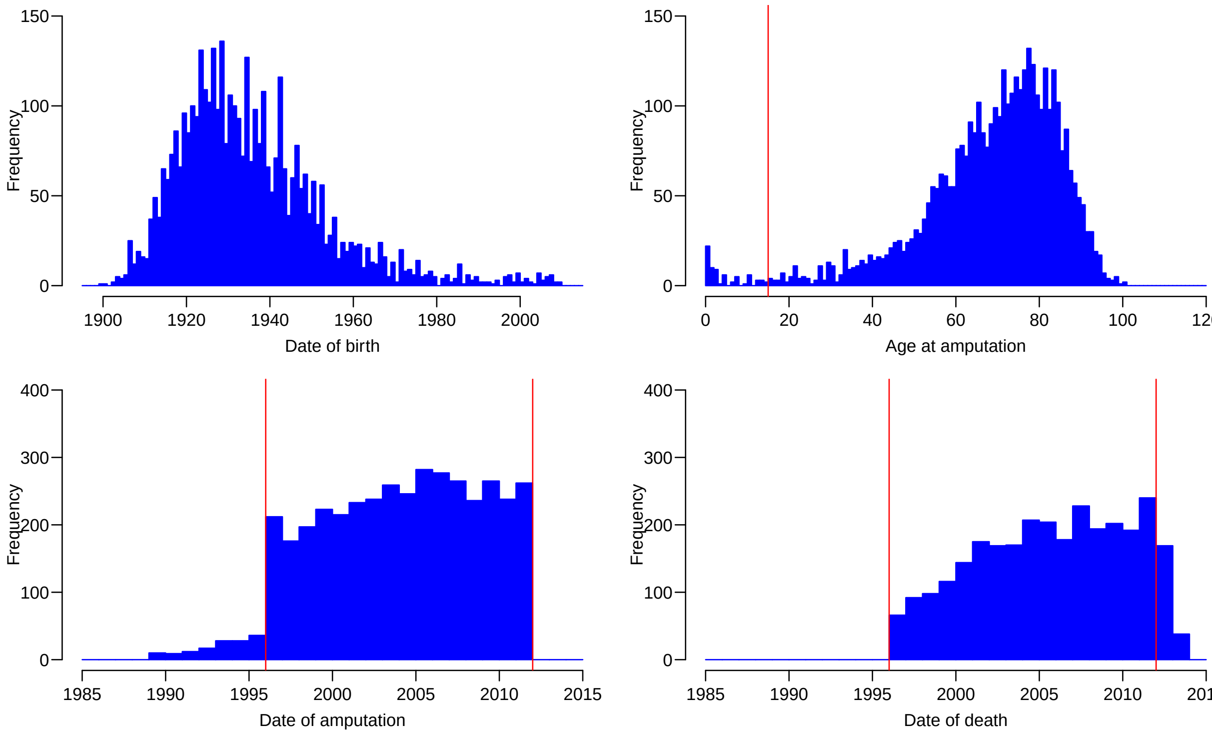

Figure 2.2: *Histograms of dates of birth and dates and ages at amputation in the Fyn amputation data.*

2.2.1 Calendar period of follow-up

From figure 2.2 we see that the amputation records can only be taken to be complete from 1996-01-01, through 2011-12-31, so this will be the calendar time window of follow-up in the study, and we shall not count deaths after this either.

2.2.2 Very young amputees

We note figure 2.2 that some very young people are amputated — is that possible? We make a list of those with amputations in ages under 5, and put in a file for inspection:

```
> write.csv2( transform( subset( amp, doa-dob < 5 ),
+                        doa = as.Date.cal.yr(doa),
+                        dob = as.Date.cal.yr(dob) ),
+             file="./data/under5.csv" )
```

Following scrutiny of these it appears that these are real amputations, many of congenital malformations such as extra toes etc. so we include these as “real” amputations only if they are after 15 year of age (note the dip in the histogram of amputation dates at this point):

```
> dim( amp )
[1] 3964   18
> amp <- subset( amp, doa-dob > 15 )
> dim( amp )
[1] 3894   18
```

2.2.3 Amputation status

In order to do the analysis of the amputation rates, and take into account previous amputations, we only consider the first amputation in each of the categories "foot", "knee" and "thigh".

We first provide an overview of the number of amputations per person:

```
> with( amp, addmargins( table(table(cpr)) ) )
   1    2    3    4    5    6    7    8    9   10   11  Sum
1585  549  173   85   30   18    8    1    1    1    1 2452
```

so some 60% of the persons are registered with one amputation, and 40% with at least 2, and 4 patients have more than 8 amputations recorded.

Of more interest is the classification of amputations by severity; we have the following distribution of amputations:

```
> options( width=90 )
> with( amp, cbind( addmargins( table(ampu.txt,amputation) ) ) )
                                                                     Ankel og fod
                                                                              429
Amputation i ankelled a.m. Syme                                                 4
Amputation paa laarben                                                          0
Amputation paa underben                                                         0
Anden amputations- eller eksartikulationssoperation paa ankel eller fod        15
Anden amputationsoperation paa hofte eller laar                                 0
Anden amputationsoperation paa knae eller underben                              0
Eksartikulation af taa i interfalangealled                                     66
Eksartikulation i hofteled                                                      0
Eksartikulation i knaeled                                                       0
Eksartikulation i talokruralled                                                 6
Intertarsal eksartikulation                                                     5
Metatarsofalangeal eksartikulation                                            120
Partiel amputation af taa                                                     296
Tarsometatarsal eksartikulation                                                36
Transmetatarsal amputation                                                    554
Sum                                                                          1531
                                                                     Hofte og laar
                                                                               25
Amputation i ankelled a.m. Syme                                                 0
Amputation paa laarben                                                       1124
Amputation paa underben                                                         0
Anden amputations- eller eksartikulationssoperation paa ankel eller fod         0
Anden amputationsoperation paa hofte eller laar                                14
Anden amputationsoperation paa knae eller underben                              0
Eksartikulation af taa i interfalangealled                                      0
Eksartikulation i hofteled                                                     50
Eksartikulation i knaeled                                                       0
Eksartikulation i talokruralled                                                 0
Intertarsal eksartikulation                                                     0
Metatarsofalangeal eksartikulation                                              0
Partiel amputation af taa                                                       0
Tarsometatarsal eksartikulation                                                 0
Transmetatarsal amputation                                                      0
Sum                                                                          1213
                                                                     Knae og underben
                                                                                  62
Amputation i ankelled a.m. Syme                                                    0
Amputation paa laarben                                                             0
Amputation paa underben                                                          736
Anden amputations- eller eksartikulationssoperation paa ankel eller fod            0
Anden amputationsoperation paa hofte eller laar                                    0
Anden amputationsoperation paa knae eller underben                                 3
Eksartikulation af taa i interfalangealled                                         0
```

```
Eksartikulation i hofteled                                                       0
Eksartikulation i knaeled                                                      349
Eksartikulation i talokruralled                                                  0
Intertarsal eksartikulation                                                      0
Metatarsofalangeal eksartikulation                                               0
Partiel amputation af taa                                                        0
Tarsometatarsal eksartikulation                                                  0
Transmetatarsal amputation                                                       0
Sum                                                                           1150
                                                                   Sum
                                                                   516
Amputation i ankelled a.m. Syme                                      4
Amputation paa laarben                                            1124
Amputation paa underben                                            736
Anden amputations- eller eksartikulationssoperation paa ankel eller fod  15
Anden amputationsoperation paa hofte eller laar                     14
Anden amputationsoperation paa knae eller underben                   3
Eksartikulation af taa i interfalangealled                          66
Eksartikulation i hofteled                                          50
Eksartikulation i knaeled                                          349
Eksartikulation i talokruralled                                      6
Intertarsal eksartikulation                                          5
Metatarsofalangeal eksartikulation                                 120
Partiel amputation af taa                                          296
Tarsometatarsal eksartikulation                                     36
Transmetatarsal amputation                                         554
Sum                                                               3894
```

We must have classification of the severity based on these classes, and we shall use:

$$\text{Ankel og fod} \ < \text{Knæ og underben} < \text{Hofte og lår}$$

For annotation of the states of amputation we define a new factor:

```
> amp <- transform( amp, agr = Relevel( amputation, list( Foot=1,
+                                                          Knee=3,
+                                                          Thig=2 ) ) )
> with( amp, table( amputation, agr ) )
                   agr
amputation          Foot Knee Thig
  Ankel og fod      1531    0    0
  Hofte og laar        0    0 1213
  Knae og underben     0 1150    0
```

However we only want the first of each of these types of amputations in the dataset, so we generate separate dates for each of these:

```
> amp <- transform( amp, doF = ifelse( agr=="Foot", doa, NA ),
+                        doK = ifelse( agr=="Knee", doa, NA ),
+                        doT = ifelse( agr=="Thig", doa, NA ),
+                        dod = ifelse( is.na(dod), Inf, dod ) )
```

Then we can use `aggregate` to generate a dataset with one record per amputee and dates for the first amputation of each category:

```
> amp1 <- aggregate( amp[,c("doF","doK","doT","dob","dod")],
+                    amp[,c("cpr"),drop=FALSE],
+                    FUN = function(x){ x<-min(x,na.rm=TRUE)
+                                       ifelse(x==Inf,NA,x) } )
> cbind( dim(amp), dim(amp1) )
     [,1] [,2]
[1,] 3894 2452
[2,]   22    6
```

We can get an overview of how many go through different amputations:

```
> with( amp1, print( ftable( addmargins( table(
+                                paste( ifelse(!is.na(doF),"F","-"),
+                                       ifelse(!is.na(doK),"K","-"),
+                                       ifelse(!is.na(doT),"T","-"),
+                                       sep="" ),
+                               "F<K"=doF<doK,
+                               "K<T"=doK<doT,
+                               "F<T"=doF<doT,
+                               useNA="ifany" ), 1 ),
+                          col.vars=1 ),
+                  zero="." ) )
                   F--  FK-  FKT  F-T  -K-  -KT  --T  Sum
F<K   K<T   F<T
FALSE FALSE FALSE    .    .    2    .    .    .    .    2
            TRUE     .    .    .    .    .    .    .    .
            NA       .    .    .    .    .    .    .    .
      TRUE  FALSE    .    .    7    .    .    .    .    7
            TRUE     .    .   10    .    .    .    .   10
            NA       .    .    .    .    .    .    .    .
      NA    FALSE    .    .    .    .    .    .    .    .
            TRUE     .    .    .    .    .    .    .    .
            NA       .   23    .    .    .    .    .   23
TRUE  FALSE FALSE    .    .    1    .    .    .    .    1
            TRUE     .    .    1    .    .    .    .    1
            NA       .    .    .    .    .    .    .    .
      TRUE  FALSE    .    .    .    .    .    .    .    .
            TRUE     .    .   50    .    .    .    .   50
            NA       .    .    .    .    .    .    .    .
      NA    FALSE    .    .    .    .    .    .    .    .
            TRUE     .    .    .    .    .    .    .    .
            NA       .  170    .    .    .    .    .  170
NA    FALSE FALSE    .    .    .    .    .    .    .    .
            TRUE     .    .    .    .    .    .    .    .
            NA       .    .    .    .    .   22    .   22
      TRUE  FALSE    .    .    .    .    .    .    .    .
            TRUE     .    .    .    .    .    .    .    .
            NA       .    .    .    .    .  183    .  183
      NA    FALSE    .    .    .    8    .    .    .    8
            TRUE     .    .    .   62    .    .    .   62
            NA     767    .    .    .  495    .  651 1913
```

We see that not all persons with more than one amputation have these in order of increasing severity, so we remove the dates of foot amputation that are after any of the more severe amputations, and the dates of knee amputations that are after thigh amputations:

```
> amp2 <- transform( amp1, doF = ifelse( doF<pmin(doK,doT,Inf,na.rm=TRUE), doF, NA ),
+                          doK = ifelse( doK<pmin(    doT,Inf,na.rm=TRUE), doK, NA ) )
> par( mfrow=c(2,3), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n" )
> with( amp1, plot( doF, doK, pch=16 ) )
> with( amp1, plot( doF, doT, pch=16 ) )
> with( amp1, plot( doK, doT, pch=16 ) )
> with( amp2, plot( doF, doK, pch=16 ) )
> with( amp2, plot( doF, doT, pch=16 ) )
> with( amp2, plot( doK, doT, pch=16 ) )
```

2.3 Merging diabetes and amputation data

Finally we merge the amputation dataset with the diabetes dataset:

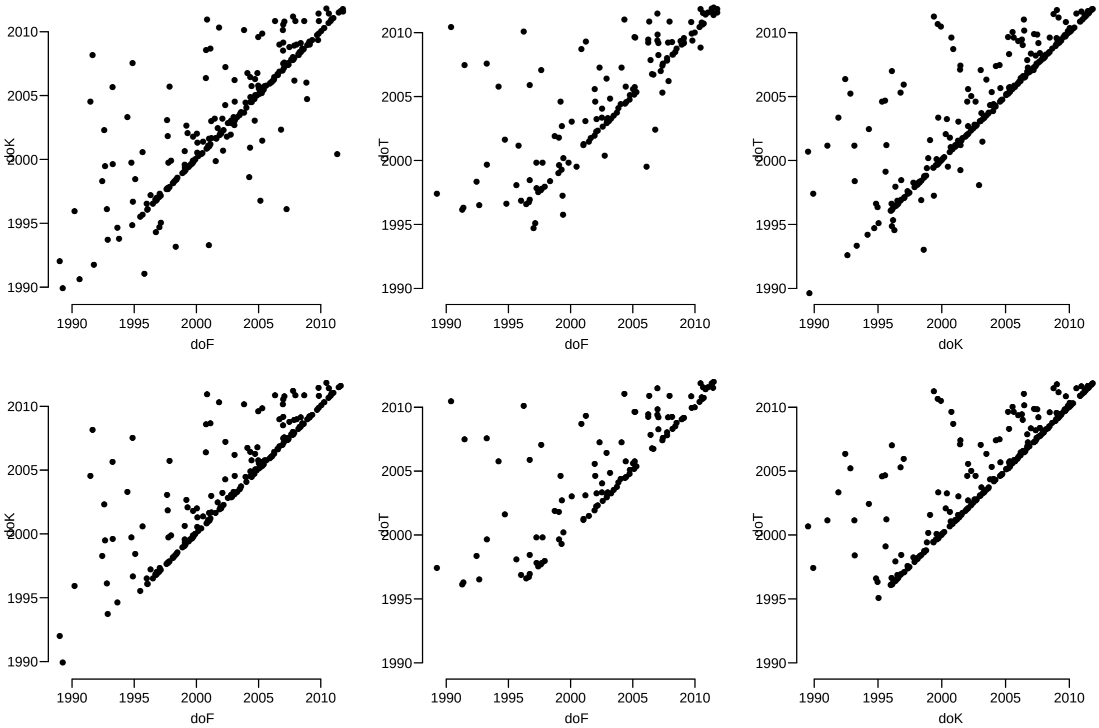

Figure 2.3: *Changes to the amputation dates, top row is the* `amp1` *data frame, bottom row the* `amp2`*.*

```
> load( file="./data/ndr.fyn.Rda" )
> intersect( names(amp2), names(ndr.fyn) )
[1] "cpr"
> str( ndr.fyn )
'data.frame':        49452 obs. of  16 variables:
 $ cpr       : num  1.01e+08 1.01e+08 1.01e+08 1.01e+08 1.01e+08 ...
 $ foddto    : num  1904 1906 1907 1907 1909 ...
 $ sex       : Factor w/ 2 levels "K","M": 1 1 1 2 1 1 2 1 1 2 ...
 $ inkldto   : num  1994 1994 1990 1991 1990 ...
 $ inklaarsag: Factor w/ 6 levels "blod2i5","blod5i1",..: 4 5 5 2 3 2 2 2 2 5 ...
 $ dodsdto   : num  1997 1996 1992 2003 2003 ...
 $ lpr       : num  NA 1994 1990 2003 1998 ...
 $ fodt      : num  NA NA NA NA 1990 ...
 $ blod2i5   : num  NA NA NA 1996 1997 ...
 $ blod5i1   : num  NA NA NA 1991 1991 ...
 $ ins       : num  1994 NA NA NA NA ...
 $ oad       : num  NA 1996 NA NA NA ...
 $ eFyn      : num  1996 1996 1996 1996 1996 ...
 $ xFyn      : num  1997 1996 1992 2003 2003 ...
 $ dox       : num  1997 1996 1992 2003 2003 ...
 $ doe       : num  1996 1996 1996 1996 1996 ...
> dmamp <- merge( amp2, ndr.fyn, all=TRUE )
```

We then define the entry and exit for all patients with DM or amputation. Note that we in this dataset only count follow-up after the earliest of a DM diagnosis or amputation, the risk time without both will subsequently be obtained from the population data by subtraction.

```
> dmamp <- transform( dmamp, sex = cpr2sex( cpr ),
+                            dod = pmin( dod, dodsdto, na.rm=TRUE ),
+                           doDM = inkldto )
```

We note that the residence records that comes with the NDR are truncated at the date of diabetes:

```
> with( dmamp, table( doDM > eFyn, useNA="ifany" ) )
FALSE  <NA>
49452  1321
```

hence we do not use the `eFyn` as entry variable, for those with a registered amputation we use their amputation date as the entry to the study, otherwise we would artificially exclude follow-up prior to DM among amputees. For the sake of convenience in the manipulation of Lexis objects to be constructed, we let persons that enter the cohort by way of amputation, enter 1 day prior to amputation, so that the amputation will be recorded as a transition, that is with a transition (`lex.Cst`≠`lex.Xst`) from "Well" to ampuation. Moreover, some dates of DM are identical to the dates of amputation; for these we put the date of diabetes *after* the date of amputation:

```
> dmamp <- transform( dmamp, doDM = ifelse( doDM==pmin(doF,doK,doT,Inf,na.rm=TRUE),
+                                           doDM+1/365,
+                                           doDM ) )
> dmamp <- transform( dmamp, dob = pmin( dob, foddto, na.rm=TRUE ),
+                            doe = pmax( 1996, # eFyn,
+                                        pmin( doDM,
+                                              doF - 1/365,
+                                              doK - 1/365,
+                                              doT - 1/365,
+                                              na.rm=TRUE ),
+                                        na.rm=TRUE ),
+                            dox = pmin( 2012, xFyn, dod, dox, na.rm=TRUE ) )
> names( dmamp )
 [1] "cpr"        "doF"        "doK"        "doT"        "dob"        "dod"
 [7] "foddto"     "sex"        "inkldto"    "inklaarsag" "dodsdto"    "lpr"
[13] "fodt"       "blod2i5"    "blod5i1"    "ins"        "oad"        "eFyn"
[19] "xFyn"       "dox"        "doe"        "doDM"
```

Before we save the dataset for further analysis we strip the person-ids and create bogus ids for person instead:

```
> dmamp <- dmamp[,-grep("cpr",names(dmamp))]
> str( dmamp )
'data.frame':        50773 obs. of  21 variables:
 $ doF       : num  NA NA NA NA NA NA NA NA NA NA ...
 $ doK       : num  NA NA NA NA NA NA NA NA NA NA ...
 $ doT       : num  NA NA NA NA NA NA NA NA NA NA ...
 $ dob       : num  1904 1906 1907 1907 1909 ...
 $ dod       : num  1997 1996 1992 2003 2003 ...
 $ foddto    : num  1904 1906 1907 1907 1909 ...
 $ sex       : Factor w/ 2 levels "M","F": 2 2 2 1 2 2 1 2 2 1 ...
 $ inkldto   : num  1994 1994 1990 1991 1990 ...
 $ inklaarsag: Factor w/ 6 levels "blod2i5","blod5i1",..: 4 5 5 2 3 2 2 2 2 5 ...
 $ dodsdto   : num  1997 1996 1992 2003 2003 ...
 $ lpr       : num  NA 1994 1990 2003 1998 ...
 $ fodt      : num  NA NA NA NA 1990 ...
 $ blod2i5   : num  NA NA NA 1996 1997 ...
 $ blod5i1   : num  NA NA NA 1991 1991 ...
 $ ins       : num  1994 NA NA NA NA ...
 $ oad       : num  NA 1996 NA NA NA ...
 $ eFyn      : num  1996 1996 1996 1996 1996 ...
```

```
 $ xFyn       : num  1997 1996 1992 2003 2003 ...
 $ dox        : num  1997 1996 1992 2003 2003 ...
 $ doe        : num  1996 1996 1996 1996 1996 ...
 $ doDM       : num  1994 1994 1990 1991 1990 ...
```

```
> summary( dmamp )
```

```
      doF             doK             doT             dob             dod       
 Min.   :1989    Min.   :1990    Min.   :1990    Min.   :1891   Min.   :1990   
 1st Qu.:2000    1st Qu.:2000    1st Qu.:2001    1st Qu.:1926   1st Qu.:1998   
 Median :2004    Median :2003    Median :2005    Median :1938   Median :2003   
 Mean   :2004    Mean   :2003    Mean   :2005    Mean   :1939   Mean   :2003   
 3rd Qu.:2008    3rd Qu.:2007    3rd Qu.:2009    3rd Qu.:1950   3rd Qu.:2008   
 Max.   :2012    Max.   :2012    Max.   :2012    Max.   :2010   Max.   :2013   
 NA's   :49723   NA's   :49835   NA's   :49776                  NA's   :31223  
     foddto      sex          inkldto        inklaarsag       dodsdto            lpr       
 Min.   :1891   M:26265   Min.   :1984   blod2i5:   54   Min.   :1990    Min.   :1984   
 1st Qu.:1926   F:24508   1st Qu.:1995   blod5i1:21897   1st Qu.:1998    1st Qu.:1995   
 Median :1939             Median :2002   fodt   : 3605   Median :2003    Median :2001   
 Mean   :1939             Mean   :2001   ins    :  595   Mean   :2003    Mean   :2001   
 3rd Qu.:1951             3rd Qu.:2008   lpr    :16272   3rd Qu.:2008    3rd Qu.:2006   
 Max.   :2010             Max.   :2012   oad    : 7029   Max.   :2012    Max.   :2012   
 NA's   :1321             NA's   :1321   NA's   : 1321   NA's   :32174   NA's   :19378  
      fodt          blod2i5         blod5i1           ins             oad       
 Min.   :1990    Min.   :1994    Min.   :1990    Min.   :1994    Min.   :1994   
 1st Qu.:1995    1st Qu.:1996    1st Qu.:1996    1st Qu.:1995    1st Qu.:1997   
 Median :2000    Median :2002    Median :2002    Median :2002    Median :2004   
 Mean   :2001    Mean   :2002    Mean   :2002    Mean   :2002    Mean   :2003   
 3rd Qu.:2005    3rd Qu.:2007    3rd Qu.:2008    3rd Qu.:2007    3rd Qu.:2008   
 Max.   :2012    Max.   :2012    Max.   :2012    Max.   :2012    Max.   :2012   
 NA's   :33361   NA's   :43001   NA's   :17814   NA's   :37672   NA's   :26316  
      eFyn           xFyn            dox             doe            doDM     
 Min.   :1996   Min.   :1990   Min.   :1990   Min.   :1996   Min.   :1984  
 1st Qu.:1996   1st Qu.:2005   1st Qu.:2005   1st Qu.:1996   1st Qu.:1995  
 Median :2002   Median :2012   Median :2012   Median :2002   Median :2002  
 Mean   :2002   Mean   :2008   Mean   :2008   Mean   :2002   Mean   :2001  
 3rd Qu.:2008   3rd Qu.:2012   3rd Qu.:2012   3rd Qu.:2008   3rd Qu.:2008  
 Max.   :2012   Max.   :2012   Max.   :2012   Max.   :2012   Max.   :2012  
 NA's   :1321   NA's   :1321                                 NA's   :1321  
```

```
> save( dmamp, file="./data/dmamp.Rda" )
```

2.4 The population data

In this chapter we read in the population data and create the follow-up in the Fyn population. The data files we read from are all extracted from Statistics Denmark's data bank.

Population data and deaths are in separate files, some in a somewhat clumsy format as extracted from the data-bank from statistics Denmark.

```
> options( width=120 )
> library( Epi )
> clear()
> print( sessionInfo(), l=F )
```

```
R version 3.2.1 (2015-06-18)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
```

```
[1] Epi_1.1.69

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7    MASS_7.3-42     parallel_3.2.1  survival_2.38-3 etm_0.6-2       splines_3.2.1
[8] lattice_0.20-31
```

2.4.1 Population

The population data are in 3 different files due to changes in the municipal changes in DK, but all have population figures in ages 0–125, that is 126 age-groups — the last ones largely empty, though. The filenames should be roughly self-explanatory in the following.

