EPIC multistate models

SDCC: BxC, IARC: P Ferrari
September 2017
http://bendixcarstensen.com/
Version 2

Compiled Tuesday 10th April, 2018, 18:36
from: /home/bendix/sdc/extn/IARC/Ferrari/r/MSsetup.tex

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

Contents

Chapter 1

Multimorbidity and mortality in EPIC

1.1 3-state modeling

We want to model occurrence of DM, CVD and Cancer in the EPIC cohorts, and assess to which extent these depend on baseline characteristics.

Fist we visualize the rates we are concerned about:

```
> states <- c("Well","DM","CVD","Ca","DM+CVD","DM+Ca","CVD+Ca","DM+CVD+Ca","Dead")
> ns <- length(states)
> TM <- matrix(NA,ns,ns)
> rownames(TM) <- colnames(TM) <- states
> TM["Well",c("DM","CVD","Ca","Dead")] <- 1
> TM[c("DM","Ca"),c("DM+Ca","Dead")] <- 1
> TM[c("DM","CVD"),c("DM+CVD","Dead")] <- 1
> TM[c("Ca","CVD"),c("CVD+Ca","Dead")] <- 1
> TM[c("DM+CVD","CVD+Ca","DM+Ca"),c("DM+CVD+Ca","Dead")] <- 1
> TM[c("DM+CVD+Ca"),c("Dead")] <- 1
> TM
          Well DM CVD Ca DM+CVD DM+Ca CVD+Ca DM+CVD+Ca Dead
Well        NA  1   1  1     NA    NA     NA         NA    1
DM          NA NA  NA NA      1     1     NA         NA    1
CVD         NA NA  NA NA      1    NA      1         NA    1
Ca          NA NA  NA NA     NA     1      1         NA    1
DM+CVD      NA NA  NA NA     NA    NA     NA          1    1
DM+Ca       NA NA  NA NA     NA    NA     NA          1    1
CVD+Ca      NA NA  NA NA     NA    NA     NA          1    1
DM+CVD+Ca   NA NA  NA NA     NA    NA     NA         NA    1
Dead        NA NA  NA NA     NA    NA     NA         NA   NA
```

With the transition matrix we can set up the graph of transitions:

```
> library(Epi)
> # x11(h=16,w=24,p=24)
> boxes.matrix(TM,boxpos=list(x=c(10,35,35,35,65,65,65,90,50),
+                             y=c(40,60,40,20,60,40,20,40,90)-10),
+              hmult=2,wmult=1.2,col.arr="gray",
+              col.border=c(rep("transparent",8),"black"))
> par(new=TRUE)
> boxes.matrix(TM[-9,-9],boxpos=list(x=c(10,35,35,35,65,65,65,90),
```

```
+                                    y=c(40,60,40,20,60,40,20,40)-10),
+              hmult=2,wmult=1.2,lwd.arr=3,col.bg=gray(0.9),
+              col.border=gray(0.9))
```