So we read the three files, reshape them and put them together, and finally we compute the risk time in Lexis triangles:

```
> b1 <- read.csv2( "./data/FynBef19962006.csv", header=FALSE, as.is=TRUE )
> str( b1 )
'data.frame':        252 obs. of  14 variables:
 $ V1 : logi  NA NA NA NA NA NA ...
 $ V2 : logi  NA NA NA NA NA NA ...
 $ V3 : int  0 1 2 3 4 5 6 7 8 9 ...
 $ V4 : int  3164 3140 3021 3136 2969 2996 2803 2717 2634 2723 ...
 $ V5 : int  3004 3215 3148 3044 3171 2987 2999 2831 2735 2660 ...
 $ V6 : int  3005 3005 3218 3161 3050 3167 2982 3008 2846 2734 ...
 $ V7 : int  2803 2974 2995 3235 3156 3064 3172 2983 3018 2847 ...
 $ V8 : int  2915 2815 2968 3001 3245 3174 3086 3185 3003 3027 ...
 $ V9 : int  2823 2889 2821 2977 3018 3268 3200 3113 3200 3029 ...
 $ V10: int  2713 2829 2900 2842 3001 3051 3282 3227 3113 3214 ...
 $ V11: int  2613 2742 2863 2927 2865 3018 3058 3300 3256 3136 ...
 $ V12: int  2653 2637 2773 2897 2926 2906 3036 3087 3314 3258 ...
 $ V13: int  2677 2677 2662 2799 2907 2942 2929 3034 3089 3343 ...
 $ V14: int  2695 2707 2708 2674 2796 2923 2949 2930 3031 3099 ...

> b1 <- transform( b1, sex = factor( rep(1:2,each=126), labels=c("M","F") ),
+                   A = as.numeric( V3 ) )
> B1 <- aggregate( b1[,4:14], b1[,c("sex","A")], FUN=sum )
> str( B1 )
'data.frame':        252 obs. of  13 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ V4 : int  3164 2917 3140 2964 3021 2873 3136 2985 2969 2904 ...
 $ V5 : int  3004 2760 3215 2935 3148 2985 3044 2907 3171 2993 ...
 $ V6 : int  3005 2879 3005 2773 3218 2958 3161 2977 3050 2918 ...
 $ V7 : int  2803 2616 2974 2921 2995 2791 3235 2970 3156 3003 ...
 $ V8 : int  2915 2765 2815 2635 2968 2940 3001 2808 3245 2990 ...
 $ V9 : int  2823 2835 2889 2785 2821 2678 2977 2951 3018 2828 ...
 $ V10: int  2713 2674 2829 2843 2900 2805 2842 2680 3001 2966 ...
 $ V11: int  2613 2494 2742 2718 2863 2864 2927 2807 2865 2690 ...
 $ V12: int  2653 2651 2637 2540 2773 2739 2897 2866 2926 2826 ...
 $ V13: int  2677 2621 2677 2678 2662 2563 2799 2768 2907 2893 ...
 $ V14: int  2695 2564 2707 2651 2708 2696 2674 2586 2796 2774 ...

> L1 <- reshape( B1, varying=3:13,
+                    times=1996:2006,
+                    v.names="N",
+                    timevar="P",
+                    direction="long" )
> str( L1 )
'data.frame':        2772 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : int  1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...
 $ N  : int  3164 2917 3140 2964 3021 2873 3136 2985 2969 2904 ...
```

```
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ N: chr  "V4" "V5" "V6" "V7" ...
  .. ..- attr(*, "v.names")= chr "N"
  .. ..- attr(*, "times")= int  1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 ...
  ..$ v.names: chr "N"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

```
> b2 <- read.csv2( "./data/FynBef20072009.csv", header=FALSE, as.is=TRUE )
> b2 <- transform( subset( b2, !is.na(V4) ),
+                  sex = factor( rep(1:2,each=1260), labels=c("M","F") ),
+                    A = as.numeric( V3 ) )
> str( b2 )
```

```
'data.frame':         2520 obs. of  8 variables:
 $ V1 : chr  " " " " " " " " ...
 $ V2 : chr  " " " " " " " " ...
 $ V3 : int  0 1 2 3 4 5 6 7 8 9 ...
 $ V4 : int  245 235 242 256 235 285 284 280 276 304 ...
 $ V5 : int  226 253 246 254 251 230 292 280 276 280 ...
 $ V6 : int  241 233 270 252 258 262 237 292 281 279 ...
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  0 1 2 3 4 5 6 7 8 9 ...
```

```
> B2 <- aggregate( b2[,4:6], b2[,c("sex","A")], FUN=sum )
> str( B2 )
```

```
'data.frame':         252 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ V4 : int  2749 2612 2735 2595 2727 2678 2745 2730 2676 2601 ...
 $ V5 : int  2834 2612 2811 2676 2754 2624 2752 2714 2752 2744 ...
 $ V6 : int  2904 2617 2863 2659 2840 2714 2782 2637 2788 2729 ...
```

```
> L2 <- reshape( B2, varying=3:5,
+                    times=2007:2009,
+                    v.names="N",
+                    timevar="P",
+                    direction="long" )
> str( L2 )
```

```
'data.frame':         756 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : int  2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
 $ N  : int  2749 2612 2735 2595 2727 2678 2745 2730 2676 2601 ...
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ N: chr  "V4" "V5" "V6"
  .. ..- attr(*, "v.names")= chr "N"
  .. ..- attr(*, "times")= int  2007 2008 2009
  ..$ v.names: chr "N"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

```
> b3 <- read.csv2( "./data/FynBef20102012.csv", header=FALSE, as.is=TRUE )
> b3 <- transform( subset( b3, !is.na(V4) ),
+                  sex = factor( rep(1:2,each=1260), labels=c("M","F") ),
+                    A = as.numeric( V3 ) )
> str( b3 )
```

```
'data.frame':         2520 obs. of  8 variables:
 $ V1 : chr  " " " " " " " " ...
 $ V2 : chr  " " " " " " " " ...
 $ V3 : int  0 1 2 3 4 5 6 7 8 9 ...
 $ V4 : int  216 251 252 279 252 267 271 233 288 283 ...
 $ V5 : int  193 228 245 253 280 260 262 269 231 285 ...
```

```
 $ V6 : int  188 202 230 248 257 280 264 264 265 234 ...
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  0 1 2 3 4 5 6 7 8 9 ...
```

```
> B3 <- aggregate( b3[,4:6], b3[,c("sex","A")], FUN=sum )
> str( B3 )
```

```
'data.frame':        252 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ V4 : int  2678 2526 2927 2662 2895 2675 2848 2720 2803 2652 ...
 $ V5 : int  2642 2435 2706 2569 2932 2666 2904 2697 2844 2706 ...
 $ V6 : int  2436 2308 2683 2463 2718 2548 2943 2690 2918 2696 ...
```

```
> L3 <- reshape( B3, varying=3:5,
+                    times=2010:2012,
+                    v.names="N",
+                    timevar="P",
+                    direction="long" )
> str( L3 )
```

```
'data.frame':        756 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ N  : int  2678 2526 2927 2662 2895 2675 2848 2720 2803 2652 ...
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ N: chr  "V4" "V5" "V6"
  .. ..- attr(*, "v.names")= chr "N"
  .. ..- attr(*, "times")= int  2010 2011 2012
  ..$ v.names: chr "N"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

```
> BF <- rbind( L1, L2, L3 )
> str( BF )
```

```
'data.frame':        4284 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : int  1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...
 $ N  : int  3164 2917 3140 2964 3021 2873 3136 2985 2969 2904 ...
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ N: chr  "V4" "V5" "V6" "V7" ...
  .. ..- attr(*, "v.names")= chr "N"
  .. ..- attr(*, "times")= int  1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 ...
  ..$ v.names: chr "N"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

```
> xtabs( N ~ P + sex, data=BF )
```

```
      sex
P           M      F
  1996 232450 238078
  1997 232883 238539
  1998 233059 238814
  1999 232976 238756
  2000 233040 238934
  2001 233089 238975
  2002 233377 239127
  2003 233903 239568
  2004 234604 240478
  2005 235469 241111
  2006 236562 241785
  2007 237768 242848
```

```
  2008 238692 243718
  2009 239936 244410
  2010 240144 244718
  2011 240192 244777
  2012 240498 244692
```

The data frame BF now contain the *number* of persons in Fyn at 1 January 1996–2012 in 1-year age-classes, separately for men and women. Based on this we then generate the risk time in Lexis triangles:

```
> YF <- rbind( cbind( N2Y( data=subset( BF, sex=="M" ) ), sex="M" ),
+              cbind( N2Y( data=subset( BF, sex=="F" ) ), sex="F" ) )
> head( YF )
          A        P        Y sex
1 0.3333333 1996.667 1493.500   M
2 1.3333333 1996.667 1599.000   M
3 2.3333333 1996.667 1572.667   M
4 3.3333333 1996.667 1518.167   M
5 4.3333333 1996.667 1579.667   M
6 5.3333333 1996.667 1490.500   M
```

However, we really only want the population in A-sets not Lexis triangles, since the death counts are in A-sets so we aggregate to this:

```
> YF <- aggregate( YF[,"Y"], cbind( sex = YF[,c("sex")],
+                                   floor( YF[,c("A","P")] ) ), FUN=sum )
> names( YF )[4] <- "Y"
```

So YF now has the population risk time in Fyn in ages 0–125 and for each of the years 1996–2011

```
> addmargins( xtabs( Y ~ P, data=YF ) )
P
     1996      1997      1998      1999      2000      2001      2002      2003      2004      2005
 470963.5  471645.2  471800.7  471847.8  472020.0  472281.7  472975.3  474264.8  475822.5  477453.5
     2008      2009      2010      2011       Sum
 483365.3  484592.7  484903.7  485068.0 7629966.3
```

2.4.2 Deaths

The death figures are a bit more tricky, as they only are available in 1, 4 and then 5-year classes for the period prior to 2006, but in 1-year classes (and by municipality) for the years 2006–2012 (but we skip the last year because we do not need it):

```
> m1 <- read.csv2( "./data/FynDod19962005.csv", header=FALSE, as.is=TRUE, skip=0 )
> str( m1 )
'data.frame':        38 obs. of  14 variables:
 $ V1 : chr  "I ALT" "I ALT" "I ALT" "I ALT" ...
 $ V2 : chr  "Fyns Amt" "Fyns Amt" "Fyns Amt" "Fyns Amt" ...
 $ V3 : chr  "M\xe6nd" "M\xe6nd" "M\xe6nd" "M\xe6nd" ...
 $ V4 : chr  "0" "1-4" "5-9" "10-14" ...
 $ V5 : int  22 1 2 6 8 16 11 27 36 56 ...
 $ V6 : int  20 2 2 5 11 17 21 26 30 54 ...
 $ V7 : int  14 3 5 1 6 9 20 28 26 63 ...
 $ V8 : int  13 4 2 3 14 14 22 24 22 55 ...
 $ V9 : int  18 3 3 4 5 15 17 18 26 35 ...
 $ V10: int  13 6 0 7 7 10 5 16 21 29 ...
 $ V11: int  15 4 4 3 8 14 11 28 45 45 ...
 $ V12: int  13 4 1 5 10 6 32 20 28 45 ...
 $ V13: int  16 0 1 4 6 12 11 20 30 42 ...
 $ V14: int  12 3 3 2 8 32 11 11 20 37 ...
```

```
> head( m1 )

      V1        V2       V3    V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14
1 I ALT Fyns Amt M\xe6nd     0 22 20 14 13 18  13  15  13  16  12
2 I ALT Fyns Amt M\xe6nd   1-4  1  2  3  4  3   6   4   4   0   3
3 I ALT Fyns Amt M\xe6nd   5-9  2  2  5  2  3   0   4   1   1   3
4 I ALT Fyns Amt M\xe6nd 10-14  6  5  1  3  4   7   3   5   4   2
5 I ALT Fyns Amt M\xe6nd 15-19  8 11  6 14  5   7   8  10   6   8
6 I ALT Fyns Amt M\xe6nd 20-24 16 17  9 14 15  10  14   6  12  32

> m2 <- read.csv2( "./data/FynDod20062012.csv", header=FALSE, as.is=TRUE, skip=4 )
> str( m2 )

'data.frame':        2021 obs. of  10 variables:
 $ V1 : chr  " " " " " " " " " " ...
 $ V2 : chr  "Assens" " " " " " " " " ...
 $ V3 : chr  "" "0" "1" "2" ...
 $ V4 : int  NA 3 1 0 0 0 0 0 0 0 ...
 $ V5 : int  NA 1 0 0 0 0 0 0 0 1 ...
 $ V6 : int  NA 1 0 1 0 0 0 0 1 0 ...
 $ V7 : int  NA 0 0 0 0 0 0 0 0 0 ...
 $ V8 : int  NA 2 0 0 0 0 0 0 0 0 ...
 $ V9 : int  NA 1 0 0 0 0 0 0 0 0 ...
 $ V10: int  NA 0 0 0 0 0 0 0 0 0 ...

> head( m2 )

  V1     V2 V3 V4 V5 V6 V7 V8 V9 V10
1    Assens    NA NA NA NA NA NA  NA
2           0  3  1  1  0  2  1   0
3           1  1  0  0  0  0  0   0
4           2  0  0  1  0  0  0   0
5           3  0  0  0  0  0  0   0
6           4  0  0  0  0  0  0   0

> m2 <- transform( subset( m2, !is.na(V4) ),
+                  sex = factor( rep(1:2,each=1000), labels=c("M","F") ),
+                    A = as.numeric( V3 ) )
> M2 <- aggregate( m2[,4:9], m2[,c("sex","A")], FUN=sum )
> M2 <- reshape( M2, varying=3:8,
+                    times=2006:2011,
+                    v.names="D",
+                    timevar="P",
+                    direction="long" )
> str( M2 )

'data.frame':        1194 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : int  2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
 $ D  : int  10 11 2 0 0 0 0 0 0 0 ...
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ D: chr  "V4" "V5" "V6" "V7" ...
  .. ..- attr(*, "v.names")= chr "D"
  .. ..- attr(*, "times")= int  2006 2007 2008 2009 2010 2011
  ..$ v.names: chr "D"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

So now we have the number of deaths in Fyn for the period 1996 through 2012 in `m1`, but the deaths are only available in irregular age-classes for the years 1996–2005 incl. In order to get a usable dataset for the mortality we redistribute the deaths in these age-classes into the single-year age-classes according to the empirical distribution of deaths within these age-classes in the years 2006–2012. This is done as follows:

Let x be the vector of number of deaths in a given year in the age classes 0, 1–4, 5-9, ..., 85+ — a vector of length 19. We really want a vector of length 100, with number of deaths in the 100 age-classes 0,1,2,...,99+. This is merely a question of pre-multiplication with a 100×19 matrix, where each column correspond to one of the 19 age-classes 55–59, say, and where the 5 entries corresponding to the 1-year classes 55 through 59 in this contains the relative distribution of deaths in age classes 55 to 59 as derived from the data for 2006–2012. The machinery needed is really twice as big as we need to do it for males and females separately, so we need a block matrix with diagonal blocks of size 100×19 and 0s outside of these.

First we take the marginal age-distribution of deaths in single-year age-classes from the years 2006–12, and devise a small function to construct the matrix:

```
> T200 <- xtabs( D ~ A + sex, data=M2 )
> Mmake <-
+ function( V )
+    {
+ v100 <- c( V[1  ], rep(0,100),
+            V[2:5], rep(0,100) )
+ for( i in 1:16 ) v100 <- c( v100, V[(i*5)+1:5], rep(0,100) )
+ v100 <- c( v100, V[86:100] )
+ M <- matrix( v100, 100, 19 )
+ sweep( M, 2, apply(M,2,sum), "/" )
+    }
> Mr <- Mmake( T200[,"M"] )
> Fr <- Mmake( T200[,"F"] )
```

We then multiply the *number* of deaths in the years 1996–2005 with the appropriate matrix to get the deaths distributed across 1-year age-classes:

```
> str( m1 )
'data.frame':        38 obs. of  14 variables:
 $ V1 : chr  "I ALT" "I ALT" "I ALT" "I ALT" ...
 $ V2 : chr  "Fyns Amt" "Fyns Amt" "Fyns Amt" "Fyns Amt" ...
 $ V3 : chr  "M\xe6nd" "M\xe6nd" "M\xe6nd" "M\xe6nd" ...
 $ V4 : chr  "0" "1-4" "5-9" "10-14" ...
 $ V5 : int  22 1 2 6 8 16 11 27 36 56 ...
 $ V6 : int  20 2 2 5 11 17 21 26 30 54 ...
 $ V7 : int  14 3 5 1 6 9 20 28 26 63 ...
 $ V8 : int  13 4 2 3 14 14 22 24 22 55 ...
 $ V9 : int  18 3 3 4 5 15 17 18 26 35 ...
 $ V10: int  13 6 0 7 7 10 5 16 21 29 ...
 $ V11: int  15 4 4 3 8 14 11 28 45 45 ...
 $ V12: int  13 4 1 5 10 6 32 20 28 45 ...
 $ V13: int  16 0 1 4 6 12 11 20 30 42 ...
 $ V14: int  12 3 3 2 8 32 11 11 20 37 ...
> Mpre <- rbind( cbind( Mr   , Mr*0 ),
+                cbind( Fr*0, Fr   ) )
> # Check:
> apply( Mpre, 2, sum )
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
> m1.r <- data.frame( round( Mpre %*% as.matrix( m1[,5:14] ) ),
+                     A = rep(0:99,2),
+                   sex = factor( rep(1:2,each=100),
+                                 labels=c("M","F") ) )
> str( m1.r )
'data.frame':        200 obs. of  12 variables:
 $ V5 : num  22 0 0 0 0 0 0 0 1 0 ...
 $ V6 : num  20 1 1 0 0 0 0 0 1 0 ...
 $ V7 : num  14 1 1 0 0 0 1 1 2 1 ...
```

```
 $ V8 : num  13 2 2 0 0 0 0 0 1 0 ...
 $ V9 : num  18 1 1 0 0 0 1 1 1 1 ...
 $ V10: num  13 3 2 0 0 0 0 0 0 0 ...
 $ V11: num  15 2 2 0 0 0 1 1 1 1 ...
 $ V12: num  13 2 2 0 0 0 0 0 0 0 ...
 $ V13: num  16 0 0 0 0 0 0 0 0 0 ...
 $ V14: num  12 1 1 0 0 0 1 1 1 1 ...
 $ A  : int  0 1 2 3 4 5 6 7 8 9 ...
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
```

Finally we can now also put the deaths in the years 1996–2005 in the long form

```
> M1 <- reshape( m1.r, varying=1:10,
+                      times=1996:2005,
+                      v.names="D",
+                      timevar="P",
+                      direction="long" )
> str( M1 )
'data.frame':         2000 obs. of  5 variables:
 $ A  : int  0 1 2 3 4 5 6 7 8 9 ...
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ P  : int  1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...
 $ D  : num  22 0 0 0 0 0 0 0 1 0 ...
 $ id : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ D: chr  "V5" "V6" "V7" "V8" ...
  .. ..- attr(*, "v.names")= chr "D"
  .. ..- attr(*, "times")= int  1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
  ..$ v.names: chr "D"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "P"
```

— and combine with the deaths from the years 2006–2011:

```
> DF <- data.frame( rbind( M1, M2 )[,c("sex","A","P","D")] )
> str( DF )
'data.frame':         3194 obs. of  4 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  0 1 2 3 4 5 6 7 8 9 ...
 $ P  : int  1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...
 $ D  : num  22 0 0 0 0 0 0 0 1 0 ...
> xtabs( D ~ A + P, data = DF )
    P
A    1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
  0    35   37   21   24   33   32   21   19   24   24   21   21   35   22   23   19
  1     1    3    4    2    2    4    4    3    3    4    2    2    1    3    2    1
  2     1    2    2    2    2    2    3    3    1    2    0    0    2    2    3    0
  3     0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    1
  4     0    0    0    0    0    0    0    0    0    0    0    0    0    1    0    0
  5     1    0    0    0    0    1    1    0    1    1    1    0    0    0    0    2
  6     0    0    1    0    1    0    1    0    0    1    1    1    0    1    0    0
  7     1    0    1    0    1    1    2    0    1    2    2    0    1    0    0    1
  8     1    1    2    1    1    0    1    0    0    1    0    1    2    1    0    0
  9     0    0    1    0    1    0    1    0    0    1    1    0    0    0    1    1
  10    0    0    0    0    0    1    0    0    0    0    1    0    0    0    0    0
  11    2    1    0    1    1    2    1    1    1    0    1    0    1    0    1    1
  12    2    1    0    1    1    2    1    1    1    0    2    0    0    1    0    1
  13    4    2    1    2    3    3    1    2    3    1    2    1    2    1    1    0
  14    2    0    1    1    1    1    0    0    1    0    1    2    0    0    0    0
  15    2    2    0    1    0    1    1    1    1    1    3    0    1    1    3    0
  16    3    3    1    3    1    1    2    2    2    2    2    2    2    5    1    2
  17    3    3    1    3    1    2    2    2    2    2    2    3    1    2    1    4
  18    3    4    2    4    2    2    2    3    3    2    4    4    5    3    2    2
```

19	3	3	2	2	2	2	2	2	3	1	1	0	3	3	4	3
20	5	5	2	3	4	3	4	2	4	9	1	3	3	1	4	3
21	5	5	2	3	4	3	5	2	4	9	5	1	3	1	4	2
22	3	3	1	2	2	2	3	2	3	5	4	0	2	3	0	0
23	4	4	2	3	3	3	4	2	3	8	1	4	1	1	2	4
24	5	4	3	3	4	3	5	3	4	8	2	4	4	3	1	1
25	2	5	4	4	3	1	3	7	2	2	3	1	1	4	1	2
26	4	7	6	6	5	2	5	10	4	3	4	2	3	4	4	3
27	2	5	3	3	3	2	3	5	1	1	3	2	1	3	2	1
28	2	5	5	4	4	1	3	7	2	2	2	4	2	2	3	1
29	4	7	6	6	6	2	6	10	4	3	2	7	1	7	2	2
30	9	8	9	9	6	6	8	7	6	4	8	3	5	3	5	5
31	8	8	8	7	6	5	8	6	5	4	5	6	2	4	4	3
32	6	5	6	6	5	4	6	4	4	3	7	1	6	2	2	2
33	6	5	5	6	5	3	5	4	4	3	3	4	5	2	3	2
34	8	7	7	7	6	5	6	6	5	3	3	5	5	4	4	5
35	11	9	9	9	8	7	11	8	8	7	6	9	5	6	5	6
36	10	8	7	7	7	6	11	8	8	6	3	5	10	5	5	8
37	11	8	7	8	7	6	11	8	8	6	2	5	6	8	8	8
38	10	7	7	8	7	7	11	7	7	5	11	4	3	6	7	3
39	16	11	11	11	11	10	16	11	11	9	8	10	4	8	14	8
40	16	15	17	14	12	9	11	11	13	12	12	7	9	6	12	10
41	12	12	13	11	9	7	10	10	10	9	12	13	5	7	4	5
42	21	20	24	19	15	12	16	16	17	16	13	13	16	21	8	9
43	16	16	18	15	12	9	12	12	13	13	10	11	10	11	6	12
44	22	22	24	21	16	13	17	17	17	16	11	13	11	12	22	15
45	17	16	12	16	16	16	14	16	15	16	7	15	14	13	17	13
46	22	20	16	21	20	21	18	21	19	20	13	16	11	13	26	23
47	23	21	16	21	21	21	19	21	20	21	15	17	25	17	9	22
48	24	21	17	23	22	21	20	22	20	22	18	15	21	17	20	18
49	32	29	24	30	29	29	27	29	28	29	27	26	27	18	19	29
50	27	34	34	33	32	31	32	36	31	26	28	35	28	20	28	21
51	23	28	28	27	26	26	27	29	26	21	17	32	22	16	21	26
52	31	37	37	36	35	35	34	39	34	29	37	28	29	30	28	24
53	37	44	44	44	42	42	41	47	41	35	38	45	29	34	31	33
54	37	45	46	44	42	42	42	47	41	36	37	45	31	37	38	27
55	34	36	35	41	43	44	43	46	42	39	28	44	38	60	41	35
56	37	38	38	44	47	47	47	50	45	43	38	45	49	41	48	44
57	38	38	38	44	46	47	45	48	44	42	49	48	38	32	54	42
58	43	44	44	50	53	54	53	56	51	49	55	52	45	63	33	56
59	46	47	46	53	57	58	57	60	54	52	61	49	49	56	52	54
60	59	55	62	56	58	54	54	56	57	55	79	69	67	57	57	51
61	59	54	62	56	58	53	53	55	56	55	57	76	57	80	52	53
62	59	55	62	56	58	53	55	57	57	55	75	69	66	73	50	47
63	68	64	73	66	68	63	64	66	66	65	72	79	78	85	68	63
64	70	66	75	68	70	65	66	68	69	66	60	77	78	87	83	77
65	94	86	91	85	82	75	72	73	70	71	60	68	73	85	83	89
66	98	89	95	89	85	78	74	76	71	73	72	74	67	83	90	86
67	97	88	94	88	85	77	74	75	71	73	60	53	74	96	92	92
68	102	93	99	92	90	81	77	80	75	77	79	88	85	73	80	90
69	105	96	102	95	92	83	80	82	78	79	83	93	87	77	92	75
70	120	114	107	107	102	98	98	91	91	87	89	88	87	98	82	95
71	132	126	117	118	113	108	108	101	100	96	96	121	99	86	100	91
72	149	142	133	133	128	122	122	114	113	107	121	125	109	100	100	116
73	149	143	134	134	128	122	123	114	114	108	128	105	104	107	130	99
74	149	144	134	134	128	122	123	115	113	108	113	98	120	118	114	111
75	166	148	145	152	138	143	146	126	131	130	134	112	129	133	123	97
76	170	153	149	156	143	146	150	129	135	134	128	110	128	145	111	126
77	190	170	166	173	158	163	166	143	150	149	158	129	151	129	130	136
78	199	179	174	182	166	170	174	150	157	155	149	156	138	153	148	127
79	181	162	158	165	150	155	158	136	142	142	138	140	122	119	128	145
80	192	181	177	180	173	181	183	168	181	164	157	172	151	181	165	149
81	186	176	172	175	169	176	178	164	176	159	164	162	154	151	173	144
82	189	178	175	178	171	179	181	165	178	162	141	164	167	174	169	146
83	205	194	190	193	186	194	196	180	194	176	176	198	176	192	159	143
84	207	196	192	196	189	197	200	183	195	179	170	181	174	188	154	189

```
  85  171  175  166  185  174  180  190  186  182  181  212  181  173  175  170  154
  86  171  175  166  185  174  179  191  185  181  181  185  181  167  193  170  168
  87  159  163  154  172  162  168  178  173  169  169  137  194  168  169  176  148
  88  152  157  149  166  156  161  171  166  163  163  154  169  151  173  146  162
  89  152  158  149  165  157  161  171  167  163  164  192  161  149  155  146  153
  90  139  143  135  151  143  146  157  152  147  149  149  144  144  126  140  166
  91  118  122  116  128  121  126  133  129  126  127  135  119  115  135  121  117
  92  111  114  109  120  114  118  125  121  118  119  116  113  110  119  115  121
  93   97   99   94  105   99  103  109  106  103  104   93   99   93   99  103  119
  94   81   84   79   89   84   86   92   89   87   89   80   80   85  101   74   92
  95   67   69   66   73   69   70   75   73   71   73   61   67   80   62   69   80
  96   54   57   54   60   57   58   63   60   59   60   55   49   66   57   53   65
  97   40   41   39   43   41   43   45   43   42   43   32   38   36   40   50   54
  98   30   31   30   33   31   32   34   33   32   33   28   35   28   34   29   35
  99    2    2    2    2    2    2    2    2    2    2    3    1    2    3    0    3
```

2.4.3 Merging deaths and PY

Finally we can merge the two datasets, but only for the ages up to 99; the deaths in ages 99 and above are pooled, and will not be in the analysis:

```
> FUfyn <- merge( subset( DF, A<99 & A>14 ), YF, all.x=TRUE )
> names( FUfyn )[match(c("D","Y"),names(FUfyn))] <- c("D.all","Y.all")
> str( FUfyn )

'data.frame':        2688 obs. of  5 variables:
 $ sex  : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
 $ A    : num  15 15 15 15 15 15 15 15 15 15 ...
 $ P    : int  1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 ...
 $ D.all: num  1 1 0 0 0 0 0 0 1 0 ...
 $ Y.all: num  2580 2409 2335 2386 2481 ...

> with( FUfyn, table(A) )

A
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 4
32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 3
55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 8
32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 3
95 96 97 98
32 32 32 32

> with( FUfyn, table(P) )

P
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
 168  168  168  168  168  168  168  168  168  168  168  168  168  168  168  168

> save( FUfyn, file="./data/FUfyn.Rda" )
```

Thus `FUfyn` now have the deaths and person-time for the *entire* Fyn population for the years 1996–2011, in age classes 15–98, incl.:

```
> table( FUfyn$A )

15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 4
32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 3
55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 8
32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 3
95 96 97 98
32 32 32 32

> round( addmargins( xtabs( cbind(D.all,Y.all) ~ P, data=FUfyn ), 1 ) )
```

```
P         D.all   Y.all
  1996     5640  388825
  1997     5464  388228
  1998     5331  387326
  1999     5537  386395
  2000     5277  385544
  2001     5262  385029
  2002     5437  385273
  2003     5212  386207
  2004     5172  387568
  2005     5034  389394
  2006     5009  391790
  2007     5095  394211
  2008     4897  396466
  2009     5123  398277
  2010     4903  399690
  2011     4869  401087
  Sum     83262 6251310
```

2.5 Data set with individual follow-up

2.5.1 Overview

The cohort data contains all persons with residence in Fyn who either had a diagnosis of diabetes and/or were amputated. The goal is to classify follow-up by DM status, amputation status (note that this has several (n) categories!), sex and age and date of follow-up. This will of course only have three ($2n-1$) of the possible 4 ($2n$) combinations of DM and amputation status (missing those without any of them).

Subtracting the total follow-up[1] in these classes from the total population follow-up in Fyn will give the follow-up among persons without DM and amputation. We can use the database to tabulate the relevant events (DM, amputations and deaths) in this part of the population.

Combination will then give the total follow-up time classified by diabetes status, amputation status, sex, age and calendar time. The amputation events will then be classified accordingly, and we can then analyze amputation rates and mortality rates by these variables. The major focus being on the effect of diabetes, calendar time and how the diabetes effect possibly changes over calendar time.

To this end we first set up a `Lexis` object for the follow-up through states:

2.5.2 A `Lexis` object, for the cohort

The ultimate goal is to follow the entire Fyn population through diabetes and amputation, so we set up a Lexis object where we for a start let everyone enter alive and exit at death. Note that we use a numerical difference of less than 3 days (0.01 year) to ascertain a death at the end of follow-up. This is because many of the status-dates for exit are just a few days before date of death, which is clearly wrong. Moreover we follow persons from their 15th birthday and censor them at their 99th birthday, because this is the extent of the mortality and population data.

[1] Note that we here use the term “follow-up” to include *both* the follow-up `time` and the *deaths* occurring during this time.

```
> library( Epi )
> clear()
> load( file="./data/dmamp.Rda" )
> dmamp <- transform( dmamp, dox = pmin( dox, dob+99 ),
+                            doe = pmax( doe, dob+15 ) )
> Lx <- Lexis( entry = list( per = doe,
+                            age = doe-dob ),
+               exit = list( per = dox ),
+        exit.status = factor( (pmin(dod,Inf,na.rm=TRUE)-dox)<0.01,
+                             labels=c("Well","Dead") ),
+               data = subset(dmamp,dox>doe) )
NOTE: entry.status has been set to "Well" for all.
```

We are interested in the transitions between amputation states separately between the diabetes and non-diabetes states, so we first cut the follow-up of all persons at the diabetes-diagnosis, and then *separately* cut the follow-up in the subsets with and without diabetes at the amputation dates.

We then cut the follow-up into two different Lexis objects, one pre- and one post-DM. Note that it is only here that we can define the duration of diabetes; it will be `NA` for follow-up before diagnosis of diabetes:

```
> LxD <- cutLexis( Lx, cut = Lx$doDM,
+                 new.state = "DM",
+                 new.scale = "dur",
+                 precursor = "Well" )
> summary( LxD )
Transitions:
     To
From   Well    DM  Dead  Records:  Events: Risk time:  Persons:
  Well  492    93   827      1412      920    4925.47      1412
  DM      0 30777 14745     45522    14745  287366.85     45522
  Sum   492 30870 15572     46934    15665  292292.32     46841

> summary( LxD$dur )

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0000  0.0000  0.0000  0.8366  0.0000 15.0000    1412

> La <- subset( LxD, lex.Cst=="Well")
> Ld <- subset( LxD, lex.Cst=="DM")
> summary( LxD ) ; summary( La ) ; summary( Ld )

Transitions:
     To
From   Well    DM  Dead  Records:  Events: Risk time:  Persons:
  Well  492    93   827      1412      920    4925.47      1412
  DM      0 30777 14745     45522    14745  287366.85     45522
  Sum   492 30870 15572     46934    15665  292292.32     46841

Transitions:
     To
From   Well DM Dead  Records:  Events: Risk time:  Persons:
  Well  492 93  827      1412      920    4925.47      1412

Transitions:
     To
From Well    DM  Dead  Records:  Events: Risk time:  Persons:
  DM    0 30777 14745     45522    14745   287366.8     45522
```

These are now separately cut at the amputation dates to keep track of the amputations separately for persons with and without diabetes. Moreover, we also include new timescales in order to be able to evaluate the effect of time since ampuation on mortality:

```
> LaF <- cutLexis( La , cut = La$doF,
+                  new.state = "BAA",
+                  new.scale = "tsF",
+                  precursor = "Well" )
> LaK <- cutLexis( LaF, cut = LaF$doK,
+                  new.state = "BKA",
+                  new.scale = "tsK",
+                  precursor = c("Well","BAA") )
> LaT <- cutLexis( LaK, cut = LaK$doT,
+                  new.state = "AKA",
+                  new.scale = "tsT",
+                  precursor = c("Well","BAA","BKA") )
> summary( LaT )
Transitions:
     To
From   Well BAA BKA AKA DM Dead  Records:  Events: Risk time:  Persons:
  Well    0 505 396 447 13    7      1368     1368       3.71      1368
  BAA     0 241  46  33 35  168       523      282    2306.51       523
  BKA     0   0 113 120 24  199       456      343    1276.73       456
  AKA     0   0   0 138 21  453       612      474    1338.52       612
  Sum     0 746 555 738 93  827      2959     2467    4925.47      1412
```

We see that some of the amputees moves on to diabetes, that is the amputations were prior to DM diagnosis. Note that the persons' follow-up after diagnosis of DM is not in the this (non-diabetic) part of the database (`LaT`). By that token the DM state should be renamed according to the state *from* which it occurs, with a "`(DM)`" appended. The state these persons is in is not `DM`, but a state of amputated with DM. So by renaming the states this way we choose not to distinguish whether or not the amputation has occurred before or after the diagnosis of diabetes.

Also note that by doing the updating of the state "by hand" this way we keep the values of the timescales `tsF`, `tsK` and `tsT`, as the times since the amputation:

```
> LaT <- transform( LaT,
+                lex.Xst = factor( ifelse( lex.Xst=="DM" & lex.Cst!="Well",
+                                          paste(as.character(lex.Cst),
+                                                "(DM)",
+                                                sep=""),
+                                          as.character(lex.Xst) ) ) )
> LaT <- Relevel( LaT )
> summary( LaT )
Transitions:
     To
From   Well BAA BKA AKA DM Dead AKA(DM) BAA(DM) BKA(DM)  Records:  Events: Risk time:  Persons:
  Well    0 505 396 447 13    7       0       0       0      1368     1368       3.71      1368
  BAA     0 241  46  33  0  168       0      35       0       523      282    2306.51       523
  BKA     0   0 113 120  0  199       0       0      24       456      343    1276.73       456
  AKA     0   0   0 138  0  453      21       0       0       612      474    1338.52       612
  Sum     0 746 555 738 13  827      21      35      24      2959     2467    4925.47      1412
```

We then cut the follow-up in the diabetic state similarly, making sure that the level names are the same as just defined above, thereby putting persons with diabetes *after* a given type of amputation in the same box as those with the same amputation subsequent to diabetes:

```
> LdF <- cutLexis( Ld , cut = Ld$doF,
+                  new.state = "BAA(DM)",
+                  new.scale = "tsF",
+                  precursor = c("Well","DM") )
> LdK <- cutLexis( LdF, cut = LdF$doK,
+                  new.state = "BKA(DM)",
```

```
+                  new.scale = "tsK",
+                  precursor = c("Well","DM","BAA(DM)") )
> LdT <- cutLexis( LdK, cut = LdK$doT,
+                  new.state = "AKA(DM)",
+                  new.scale = "tsT",
+                  precursor = c("Well","DM","BAA(DM)","BKA(DM)") )
> summary( Ld ) ; summary( LdT )
Transitions:
     To
From Well    DM  Dead  Records:  Events: Risk time:  Persons:
  DM    0 30777 14745     45522    14745   287366.8     45522
Transitions:
     To
From      Well    DM BAA(DM) BKA(DM) AKA(DM)  Dead  Records:  Events: Risk time:  Persons:
  DM         0 30468     483     280     206 13959     45396    14928  284146.23     45396
  BAA(DM)    0     0     143     162      28   207       540      397    1336.68       540
  BKA(DM)    0     0       0      83     126   276       485      402    1220.48       485
  AKA(DM)    0     0       0       0      83   303       386      303     663.46       386
  Sum        0 30468     626     525     443 14745     46807    16030  287366.85     45522
```

We can now join these two datasets; recall the initial set-up of `Lx` defined the `lex.id` to keep track of who is who, and the cutting of the follow-up preserves all follow-up time, so when we join the datasets by `rbind`-ing them we get the total follow-up. Note that we use `Relevel` to fix up the states and their order in `lex.Cst` and `lex.Xst`.

```
> levels( LaT )
[1] "Well"    "BAA"     "BKA"     "AKA"     "DM"      "Dead"    "AKA(DM)" "BAA(DM)" "BKA(DM)"
> levels( LdT )
[1] "Well"    "DM"      "BAA(DM)" "BKA(DM)" "AKA(DM)" "Dead"
> levels( LA <- rbind( LaT, LdT ) )
[1] "Well"    "BAA"     "BKA"     "AKA"     "DM"      "Dead"    "AKA(DM)" "BAA(DM)" "BKA(DM)"
> LA <- Relevel( LA,
+                 match( c("Well", "DM",
+                          "BAA"    ,"BKA"    ,"AKA",
+                          "BAA(DM)","BKA(DM)","AKA(DM)",
+                          "Dead"),
+                        levels(LA) ) )
> summary( LA )
Transitions:
     To
From      Well    DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM)  Dead  Records:  Events: Risk time:  Person
  Well       0    13 505 396 447       0       0       0     7      1368     1368       3.71      13
  DM         0 30468   0   0   0     483     280     206 13959     45396    14928  284146.23     453
  BAA        0     0 241  46  33      35       0       0   168       523      282    2306.51       5
  BKA        0     0   0 113 120       0      24       0   199       456      343    1276.73       4
  AKA        0     0   0   0 138       0       0      21   453       612      474    1338.52       6
  BAA(DM)    0     0   0   0   0     143     162      28   207       540      397    1336.68       5
  BKA(DM)    0     0   0   0   0       0      83     126   276       485      402    1220.48       4
  AKA(DM)    0     0   0   0   0       0       0      83   303       386      303     663.46       3
  Sum        0 30481 746 555 738     661     549     464 15572     49766    18497  292292.32     468
> data.frame( sapply( attributes(LA)[grep("time",names(attributes(LA)))], cbind ) )
  time.scales time.since
1         per
2         age
3         dur         DM
4         tsF        BAA
5         tsK        BKA
6         tsT        AKA
> save( LA, file="./data/LA.Rda" )
```

Note that we have no persons in the "Well" state any more, because the cohort we are working with are defined by entry *either* at diabetes diagnosis *or* amputation, so the disappearance of the "Well" state is a check on this.

2.5.3 Cohort follow-up

We can get an overview of the transitions here:

```
> load( file="./data/LA.Rda" )
> # windows(pointsize=8)
> boxes.Lexis( LA, boxpos=list(x=c(10,10,25,60,92,25,60,92,77.5),
+                              y=c(60,40,85,95,80,
+                                  100-c(85,95,80),50)),
+                  wm=1.02, hm=1.2, show.BE=T, scale.R=100,
+                  show.D=TRUE, show.R=FALSE, lwd.arr=2 )
```

There are a lot of transitions between states, but still we are missing the state and the transitions from the non-DM, non-amputated state (not in the picture), that is the population in Fyn, however we can use the picture to conceptualize what is of interest.

First of all we are not interested in the *absolute* rates of amputation *per se*, but mostly the *relative* size of rates. We will also be looking at the mortality rates.

We are primarily interested in rates of transition *into* states:

- Rates of foot-amputation, how they depend on DM status, and how these rates have developed over calendar time.
- Rates of knee-amputation, how they depend on DM status, previous foot-amputation and how these rates have developed over calendar time.
- Rates of thigh-amputation, how they depend on DM status, previous foot- and knee-amputation and how these rates have developed over calendar time.
- Mortality rates, how they depend on DM status, previous foot-, knee- and thigh-amputation and how these rates have developed over calendar time.

Thus we see that the blurred picture conveyed in figure 2.4 boils down to 4 analyses of rates and description of how these depend on disease status and calendar time.

2.5.3.1 The non-cohort follow-up

In order to get the follow-up for the part of the population not known to have either DM or amputation we must subtract the PY and deaths in this group from the total PY and deaths in Fyn. And we must also tabulate the amputation events among those outside the cohort, that is amputations among persons without DM and with no prior amputation. This requires that we split the follow-up in the cohort by age and period and tabulate risk time and deaths:

```
> timeScales( LA )
[1] "per" "age" "dur" "tsF" "tsK" "tsT"
```

We will split the follow-up along age and calendar time in 1-year intervals, so roughly every year of follow-up will contribute two records, so we shall expect the final dataset to have some 600,000 records:

```
> system.time( La <- splitLexis( LA, 0:120, time.scale="age" ) )
   user  system elapsed
  5.804   0.520   6.322
> nrow( La )
```

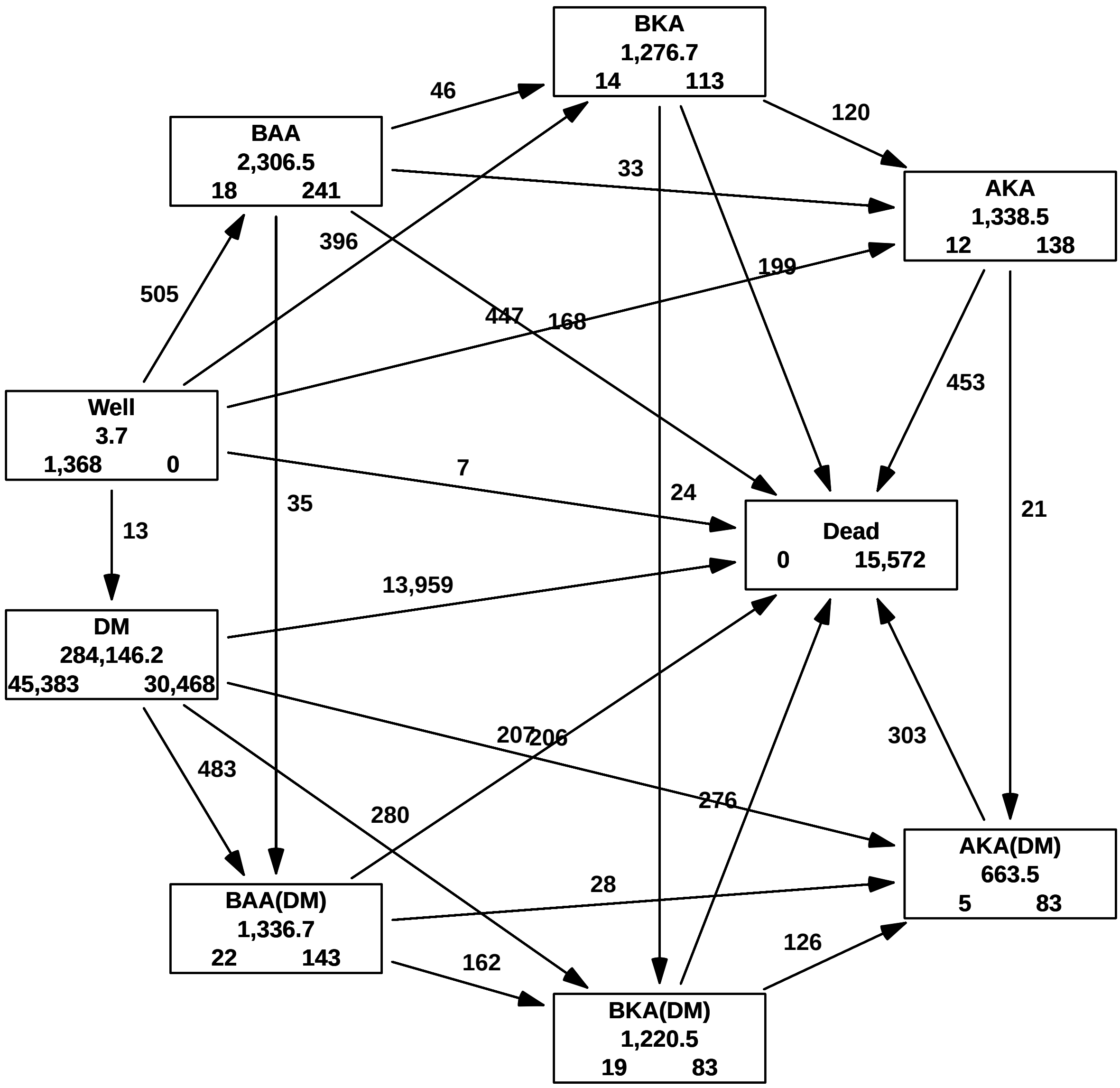

Figure 2.4: *Transitions between states — preliminary!. The numbers in the boxes are the number of person-years (center) and the number of persons starting in each state (lower left), respectively ending in each state (lower right). The numbers on the arrows are the number of observed transitions; they are all printed to the* left *of the arrows. The number of persons starting in "BAA", "BKA" and "AKA" are not correct, most of these represent transitions* into *the states from the "Well" state (which, incidentally, is not in the display.*

```
[1] 341486
> system.time( Lap <- splitLexis( La, 1990+0:30, time.scale="per" ) )
   user  system elapsed
 19.645   1.392  21.033
> nrow( Lap )
[1] 614138
```

...so the estimated 600,000 records were not quite off.

We then tabulate the person-years and the deaths by sex, age and calendar time:

```
> system.time(
+ FUtab <- aggregate( cbind( Y.coh = Lap$lex.dur,
+                            D.coh = (Lap$lex.Xst=="Dead") ),
+                     data.frame( A = timeBand( Lap, "age", "left" ),
+                                 P = timeBand( Lap, "per", "left" ),
+                               sex = Lap$sex ),
+                     FUN = sum ) )
   user  system elapsed
  2.490   0.012   2.500
> str( FUtab )
'data.frame':        2685 obs. of  5 variables:
 $ A    : num  15 16 17 18 19 20 21 22 23 24 ...
 $ P    : num  1996 1996 1996 1996 1996 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ Y.coh: num  13.9 13.5 13.4 17.3 16.7 ...
 $ D.coh: num  0 0 0 0 0 0 0 0 0 1 ...
```

Thus FUtab has the follow-up time and deaths in the cohort tabulated by sex, age and period:

```
> round(
+ ftable( addmargins(
+         xtabs( cbind(D.coh,Y.coh) ~ A10 + P2 + sex,
+                data = transform( FUtab, A10= floor(A/10)*10,
+                                          P2 = floor(P/2)*2 ) ),
+                    margin = 1:2 ),
+         row.vars=c(4,3,1) ), 0 )
              P2   1996   1998   2000   2002   2004   2006   2008   2010    Sum
      sex A10
D.coh M   10          1      0      0      0      1      0      0      0      2
          20          2      1      1      0      3      1      1      3     12
          30          7     11      6      5      7      2      3      5     46
          40         20     20     21     26     31     16     18     28    180
          50         43     59     73    110     75     87    109    100    656
          60        131    171    161    181    193    221    243    240   1541
          70        285    282    306    298    319    340    367    419   2616
          80        250    253    251    332    334    340    395    405   2560
          90         39     53     51     62     64     76     78     95    518
          Sum       778    850    870   1014   1027   1083   1214   1295   8131
      F   10          0      0      0      0      0      0      1      1      2
          20          1      2      1      3      1      0      0      2     10
          30          1      3      4      2      3      1      6      5     25
          40          8      6     14     15     10     11      8     11     83
          50         16     30     31     39     37     43     42     49    287
          60         83    103     83    102    102    105    111    128    817
          70        221    217    199    234    239    233    279    274   1896
          80        277    297    352    394    376    429    434    447   3006
          90        113    119    128    148    177    189    206    235   1315
          Sum       720    777    812    937    945   1011   1087   1152   7441
Y.coh M   10        143    125    110    142    185    179    166    156   1206
          20        445    433    429    455    433    408    413    371   3387
          30        713    792    862    904    905    848    785    790   6599
          40       1432   1593   1712   1880   2093   2163   2339   2522  15735
          50       2378   2873   3507   4066   4393   4513   4731   4976  31438
          60       2561   3005   3582   4384   5482   6510   7332   8425  41281
          70       2455   2824   3203   3556   4120   4664   5501   6367  32691
          80       1121   1264   1471   1644   1837   2091   2341   2658  14428
          90         94    111    119    159    200    249    288    337   1556
          Sum     11344  13020  14995  17190  19648  21624  23896  26602 148320
      F   10        129    118    105    114    147    170    182    158   1123
          20        409    381    398    372    368    373    417    416   3135
```

```
30         702   789   879   949  1005   975   904   914   7117
40        1060  1188  1346  1564  1741  1898  2078  2188  13063
50        1502  1908  2381  2776  3134  3251  3446  3745  22143
60        2274  2537  2874  3393  4067  4739  5460  6411  31755
70        3058  3312  3706  4146  4579  4970  5473  6038  35283
80        1995  2376  2733  2980  3345  3715  4036  4388  25567
90         329   358   449   582   668   726   792   883   4787
Sum      11458 12966 14871 16876 19055 20817 22789 25141 143973
```

For the overview, we derive the number of occurrences of the three types of amputation in both the diabetic and the non-diabetic part of the population; that is the *entry* state for persons starting because of an amputation in the cohort. However, we are not interested in transitions from say LOA to LOA(DM) because this type of transition represents a diagnosis of diabetes and not an amputation, hence:

```
> with( subset(LA ,        lex.Cst      !=       lex.Xst      ), table( lex.Xst ) )
lex.Xst
   Well      DM     BAA     BKA     AKA BAA(DM) BKA(DM) AKA(DM)    Dead
      0      13     505     442     600     518     466     381   15572

> ( tt <-
+ with( subset(LA ,substr(lex.Cst,1,3)!=substr(lex.Xst,1,3)), table( lex.Xst ) ) )
lex.Xst
   Well      DM     BAA     BKA     AKA BAA(DM) BKA(DM) AKA(DM)    Dead
      0      13     505     442     600     483     442     360   15572

> tt <- cbind( tt[3:5], tt[6:8] )
> colnames( tt ) <- c("noDM","DM")
> tt <- addmargins( tt )
> round( cbind( tt, sweep( tt[,1:2], 2, tt[4,1:2], "/" )*100 ), 1 )

    noDM   DM  Sum  noDM    DM
BAA  505  483  988  32.6  37.6
BKA  442  442  884  28.6  34.4
AKA  600  360  960  38.8  28.0
Sum 1547 1285 2832 100.0 100.0
```

Since the analysis will be based on split records of follow-up among diabetes patients, but on tabulated data by age, period and sex for the non-diabetes patients, we aggregate the number of amputations among non-diabetics by age, calendar time and sex:

```
> system.time(
+ FUamp <- with( LA,
+           aggregate( cbind( F.noDM = (lex.Cst=="Well" & lex.Xst=="BAA"),
+                             K.noDM = (lex.Cst=="Well" & lex.Xst=="BKA"),
+                             T.noDM = (lex.Cst=="Well" & lex.Xst=="AKA") ),
+                      data.frame( A = floor( age ),
+                                  P = floor( per ),
+                                sex =        sex ),
+                      FUN = sum ) ) )
   user  system elapsed
  0.274   0.000   0.274

> str( FUamp )
'data.frame':        2553 obs. of  6 variables:
 $ A     : num  15 16 17 18 19 20 21 22 23 24 ...
 $ P     : num  1996 1996 1996 1996 1996 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ F.noDM: int  0 0 0 0 0 1 0 0 0 0 ...
 $ K.noDM: int  0 0 0 0 0 0 0 0 0 0 ...
 $ T.noDM: int  0 0 0 0 0 0 1 0 0 0 ...
```

```
> ftable( addmargins(
+         xtabs( cbind(F.noDM,K.noDM,T.noDM) ~ A10 + P2 + sex,
+                data = transform( FUamp, A10= floor(A/10)*10,
+                                         P2 = floor(P/2)*2 ) ),
+                    margin = 1 ),
+         row.vars=c(4,1), col.vars=c(3,2) )

           sex    M                                       F
           P2  1996 1998 2000 2002 2004 2006 2008 2010 1996 1998 2000 2002 2004 2006 2008 2010
       A10
F.noDM 10         0    1    2    1    1    3    0    1    0    0    1    2    0    1    1    0
       20         3    4    1    2    0    0    0    0    1    0    1    2    0    2    1    2
       30         1    4    5    5    3    2    1    5    0    2    2    1    1    1    1    0
       40         2    5    3    5    4    7    4    4    2    0    1    2    3    1    1    5
       50         4    3    3    4   11    4    2   11    0    5    3    3    1    7    4    2
       60         5    6    8    7    8    5    3    6    3    4    3    6    4    1    4    3
       70         6    7    9    8    8   11    6    7    5    3    8    9    6    3   10    8
       80         5    6    9    8    6   12    6    4    7   11    7    5    4   10    9    2
       90         0    0    1    0    0    4    0    1    1    1    2    1    4    1    3    2
       Sum       26   36   41   40   41   48   22   39   19   26   28   31   23   27   34   24
K.noDM 10         2    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
       20         1    2    5    0    1    2    4    2    0    0    1    0    1    0    1    0
       30         1    0    1    0    1    2    1    2    0    0    1    2    0    2    0    0
       40         1    1    3    4    3    0    9    5    0    0    1    1    1    2    0    0
       50         1    4    4    7    6    6    4    8    1    0    1    0    2    0    1    0
       60         6    6    5    6    6    8    1    6    0    4    4    3    7    4    2    5
       70         6   14    9    5    8   10    7    4    9    7    3    3    4    5    2    3
       80         7    8    3    4    5    9    4    1    9   11   11    7    5    3    5    0
       90         2    0    1    0    2    0    1    0    4    0    6    2    1    1    1    0
       Sum       27   35   31   26   32   37   31   28   23   22   28   18   21   17   12    8
T.noDM 10         0    0    0    1    0    0    0    0    0    0    0    0    0    0    0    0
       20         1    0    0    0    0    0    0    2    0    0    0    0    0    0    1    0
       30         0    0    0    0    0    0    0    0    0    1    0    0    1    0    1    0
       40         0    2    2    2    4    0    1    3    0    0    2    1    1    1    1    0
       50         0    2    8    1    2    2    4    2    0    3    1    0    0    1    1    2
       60         2    3    3    4    5    4    7    6    6    5    2    6    2    1    2    2
       70        11   14    5    6    4    8   11    5    7    8    8   11    6   12    7   10
       80         9    8    5    6    5    9   10   12    9    8    8    8   12   16   11   17
       90         1    1    0    1    2    2    1    0    3    2    5    6    7    4    2    7
       Sum       24   30   23   21   22   25   34   30   25   27   26   32   29   35   26   38
```