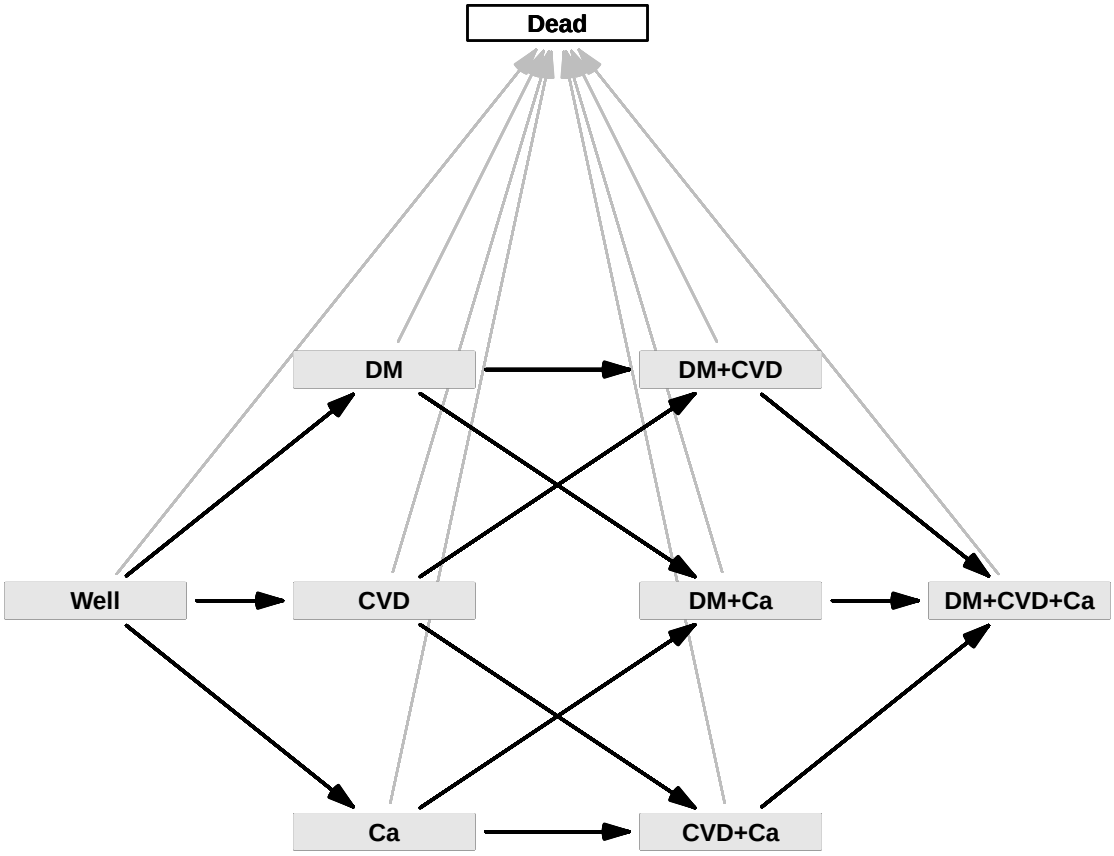

Figure 1.1: *States of interest in the MM study. Note that the NE arraows correspnd to T2D occurrence, the E arrows to CVD occurrence and the SE arrows to Ca occurrence. Also note that we have not separated persons with CVD before Ca from persons with Ca before CVD; they are all in the state CVD+Ca. Implicitly this assumes thet the T2D occurrence rates are the same regardless of which of the two conditions occurred first.* `MS-boxes`

There are two sets of analyses of immediate interest, one addressing the disease rates — the black rates in figure 1.1 (there are 12 of these), and one addressing the mortality rates — the gray lines (there are 8 of these rates).

1.2 Setting up data for analysis

The base dataset contains the relevant dates for each person:

- `doBth` — date of birth
- `doExm` — date of examination (entry to EPIC)
- `doDM` — date of DM
- `doCVD` — date of CVD
- `doCa` — date of Cancer
- `doDth` — date of death
- `doX` — date of exit, *i.e.* censoring, last date of follow-up for *all* events.

Suppose the data set is called `EPIC`, the we set up a Lexis object by first using death:

```
> L0 <- Lexis( entry = list(per=doExm,
+                           age=doExm-doBth,
+                           tfx=0),
+              exit = list(per=doX),
+       exit.status = factor(!is.na(doDth),labels=c("Well","Dead")),
+              data = EPIC)
```

In order to get the follow-up split between the various other states we cut follow-up in different states:

```
> LD <- cutLexis( subset(L0,!is.na(doDM) &
+                           (doDM<doCVD|is.na(doCVD)) &
+                           (doDM<doCa |is.na(doCa )) ),
+                 cut=doDM,
+                 new.state="DM",
+                 pre="Well" )
> LV <- cutLexis( subset(L0,!is.na(doCVD) &
+                           (doCVD<doDM|is.na(doDM)) &
+                           (doCVD<doCa|is.na(doCa)) ),
+                 cut=doDM,
+                 new.state="CVD",
+                 pre="Well" )
> LC <- cutLexis( subset(L0,!is.na(doCa) &
+                           (doCa<doDM |is.na(doDM )) &
+                           (doCa<doCVD|is.na(doCVD)) ),
+                 cut=doDM,
+                 new.state="Ca",
+                 pre="Well" )
> L0 <- subset( L0, is.na(doDM) & is.na(doCVD) & is.na(doCa) )
```

The three data sets `LD`, `LV` and `LC` now has the follow-up from all person with at least one data of diagnosis, subdivided at the first diagnosis, and `L0` that of the remaining persons (*i.e* persons without any disease event).