```
> addmargins( xtabs( cbind(F.noDM,K.noDM,T.noDM) ~ sex,
+                    data = FUamp ),
+             margin = 1 )

sex   F.noDM K.noDM T.noDM
  M      293    247    209
  F      212    149    238
  Sum    505    396    447
```

The dataset `FUamp` now contains the number of amputations by sex, age and calendar time among non-diabetics in Fyn.

This is now merged with the follow-up for death to provide a dataset with the follow-up time, deaths and the number of amputation events occurring in persons without diabetes — note that there are necessarily units where there is only follow-up time but no amputation events, but not vice versa, as seen here:

```
> FUpop <- merge( FUamp, FUtab, all=TRUE )
> c( nrow(FUtab),
+    nrow(FUamp),
+    nrow(FUpop) )
[1] 2685 2553 2685
> FUpop[is.na(FUpop)] <- 0
```

This data frame is now further merged with the total follow-up for death in Fyn, `FUfyn`, constructed previously in order to provide the person-years and the number of deaths among those not in the cohort (the "Well" state):

```
> load( file="./data/FUfyn.Rda" )
> FUnoDM <- merge( FUpop, FUfyn, all=TRUE )
> c( nrow(FUpop),
+    nrow(FUfyn),
+    nrow(FUnoDM) )
[1] 2685 2688 2688
```

```
> FUnoDM[is.na(FUnoDM)] <- 0
> FUnoDM <- transform( FUnoDM, D = pmax( D.all - D.coh, 0 ),
+                              Y = pmax( Y.all - Y.coh, 0 ) )
> cbind( round( addmargins( xtabs( cbind( D, D.coh )      ~ P, data=FUnoDM ), 1 ) ),
+        round( addmargins( xtabs( cbind( Y, Y.coh )/1000 ~ P, data=FUnoDM ), 1 ), 1 ) )
```

```
         D D.coh      Y Y.coh
1996  4914   726  377.8  11.0
1997  4692   772  376.5  11.8
1998  4568   764  374.8  12.5
1999  4674   863  372.9  13.5
2000  4469   809  371.1  14.4
2001  4389   873  369.6  15.4
2002  4476   961  368.8  16.5
2003  4222   990  368.6  17.6
2004  4197   975  368.7  18.8
2005  4037   997  369.5  19.9
2006  4001  1008  371.0  20.8
2007  4009  1086  372.6  21.6
2008  3813  1084  373.7  22.7
2009  3906  1217  374.3  24.0
2010  3695  1208  374.5  25.2
2011  3630  1239  374.6  26.5
Sum  67692 15572 5959.0 292.3
```

The dataset `FUnoDM` is classified by sex, age and calendar time, and contains the deaths and person-years in the non-cohort part of the Fyn population in `D` and `Y`, and the number of amputations in this population in `F.noDM`, `K.noDM` and `T.noDM`, respectively.

Finally the relevant analysis datasets are saved:

```
> save( FUnoDM, Lap, LA, file="./data/AmpAna.Rda" )
```

2.5.4 Overview of analysis dataset

2.5.4.1 Amputations by sex, DM status and time

We first fish out the number of amputations by sex and diabetes status, and show them in compact tabular form so that they can be used in the paper:

```
> summary( LA )
Transitions:
     To
From      Well    DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM)  Dead  Records:  Events: Risk time:  Person
  Well       0    13 505 396 447       0       0       0     7      1368     1368       3.71      13
  DM         0 30468   0   0   0     483     280     206 13959     45396    14928  284146.23     453
  BAA        0     0 241  46  33      35       0       0   168       523      282    2306.51       5
  BKA        0     0   0 113 120       0      24       0   199       456      343    1276.73       4
  AKA        0     0   0   0 138       0       0      21   453       612      474    1338.52       6
  BAA(DM)    0     0   0   0   0     143     162      28   207       540      397    1336.68       5
  BKA(DM)    0     0   0   0   0       0      83     126   276       485      402    1220.48       4
  AKA(DM)    0     0   0   0   0       0       0      83   303       386      303     663.46       3
  Sum        0 30481 746 555 738     661     549     464 15572     49766    18497  292292.32     468
```

```
> sL <- summary( LA, by="sex" )
> tm <- sL[["M"]][[1]][1:8,1:8]
> tf <- sL[["F"]][[1]][1:8,1:8]
> dn <- rownames( tf )
> diag( tm ) <- diag( tf ) <- 0
> tm[dn[3:5],dn[6:8]] <- 0
> tf[dn[3:5],dn[6:8]] <- 0
> ta <- NArray( list( rownames(tf)[3:5],
+                     c("noDM","DM"),
+                     c("M","F") ) )
> ftable( ta, row.vars=1 )
     noDM    DM
        M  F  M  F

BAA     NA NA NA NA
BKA     NA NA NA NA
AKA     NA NA NA NA
> ta[,,"M"] <- apply( tm, 2, sum )[3:8]
> ta[,,"F"] <- apply( tf, 2, sum )[3:8]
> ftable( tt <- addmargins(ta), col.vars=3:2 )
        M              F              Sum
     noDM   DM  Sum noDM   DM  Sum noDM   DM  Sum

BAA   293  351  644  212  132  344  505  483  988
BKA   277  292  569  165  150  315  442  442  884
AKA   300  193  493  300  167  467  600  360  960
Sum   870  836 1706  677  449 1126 1547 1285 2832
> round(
+ ftable( sweep( tt, 2:3, apply(tt,2:3,sum)/2, "/" )*100,
+         col.vars=3:2 ), 1 )
         M                 F                 Sum
      noDM    DM   Sum  noDM    DM   Sum  noDM    DM   Sum

BAA   33.7  42.0  37.7  31.3  29.4  30.6  32.6  37.6  34.9
BKA   31.8  34.9  33.4  24.4  33.4  28.0  28.6  34.4  31.2
AKA   34.5  23.1  28.9  44.3  37.2  41.5  38.8  28.0  33.9
Sum  100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0
```

Here are all amputations for they all have `lex.Cst`≠`lex.Xst` at their amputation, but also recalling the transitions from `LOA` to `LOA(DM)` should not be counted

```
> Aamp <- subset( LA, lex.Xst %in% levels(LA)[grep("A",levels(LA))] &
+                     substr(lex.Cst,1,3) != substr(lex.Xst,1,3) )
> Aamp <- data.frame( transform( Aamp, per=per+lex.dur )[,c("per","sex","lex.Xst")] )
> Aamp <- transform( Aamp, DM = factor( Relevel( lex.Xst,
+                               list(noDM=3:5,DM=6:8,x=c(1:2,9)) ) ),
+                        Amp = factor( Relevel( lex.Xst,
+                               list(BAA=c(3,6),BKA=c(4,7),AKA=c(5,8),x=c(1:2,9) ) ) ) )
> with( Aamp, table( DM,lex.Xst ) )
      lex.Xst
DM      Well  DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM) Dead
  noDM     0   0 505 442 600       0       0       0    0
  DM       0   0   0   0   0     483     442     360    0
> with( Aamp, table( Amp,lex.Xst ) )
     lex.Xst
Amp    Well  DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM) Dead
  BAA     0   0 505   0   0     483       0       0    0
  BKA     0   0   0 442   0       0     442       0    0
  AKA     0   0   0   0 600       0       0     360    0
```

Now we have the total number of amputations each year by sex, date and amputation type, so we can create a simple overview by sex and diabetes status:

```
> tt <- with( Aamp, addmargins( table( Amp, DM, sex ) ) )
> ftable( tt, col.vars=3:2 )
    sex    M              F              Sum
    DM   noDM   DM  Sum noDM   DM  Sum noDM   DM  Sum
Amp
BAA       293  351  644  212  132  344  505  483  988
BKA       277  292  569  165  150  315  442  442  884
AKA       300  193  493  300  167  467  600  360  960
Sum       870  836 1706  677  449 1126 1547 1285 2832
> round(
+ ftable( sweep( tt, 2:3, apply(tt,2:3,sum)/2, "/" )*100,
+         col.vars=3:2 ), 1 )
    sex     M                   F                   Sum
    DM     noDM    DM   Sum  noDM    DM   Sum  noDM    DM   Sum
Amp
BAA        33.7  42.0  37.7  31.3  29.4  30.6  32.6  37.6  34.9
BKA        31.8  34.9  33.4  24.4  33.4  28.0  28.6  34.4  31.2
AKA        34.5  23.1  28.9  44.3  37.2  41.5  38.8  28.0  33.9
Sum       100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0
```

Then we use the table to make a graph of the absolute number of amputation among DM and non-DM per year

```
> TT <- with( Aamp, table( Amp, floor(per), DM ) )
> ftable( TT, col.vars=c(3,1) )
     DM   noDM            DM
     Amp   BAA BKA AKA BAA BKA AKA

1996        22  28  42  24  29  12
1997        23  29  30  25  26  12
1998        39  24  35  19  30   8
1999        23  38  40  28  36  12
2000        37  40  25  24  25  17
2001        32  26  38  32  31  20
2002        35  28  23  35  32  22
2003        36  20  42  31  30  30
2004        35  25  38  33  21  22
2005        28  34  35  35  39  30
2006        39  35  41  26  28  25
2007        37  26  41  43  21  25
2008        30  15  33  25  34  27
2009        26  35  48  28  18  34
2010        35  14  40  37  28  32
2011        28  25  49  38  14  32
> CT <- apply(TT,2:3,cumsum)
> pla <- function(){
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
+ matplot( as.numeric(dimnames(TT)[[2]]),
+          t(rbind( CT[,,1], CT[,,2] )),
+          xlab="", ylab="Annual number of amputations",
+          ylim=c(0,120), yaxs="i", xaxt="n",
+          type="l", lty=rep(c(2,1),each=3), lwd=c(3,3,5), col="black" )
+ axis( side=1, at=seq(1995,2010,5) )
+ text( rep(2011,3), c(17,44,72), dimnames(TT)[[1]], adj=1 )
+ axis( side=1, at=seq(1995,2011,1), labels=NA )}
> pla()
> postscript( "./graph/Fig1.eps", height=6, width=6 )
> pla()
> dev.off()
pdf
  2
> # win.metafile( "Fig1.emf",height=6,width=6)
> # pla()
> # dev.off()
```

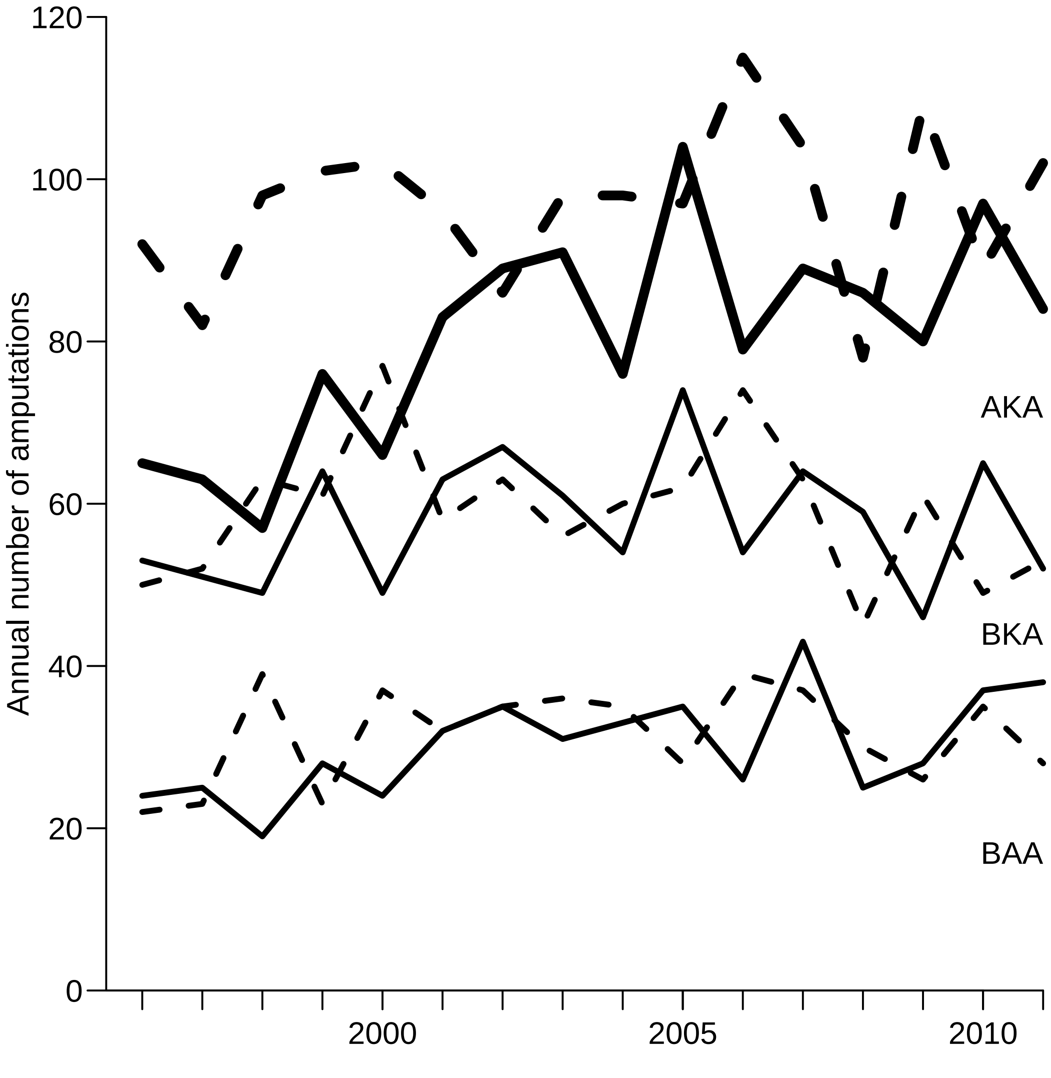

Figure 2.5: *Number of amputations in diabetes patients (full lines) and non-diabetes persons (broken lines). The lower curves are foot amputations, the middle foot+knee and the thick curve the total no. of amputations.*

2.5.4.2 Transitions used in modelling

For graphical overview of the entire flow of persons we show all transitions in the population with colouring of those we considered; first the standard

```
> bb <-
+ boxes.Lexis( LA,
+               boxpos=list(x=c(10,10,25,60,92,25,60,92,77.5),
+                           y=c(60,40,87,95,78,
+                               100-c(87,95,78),50)),
+               wm=1.1, hm=1.2, show.BE=T, scale.R=100, font=1,
+               show.D=TRUE, show.R=FALSE, lwd.arr=2, pos.arr=0.55, font.arr=1 )
> bb
```

```
$Boxes
    xx yy      wd     ht font lwd col.txt col.border      col.bg
1 10.0 60 20.7944 8.0949    1   2   black      black transparent
2 10.0 40 20.7944 8.0949    1   2   black      black transparent
```

```
3 25.0 87 20.7944 8.0949    1   2   black      black transparent
4 60.0 95 20.7944 8.0949    1   2   black      black transparent
5 92.0 78 20.7944 8.0949    1   2   black      black transparent
6 25.0 13 20.7944 8.0949    1   2   black      black transparent
7 60.0  5 20.7944 8.0949    1   2   black      black transparent
8 92.0 22 20.7944 8.0949    1   2   black      black transparent
9 77.5 50 20.7944 8.0949    1   2   black      black transparent

$State.names
[1] "Well\n3.7\n1,368          0"           "DM\n284,146.2\n45,383           30,468"
[3] "BAA\n2,306.5\n18          241"         "BKA\n1,276.7\n14           113"
[5] "AKA\n1,338.5\n12          138"         "BAA(DM)\n1,336.7\n22          143"
[7] "BKA(DM)\n1,220.5\n19          83"      "AKA(DM)\n663.5\n5           83"
[9] "Dead\n0            15,572"

$Tmat

          Well DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM)  Dead
  Well      NA 13 505 396 447      NA      NA      NA     7
  DM        NA NA  NA  NA  NA     483     280     206 13959
  BAA       NA NA  NA  46  33      35      NA      NA   168
  BKA       NA NA  NA  NA 120      NA      24      NA   199
  AKA       NA NA  NA  NA  NA      NA      NA      21   453
  BAA(DM)   NA NA  NA  NA  NA      NA     162      28   207
  BKA(DM)   NA NA  NA  NA  NA      NA      NA     126   276
  AKA(DM)   NA NA  NA  NA  NA      NA      NA      NA   303
  Dead      NA NA  NA  NA  NA      NA      NA      NA    NA

$Arrows
   lwd.arr col.arr pos.arr col.txt.arr font.arr offset.arr
1        2   black    0.55       black        1          2
2        2   black    0.55       black        1          2
3        2   black    0.55       black        1          2
4        2   black    0.55       black        1          2
5        2   black    0.55       black        1          2
6        2   black    0.55       black        1          2
7        2   black    0.55       black        1          2
8        2   black    0.55       black        1          2
9        2   black    0.55       black        1          2
10       2   black    0.55       black        1          2
11       2   black    0.55       black        1          2
12       2   black    0.55       black        1          2
13       2   black    0.55       black        1          2
14       2   black    0.55       black        1          2
15       2   black    0.55       black        1          2
16       2   black    0.55       black        1          2
17       2   black    0.55       black        1          2
18       2   black    0.55       black        1          2
19       2   black    0.55       black        1          2
20       2   black    0.55       black        1          2
21       2   black    0.55       black        1          2
22       2   black    0.55       black        1          2
23       2   black    0.55       black        1          2
24       2   black    0.55       black        1          2

$Arrowtext
 [1] "13"     "505"    "396"    "447"    "7"      "483"    "280"    "206"    "13,959" "46"     "33"
[12] "35"     "168"    "120"    "24"     "199"    "21"     "453"    "162"    "28"     "207"    "126"
[23] "276"    "303"

attr(,"class")
[1] "MS"
```

In order to get things right we must update the `bb` object with the missing information on transitions from the "Well" state.

```
> bb$Tmat["Well","Dead"] <- sum( FUnoDM$D )
> bb$Tmat["Well","DM"] <- with( subset( LA, lex.Cst=="DM" ), sum(doDM>1996) )
> bb$Arrowtext <- formatC( t(bb$Tmat)[!is.na(t(bb$Tmat))],
+                          format="f", digits=0, big.mark="," )
> bb$State.names[1] <- paste("non-DM\n",
+          formatC( sum( FUnoDM$Y ) , format="f", digits=1, big.mark="," ) )
> bb$State.names[9] <- paste("Dead\n          ",
+         formatC( sum(FUnoDM$D.all), format="f", digits=0, big.mark="," ) )
> # Exercise just to see which elements of the arrows to update:
> ano <- as.vector( t(bb$Tmat) )
> ano[!is.na(ano)] <- 1:sum(!is.na(ano))
> dim( ano ) <- dim( bb$Tmat )
> dimnames( ano ) <- dimnames( bb$Tmat )
> t(ano)

        Well DM BAA BKA AKA BAA(DM) BKA(DM) AKA(DM) Dead
Well      NA  1   2   3   4      NA      NA      NA    5
DM        NA NA  NA  NA  NA       6       7       8    9
BAA       NA NA  NA  10  11      12      NA      NA   13
BKA       NA NA  NA  NA  14      NA      15      NA   16
AKA       NA NA  NA  NA  NA      NA      NA      17   18
BAA(DM)   NA NA  NA  NA  NA      NA      19      20   21
BKA(DM)   NA NA  NA  NA  NA      NA      NA      22   23
AKA(DM)   NA NA  NA  NA  NA      NA      NA      NA   24
Dead      NA NA  NA  NA  NA      NA      NA      NA   NA

> bb$Arrows[,"col.arr"]    <-
+ bb$Arrows[,"col.txtarr"] <- gray(0.45)
> bb$Arrows[c(2,6)]              ,"col.arr"]     <-
+ bb$Arrows[c(2,6)]              ,"col.txt.arr"] <- "limegreen"
> bb$Arrows[c(3,10,7,19)]        ,"col.arr"]     <-
+ bb$Arrows[c(3,10,7,19)]        ,"col.txt.arr"] <- "blue"
> bb$Arrows[c(4,11,14,8,20,22),"col.arr"]     <-
+ bb$Arrows[c(4,11,14,8,20,22),"col.txt.arr"] <- "red"
> bb$Arrows[c(1,12,15,17)]       ,"col.arr"]     <-
+ bb$Arrows[c(1,12,15,17)]       ,"col.txt.arr"] <- gray(0.75)
> bb$Arrows[,"lwd.arr"] <- 3
> bb$Arrows[c(6,7,19,8,20,22),"lwd.arr"] <- 5
> bb$Arrows[c(3,7,13,15,16,20),"pos.arr"] <- c(0.63,0.66,0.57,0.57,0.59,0.50)
> boxes( bb, wmult=0.9 )
> BB <- bb
> postscript( "./graph/Fig1.eps", height=12, width=12 )
> boxes( BB )
> dev.off()

pdf
  2

> # win.metafile("Fig1.emf",height=12,width=12)
> # boxes( BB )
> # dev.off()
```

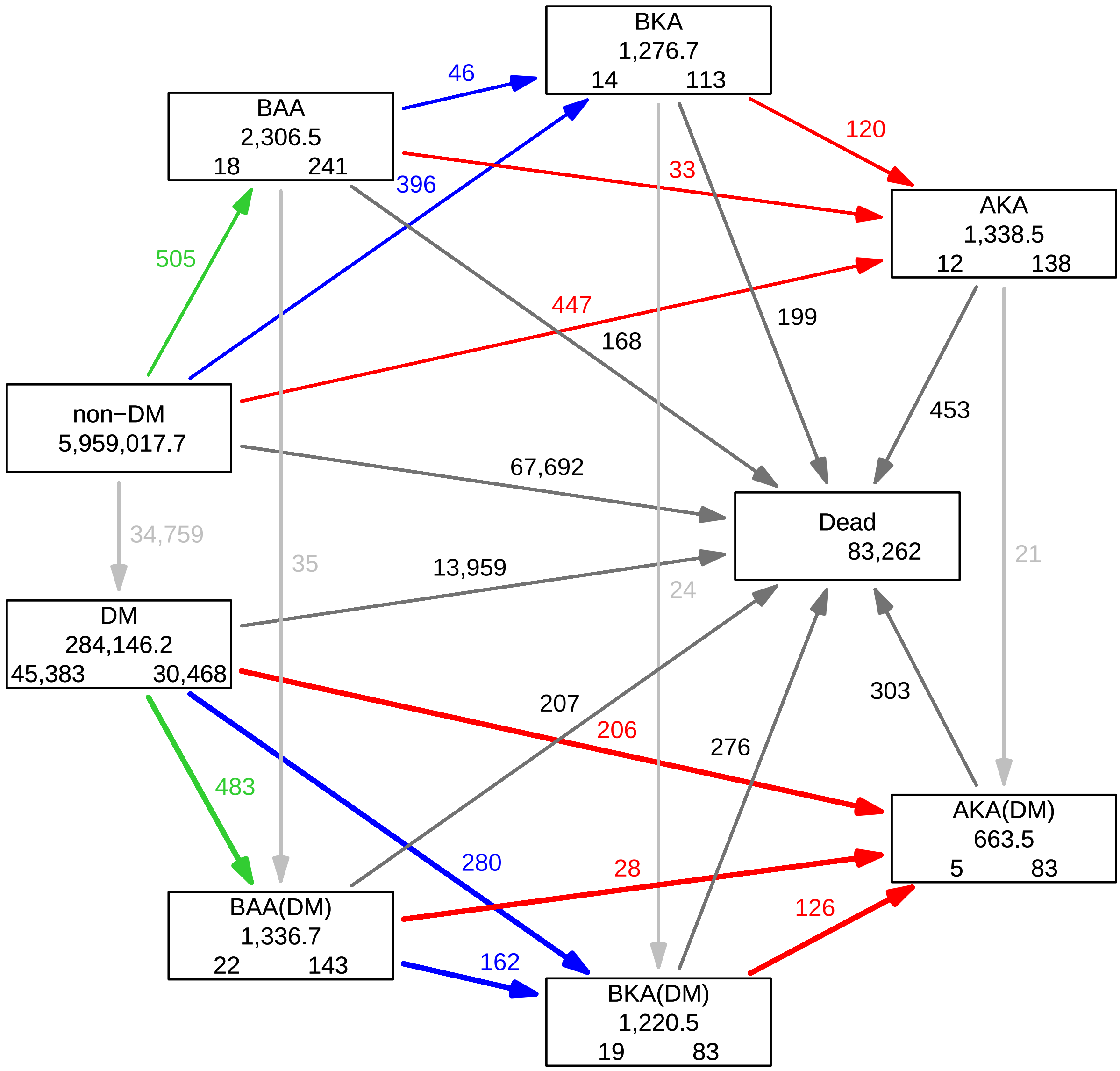

Figure 2.6: *Transitions between states. The numbers in the boxes are the number of person-years (center) and the number of persons starting in each state (lower left), respectively ending in each state (lower right). The numbers on the arrows are the number of observed transitions; they are all printed to the* left *of the arrows. In the "Well" state is only given the person-years because this state represents the non-affected part of the Fyn population. The core analyses are comparison of the thick coloured arrows with the correponding thin arrows; green for below ankle amputations (BAA), blue for below knee amputations (BKA) and red for above knee amputations (AKA).*

Chapter 3

Analysis of rates

Recall from figure 2.6 that we have 4 different types of rates to analyze — each colored differently:

- Below ankle amputation rates (BAA) — green
- Below knee amputation rates (BKA) — blue
- Above knee amputation rates (AKA) — red
- Mortality rates — black — outside the scope of this report.

3.1 Datasets

We construct datasets for analysis of these rates by using the population dataset, `FUnoDM` and combining it with suitable subsets of the time-split cohort dataset `Lap`.