The three should now be further subdivided according to the next event, for example for those with diabetes as first event, into those with no further event (`LD0`), those with a CVD event (`LDV`) and those with a cancer event (`LDC`):

```
> LD0 <-           subset(LD, is.na(doCVD) & is.na(doCa) )
> LDV <- cutLexis( subset(LD,!is.na(doCVD) &
+                            doCVD<doCa|is.na(doCa)),
+                  cut=doCVD,
+                  new.state="DM+CVD" )
> LDC <- cutLexis( subset(LD,!is.na(doCa) &
+                            doCa<doCVD|is.na(doCVD)),
+                  cut=doCa,
+                  new.state="DM+Ca" )
```

Similar subdivisions is then made for persons with CVD, resp. cancer as first event, and finally those with a third event are considered from the 6 subsets with two initial events.

Finally all data sets (Lexis objects) are collected in a single object, `LE`, say, and using this we can analyze rates (as for example outlined in [1]) by a simple `glm` with Poisson error:

```
> glm( (lex.Xst=="Ca") ~ Ns(age) + sex,
+      family = poisson,
+      offset = log(lex.dur),
+        data = subset(LE,lex.Cst=="Well") )
```

1.3 4 states

Now suppose that we instead have 4 disease states of interest, DM, Ca, CHD and Str. The we must use these 4 disease states and all possible combinations of these as possible states:

```
> dn1 <- c("DM","Ca","CHD","Str")
> dn2 <- outer(dn1,dn1,paste,sep="+")
> dn2 <- dn2[upper.tri(dn2)]
> dn3 <- NULL
> for( i in 4:1 )
+ dn3  <- c(dn3,paste( dn1[-i], collapse="+" ))
> dn4 <- paste( dn1, collapse="+" )
> dn1
[1] "DM"  "Ca"  "CHD" "Str"
> dn2
[1] "DM+Ca"   "DM+CHD"  "Ca+CHD"  "DM+Str"  "Ca+Str"  "CHD+Str"
> dn3
[1] "DM+Ca+CHD"  "DM+Ca+Str"  "DM+CHD+Str" "Ca+CHD+Str"
> dn4
[1] "DM+Ca+CHD+Str"
> TM <- matrix(NA,15,15)
> rownames(TM) <- colnames(TM) <- c(dn1,dn2,dn3,dn4)
> TM[1,c(5,6,8)] <-
+ TM[2,c(5,7,9)] <-
+ TM[3,c(6,7,10)] <-
+ TM[4,c(8,9,10)] <-
+ TM[5,c(11,12)] <-
+ TM[6,c(11,13)] <-
+ TM[7,c(11,14)] <-
+ TM[8,c(12,13)] <-
+ TM[9,c(12,14)] <-
```

```
+ TM[10,c(13,14)] <-
+ TM[11,15] <-
+ TM[12,15] <-
+ TM[13,15] <-
+ TM[14,15] <- 1
```

We also need an initial "Well" state (without recorded disease), from which only the 4 states with one disease present can be reached:

```
> TM <- rbind( c(NA,rep(1,4),rep(NA,11)),
+              cbind( NA, TM ) )
> rownames(TM)[1] <-
+ colnames(TM)[1] <- "Well"
```

Finally we should expand with a death state which can be reached from any other state:

```
> TM <- rbind( cbind( TM, 1 ), NA )
> rownames(TM)[nrow(TM)] <-
+ colnames(TM)[ncol(TM)] <- "Dead"
> TM
```

	Well	DM	Ca	CHD	Str	DM+Ca	DM+CHD	Ca+CHD	DM+Str	Ca+Str	CHD+Str	DM+Ca+CHD	DM+Ca+
Well	NA	1	1	1	1	NA	NA	NA	NA	NA	NA	NA	
DM	NA	NA	NA	NA	NA	1	1	NA	1	NA	NA	NA	
Ca	NA	NA	NA	NA	NA	1	NA	1	NA	1	NA	NA	
CHD	NA	NA	NA	NA	NA	NA	1	1	NA	NA	1	NA	
Str	NA	NA	NA	NA	NA	NA	NA	NA	1	1	1	NA	
DM+Ca	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	
DM+CHD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	
Ca+CHD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	
DM+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Ca+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
CHD+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
DM+Ca+CHD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
DM+Ca+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
DM+CHD+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Ca+CHD+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
DM+Ca+CHD+Str	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Dead	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