```
> library( Epi )
> library( splines )
> clear()
> load( file="./data/AmpAna.Rda" )
```

The data for the analysis will consist of tabulated follow-up in the non-DM, non-amputated population (the dataframe `FUnoDM`), in order to get the amputation rates among persons without diabetes. This is a dataset classified by sex and age and date at follow-up, where we use the person-years and the relevant event indicator as outcome. The other part of data are the individually split records from the cohort, where we select persons in the relevant states (using `lex.Cst`) and define the relevant outcome (using `lex.Xst`). All that is needed is to define the variable names from the two datasources to be the same. Since the event variable differs between analyses of different types of amputation, and since the subset of the cohort needed also differs, the analysis dataset need to be defined separately for each analysis.

3.2 Statistical models

The models will be models for the amputation rates, describing how amputation rates depend on age, sex, diabtes and amputaion status, and on calendar time.

So for all three types of amputations we fit a simple model with a continuous age effect, and separate effects of sex and diabetes/amputation status. Note that for the "first" amputations, foot, there is only a diabetes status, as all foot amputations as defined are among persons not previously amputated.

However, we would also like a more elaborate model where:

- rates depend on duration of diabetes,
- there is an interaction between diabetes/amputation status and
  - sex
  - age
  - calendar time

In order to be able to give summary figures for diabetes and amputation status effects for each sex, we first fit the intermediate model with only sex-interaction, and the subsequently the model with the age- and calendar time interactions.

So we shall fit and report from three different models for each type of amputation, the first one giving the raw amputaion RRs by sex and diabetes status, controlled for age, the second sex×status classified effects also controlled for DM duration, and the third giving a more detailed description of the amputation rates, that we report graphically.

3.3 BAA amputations

In the no DM dataset the response variables are `F.noDM` and `Y`, and the explanatory variables are age (`A`), calendar time (`P`), sex and of course diabetes status (None). However `A` and `P` are coded as the left endpoint of the interval so we add 0.5.

From the cohort dataset we extract the same variables, but in this case age and period represent the *actual* age at the start of an individual piece of follow-up, so here we add half of the interval length, `lex.dur`.

The two parts of the dataset are then merged:

```
> anaF <- rbind( with( FUnoDM,
+                 data.frame( A = A+0.5,
+                             P = P+0.5,
+                         DMdur = 0,
+                           sex = sex,
+                            DM = "No",
+                             D = F.noDM,
+                             Y = Y ) ),
+                 with( subset( Lap, lex.Cst=="DM" ),
+                 data.frame( A = age+lex.dur/2,
+                             P = per+lex.dur/2,
+                         DMdur = dur+lex.dur/2,
+                           sex = sex,
+                            DM = "Yes",
+                             D = as.numeric( lex.Xst=="BAA(DM)" ),
+                             Y = lex.dur ) ) )
> str( anaF )
'data.frame':        596643 obs. of  7 variables:
 $ A    : num  15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 ...
 $ P    : num  1996 1996 1998 1998 1998 ...
 $ DMdur: num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ DM   : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ D    : num  0 0 0 0 0 0 0 0 1 0 ...
 $ Y    : num  2649 2568 2516 2402 2434 ...

> round( ftable( addmargins(
+ abind( with( anaF, table( sex,  DM, D ) ),
+        Ftab <- xtabs( cbind(D,Y) ~ sex + DM, data=anaF ) ), 1 ) ) )
                0       1       2       3       D       Y

M   No       1081     235      26       2     293 2919600
    Yes    299515     351       0       0     351  143240
F   No       1159     158      27       0     212 3039417
    Yes    293957     132       0       0     132  140906
Sum No       2240     393      53       2     505 5959018
    Yes    593472     483       0       0     483  284146

> PFtab <- xtabs( cbind(D,Y) ~ floor(P) + floor(A) + DM, data=anaF )
> save( PFtab, file="./data/PFtab.Rda" )
```

Note that for the non-diabetic part of the population we have coded the variable `DMdur` as 0, so that if we define a natural spline for this with a knot in 0, the spline effect will be 0 in 0, and hence the effect of `DM` will be the effect of DM at the time of diagnosis. This an important quirk to observe when a timescale is only defined for only a subset of the data (the diabetes patients in this case).

We now model the foot-amputation occurrence (in `D`, `Y`) using natural splines for age and diabetes duration, so we need to look at the distribution of events along age and duration:

```
> with( subset(anaF,D>0 & DM=="No"),
+       quantile( rep(A,D), probs=0:5/5 ) )
```

```
  0%  20%  40%  60%  80% 100% 
15.5 47.3 62.5 74.5 82.5 96.5 

> with( subset(anaF,D>0 & DM=="Yes"),
+       quantile( rep(A,D), probs=0:5/5 ) )

      0%      20%      40%      60%      80%     100% 
30.06229 58.34565 66.63901 73.16831 80.23107 98.16496 

> with( subset(anaF,D>0 & DM=="Yes"),
+       quantile( rep(DMdur,D), probs=0:10/10 ) )

         0%         10%         20%         30%         40%         50%         60% 
 0.001368925  0.614442163  3.301916496  5.236960986  6.909924709  8.229295003  9.656536619 11.830527
        80%         90%        100% 
14.085010267 15.978165640 21.508213552 
```

Hence we define knots and contrast matrices for age and duration splines:

```
> np <- 100
> a.kn <- c(45,60,75,80)
> a.pt <- seq(30,90,,np)
> a.rf <- 70
> a.ct <- Ns(     a.pt     , knots=a.kn )
> a.cr <- Ns( rep(a.rf,np), knots=a.kn )
> d.kn <- c(0,3,8,15)
> d.pt <- seq(0,20,,np)
> d.rf <- 8
> d.ct <- Ns(     d.pt     , knots=d.kn )
> d.cr <- Ns( rep(d.rf,np), knots=d.kn )
```

With this in place we can now model the foot-amputation rates, initially with a proportional hazards model where the diabetes effect is the same regardless of sex, calendar time and age.

```
> system.time(
+ F0 <- glm( D ~ Ns(A,kn=a.kn) +
+                DM + sex,
+            offset = log(Y/1000),
+            family = poisson,
+              data = anaF ) )

   user  system elapsed 
 10.278   0.977  11.275 

> round( pF0 <- ci.exp( F0 ), 3 )

                  exp(Est.)  2.5%  97.5%
(Intercept)           0.077 0.068  0.088
Ns(A, kn = a.kn)1     3.252 2.596  4.075
Ns(A, kn = a.kn)2     7.431 6.046  9.135
Ns(A, kn = a.kn)3     3.332 2.752  4.035
DMYes                10.675 9.353 12.184
sexF                  0.449 0.393  0.513
```

We see a substantial sex-effect — women have an amputation incidence which is half that of men, and persons with diabetes have an about 10-fold increased rate of amputations.

Then we expand the model with the DM duration and the interaction between sex and DM:

```
> F1 <- update( F0, . ~ . + Ns(DMdur,kn=d.kn)
+                       - DM + DM:sex )
> round( pF1 <- ci.exp( F1 ), 3 )
```

```
                         exp(Est.)   2.5%  97.5%
(Intercept)                  0.070  0.061  0.080
Ns(A, kn = a.kn)1            3.185  2.541  3.992
Ns(A, kn = a.kn)2            7.192  5.850  8.842
Ns(A, kn = a.kn)3            3.241  2.676  3.925
sexF                         0.604  0.506  0.722
Ns(DMdur, kn = d.kn)1        2.165  1.572  2.983
Ns(DMdur, kn = d.kn)2        0.605  0.294  1.246
Ns(DMdur, kn = d.kn)3        3.762  2.977  4.755
sexM:DMYes                  14.660 10.510 20.448
sexF:DMYes                   7.525  5.203 10.883
```

We see that the diabetes effect among men is about twice as large than among women, and this is on top of the quite large effect of sex among non-diabetics.

```
> CM <- rbind( " NoDM, M vs. F" = c(-1,0, 0),
+              "   DM, M vs. F" = c(-1,1,-1),
+              "F, DM vs. noDM" = c( 0,0, 1),
+              "M, DM vs. noDM" = c( 0,1, 0) )
> colnames( CM ) <- rownames( ci.exp( F1, subset="sex" ) )
> CM
               sexF sexM:DMYes sexF:DMYes
 NoDM, M vs. F   -1          0          0
   DM, M vs. F   -1          1         -1
F, DM vs. noDM    0          0          1
M, DM vs. noDM    0          1          0
```

```
> round( rrf <- ci.exp( F1, subset="sex", ctr.mat=CM ), 3 )
               exp(Est.)   2.5%  97.5%
 NoDM, M vs. F     1.655  1.385  1.976
   DM, M vs. F     3.224  2.633  3.947
F, DM vs. noDM     7.525  5.203 10.883
M, DM vs. noDM    14.660 10.510 20.448
```

From these estimates it looks as if there is an interaction, which indeed is the case even if the estimates are quite strongly correlated:

```
> round( Wald( F1, subset="DMYes", ctr.mat=cbind(-1,1) ), 3 )
 Chisq   d.f.      P
23.849  1.000  0.000
```

```
> round( cor( ci.lin( F1, subset="sex", ctr.mat=CM, vcov=TRUE )$vcov ), 3 )
                NoDM, M vs. F    DM, M vs. F F, DM vs. noDM M, DM vs. noDM
 NoDM, M vs. F          1.000         -0.471         -0.034         -0.603
   DM, M vs. F         -0.471          1.000         -0.866         -0.420
F, DM vs. noDM         -0.034         -0.866          1.000          0.818
M, DM vs. noDM         -0.603         -0.420          0.818          1.000
```

From the estimates of the overall rates we see that non-diabetic men have amputation rates some 60% larger than those among non-diabetic women, but men with diabetes have BAA amputation rates about three the rates among women with diabetes. The RR of amputation between persons with and without diabetes is 14.7, (95% c.i.: 10.5 – 20.4) for men and 7.5, (95% c.i.: 5.2 – 10.9) for women. So not only are the general amputation rates higher among men, but the diabetes-effect is also larger.

We now further update the model with the interactions with age and calendar time, and then save the models for retrieval for final reporting:

```
> F2 <- update( F1, . ~ . + DM:I(P-2000) + DM:I(A-60) )
> round( pF2 <- ci.exp( F2 ), 3 )
```

```
                        exp(Est.)   2.5%  97.5%
(Intercept)                 0.089  0.075  0.106
Ns(A, kn = a.kn)1           2.089  1.631  2.676
Ns(A, kn = a.kn)2           2.846  2.062  3.928
Ns(A, kn = a.kn)3           2.271  1.834  2.811
sexF                        0.578  0.483  0.691
Ns(DMdur, kn = d.kn)1       2.213  1.605  3.052
Ns(DMdur, kn = d.kn)2       0.714  0.348  1.467
Ns(DMdur, kn = d.kn)3       6.062  4.652  7.901
sexM:DMYes                 27.436 19.326 38.950
sexF:DMYes                 15.309 10.382 22.576
DMNo:I(P - 2000)            1.005  0.986  1.024
DMYes:I(P - 2000)           0.902  0.881  0.924
DMNo:I(A - 60)              1.028  1.020  1.037
DMYes:I(A - 60)             1.000  1.000  1.000
```

```
> round( (ci.exp( F2, subset="P" )-1)*100, 1 )
```

```
                    exp(Est.)  2.5% 97.5%
DMNo:I(P - 2000)          0.5  -1.4   2.4
DMYes:I(P - 2000)        -9.8 -11.9  -7.6
```

so we see there is virtually no change in BAA rates among persons without diabetes, but a 10% annual decrease among persons with diabetes.

For later summary of reporting we extract the RR associated with DM, both overall and by sex

```
> ci.exp( F0 )
```

```
                      exp(Est.)      2.5%       97.5%
(Intercept)          0.07748924 0.0684494  0.08772292
Ns(A, kn = a.kn)1    3.25229052 2.5958715  4.07469840
Ns(A, kn = a.kn)2    7.43149141 6.0458420  9.13471850
Ns(A, kn = a.kn)3    3.33224141 2.7520136  4.03480305
DMYes               10.67548411 9.3534830 12.18433395
sexF                 0.44932483 0.3934424  0.51314451
```

```
> save( Ftab, pF0, pF1, pF2, file="./data/Fmod.Rda" )
```

From the estimates we see that there is no change in the foot amputation rates for persons without diabetes (annual change 0.5%(-1.4; $+2.4$)%), whereas the the change for diabetic patients is dramatic, an annual decline of 9.8% (7.6–11.9)%.

There is also a significant age-interaction; the amputation rates among persons without diabetes increase 2.8% steeper per year of age than among persons with diabetes:

```
> round( (ci.exp( F2, subset="I\\(A" )-1)*100, 1 )
```

```
                exp(Est.) 2.5% 97.5%
DMNo:I(A - 60)        2.8    2   3.7
DMYes:I(A - 60)       0.0    0   0.0
```

Noter taht there reallis only is one effect here, the other linear effect is aliased with the natural spline in age.

In order to show these effects we can show the amputation rates for men, resp. women for persons diagnosed in ages 40, 50, 60 and 70 in 2000, with and without diabetes, and as compared to the rates among persons without diabetes.

```
> CM <-
+ function( A, sex, DM=TRUE, durt=d.pt )
+ cbind(1,Ns(A+durt,kn=a.kn),sex=="F",
+         Ns(  durt,kn=d.kn)*DM,
+        (sex=="M")*  DM , (sex=="F")*DM, # DM:sex
+             durt *(!DM),      durt *DM, # DM:P
+       (A-60+durt)*(!DM),(A-60+durt)*DM  # DM:A
+                )
> plf <- function( ylrt=c(0.01,10),
+                  ylrr=c(1,100) )
+ {
+ par( mfrow=c(1,2), mar=c(3,3,1,1), oma=c(0,1,0,0), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
+
+ d40m <- ci.exp( F2, ctr.mat=CM(40,"M") )
+ d50m <- ci.exp( F2, ctr.mat=CM(50,"M") )
+ d60m <- ci.exp( F2, ctr.mat=CM(60,"M") )
+ d70m <- ci.exp( F2, ctr.mat=CM(70,"M") )
+ ndmm <- ci.exp( F2, ctr.mat=CM(a.pt,"M",FALSE,rep(0,length(a.pt))) )
+ d40f <- ci.exp( F2, ctr.mat=CM(40,"F") )
+ d50f <- ci.exp( F2, ctr.mat=CM(50,"F") )
+ d60f <- ci.exp( F2, ctr.mat=CM(60,"F") )
+ d70f <- ci.exp( F2, ctr.mat=CM(70,"F") )
+ ndmf <- ci.exp( F2, ctr.mat=CM(a.pt,"F",FALSE,rep(0,length(a.pt))) )
+
+ matplot( a.pt, ndmm,
+          type="l", lty="11", lend=1, lwd=c(4,1,1), col="blue",
+          log="y", xlim=c(40,90), ylim=ylrt,
+          xlab="Age (years)", ylab="", yaxt="n" )
+ ylb  <- 10^c(-2:4)
+ wlb  <- ( ylb>=ylrt[1] & ylb<=ylrt[2] )
+ ytc  <- as.vector(outer(1:9,10^c(-2:1),"*"))
+ wtc  <- ( ytc>=ylrt[1] & ytc<=ylrt[2] )
+ matlines(    a.pt, ndmf, type="l", lty="11", lend=1, lwd=c(4,1,1), col="red" )
+ matlines( 40+d.pt, d40m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 50+d.pt, d50m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 60+d.pt, d60m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 70+d.pt, d70m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 40+d.pt, d40f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ matlines( 50+d.pt, d50f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ matlines( 60+d.pt, d60f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ matlines( 70+d.pt, d70f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ axis( side=2, at=ylb[wlb],
+       labels=c("0.01","0.1","1","10","100","1000","10,000")[wlb] )
+ axis( side=2, at=ytc[wtc], labels=NA, tcl=-0.3 )
+ mtext( "BAA rates (per 1,000 PY)", side=2, outer=F, line=2.5, las=0 )
+ mtext( "a", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+
+ # The second plot of the RRs
+ d40m <- ci.exp( F2, ctr.mat=CM(40,"M")-CM(40,"M",FALSE) )
+ d50m <- ci.exp( F2, ctr.mat=CM(50,"M")-CM(50,"M",FALSE) )
+ d60m <- ci.exp( F2, ctr.mat=CM(60,"M")-CM(60,"M",FALSE) )
+ d70m <- ci.exp( F2, ctr.mat=CM(70,"M")-CM(70,"M",FALSE) )
+ d40f <- ci.exp( F2, ctr.mat=CM(40,"F")-CM(40,"F",FALSE) )
+ d50f <- ci.exp( F2, ctr.mat=CM(50,"F")-CM(50,"F",FALSE) )
+ d60f <- ci.exp( F2, ctr.mat=CM(60,"F")-CM(60,"F",FALSE) )
+ d70f <- ci.exp( F2, ctr.mat=CM(70,"F")-CM(70,"F",FALSE) )
+ matplot( a.pt, ndmm, type="n",
+          log="y", xlim=c(40,90), ylim=ylrr,
+          xlab="Age (years)", ylab="" )
+ matlines( 40+d.pt, d40m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 50+d.pt, d50m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 60+d.pt, d60m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 70+d.pt, d70m, type="l", lty=1, lwd=c(4,1,1), col="blue" )
+ matlines( 40+d.pt, d40f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ matlines( 50+d.pt, d50f, type="l", lty=1, lwd=c(4,1,1), col="red" )
```

```
+ matlines( 60+d.pt, d60f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ matlines( 70+d.pt, d70f, type="l", lty=1, lwd=c(4,1,1), col="red" )
+ axis( side=2, at=outer(1:9,10^c(0:1),"*"), labels=NA, tcl=-0.3 )
+ mtext( "DM vs non-DM HR of BAA", side=2, outer=F, line=2.5, las=0 )
+ mtext( "b", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+ }
> plf( ylrt=c(0.02,6),
+      ylrr=c(1,100) )


> plf( ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> postscript( "./graph/Fig3-BAA.eps", height=5, width=7.5 )
> plf( ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> dev.off()

pdf
  2
```

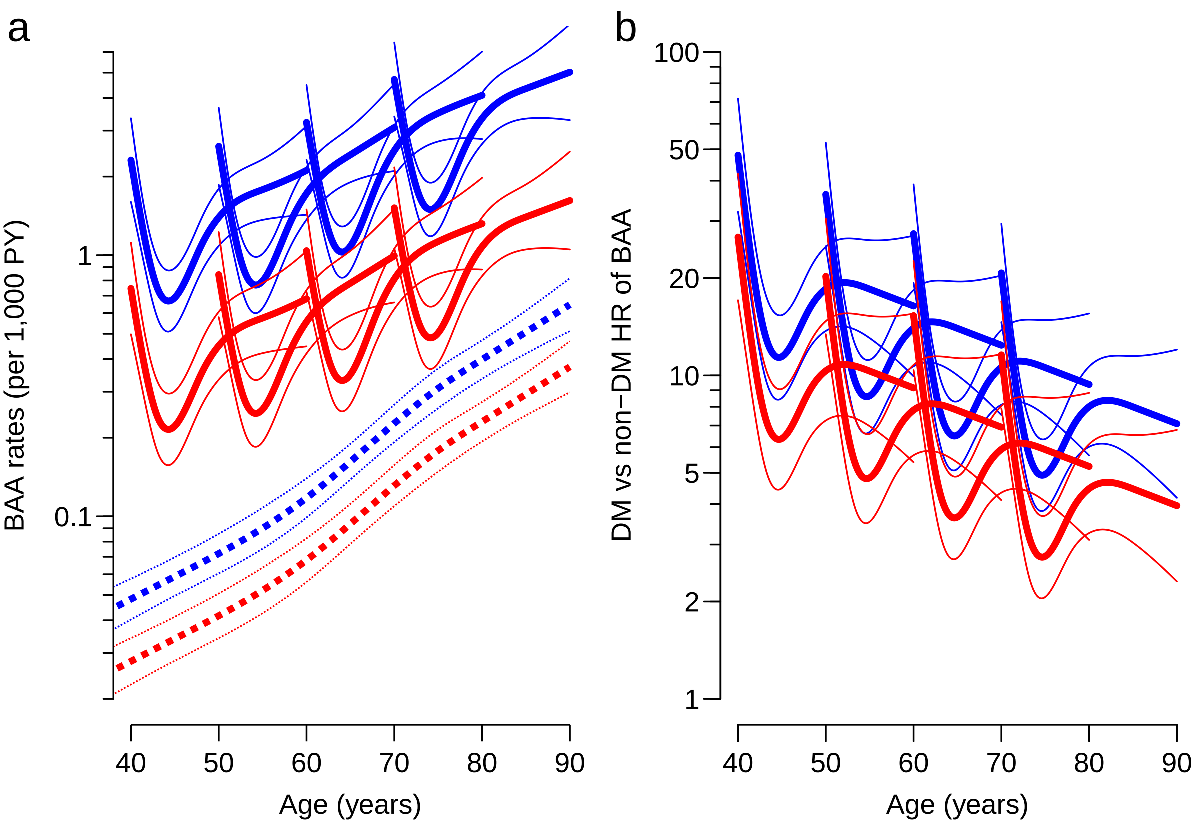

Figure 3.1: *Left: BAA amputation rates among persons without diabetes (broken lines) and persons with diabetes diagnosed in ages 40, 50, 60 and 70 (full lines). Right: Rate ratio of foot amputation between persons with and without diabetes for ages at diabetes diagnosis 40,...,70.*
Blue: men, red: women.

In figure 3.1 is seen that overall foot amputation rates increase by age; that rates among diabetes patients are 10–20 times higher than among non-diabetic patients, but less steeply increasing by age; the RR relative to persons without diabetes is generally decreasing with age. The same phenomenon as in many other co-morbidity studies is also seen, namely

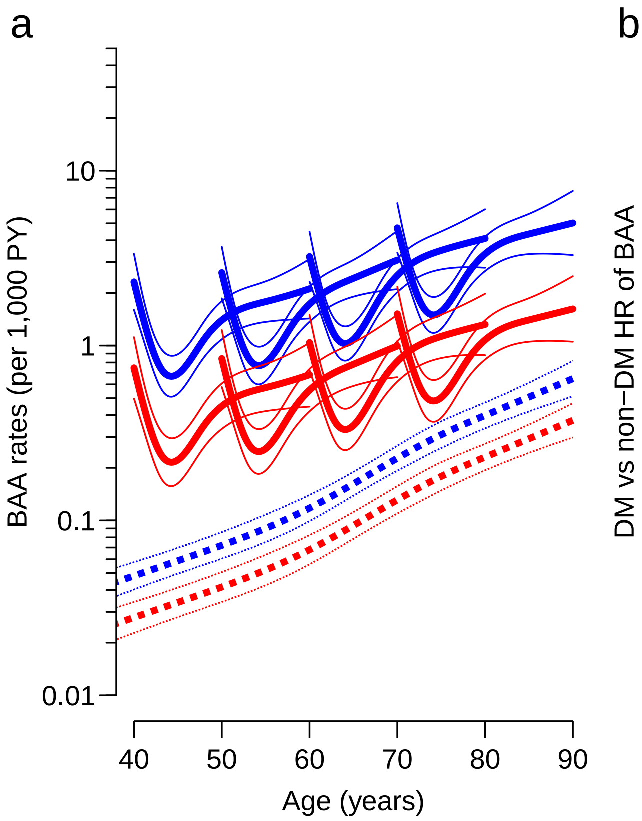

Figure 3.2: *Same figure as 3.1 but with axes scaled as the figures for BKA and AKA.*

that the frequency of amputations is high just after diagnosis of diabetes. It looks as if the amputation rates among diabetes patients after 10+ years of duration are slightly higher the younger the age at diagnosis.

3.4 BKA amputations

In the "no DM" dataset the response variables are now `K.noDM` and `Y`, and the explanatory variables are as in the knee case.

From the cohort dataset we extract the same variables, but in this case age and period represent the *actual* age at the start of an individual piece of follow-up, so here we add half of the interval length, `lex.dur`. But we must also include those who are already in the "BAA" states, and include another variable, namely the below ankle (BAA) amputation status. Also note that the dataset `Lap` contains some follow-up without diabetes (among the non-diabetic amputees), where the value of the timescale `dur` (diabetes duration) is `NA`, so we also set these to 0 in the analysis dataset.

The two parts of the dataset are then joined:

```
> anaK <- rbind( with( FUnoDM,
+                data.frame( A = A+0.5,
+                            P = P+0.5,
+                        DMdur = 0,
+                          sex = sex,
+                          amp = "No",
+                            D = K.noDM,
+                            Y = Y ) ),
+                with( subset( Lap, lex.Cst %in% c("DM","BAA","BAA(DM)") ),
+                data.frame( A = age+lex.dur/2,
+                            P = per+lex.dur/2,
+                        DMdur = pmax( 0, dur+lex.dur/2, na.rm=TRUE ),
+                          sex = sex,
+                          amp = lex.Cst,
+                            D = as.numeric(lex.Xst %in% c("BKA","BKA(DM)") ),
+                            Y = lex.dur ) ) )
> anaK$amp <- factor( anaK$amp )
> str( anaK )
'data.frame':        604760 obs. of  7 variables:
 $ A    : num  15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 ...
 $ P    : num  1996 1996 1998 1998 1998 ...
 $ DMdur: num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ amp  : Factor w/ 4 levels "No","DM","BAA",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ D    : num  0 0 1 0 0 0 0 0 0 0 ...
 $ Y    : num  2649 2568 2516 2402 2434 ...

> round( ftable( addmargins(
+ abind( with( anaK, table( sex, amp, D ) ),
+        Ktab <- xtabs( cbind(D,Y) ~ sex + amp, data=anaK ) ), 1 ) ) )

                   0       1       2       3       D       Y

M   No          1125     195      20       4     247 2919600
    DM        299696     170       0       0     170  143240
    BAA         2888      30       0       0      30    1350
    BAA(DM)     2181     122       0       0     122     989
F   No          1210     120      13       1     149 3039417
    DM        293979     110       0       0     110  140906
    BAA         2054      16       0       0      16     957
    BAA(DM)      786      40       0       0      40     348
Sum No          2335     315      33       5     396 5959018
    DM        593675     280       0       0     280  284146
    BAA         4942      46       0       0      46    2307
    BAA(DM)     2967     162       0       0     162    1337

> PKtab <- xtabs( cbind(D,Y) ~ floor(P) + floor(A) + DM,
+                 data = transform( anaK,
+                                   DM=Relevel( amp,
+                                               list("No"=c(1,3),
```

```
+                                                     "Yes"=c(2,4))) ) )
> save( PKtab, file="./data/PKtab.Rda" )
```

With this in place we can now model the BKA-amputation rates, initially with a proportional hazards model where the diabetes/amputation effect is the same regardless of sex, calendar time and age.