	DM+CHD+Str	Ca+CHD+Str	DM+Ca+CHD+Str	Dead
Well	NA	NA	NA	1
DM	NA	NA	NA	1
Ca	NA	NA	NA	1
CHD	NA	NA	NA	1
Str	NA	NA	NA	1
DM+Ca	NA	NA	NA	1
DM+CHD	1	NA	NA	1
Ca+CHD	NA	1	NA	1
DM+Str	1	NA	NA	1
Ca+Str	NA	1	NA	1
CHD+Str	1	1	NA	1
DM+Ca+CHD	NA	NA	1	1
DM+Ca+Str	NA	NA	1	1
DM+CHD+Str	NA	NA	1	1
Ca+CHD+Str	NA	NA	1	1
DM+Ca+CHD+Str	NA	NA	NA	1
Dead	NA	NA	NA	NA

Drawing the transitions it is relevant to color the arrows to get a proper overview — transition rates are colored by a color corresponding to the disease *added* to the current state. Technically, we draw the same diagram twice; first with all arrows in gray, and subsequently (on top of the first one) with the colored incidence arrows, so that we have the mortality arrows underneath.

```
> clr <- c("limegreen","blue","brown","orange",gray(0.7))
> # Order colors according to the ordering convention of transitions:
> o.clr <- c(1:4,
+
+            2:4,
+          1,3,4,
+          1,2,4,
+            1:3,
+
+            3,4,
+            2,4,
+            1,4,
+            2,3,
+            1,3,
+            1,2,
+
+              4,
+              3,
+              2,
+              1)
> # Position of the boxes
> bp <- list(x=c(3,
+                rep(20,4),
+                c(50,50,43,57,50,50)-5,
+                rep(73,4),92,
+                95),
+            y=c(c(50,
+                  seq(78,22,,4),
+                  seq(97,03,,5)[c(1,2,3,3,4,5)],
+                  seq(78,22,,4),50)*0.95,
+                95))
> boxes.matrix( TM, boxpos=bp,
+               hmult=2, wmult=1.3, eq.wd=FALSE,
+               col.bg=gray(0.95),col.border=gray(0.95),
+               col.arr=clr[5] )
> par(new=TRUE)
> boxes.matrix( TM[-17,-17], boxpos=lapply(bp,function(x)x[-length(x)]),
+               hmult=2, wmult=1.3, eq.wd=FALSE,
+               col.bg=gray(0.9),col.border=gray(0.9),
+               col.arr=clr[o.clr], lwd.arr=4 )
```

In principle it should be possible to model all 32 incidence rates (8 of each kind) and all 16 mortality rates shown in figure 1.2 separately, but hardly feasible to report, let alone summarize in a sensible form. The relevant approach would be to fit separate models for each *color* of rates, using the originating state for each rate as covariate. The rates would in addition depend on sex, age, calendar time, country and population/EPIC status.

The big task is setting up the follow-up; fairly simple for the EPIC data — just as outlined for the 3-event-type situation above. But this is hardly doable for the population data for want of data — it is not realistic to get population rates corresponding to all

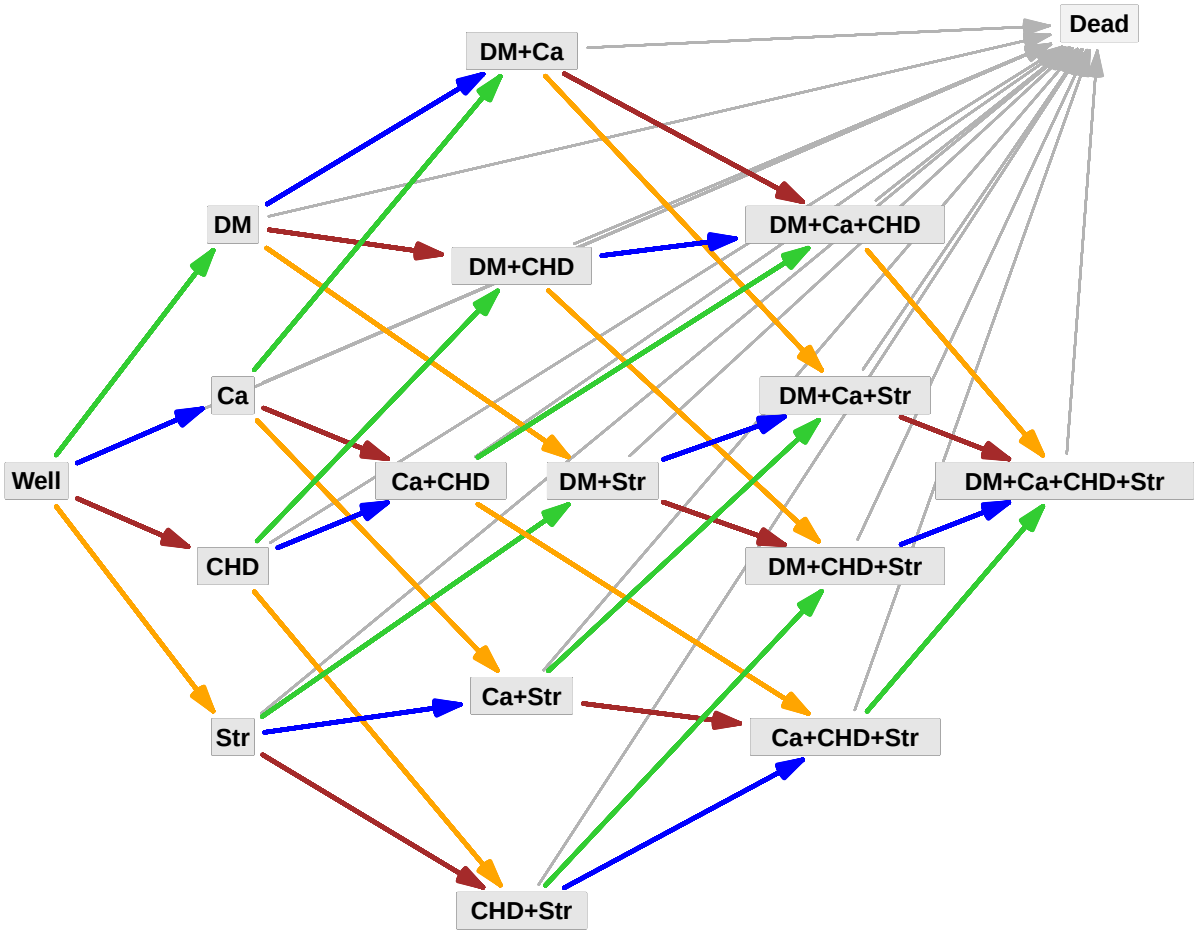

Figure 1.2: *Expanded model with both stroke and CHD as possible diagnoses of interest. Green arrows are DM occurrence, blue cancer occurrence, orange stroke occurrence, maroon arrows CHD occurrence, while gray arrows represent deaths.* `MS-tbox4`

arrows in figure 1.2; population rates are at the best available as sex-, age- and period-specific rates; that is weighted averages of arrows of the same color.

Thus assuming that only EPIC data can be used for the modeling illustrated in figure 1.2, each set of rates can be related to crude population rates of DM, cancer, stroke and CHD.

Likewise, mortality rates *from* each of the 16 states can be related to the overall mortality rates.

Note that cause-specific mortality rates are not discussed at all here; only mortality rates *from* different disease states. In the vein of relative survival it may be deliberated whether additive rather than multiplicative models might be relevant for description of the differences between mortalities in the EPIC cohort (gray lines).