```
> system.time(
+ KO <- glm( D ~ Ns(A,kn=a.kn) +
+                 amp + sex,
+             offset = log(Y/1000),
+             family = poisson,
+               data = anaK ) )
   user  system elapsed
 11.747   1.564  13.351
> round( pKO <- ci.exp( KO ), 3 )
                  exp(Est.)    2.5%   97.5%
(Intercept)           0.045   0.038   0.052
Ns(A, kn = a.kn)1     5.311   4.167   6.770
Ns(A, kn = a.kn)2    13.696  10.462  17.929
Ns(A, kn = a.kn)3     4.520   3.660   5.582
ampDM                 7.064   6.029   8.277
ampBAA              155.687 114.394 211.886
ampBAA(DM)          671.250 554.423 812.695
sexF                  0.533   0.463   0.614
```

We see a very dramatic effect of previous BAA amputation, almost 100-fold in persons with DM and even more in persons with:

```
> CA <- rbind( c(1,0,0), c(0,-1,1), c(0,1,0), c(-1,0,1) )
> CA <- rbind( CA[1:2,], CA[2,]-CA[1,],
+              CA[3:4,], CA[4,]-CA[3,] )
> rownames( CA ) <- c("No Amp:  DM vs. no",
+                     "    Amp:  DM vs. no",
+                     "             Ratio:",
+                     " No DM: Amp vs. no",
+                     "    DM: Amp vs. no",
+                     "             Ratio:" )
> CA
                   [,1] [,2] [,3]
No Amp:  DM vs. no    1    0    0
   Amp:  DM vs. no    0   -1    1
            Ratio:   -1   -1    1
 No DM: Amp vs. no    0    1    0
    DM: Amp vs. no   -1    0    1
            Ratio:   -1   -1    1
> round( ci.exp( KO, subset="amp", ctr.mat=CA ), 3 )
                   exp(Est.)    2.5%   97.5%
No Amp:  DM vs. no     7.064   6.029   8.277
   Amp:  DM vs. no     4.312   3.103   5.990
            Ratio:     0.610   0.424   0.878
 No DM: Amp vs. no   155.687 114.394 211.886
    DM: Amp vs. no    95.025  78.151 115.542
            Ratio:     0.610   0.424   0.878
```

We see there is an interaction between previous amputation and diabetes status, previous amputation carries a 150-fold risk for persons without diabetes, but only a 100-fold risk for persons with diabetes; so the combined effect of diabetes and previous amputation is not the product of the two effects.

As before we also see a substantial sex-effect — women have an amputation incidence which is half that of men, and persons with diabetes have an about 10-fold increased rate of amputations.

Then we expand the model with the DM duration and the interaction between sex and DM/amputation:

```
> K1 <- update( K0, . ~ . + Ns(DMdur,kn=d.kn)
+                        - amp + amp:sex )
> round( pK1 <- ci.exp( K1 ), 3 )
                      exp(Est.)    2.5%    97.5%
(Intercept)               0.046   0.039    0.055
Ns(A, kn = a.kn)1         5.250   4.116    6.697
Ns(A, kn = a.kn)2        13.610  10.392   17.825
Ns(A, kn = a.kn)3         4.439   3.592    5.486
sexF                      0.489   0.398    0.599
Ns(DMdur, kn = d.kn)1     2.157   1.529    3.043
Ns(DMdur, kn = d.kn)2     0.621   0.273    1.411
Ns(DMdur, kn = d.kn)3     2.226   1.713    2.893
sexM:ampDM                7.650   5.181   11.295
sexF:ampDM                8.389   5.508   12.778
sexM:ampBAA             156.297 106.810  228.713
sexF:ampBAA             154.960  92.276  260.224
sexM:ampBAA(DM)         572.012 372.969  877.279
sexF:ampBAA(DM)         807.634 488.738 1334.605
```

In this case there does not seem to be any sex-interaction

We see that the diabetes effect among men is about twice as large than among women, and this is on top of the quite large effect of sex among non-diabetics.

```
> CM <- rbind( "   NoDM, M vs. F" = c(-1,0,0,0,0,0,0),
+              "     DM, M vs. F" = c(0,1,-1,0,0,0,0),
+              "    BAA, M vs. F" = c(0,0,0,1,-1,0,0),
+              "BAA(DM), M vs. F" = c(0,0,0,0,0,1,-1) )
> colnames( CM ) <- rownames( ci.exp( K1, subset="sex" ) )
> CM
                 sexF sexM:ampDM sexF:ampDM sexM:ampBAA sexF:ampBAA sexM:ampBAA(DM) sexF:ampBAA(DM)
   NoDM, M vs. F   -1          0          0           0           0               0               0
     DM, M vs. F    0          1         -1           0           0               0               0
    BAA, M vs. F    0          0          0           1          -1               0               0
BAA(DM), M vs. F    0          0          0           0           0               1              -1
```

```
> round( ci.exp( K1, subset="sex", ctr.mat=CM ), 3 )

                 exp(Est.)  2.5% 97.5%
   NoDM, M vs. F     2.047 1.669 2.511
     DM, M vs. F     0.912 0.666 1.249
    BAA, M vs. F     1.009 0.532 1.913
BAA(DM), M vs. F     0.708 0.469 1.069
```

From these estimates it looks as if there is only a difference between sexes among persons with neither DM nor previous amputation:

```
> round( Wald( K1, subset="sex", ctr.mat=rbind(CM[2,]-CM[3,],
+                                              CM[3,]-CM[4,]) ), 3 )
Chisq  d.f.     P
1.624 2.000 0.444
```

```
> round( Wald( K1, subset="sex", ctr.mat=CM[2:4,] ), 3 )

Chisq  d.f.     P
2.795 3.000 0.424
```

Thus neither are the sex differences different between the three levels, nor are they different from 1, so in principle we could model data using only a sex-effect among persons without DM and amputation. However, that is a model that seriously lacks a biological, let alone clinical, underpinning, so we will keep the interaction model.

We now further update the model with the interactions with age and calendar time:

```
> K2 <- update( K1, . ~ . + amp:I(P-2000) + amp:I(A-60) )
> round( pK2 <- ci.exp( K2 ), 3 )
                          exp(Est.)     2.5%    97.5%
(Intercept)                   0.101    0.078    0.131
Ns(A, kn = a.kn)1             1.639    1.132    2.375
Ns(A, kn = a.kn)2             1.308    0.675    2.535
Ns(A, kn = a.kn)3             1.497    1.050    2.136
sexF                          0.463    0.377    0.569
Ns(DMdur, kn = d.kn)1         1.967    1.384    2.796
Ns(DMdur, kn = d.kn)2         0.746    0.331    1.684
Ns(DMdur, kn = d.kn)3         5.016    3.666    6.864
sexM:ampDM                   14.890    9.831   22.552
sexF:ampDM                   16.460   10.415   26.015
sexM:ampBAA                 359.929  218.871  591.896
sexF:ampBAA                 383.235  196.411  747.765
sexM:ampBAA(DM)            1785.227 1114.582 2859.400
sexF:ampBAA(DM)            2722.046 1570.112 4719.114
ampNo:I(P - 2000)             0.977    0.956    0.998
ampDM:I(P - 2000)             0.849    0.824    0.875
ampBAA:I(P - 2000)            0.860    0.803    0.922
ampBAA(DM):I(P - 2000)        0.833    0.797    0.871
ampNo:I(A - 60)               1.061    1.045    1.077
ampDM:I(A - 60)               1.043    1.026    1.061
ampBAA:I(A - 60)              1.037    1.013    1.061
ampBAA(DM):I(A - 60)          1.000    1.000    1.000
> round( (ci.exp( K2, subset="P" )-1)*100, 1 )

                          exp(Est.)  2.5% 97.5%
ampNo:I(P - 2000)              -2.3  -4.4  -0.2
ampDM:I(P - 2000)             -15.1 -17.6 -12.5
ampBAA:I(P - 2000)            -14.0 -19.7  -7.8
ampBAA(DM):I(P - 2000)        -16.7 -20.3 -12.9
> save( Ktab, pK0, pK1, pK2, file="./data/Kmod.Rda" )
```

From the estimates we see that there is only a small change in the BKA amputation rates for persons without diabetes and previous amputation (annual change $-3.5\%(-5.5 - - - 1.4)\%$), whereas the the change for patients with either diabetes or previous amputation is dramatic, an annual decline of about 15% in all three groups.

There is also a significant age-interaction; the amputation rates among persons without diabetes increase steeper (3% per year) than among persons with diabetes:

```
> round( (ci.exp( K2, subset="I\\(A" )-1)*100, 1 )
                        exp(Est.) 2.5% 97.5%
ampNo:I(A - 60)               6.1  4.5   7.7
ampDM:I(A - 60)               4.3  2.6   6.1
ampBAA:I(A - 60)              3.7  1.3   6.1
ampBAA(DM):I(A - 60)          0.0  0.0   0.0
> ( CM <- rbind( "      DM vs noDM"=c(-1,1,0,0),
+                "BAA     vs noDM"=c(-1,0,1,0),
+                "BAA(DM) vs noDM"=c(-1,0,0,1) ) )
                  [,1] [,2] [,3] [,4]
     DM vs noDM     -1    1    0    0
BAA     vs noDM     -1    0    1    0
BAA(DM) vs noDM     -1    0    0    1
```

```
> round( (ci.exp( K2, subset="I\\(A", ctr.mat=CM )-1)*100, 1 )
```

```
                exp(Est.) 2.5% 97.5%
     DM vs noDM      -1.7 -2.8  -0.5
BAA     vs noDM      -2.3 -4.2  -0.3
BAA(DM) vs noDM      -5.7 -7.1  -4.3
```

There are only 3 effects here, the 4th linear effect is aliased with the natural spline in age. We see that the more severe the condition, the steeper the descent by age.

In order to show these effects we can show the amputation rates for men, resp. women for persons diagnosed in ages 40, 50, 60 and 70 in 2000, with and without diabetes and with and without previous BAA amputation, and as compared to the rates among persons without diabetes.

```
> col.int <-
+ function( clr, n ) # color-interpolation
+ rgb( cbind( seq(1,0,,n),
+             seq(0,1,,n) ) %*% t(col2rgb(clr[1:2])),
+      maxColorValue = 255 )
> clr <- c("forestgreen",col.int(c("orange","red"),3))
> CM <-
+ function( A, sex, DM=TRUE, BAA=FALSE, durt=d.pt )
+ cbind(1,Ns(A+durt,kn=a.kn), sex=="F",
+         Ns(  durt,kn=d.kn)*DM,
+        (sex=="M")*( DM)*(!BAA), (sex=="F")*( DM)*(!BAA), # DM:amp
+        (sex=="M")*(!DM)*( BAA), (sex=="F")*(!DM)*( BAA), # DM:amp
+        (sex=="M")*( DM)*( BAA), (sex=="F")*( DM)*( BAA), # DM:amp
+          0* durt *(!DM)*(!BAA), # P:amp
+          0* durt *( DM)*(!BAA), # P:amp
+          0* durt *(!DM)*( BAA), # P:amp
+          0* durt *( DM)*( BAA), # P:amp
+       (A-60+durt)*(!DM)*(!BAA), # A:amp
+       (A-60+durt)*( DM)*(!BAA), # A:amp
+       (A-60+durt)*(!DM)*( BAA), # A:amp
+       (A-60+durt)*( DM)*( BAA)  # A:amp
+       )
> az <- rep(0,length(a.pt))
> plk <- function( allc=TRUE,
+                  ylrt=c(0.01,1000),
+                  ylrr=c(1,100) )
+ {
+ par( mfrow=c(1,2), mar=c(3,3,1,1), oma=c(0,1,0,0), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
+
+ # First plot of the amputation rates in the 4 groups:
+ matplot( NA,
+          type="n", lty=1, lwd=c(4,1,1), col="forestgreen",
+          log="y", xlim=c(40,90), ylim=ylrt,
+          xlab="Age (years)", ylab="", yaxt="n" )
+ ylb  <- 10^c(-2:4)
+ wlb  <- ( ylb>=ylrt[1] & ylb<=ylrt[2] )
+ ytc  <- as.vector(outer(1:9,10^c(-2:1),"*"))
+ wtc  <- ( ytc>=ylrt[1] & ytc<=ylrt[2] )
+ axis( side=2, at=ylb[wlb],
+       labels=c("0.01","0.1","1","10","100","1000","10,000")[wlb] )
+ axis( side=2, at=ytc[wtc], labels=NA, tcl=-0.3 )
+ mtext( "BKA rates (per 1,000 PY)", side=2, outer=F, line=2.5, las=0 )
+ for( A in 4:7*10 )
+ matlines( A+d.pt, cbind( ci.exp( K2, ctr.mat=CM(A,"M",1,0,d.pt) ),
+                if( allc ) ci.exp( K2, ctr.mat=CM(A,"M",1,1,d.pt) ),
+                           ci.exp( K2, ctr.mat=CM(A,"F",1,0,d.pt) ),
+                if( allc ) ci.exp( K2, ctr.mat=CM(A,"F",1,1,d.pt) ) ),
+           type="l", lty=1, lwd=c(4,1,1),
+           col=rep(c("blue","red"),each=6/(2-allc)) )
```

```
+ matlines(   a.pt, cbind( ci.exp( K2, ctr.mat=CM(a.pt,"M",0,0,az) ),
+               if( allc ) ci.exp( K2, ctr.mat=CM(a.pt,"M",0,1,az) ),
+                          ci.exp( K2, ctr.mat=CM(a.pt,"F",0,0,az) ),
+               if( allc ) ci.exp( K2, ctr.mat=CM(a.pt,"F",0,1,az) ) ),
+           type="l", lty="11", lend=1, lwd=c(4,1,1),
+           col=rep(c("blue","red"),each=6/(2-allc)) )
+ mtext( "a", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+
+ # The second plot of the RRs reltive to non-dm
+ matplot( a.pt, a.pt, type="n",
+          log="y", xlim=c(40,90), ylim=ylrr,
+          xlab="Age (years)", ylab="" )
+ axis( side=2, at=outer(1:9,10^c(0:1),"*"), labels=NA, tcl=-0.3 )
+ mtext( "DM vs non-DM HR of BKA",
+        side=2, outer=F, line=2.5, las=0 )
+ for( A in 4:7*10 )
+ matlines( A+d.pt, cbind( ci.exp( K2, ctr.mat=CM(A,"M",1,0,d.pt)-
+                                              CM(A,"M",0,0,d.pt)),
+               if( allc ) ci.exp( K2, ctr.mat=CM(A,"M",1,1,d.pt)-
+                                              CM(A,"M",0,1,d.pt) ),
+                          ci.exp( K2, ctr.mat=CM(A,"F",1,0,d.pt)-
+                                              CM(A,"F",0,0,d.pt) ),
+               if( allc ) ci.exp( K2, ctr.mat=CM(A,"F",1,1,d.pt)-
+                                              CM(A,"F",0,1,d.pt) ) ),
+           type="l", lty=1, lwd=c(4,1,1),
+           col=rep(c("blue","red"),each=6/(2-allc)) )
+ axis( side=2, at=outer(1:9,10^c(0:1),"*"), labels=NA, tcl=-0.3 )
+ mtext( "b", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+ # abline( h=1 )
+ }
> plk()
```

```
> plk( allc=FALSE,
+      ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> postscript( "./graph/Fig4-BKA.eps", height=5, width=7.5 )
> plk( allc=FALSE,
+      ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> dev.off()
```

```
pdf
  2
```

In figure 3.3 is seen that overall knee amputation rates increase by age in all four groups; that rates among diabetes patients are 5–10 times higher than among non-diabetic patients, but more steeply increasing by age/duration; the RR relative to persons without diabetes is generally decreasing with age at onset of diabetes. As for the foot amputation rates we also see the same phenomenon as in many other co-morbidity studies is also seen, namely that the frequency of amputations is high just after diagnosis of diabetes.

A previous foot amputation increases the rates enormously, but the diabetes-associated RR is smaller for persons with a previous foot amputation, than for people without.

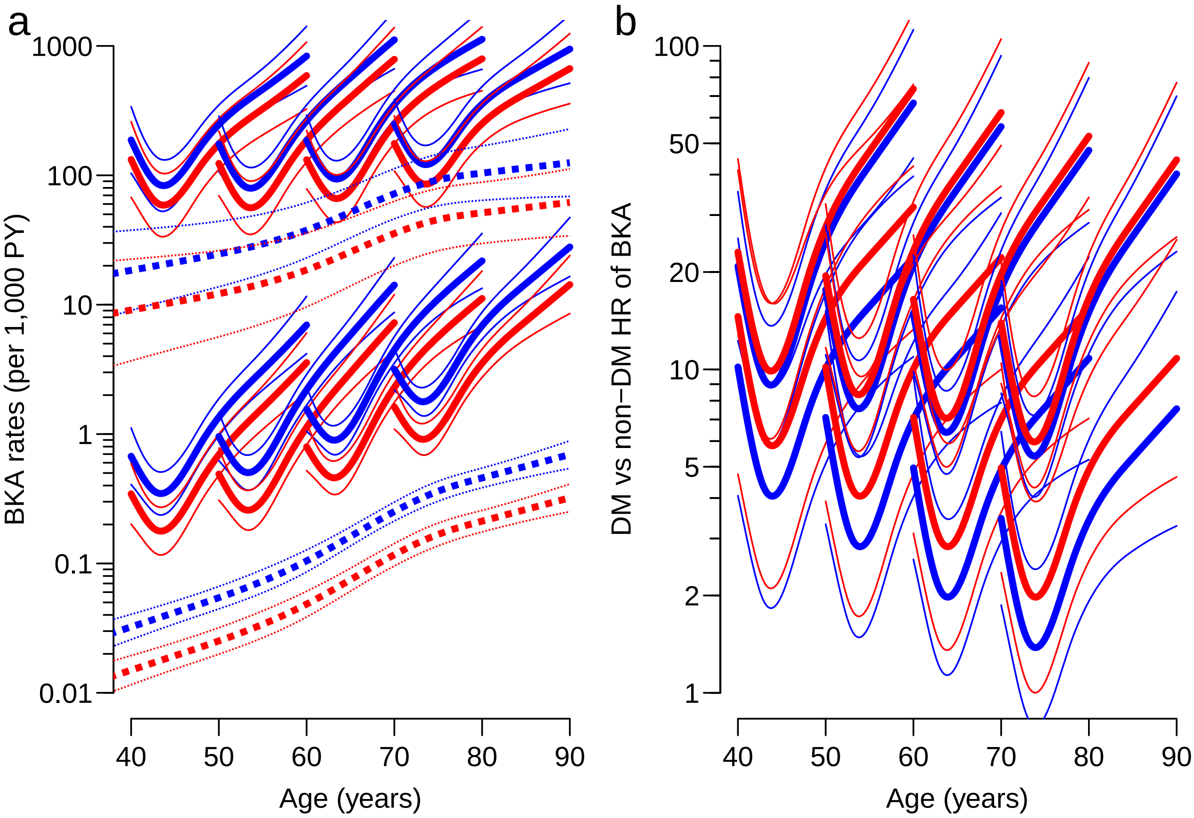

Figure 3.3: *Left: BKA amputation rates among persons without diabetes (broken lines) and persons with diabetes diagnosed in ages 40, 50, 60 and 70 (full lines). The upper set of curves are persons with a previous BAA amputation, the lower those without. Right: Rate ratio of knee amputation between persons with and without diabetes for ages at diabetes diagnosis 40,...,70. The* lower *set of curves are for persons with previous lower amputation, the* upper *for those without.*
Blue lines: men, red lines: women.

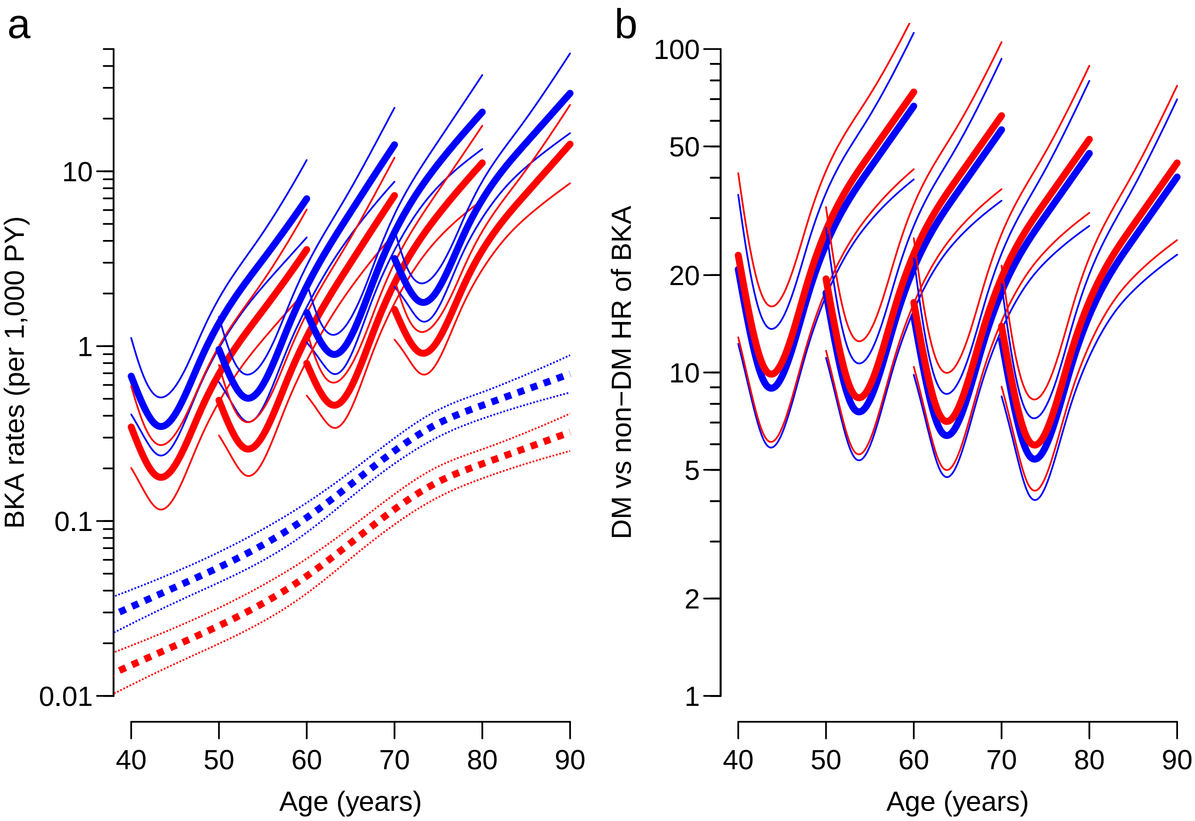

Figure 3.4: *As figure 3.3, but only for BKA without prior BAA.*

3.5 AKA amputations

In the no DM dataset (`FUnoDM`) the response variables are now `T.noDM` and `Y`, and the explanatory variables are as in the BKA case.

From the cohort dataset we extract the same variables, but in the cohort dataset, age (`A`) and period (`P`) represent the age and period at the start of each individual piece of follow-up, so here we add half of the interval length, `lex.dur`. But we must also include those who are already in the "BAA" or "BKA" states, and include two extra variables, namely the BAA and BKA amputation status. Also note that the dataset `Lap` contains follow-up without diabetes (among the non-diabetic amputees), where the value of the timescale `dur` (diabetes duration) is `NA`, so we also set these to 0 in the analysis dataset.

These two parts of the data sets are then merged (well, stacked) after aligning the variable names:

```
> anaT <- rbind( with( FUnoDM,
+                data.frame( A = A+0.5,
+                            P = P+0.5,
+                        DMdur = 0,
+                          sex = sex,
+                          amp = "No",
+                            D = T.noDM,
+                            Y = Y ) ),
+                with( subset( Lap, lex.Cst %in% c("DM","BAA","BAA(DM)",
+                                                  "BKA","BKA(DM)") ),
+                data.frame( A = age+lex.dur/2,
+                            P = per+lex.dur/2,
+                        DMdur = pmax( 0, dur+lex.dur/2, na.rm=TRUE ),
+                          sex = sex,
+                          amp = lex.Cst,
+                            D = as.numeric(lex.Xst %in% c("AKA","AKA(DM)") ),
+                            Y = lex.dur ) ) )
> anaT$amp <- factor( anaT$amp )
> round( ftable( addmargins(
+ abind( with( anaT, table( sex, amp, D ) ),
+        Ttab <- xtabs( cbind(D,Y) ~ sex + amp, data=anaT ) ), 1 ) ) )
                    0       1       2       3       D       Y

M   No           1162     160      17       5     209 2919600
    DM         299765     101       0       0     101  143240
    BAA          2900      18       0       0      18    1350
    BKA          1879      73       0       0      73     864
    BAA(DM)      2288      15       0       0      15     989
    BKA(DM)      1855      77       0       0      77     817
F   No           1138     175      30       1     238 3039417
    DM         293984     105       0       0     105  140906
    BAA          2055      15       0       0      15     957
    BKA           940      47       0       0      47     413
    BAA(DM)       813      13       0       0      13     348
    BKA(DM)       905      49       0       0      49     403
Sum No           2300     335      47       6     447 5959018
    DM         593749     206       0       0     206  284146
    BAA          4955      33       0       0      33    2307
    BKA          2819     120       0       0     120    1277
    BAA(DM)      3101      28       0       0      28    1337
    BKA(DM)      2760     126       0       0     126    1220

> anaT$Amp <- Relevel( anaT$amp,
+                      list( BAA=c("BAA","BAA(DM)"),
+                            BKA=c("BKA","BKA(DM)") ),
+                      first = FALSE )
> round( ftable( addmargins( xtabs( cbind(D,Y) ~ amp + Amp + sex,
+                                   data=anaT ),
```

```
+                      1:2 ),
+                col.vars=c(4,2),row.vars=c(3,1)) )
```

```
                       D                                        Y
             Amp      No       DM      BAA      BKA      Sum      No       DM      BAA      BKA      Sum
sex amp
M   No               209        0        0        0      209 2919600        0        0        0 2919600
    DM                 0      101        0        0      101       0   143240        0        0  143240
    BAA                0        0       18        0       18       0        0     1350        0    1350
    BKA                0        0        0       73       73       0        0        0      864     864
    BAA(DM)            0        0       15        0       15       0        0      989        0     989
    BKA(DM)            0        0        0       77       77       0        0        0      817     817
    Sum              209      101       33      150      493 2919600   143240     2338     1681 3066860
F   No               238        0        0        0      238 3039417        0        0        0 3039417
    DM                 0      105        0        0      105       0   140906        0        0  140906
    BAA                0        0       15        0       15       0        0      957        0     957
    BKA                0        0        0       47       47       0        0        0      413     413
    BAA(DM)            0        0       13        0       13       0        0      348        0     348
    BKA(DM)            0        0        0       49       49       0        0        0      403     403
    Sum              238      105       28       96      467 3039417   140906     1305      816 3182445
```

```
> PTtab <- xtabs( cbind(D,Y) ~ floor(P) + floor(A) + DM,
+                 data = transform( anaT,
+                                   DM=Relevel( amp,
+                                         list("No"=c(1,3,4),
+                                             "Yes"=c(2,5,6))) ) )
> save( PTtab, file="./data/PTtab.Rda" )
```

We now model the AKA occurrence (in `D`, `Y`) using natural splines for age and diabetes duration, using the same age and duration knots as previously. Note that we now use the `amp` factor so that we compare the AKA rates not only between persons with and without diabetes but also between different states of previous amputation:

```
> system.time(
+ t0 <- glm( D ~ Ns(A,kn=a.kn) +
+                amp + sex,
+            offset = log(Y/1000),
+            family = poisson,
+              data = anaT ) )
   user  system elapsed
 13.383   2.269  15.655
> round( ci.exp( t0 ), 3 )
                  exp(Est.)    2.5%   97.5%
(Intercept)           0.023   0.019   0.029
Ns(A, kn = a.kn)1     8.973   6.946  11.591
Ns(A, kn = a.kn)2    43.189  29.007  64.305
Ns(A, kn = a.kn)3     6.939   5.528   8.710
ampDM                 4.026   3.405   4.761
ampBAA               85.623  60.011 122.165
ampBKA              465.088 378.104 572.083
ampBAA(DM)           98.010  66.710 143.997
ampBKA(DM)          387.164 315.549 475.032
sexF                  0.872   0.766   0.994
```

We see a very dramatic effect of previous amputations, regardless of whether DM is present or not:

```
> round( ci.exp( t0, subset="BAA" ), 3 )
           exp(Est.)   2.5%   97.5%
ampBAA        85.623 60.011 122.165
ampBAA(DM)    98.010 66.710 143.997
> round(   Wald( t0, subset="BAA", ctr.mat=rbind(c(1,-1)) ), 3 )
```

```
Chisq  d.f.     P
0.275 1.000 0.600
```

```
> round( ci.exp( t0, subset="BKA" ), 3 )
```

```
           exp(Est.)    2.5%   97.5%
ampBKA       465.088 378.104 572.083
ampBKA(DM)   387.164 315.549 475.032
```

```
> round(   Wald( t0, subset="BKA", ctr.mat=rbind(c(1,-1)) ), 3 )
```

```
Chisq  d.f.     P
2.055 1.000 0.152
```

So we can use the `Amp` factor instead, so we provide a simplified overview of the dataset:

```
> round( ftable( addmargins( xtabs( cbind(D,Y) ~ Amp + sex,
+                                   data=anaT ),
+                            c(1,2) ),
+                col.vars=3:2) )
```

```
             D                         Y
    sex      M       F     Sum         M       F     Sum
Amp
No         209     238     447 2919600 3039417 5959018
DM         101     105     206  143240  140906  284146
BAA         33      28      61    2338    1305    3643
BKA        150      96     246    1681     816    2497
Sum        493     467     960 3066860 3182445 6249304
```

So we see our data has in total 965 events, half of which are among persons without either DM or previous amputations, and 40% of the rest among persons with diabetes alone. Among those with previous amputations it is only 20% that only have a BAA amputation.