Chapter 2

Reading data

First we read the supplied data:

```
> library(Epi)
> system.time( mm <- read.csv( "../data/MM.csv" ) )
   user  system elapsed
 15.318   0.133  15.446
> names( mm )
  [1] "IDEPIC"             "sex"                "D_Recrui"           "cancer"
  [5] "date_cancer"        "DIA"                "date_dia"           "CHD"
  [9] "date_chd"           "death"              "date_death"         "Country"
 [13] "Center"             "Age_Recr"           "Height_Adj"         "Weight_Adj"
 [17] "Hip_Adj"            "Waist_Adj"          "Bmi_Adj"            "Bmi_C"
 [21] "Smoke_Stat"         "A_Sta_Cig_Aggr"     "A_Giv_Cig_Aggr"     "Dur_Cig"
 [25] "Pa_Work"            "Pa_Vig"             "Pa_Index"           "Alc_Re"
 [29] "Alc_Lifetime"       "Alc_Age_Start"      "Alc_Age_Stop"       "Alc_Drinker"
 [33] "Alc_Lifetime_Wine"  "Alc_Lifetime_Spir"  "Alc_Lifetime_Beer"  "QE_ENERGY"
 [37] "Menopause"          "Use_Phrt"           "A_Menopause"        "Ever_Pill"
 [41] "Ever_Horm"          "Cvd_Prob"           "A_L_Heart"          "A_L_Stroke"
 [45] "A_Stroke"           "A_Heart"            "T_Hypert"           "A_Hypert"
 [49] "Diabet"             "Hypert"             "Stroke"             "Angina"
 [53] "Heart"              "RMed_Score_C"       "MRMed_Score_C"      "L_School"
 [57] "A_School_C"         "School"             "A_School"           "Cntr_A"
 [61] "Cntr_D"             "Cntr_F"             "Whr_Adj"            "Height_C"
 [65] "Weight_C"           "Hip_C"              "Waist_C"            "Whr_C"
 [69] "A_Sta_Smok_Aggr"    "A_Giv_Smok_Aggr"    "Dur_Smok"           "Dur_Smok_C"
 [73] "Timq_Smok"          "Timq_Cig"           "N_Cigret_Aggr"      "N_Cigret_Lifetime"
 [77] "Smoke_Intensity"    "Ftp"                "N_Ftp"              "Use_Horm"
 [81] "Use_Pill"           "Pa_Score"           "Pa_Score_M"         "M_Diy"
 [85] "M_Houswrk"          "M_Floors"           "M_Vigpa"            "M_Walk"
 [89] "M_Cycl"             "M_Gard"             "M_Sport"            "Pa_Mets"
 [93] "Pa_Recr"            "Pa_Hhld"            "Pa_Score_C"         "Pa_Score_M_C"
 [97] "Pa_Mets_C"          "Pa_Mets_C_Sx"       "Pa_Total"           "Pa_Total_Sx"
[101] "Alc_Drinktime"      "Alc_Drinktime_Beer" "Alc_Drinktime_Wine" "Alc_Drinktime_Spir"
[105] "Alc_Lifetime_Fwin"  "Alc_Drinktime_Fwin" "Alc_Re_C"           "Alc_Lifetime_C"
[109] "Alc_Pattern"        "Qe_Alc"             "Systol_1"           "Diastol_1"
[113] "Systol_2"           "Diastol_2"          "Sit_Tumo"           "Mor_Tumo"
[117] "D_Dg_Tumo"          "Cncr_Evt_Anyc"      "Cancer_Nb_Inc"      "Diabet_IA_Status"
[121] "Diabet_IA_Diag"     "Idcvd"              "smoke_new"          "alcohol"
[125] "pamets"             "bmic"               "Dietscore_Cat"      "Indexscore"
[129] "Indexgroup"         "MMed_Score_Sx"      "MMed_Score_SxCty"   "MMed_Score_SxCnt"
[133] "RMed_Score"         "MRMed_Score"
```

```
> dim( mm )
[1] 176142    134
```

Some persons actually die on the date recruited, so we remove these:

```
> with( mm,  table( as.character(D_Recrui) == as.character(date_death) ) )
 FALSE   TRUE
175894    248
> mm <- subset( mm, as.character(D_Recrui) != as.character(date_death) )
> dim( mm )
[1] 175894    134
```