```
> T0 <- update( t0, . ~ . - amp + Amp )
> round( pT0 <- ci.exp( T0 ), 3 )
```

```
                    exp(Est.)    2.5%   97.5%
(Intercept)             0.023   0.019   0.029
Ns(A, kn = a.kn)1       8.888   6.883  11.478
Ns(A, kn = a.kn)2      43.021  28.888  64.069
Ns(A, kn = a.kn)3       6.893   5.492   8.652
sexF                    0.873   0.766   0.994
AmpDM                   4.028   3.407   4.764
AmpBAA                 90.935  69.400 119.151
AmpBKA                421.829 358.273 496.659
```

So we see there is a 4-fold risk of AKA associated with diabetes alone, but a 90-fold increase with a previous BAA, and a 400-fold increase associated with a previous BKA.

As before we also see a sex-effect — but smaller than for the other amputation types, women only have a 13% smaller rate, just significant.

Then we expand the model with the DM duration and the interaction between sex and DM/amputation. However, since we pooled the amputation groups we should not include the diabetes duration as a predictor for persons in these groups. This is most effectively handled by updating the frame:

```
> with( anaT, tapply( DMdur, Amp, range ) )
```

```
$No
[1] 0 0

$DM
[1] 3.422313e-04 2.578816e+01
```

```
$BAA
[1]  0.00000 21.75291

$BKA
[1]  0.00000 21.75599

> anaT <- transform( anaT, DMdur = DMdur * (Amp=="DM") )
> with( anaT, tapply( DMdur, Amp, range ) )

$No
[1] 0 0

$DM
[1] 3.422313e-04 2.578816e+01

$BAA
[1] 0 0

$BKA
[1] 0 0

> T1 <- update( T0, . ~ . + Ns(DMdur,kn=d.kn)
+                         - Amp + Amp:sex,
+               data = anaT )
> round( pT1 <- ci.exp( T1 ), 3 )

                      exp(Est.)    2.5%   97.5%
(Intercept)               0.024   0.019   0.030
Ns(A, kn = a.kn)1         8.905   6.894  11.502
Ns(A, kn = a.kn)2        42.890  28.794  63.888
Ns(A, kn = a.kn)3         6.826   5.437   8.570
sexF                      0.857   0.711   1.033
Ns(DMdur, kn = d.kn)1     1.685   1.029   2.760
Ns(DMdur, kn = d.kn)2     0.932   0.300   2.894
Ns(DMdur, kn = d.kn)3     3.042   2.138   4.328
sexM:AmpDM                3.969   2.346   6.715
sexF:AmpDM                3.667   2.162   6.220
sexM:AmpBAA              85.096  58.855 123.037
sexF:AmpBAA              99.112  66.869 146.901
sexM:AmpBKA             407.682 328.995 505.188
sexF:AmpBKA             446.819 350.914 568.935
```

In this case there does not seem to be any sex-interaction with amputation status:

```
> ( CM <- rbind( "NoDM, M vs. F" = c(-1,0,0,0,0,0,0),
+                "  DM, M vs. F" = c(0,1,-1,0,0,0,0),
+                "BAA, M vs. F" = c(0,0,0,1,-1,0,0),
+                "BKA, M vs. F" = c(0,0,0,0,0,1,-1) ) )
              [,1] [,2] [,3] [,4] [,5] [,6] [,7]
NoDM, M vs. F   -1    0    0    0    0    0    0
  DM, M vs. F    0    1   -1    0    0    0    0
BAA, M vs. F     0    0    0    1   -1    0    0
BKA, M vs. F     0    0    0    0    0    1   -1

> round( ci.exp( T1, subset="sex", ctr.mat=CM ), 3 )

              exp(Est.)  2.5% 97.5%
NoDM, M vs. F     1.167 0.968 1.407
  DM, M vs. F     1.082 0.777 1.507
BAA, M vs. F      0.859 0.502 1.469
BKA, M vs. F      0.912 0.665 1.253
```

From these estimates it does not look as if there is any difference between sexes among persons with neither DM nor previous amputation; here are the tests for differences between the sex-effects and the test for all sex-effects being 1.

```
> round( w.eq <- Wald( T1, subset="sex", ctr.mat=rbind(CM[1,]-CM[2,],
+                                                      CM[2,]-CM[3,],
+                                                      CM[3,]-CM[4,]) ), 3 )
Chisq  d.f.     P
1.827 3.000 0.609
> round( w.a0 <- Wald( T1, subset="sex", ctr.mat=CM ), 3 )
Chisq  d.f.     P
5.559 4.000 0.235
> w.0 <- w.a0 - w.eq
> w.0[3] <- 1-pchisq( w.0[1], w.0[2] )
> round( w.0, 3 )
Chisq  d.f.     P
3.732 1.000 0.053
```

Thus neither are the sex differences different between the four levels, nor are they different from 1, so in principle we could model data without a sex-effect, however we shall keep a separate sex effect in the model, because the *difference* in test statistics is (almost) significant, and that is essentially the test of whether the *pooled* sex effect is significantly different from 0:

```
> T1 <- update( T0, . ~ . + Ns(DMdur,kn=d.kn) )
> round( ci.exp( T1 ), 3 )
                      exp(Est.)    2.5%   97.5%
(Intercept)               0.024   0.019   0.029
Ns(A, kn = a.kn)1         8.889   6.883  11.480
Ns(A, kn = a.kn)2        42.863  28.779  63.838
Ns(A, kn = a.kn)3         6.822   5.434   8.564
sexF                      0.869   0.763   0.990
AmpDM                     3.826   2.320   6.310
AmpBAA                   91.147  69.561 119.432
AmpBKA                  423.094 359.331 498.173
Ns(DMdur, kn = d.kn)1     1.680   1.026   2.751
Ns(DMdur, kn = d.kn)2     0.924   0.298   2.869
Ns(DMdur, kn = d.kn)3     3.027   2.128   4.305
```

We now further update the model with the interactions with age and calendar time:

```
> T2 <- update( T1, . ~ . + Amp:I(P-2000) + Amp:I(A-60) )
> round( pT2 <- ci.exp( T2 ), 3 )
                      exp(Est.)     2.5%    97.5%
(Intercept)               0.042    0.033    0.054
Ns(A, kn = a.kn)1         1.913    1.424    2.571
Ns(A, kn = a.kn)2         2.333    1.434    3.795
Ns(A, kn = a.kn)3         1.551    1.147    2.098
sexF                      0.884    0.775    1.009
AmpDM                     6.796    3.830   12.056
AmpBAA                  248.809  144.642  427.993
AmpBKA                 1758.288 1331.661 2321.595
Ns(DMdur, kn = d.kn)1     1.664    1.015    2.729
Ns(DMdur, kn = d.kn)2     0.953    0.307    2.960
Ns(DMdur, kn = d.kn)3     3.325    2.282    4.846
AmpNo:I(P - 2000)         1.017    0.996    1.037
AmpDM:I(P - 2000)         0.974    0.941    1.008
AmpBAA:I(P - 2000)        0.910    0.858    0.965
AmpBKA:I(P - 2000)        0.976    0.947    1.005
AmpNo:I(A - 60)           1.086    1.074    1.099
AmpDM:I(A - 60)           1.057    1.040    1.074
AmpBAA:I(A - 60)          1.053    1.028    1.078
AmpBKA:I(A - 60)          1.000    1.000    1.000
> round( (ci.exp( T2, subset="P" )-1)*100, 1 )
```

```
                    exp(Est.)  2.5% 97.5%
AmpNo:I(P - 2000)        1.7  -0.4   3.7
AmpDM:I(P - 2000)       -2.6  -5.9   0.8
AmpBAA:I(P - 2000)      -9.0 -14.2  -3.5
AmpBKA:I(P - 2000)      -2.4  -5.3   0.5
```

```
> save( Ttab, pT0, pT1, pT2, file="./data/Tmod.Rda" )
```

From the estimates we see that there is essentially no change in the AKA rates for persons without previous amputation, whereas the the change for patients with previous amputation is dramatic, an annual decline of about 5-10% in both groups — they are formally not significantly different (P=0.066):

```
> round( ci.lin( T2, subset="P", ctr.mat=rbind(c(0,0,1,-1)) ), 3 )
     Estimate StdErr      z     P   2.5%  97.5%
[1,]   -0.069  0.033 -2.072 0.038 -0.135 -0.004
> round( ci.exp( T2, subset="P", ctr.mat=rbind(c(0,0,1,-1)) ), 3 )
     exp(Est.)  2.5% 97.5%
[1,]     0.933 0.874 0.996
```

There is also a significant age-interaction; the amputation rates among persons without diabetes increase steeper (9% per year) than among persons with diabetes:

```
> round( (ci.exp( T2, subset="I\\(A" )-1)*100, 1 )
                  exp(Est.) 2.5% 97.5%
AmpNo:I(A - 60)         8.6  7.4   9.9
AmpDM:I(A - 60)         5.7  4.0   7.4
AmpBAA:I(A - 60)        5.3  2.8   7.8
AmpBKA:I(A - 60)        0.0  0.0   0.0
> ( CM <- rbind( " DM vs noDM"=c(-1,1,0,0),
+                "BAA vs noDM"=c(-1,0,1,0),
+                "BKA vs noDM"=c(-1,0,0,1) ) )
            [,1] [,2] [,3] [,4]
 DM vs noDM   -1    1    0    0
BAA vs noDM   -1    0    1    0
BKA vs noDM   -1    0    0    1
> round( (ci.exp( T2, subset="I\\(A", ctr.mat=CM )-1)*100, 1 )
            exp(Est.) 2.5% 97.5%
 DM vs noDM      -2.7 -4.1  -1.3
BAA vs noDM      -3.1 -5.3  -0.9
BKA vs noDM      -8.0 -9.0  -6.9
```

There are only 3 effects here, the 4th linear effect is aliased with the natural spline in age. We see that the more severe the condition, the steeper the descent by age relative to the no DM group.

In order to show these effects we can show the amputation rates for men, resp. women for persons diagnosed in ages 40, 50, 60 and 70 in 2000, with and without diabetes and with and without previous BAA, and as compared to the rates among persons without diabetes.

```
> col.int <-
+ function( clr, n ) # color-interpolation
+ rgb( cbind( seq(1,0,,n),
+             seq(0,1,,n) ) %*% t(col2rgb(clr[1:2])),
+      maxColorValue = 255 )
> clr <- c("forestgreen",col.int(c("orange","red"),3))
> round( ci.exp(T2), 3 )
```

```
                         exp(Est.)     2.5%    97.5%
(Intercept)                  0.042    0.033    0.054
Ns(A, kn = a.kn)1            1.913    1.424    2.571
Ns(A, kn = a.kn)2            2.333    1.434    3.795
Ns(A, kn = a.kn)3            1.551    1.147    2.098
sexF                         0.884    0.775    1.009
AmpDM                        6.796    3.830   12.056
AmpBAA                     248.809  144.642  427.993
AmpBKA                    1758.288 1331.661 2321.595
Ns(DMdur, kn = d.kn)1        1.664    1.015    2.729
Ns(DMdur, kn = d.kn)2        0.953    0.307    2.960
Ns(DMdur, kn = d.kn)3        3.325    2.282    4.846
AmpNo:I(P - 2000)            1.017    0.996    1.037
AmpDM:I(P - 2000)            0.974    0.941    1.008
AmpBAA:I(P - 2000)           0.910    0.858    0.965
AmpBKA:I(P - 2000)           0.976    0.947    1.005
AmpNo:I(A - 60)              1.086    1.074    1.099
AmpDM:I(A - 60)              1.057    1.040    1.074
AmpBAA:I(A - 60)             1.053    1.028    1.078
AmpBKA:I(A - 60)             1.000    1.000    1.000
```

```
> CM <-
+ function( A, sex, DM=TRUE, BAA=FALSE, BKA=FALSE, durt=d.pt )
+ cbind(1,Ns(A+durt,kn=a.kn), sex=="F",
+        ( DM)*(!BAA)*(!BKA), # DM:amp
+        (!DM)*( BAA)*(!BKA), # DM:amp
+        (!DM)*(!BAA)*( BKA), # DM:amp
+         Ns( durt,kn=d.kn)*DM,
+          0* durt *(!DM)*(!BAA)*(!BKA), # P:amp
+          0* durt *( DM)*(!BAA)*(!BKA), # P:amp
+          0* durt *(!DM)*( BAA)*(!BKA), # P:amp
+          0* durt *(!DM)*(!BAA)*( BKA), # P:amp
+       (A-60+durt)*(!DM)*(!BAA)*(!BKA), # A:amp
+       (A-60+durt)*( DM)*(!BAA)*(!BKA), # A:amp
+       (A-60+durt)*(!DM)*( BAA)*(!BKA), # A:amp
+       (A-60+durt)*(!DM)*(!BAA)*( BKA)  # A:amp
+       )
> az <- rep(0,length(a.pt))
> pla <- function( allc=TRUE,
+                  ylrt=c(0.01,200),
+                  ylrr=c(1,20000) )
+ {
+ par( mfrow=c(1,2), mar=c(3,3,1,1), oma=c(0,1,0,0), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
+
+ # First plot of the amputation rates in the 4 groups:
+ matplot( NA, type="n",
+          log="y", xlim=c(40,90), ylim=ylrt,
+          xlab="Age (years)", ylab="", yaxt="n" )
+ ylb  <- 10^c(-2:4)
+ wlb  <- ( ylb>=ylrt[1] & ylb<=ylrt[2] )
+ ytc  <- as.vector(outer(1:9,10^c(-2:1),"*"))
+ wtc  <- ( ytc>=ylrt[1] & ytc<=ylrt[2] )
+ axis( side=2, at=ylb[wlb],
+       labels=c("0.01","0.1","1","10","100","1000","10,000")[wlb] )
+ axis( side=2, at=ytc[wtc], labels=NA, tcl=-0.3 )
+ mtext( "AKA rates (per 1,000 PY)",
+        side=2, outer=F, line=2.5, las=0 )
+ mtext( "a", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+
+ # Non-diabetics
+ matlines(   a.pt, cbind( ci.exp( T2, ctr.mat=CM(a.pt,"M",0,0,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"F",0,0,0,az) ) ),
+           type="l", lty="11", lend=1, lwd=c(4,1,1),
+           col=rep(c("blue","red"),each=3) )
+ # Diabetes no previous amputation
+ for( A in 4:7*10 )
```

```
+ matlines( A+d.pt, cbind( ci.exp( T2, ctr.mat=CM(A,"M",1,0,0,d.pt) ),
+                          ci.exp( T2, ctr.mat=CM(A,"F",1,0,0,d.pt) ) ),
+           type="l", lty=1, lwd=c(4,1,1),
+           col=rep( c("blue","red"),each=3 ) )
+ # previous lower and middle amputation
+ if( allc )
+ matlines(   a.pt, cbind( ci.exp( T2, ctr.mat=CM(a.pt,"M",0,1,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"M",0,0,1,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"F",0,1,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"F",0,0,1,az) ) ),
+           type="l", lty=1, lwd=c(4,1,1),
+           col=rep( c("blue","red"),each=6 ) )
+
+ # The second plot of the RRs relative to non-dm
+ matplot( a.pt, a.pt,
+          type="n", lty=1, lwd=c(4,1,1), col="forestgreen",
+          log="y", xlim=c(40,90), ylim=ylrr,
+          xlab="Age (years)", ylab="" )
+ axis( side=2, at=outer(1:9,10^c(0:1),"*"), labels=NA, tcl=-0.3 )
+ mtext( "DM vs non-DM HR of AKA",
+        side=2, outer=F, line=2.5, las=0, adj=0.35 )
+ mtext( "b", side=2, at=10^par("usr")[4]*1.00, line=2.5, cex=1.5 )
+ for( A in 4:7*10 )
+ matlines( A+d.pt, cbind( ci.exp( T2, ctr.mat=CM(A,"M",1,0,0,d.pt)-
+                                             CM(A,"M",0,0,0,d.pt) ),
+                          ci.exp( T2, ctr.mat=CM(A,"F",1,0,0,d.pt)-
+                                             CM(A,"F",0,0,0,d.pt) ) ),
+           type="l", lty=1, lwd=c(4,1,1), col=gray(0.3) )
+ if( allc )
+ matlines(   a.pt, cbind( ci.exp( T2, ctr.mat=CM(a.pt,"M",0,1,0,az)-
+                                             CM(a.pt,"M",0,0,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"M",0,0,1,az)-
+                                             CM(a.pt,"M",0,0,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"F",0,1,0,az)-
+                                             CM(a.pt,"F",0,0,0,az) ),
+                          ci.exp( T2, ctr.mat=CM(a.pt,"F",0,0,1,az)-
+                                             CM(a.pt,"F",0,0,0,az) ) ),
+           type="l", lty=1, lwd=c(4,1,1), col=gray(0.3) )
+ # abline( h=1 )
+ }
> pla()
> # win.metafile("Fig5r.emf",height=6,width=9)
> # pla()
> # dev.off()
```

```
> pla( allc=FALSE,
+      ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> postscript( "./graph/Fig5-AKA.eps", height=5, width=7.5 )
> pla( allc=FALSE,
+      ylrt=c(0.01,50),
+      ylrr=c(1,100) )
> dev.off()
```

```
pdf
  2
```

```
> # win.metafile("Fig5.emf",height=6,width=9)
> # pla( allc=FALSE,
> #      ylrt=c(0.01,10),
> #      ylrr=c(1,100) )
> # dev.off()
```

In figure 3.6 is seen that overall AKA rates increase by age in all four groups; that rates among diabetes patients are 5–10 times higher than among non-diabetic patients, the RR

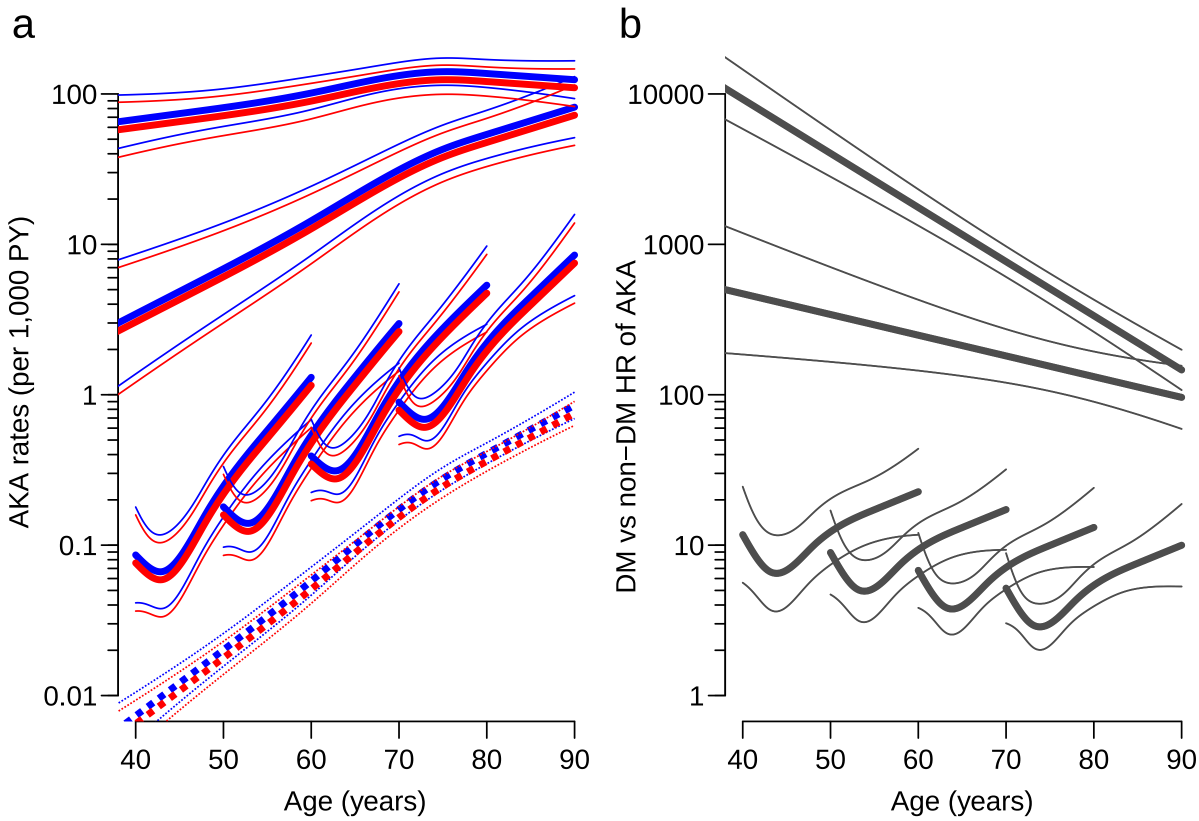

Figure 3.5: *Left: AKA amputation rates among persons without diabetes (broken lines), persons with diabetes diagnosed in ages 40, 50, 60 and 70 (orange), and persons with previous BAA or BKA (upper two curves). Right: Rate ratio of AKA between persons with and without diabetes for ages at diabetes diagnosis 40,...,70.*
Blue lines: men, red: women. The model assumes that the RRs for men and women are identical, hence the r.h.s. refer both *to men and women.*

relative to persons without diabetes is generally decreasing with age at onset of diabetes, but increasing by age *within* groups of similar onset age.

As seen for BKA previous BAA increases the rates enormously, and previous BKA even more so.

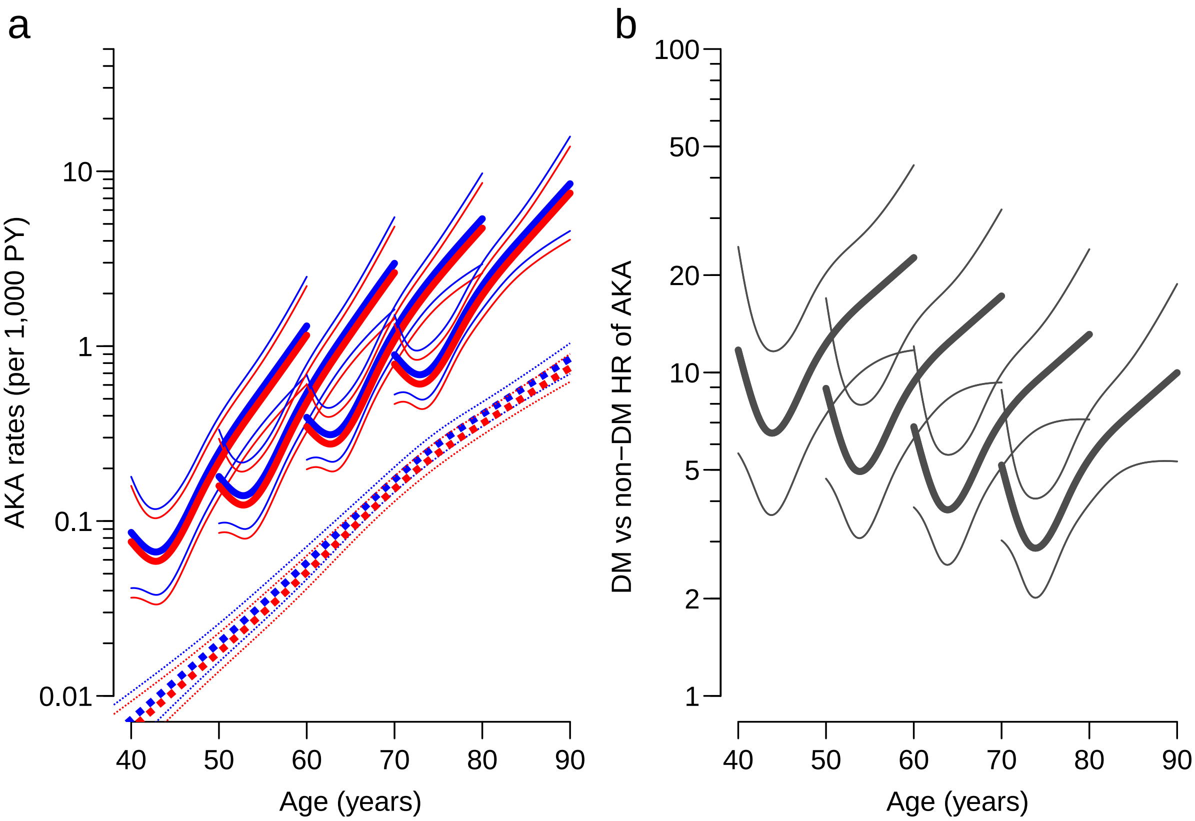

Figure 3.6: *As figure 3.5, but only for AKA without any prior amputation.*

3.6 Summary of models

Besides the number of events (amputations) and person-time in tabular form (it is in figure **??**, the two things are of interest to show together across amputations:

1. The RR between DM and non-DM from the model with DM×sex interaction.

2. The annual change in amputation rates for different groups of persons.

First we load the models so we have the relevant quantities at our disposal:

```
> library( Epi )
> clear()
> load( file = "./data/Fmod.Rda" )
> load( file = "./data/Kmod.Rda" )
> load( file = "./data/Tmod.Rda" )
```

3.6.1 Number of amputations and PY — table 1

Then we set up an arrays for the number of amputations and no. of events by sex, status ad outcome type. This is used to print the contents of table 1 here and also to output in `.csv`-files in different locales:

```
> Narr <- NArray( c( dimnames(Ttab),
+                    list(c("BAA","BKA","AKA")) ) )
> str( Narr )
 logi [1:2, 1:6, 1:2, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ sex: chr [1:2] "M" "F"
  ..$ amp: chr [1:6] "No" "DM" "BAA" "BKA" ...
  ..$    : chr [1:2] "D" "Y"
  ..$    : chr [1:3] "BAA" "BKA" "AKA"
> Narr[,  1:2   ,,"BAA"] <- Ftab
> Narr[,c(1:3,5),,"BKA"] <- Ktab
> Narr[,         ,,"AKA"] <- Ttab
> Narr[is.na(Narr)] <- 0
> Narr <- Narr[,c(1:3,5,4,6),,]
> ( dd <- ftable( addmargins(Narr[,,"D",]), col.vars=c(1,3) ) )
        sex  M                    F                    Sum
             BAA  BKA  AKA  Sum  BAA  BKA  AKA  Sum  BAA  BKA  AKA  Sum
amp
No           293  247  209  749  212  149  238  599  505  396  447 1348
DM           351  170  101  622  132  110  105  347  483  280  206  969
BAA            0   30   18   48    0   16   15   31    0   46   33   79
BAA(DM)        0  122   15  137    0   40   13   53    0  162   28  190
BKA            0    0   73   73    0    0   47   47    0    0  120  120
BKA(DM)        0    0   77   77    0    0   49   49    0    0  126  126
Sum          644  569  493 1706  344  315  467 1126  988  884  960 2832
> ( yy <- ftable( round( addmargins(Narr[,,"Y","AKA"]), 1 ), col.vars=1 ) )
        sex          M         F       Sum
amp
No         2919600.5 3039417.2 5959017.7
DM          143239.9  140906.4  284146.2
BAA           1349.5     957.0    2306.5
BAA(DM)        988.6     348.1    1336.7
BKA            863.8     412.9    1276.7
BKA(DM)        817.4     403.0    1220.5
Sum        3066859.7 3182444.7 6249304.3
```

```
> tab0 <- cbind(dd,yy)[,c(1:4,13,5:8,14,9:12,15)]
> colnames( tab0 ) <- paste( c("M:","","","","",
+                              "F:","","","","",
+                            "M+F:","","","",""),
+                            rep( c("BAA","BKA","AKA","Sum","P.Y"), 3 ),
+                            sep=""  )
> rownames( tab0 ) <- c( dimnames( Narr )[["amp"]], "Sum" )
> tab0

        M:BAA BKA AKA  Sum       P.Y F:BAA BKA AKA  Sum       P.Y M+F:BAA BKA AKA  Sum       P.Y
No        293 247 209  749 2919600.5   212 149 238  599 3039417.2     505 396 447 1348 5959017.7
DM        351 170 101  622  143239.9   132 110 105  347  140906.4     483 280 206  969  284146.2
BAA         0  30  18   48    1349.5     0  16  15   31     957.0       0  46  33   79    2306.5
BAA(DM)     0 122  15  137     988.6     0  40  13   53     348.1       0 162  28  190    1336.7
BKA         0   0  73   73     863.8     0   0  47   47     412.9       0   0 120  120    1276.7
BKA(DM)     0   0  77   77     817.4     0   0  49   49     403.0       0   0 126  126    1220.5
Sum       644 569 493 1706 3066859.7   344 315 467 1126 3182444.7     988 884 960 2832 6249304.3

> tabD <- rbind( apply( tab0[c(1,3,5),], 2, sum ),
+                apply( tab0[c(2,4,6),], 2, sum ) )
> tab1 <- rbind( tab0[-7,], tabD, tab0[7,] )
> rownames( tab1 )[c(1,7:9)] <- c("Well","No DM","Any DM","Total")
> tab1

        M:BAA BKA AKA  Sum       P.Y F:BAA BKA AKA  Sum       P.Y M+F:BAA BKA AKA  Sum       P.Y
Well      293 247 209  749 2919600.5   212 149 238  599 3039417.2     505 396 447 1348 5959017.7
DM        351 170 101  622  143239.9   132 110 105  347  140906.4     483 280 206  969  284146.2
BAA         0  30  18   48    1349.5     0  16  15   31     957.0       0  46  33   79    2306.5
BAA(DM)     0 122  15  137     988.6     0  40  13   53     348.1       0 162  28  190    1336.7
BKA         0   0  73   73     863.8     0   0  47   47     412.9       0   0 120  120    1276.7
BKA(DM)     0   0  77   77     817.4     0   0  49   49     403.0       0   0 126  126    1220.5
No DM     293 277 300  870 2921813.8   212 165 300  677 3040787.1     505 442 600 1547 5962600.9
Any DM    351 292 193  836  145045.9   132 150 167  449  141657.5     483 442 360 1285  286703.4
Total     644 569 493 1706 3066859.7   344 315 467 1126 3182444.7     988 884 960 2832 6249304.3

> write.csv2( tab1, file="./data/Tab1.csv" )
```

3.6.2 Estimates of RR — Table 2

We now retrieve the RR estimates from the different models and show then in tabular form for table 2:

```
> DMRR <- NArray( list( amp = c("BAA","BKA","AKA"),
+                       sex = c("M","F","M+F"),
+                      from = c("DM","BAA","BAA(DM)","BKA","BKA(DM)"),
+                      what = c("RR","Lo","Up") ) )
> DMRR["BAA","M+F","DM",] <- pF0[     "DMYes",                  ]
> DMRR["BAA",   -3,"DM",] <- pF1[grep("DMYes",rownames(pF1)),]
> DMRR["BKA","M+F",1:3,] <- pK0[grep("amp",rownames(pK0)),]
> DMRR["BKA","M"  ,1:3,] <- pK1[grep("M:amp",rownames(pK1)),]
> DMRR["BKA","F"  ,1:3,] <- pK1[grep("F:amp",rownames(pK1)),]
> DMRR["AKA","M+F",c(1,2,4),] <- pT0[grep(  "Amp",rownames(pT0)),]
> DMRR["AKA","M"  ,c(1,2,4),] <- pT1[grep("M:Amp",rownames(pT1)),]
> DMRR["AKA","F"  ,c(1,2,4),] <- pT1[grep("F:Amp",rownames(pT1)),]
> DMRR["AKA",,c(3,5),] <- DMRR["AKA",,c(2,4),]
> ( tab2 <- round( ftable( DMRR, col.vars=c(1,4) ), 1 ) )

            amp     BAA                  BKA                  AKA
            what     RR    Lo    Up     RR    Lo    Up     RR    Lo    Up
sex from
M   DM             14.7  10.5  20.4    7.6   5.2  11.3    4.0   2.3   6.7
    BAA              NA    NA    NA  156.3 106.8 228.7   85.1  58.9 123.0
    BAA(DM)          NA    NA    NA  572.0 373.0 877.3   85.1  58.9 123.0
    BKA              NA    NA    NA     NA    NA    NA  407.7 329.0 505.2
    BKA(DM)          NA    NA    NA     NA    NA    NA  407.7 329.0 505.2
F   DM              7.5   5.2  10.9    8.4   5.5  12.8    3.7   2.2   6.2
```

```
    BAA               NA      NA      NA  155.0   92.3  260.2   99.1   66.9  146.9
    BAA(DM)           NA      NA      NA  807.6  488.7 1334.6   99.1   66.9  146.9
    BKA               NA      NA      NA     NA     NA     NA  446.8  350.9  568.9
    BKA(DM)           NA      NA      NA     NA     NA     NA  446.8  350.9  568.9
M+F DM              10.7     9.4    12.2    7.1    6.0    8.3    4.0    3.4    4.8
    BAA               NA      NA      NA  155.7  114.4  211.9   90.9   69.4  119.2
    BAA(DM)           NA      NA      NA  671.3  554.4  812.7   90.9   69.4  119.2
    BKA               NA      NA      NA     NA     NA     NA  421.8  358.3  496.7
    BKA(DM)           NA      NA      NA     NA     NA     NA  421.8  358.3  496.7
> write.csv2( tab2, file="./data/Tab2.csv" )
```

3.6.3 Estimates of trends in rates — Table 3

We now retrieve the RR estimates from the models where we modelled the annual change over the period:

```
> Atr <- NArray( list( amp = c("BAA","BKA","AKA"),
+                     from = c("NoDM","DM","BAA","BAA(DM)","BKA","BKA(DM)"),
+                     what = c("RR","Lo","Up") ) )
> str( Atr )
 logi [1:3, 1:6, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ amp : chr [1:3] "BAA" "BKA" "AKA"
  ..$ from: chr [1:6] "NoDM" "DM" "BAA" "BAA(DM)" ...
  ..$ what: chr [1:3] "RR" "Lo" "Up"
> Atr["BAA",1:2,] <- pF2[grep("I\\(P",rownames(pF2)),]
> Atr["BKA",1:4,] <- pK2[grep("I\\(P",rownames(pK2)),]
> Atr["AKA",c(1:3,5),] <- pT2[grep("I\\(P",rownames(pT2)),]
> Atr["AKA",c(4,6),]   <- Atr["AKA",c(3,5),]
> round( ftable( Atr, col.vars=c(1,3) ), 2 )
        amp   BAA             BKA             AKA
        what   RR   Lo   Up   RR   Lo   Up   RR   Lo   Up
from
NoDM         1.00 0.99 1.02 0.98 0.96 1.00 1.02 1.00 1.04
DM           0.90 0.88 0.92 0.85 0.82 0.87 0.97 0.94 1.01
BAA            NA   NA   NA 0.86 0.80 0.92 0.91 0.86 0.97
BAA(DM)        NA   NA   NA 0.83 0.80 0.87 0.91 0.86 0.97
BKA            NA   NA   NA   NA   NA   NA 0.98 0.95 1.00
BKA(DM)        NA   NA   NA   NA   NA   NA 0.98 0.95 1.00
> dimnames( Atr )[["from"]][1] <- "Well"
> ( tab3 <- round( ftable( (Atr-1)*100, col.vars=c(1,3) ), 1 ) )
        amp     BAA                   BKA                   AKA
        what     RR     Lo     Up     RR     Lo     Up     RR     Lo     Up
from
Well            0.5   -1.4    2.4   -2.3   -4.4   -0.2    1.7   -0.4    3.7
DM             -9.8  -11.9   -7.6  -15.1  -17.6  -12.5   -2.6   -5.9    0.8
BAA              NA     NA     NA  -14.0  -19.7   -7.8   -9.0  -14.2   -3.5
BAA(DM)          NA     NA     NA  -16.7  -20.3  -12.9   -9.0  -14.2   -3.5
BKA              NA     NA     NA     NA     NA     NA   -2.4   -5.3    0.5
BKA(DM)          NA     NA     NA     NA     NA     NA   -2.4   -5.3    0.5
> write.csv2( tab3, file="./data/Tab3.csv" )
```

3.6.4 Estimated rates and HRs of first amputations

We can compare the rates and HRs of first amputations of the three kinds — this is basically subsets of the curves shown in figures 3.1, 3.3 and 3.5 where we only show the rates and HRs for the first amputations:

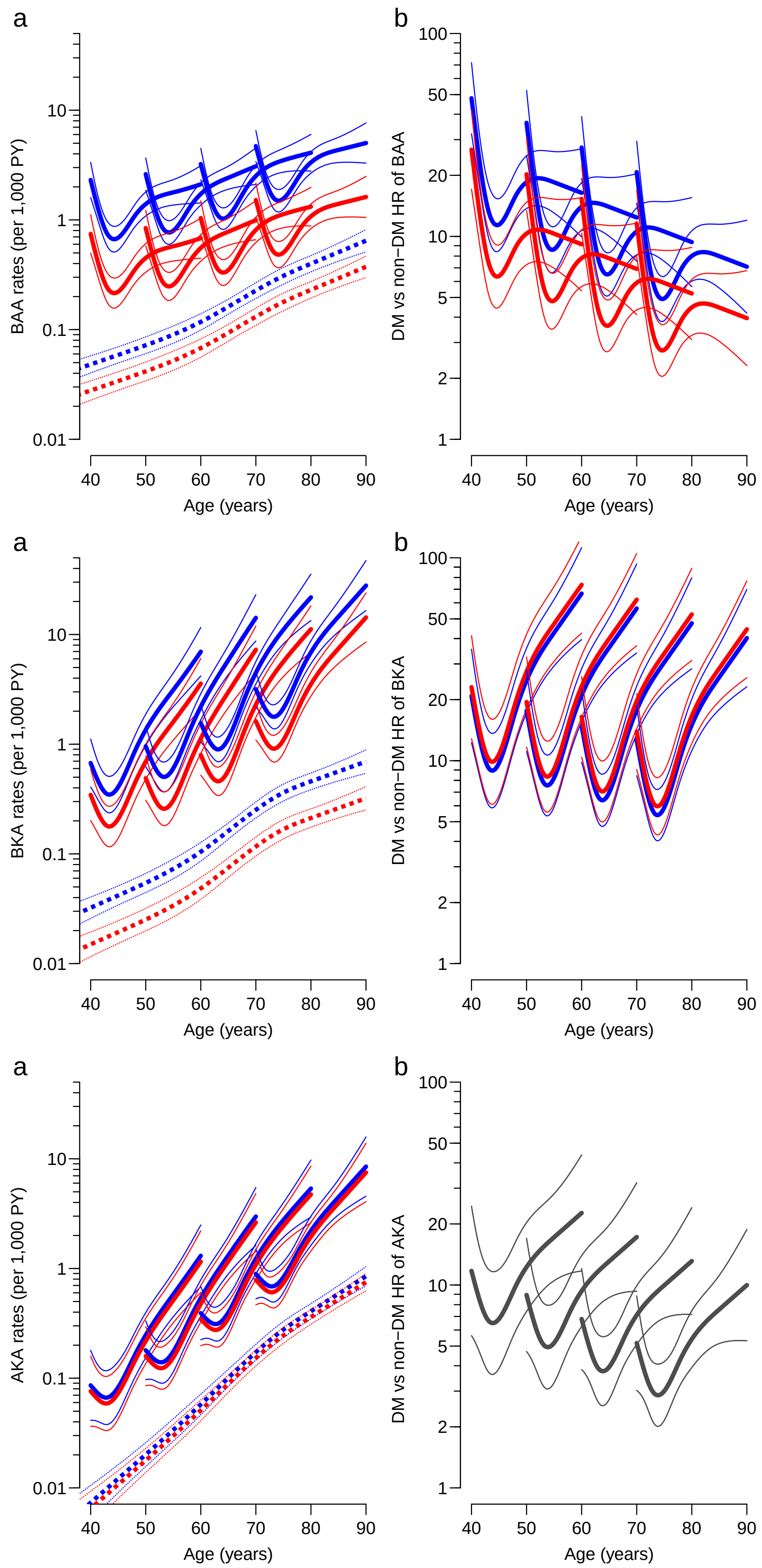

Figure 3.7: *Rates and hazard (rate) ratios for the three types of ampuations — only first amputations. These are the graphs used in the paper.*

3.7 Empirical amputation rates

In order to create a blunt overview of rates we compute directly standardized rates for each year, standardized to the age-distribution of (the follow-up time of) the non-DM part of the population in 2011. To this end we retrieve the tabulations of cases and person-years by age, year and diabetes status for each amputation type:

```
> library(Epi)
> clear()
> load( file="./data/PFtab.Rda" )
> load( file="./data/PKtab.Rda" )
> load( file="./data/PTtab.Rda" )
> str( PKtab )
 xtabs [1:16, 1:84, 1:2, 1:2] 0 1 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 4
  ..$ floor(P): chr [1:16] "1996" "1997" "1998" "1999" ...
  ..$ floor(A): chr [1:84] "15" "16" "17" "18" ...
  ..$ DM      : chr [1:2] "No" "Yes"
  ..$         : chr [1:2] "D" "Y"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = cbind(D, Y) ~ floor(P) + floor(A) + DM, data = transfor
> astd <- PFtab["2011",,"No","Y"]
> astd <- astd/sum(astd)
```

We can now compute the empirical rates by age, period and diabetes status, and from those the directly standardized rates by period and diabetes status:

```
> RF <- PFtab[,,,"D"]/PFtab[,,,"Y"]
> RK <- PKtab[,,,"D"]/PKtab[,,,"Y"]
> RT <- PTtab[,,,"D"]/PTtab[,,,"Y"]
> str( RF )
 num [1:16, 1:84, 1:2] 0 0 0 0 0.000199 ...
 - attr(*, "dimnames")=List of 3
  ..$ floor(P): chr [1:16] "1996" "1997" "1998" "1999" ...
  ..$ floor(A): chr [1:84] "15" "16" "17" "18" ...
  ..$ DM      : chr [1:2] "No" "Yes"
> length( astd )
[1] 84
> SF <- apply( RF, c(1,3), function(x) sum(x*astd) )
> SK <- apply( RK, c(1,3), function(x) sum(x*astd) )
> ST <- apply( RT, c(1,3), function(x) sum(x*astd) )
> round( cbind( SF, SK, ST )*1000, 3 )
        No   Yes    No   Yes    No   Yes
1996 0.066 1.419 0.079 1.420 0.115 0.525
1997 0.060 1.235 0.080 1.134 0.082 0.430
1998 0.108 0.918 0.070 1.470 0.100 0.455
1999 0.063 1.385 0.103 1.373 0.111 0.458
2000 0.103 1.167 0.110 0.805 0.069 0.522
2001 0.087 1.271 0.070 1.072 0.106 0.596
2002 0.098 1.125 0.079 1.185 0.068 0.686
2003 0.099 0.939 0.057 0.975 0.118 0.843
2004 0.096 0.953 0.071 0.622 0.106 0.518
2005 0.080 1.083 0.093 0.950 0.100 0.781
2006 0.106 0.953 0.098 0.646 0.112 0.571
2007 0.097 1.172 0.070 0.490 0.110 0.581
2008 0.082 1.065 0.039 1.112 0.090 0.836
2009 0.070 0.851 0.092 0.324 0.129 0.560
2010 0.094 0.774 0.038 0.692 0.107 0.605
2011 0.075 1.148 0.067 0.281 0.131 0.486
```

Then we can plot the age-standardized rates:

```
> pls <- function(){
+ par( mar=c(3,4,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
+ matplot( 1996:2011, cbind( SF, SF+SK, SF+SK+ST )*1000,
+          type="l", lty=1, lwd=rep(c(2,4,6),each=2), col=rep(c("black","red"),3),
+          log="y", xlab="Year of amputation", ylim=c(0.05,4), ylab="" )
+ axis( side=1, at=1996:2011, labels=NA )
+ axis( side=2, at=c(5:9/100,1:9/10,1:4), labels=NA )
+ mtext( "Standardised amputation rates per 1000 PY", side=2, line=3, las=0 )
+ }
> pls()
> # win.metafile( "Fig0-log.emf", height=6, width=6 )
> # pls()
> # dev.off()
```

```
> pls <- function(){
+ par( mar=c(3,4,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
+ matplot( 1996:2011, cbind( SF, SF+SK, SF+SK+ST )*1000,
+          type="l", lty=1, lwd=rep(c(2,4,6),each=2), col=rep(c("black","red"),3),
+          xlab="Year of amputation", yaxt="n", ylim=c(0,3.5), ylab="", yaxs="i" )
+ axis( side=1, at=1996:2011, labels=NA )
+ axis( side=2, at=seq(0,3.5,0.5),
+       labels=c("0",formatC(seq(0.5,3.5,0.5),format="f",digits=1)) )
+ mtext( "Standardised amputation rates per 1,000 PY", side=2, line=3, las=0 )
+ }
> pls()
> postscript( "Fig2.eps", height=6, width=6 )
> pls()
> dev.off()
```

```
pdf
  2
```

From the figures **??** and **??** (which only differ by the scaling of the y-axis) it is seen that that the largest decrease in rates is in BKA amputations among persons with diabetes.

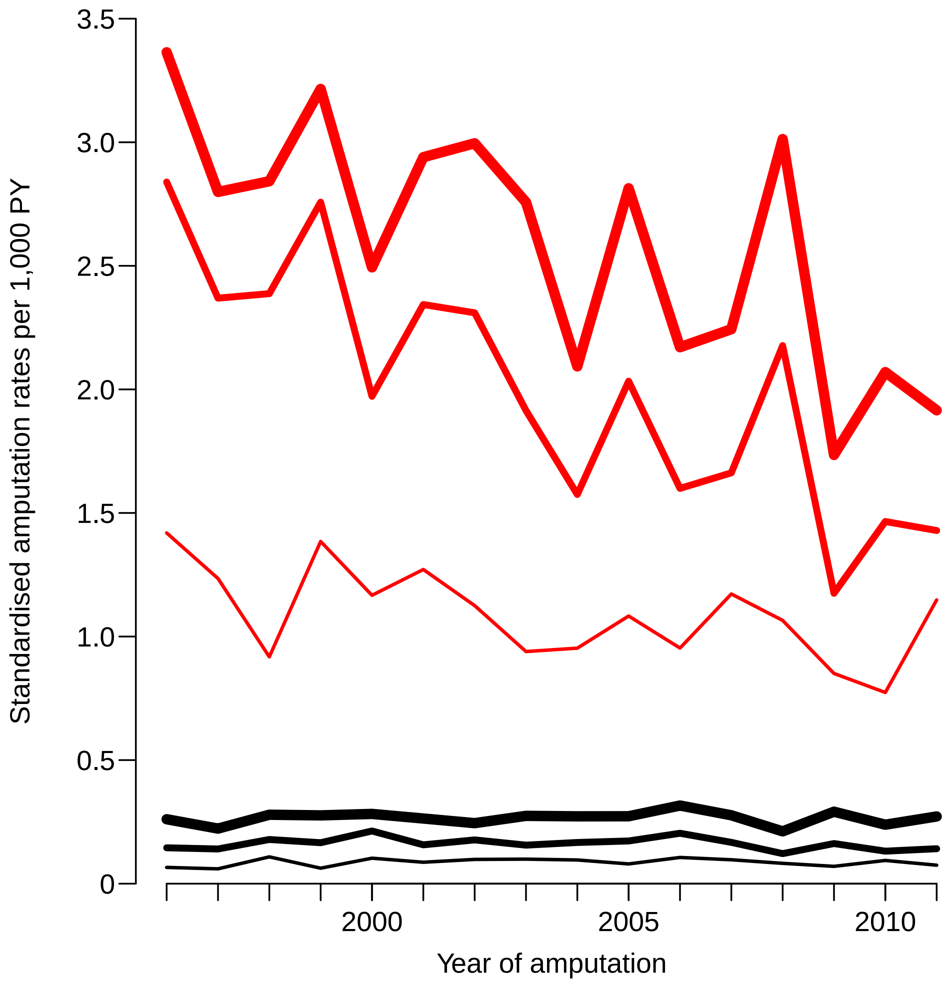

Figure 3.8: *Age-standardized amputation rates, standardized to the age-distribution of the non-diabetic part of the population in 2011. Black curves are for the part of the population without diabetes, the red ones for those with. In order of thickness the lines refer to BAA, BAA+BKA and all amputations (BAA+BKA+AKA), respectively.*

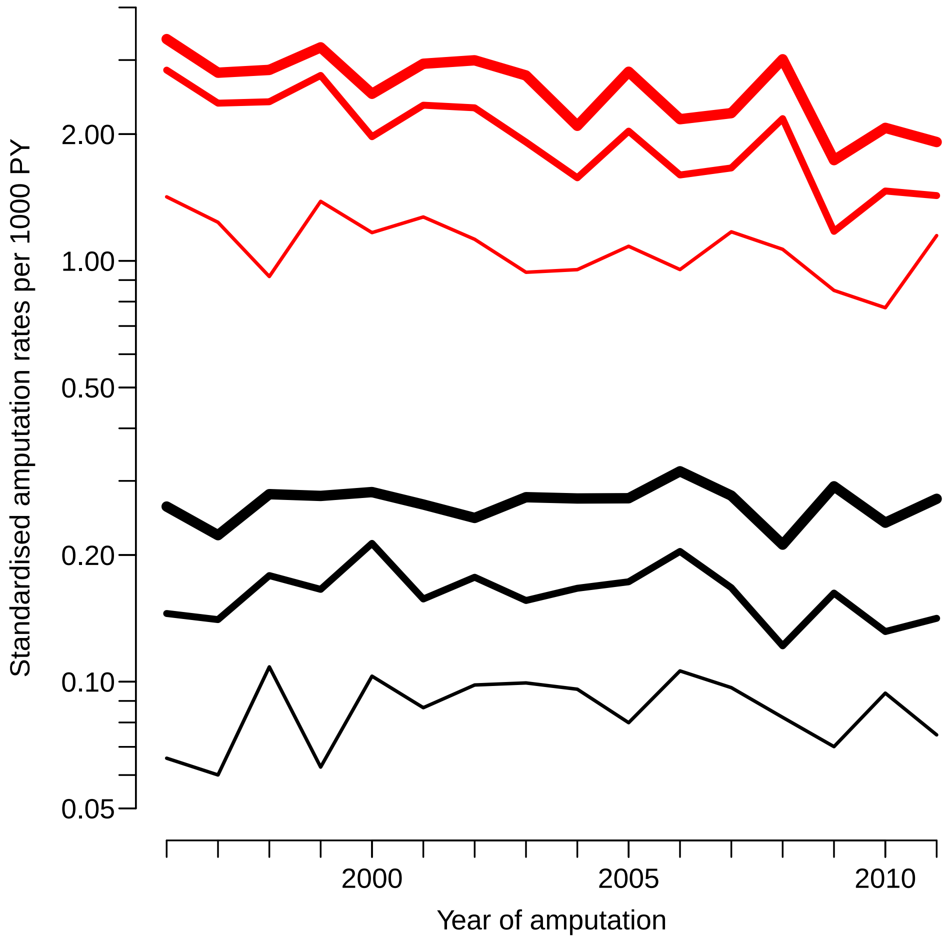

Figure 3.9: *Age-standardized amputation rates, standardized to the age-distribution of the non-diabetic part of the population in 2011. Note the log-scale of the rates. Black curves are for the part of the population without diabetes, the red ones for those with. In order of thickness the lines refer to BAA, BAA+BKA and all amputations (BAA+BKA+AKA), respectively.*