Then define the relevant date variables to look after

```
> ( dvar <- c("D_Recrui",fgrep( "ate", names(mm) )) )
[1] "D_Recrui"    "date_cancer" "date_dia"    "date_chd"    "date_death"
> str( mm[,dvar] )
'data.frame':        175894 obs. of  5 variables:
 $ D_Recrui   : Factor w/ 2211 levels "01/01/1997","01/01/1998",..: 863 1843 1005 1154 1154
 $ date_cancer: Factor w/ 4244 levels "01/01/1994","01/01/1995",..: 4244 4244 2050 4244 424
 $ date_dia   : Factor w/ 3918 levels "01/01/1994","01/01/1995",..: 3918 3918 1894 3918 391
 $ date_chd   : Factor w/ 4129 levels "01/01/1996","01/01/1998",..: 4129 4129 1991 4129 412
 $ date_death : Factor w/ 3225 levels "01/01/1996","01/01/1998",..: 3225 3225 1541 3225 322
```

Unfortunately these variables are coded to be equal to the exit time (censoring at end of 2007 or date of death). Therefore we remove these from the date variables, so only event dates are left — the only real censoring taking place here is by end of study at 31 Dec 2007 (or 1 Jan 2008). As mentioned, also the death dates do not belong in the event date variable:

```
> for( v in dvar )
+    {
+    zz <- as.character(mm[,v])
+    if( v != "date_death" ) zz <- ifelse( zz==mm$date_death, NA, zz )
+                            zz <- ifelse( zz== "31/12/2007", NA, zz )
+    mm[,v] <- as.Date( zz, format="%d/%m/%Y" )
+    }
> xdate <- 2008
```

Finally we convert the dates to `cal.yr` format for convenience:

```
> mm <- cal.yr( mm )
> summary( mm[,c(dvar,"Age_Recr")] )
    D_Recrui     date_cancer        date_dia         date_chd        date_death
 Min.   :1991   Min.   :1992     Min.   :1993     Min.   :1993     Min.   :1993
 1st Qu.:1995   1st Qu.:2000     1st Qu.:2000     1st Qu.:2001     1st Qu.:2005
 Median :1996   Median :2003     Median :2003     Median :2004     Median :2006
 Mean   :1996   Mean   :2002     Mean   :2002     Mean   :2004     Mean   :2005
 3rd Qu.:1997   3rd Qu.:2006     3rd Qu.:2005     3rd Qu.:2007     3rd Qu.:2008
 Max.   :2000   Max.   :2008     Max.   :2008     Max.   :2008     Max.   :2008
                NA's   :159870   NA's   :169009   NA's   :168181   NA's   :159353
    Age_Recr
 Min.   :17.84
 1st Qu.:43.58
 Median :51.65
 Mean   :50.38
 3rd Qu.:57.85
 Max.   :98.50
```

```
>     str( mm[,c(dvar,"Age_Recr")] )
'data.frame':        175894 obs. of  6 variables:
 $ D_Recrui   :Classes 'cal.yr', 'numeric'  num [1:175894] 1993 1993 1993 1993 1993 ...
 $ date_cancer:Classes 'cal.yr', 'numeric'  num [1:175894] NA NA NA NA NA NA NA NA NA NA ..
 $ date_dia   :Classes 'cal.yr', 'numeric'  num [1:175894] NA NA NA NA NA NA NA NA NA NA ..
 $ date_chd   :Classes 'cal.yr', 'numeric'  num [1:175894] NA NA NA NA NA NA NA NA NA NA ..
 $ date_death :Classes 'cal.yr', 'numeric'  num [1:175894] NA NA 1998 NA NA ...
 $ Age_Recr   : num  50.1 44 35 39 52.1 ...
```

Apparently there is one person with diagnosis of T2D and CHD on the same day, so we postpone the diabetes date of this person 2 days:

```
> with( mm, ftable( DiCa = date_dia==date_cancer,
+                   DiCV = date_dia==date_chd,
+                   CVCa = date_chd==date_cancer,
+                   exclude=NULL ) )
            CVCa  FALSE     NA
DiCa  DiCV
FALSE FALSE          81      0
      TRUE            0      0
      NA              0    769
NA    FALSE           0    472
      TRUE            0      1
      NA            922 173649
> mm$date_dia <- mm$date_dia + (mm$date_dia==mm$date_chd)/180
```

2.1 Defining a Lexis object

Finally we are in a position to set up a Lexis object showing the transitions, note we define not only calendar time and current age as time scales but also time since baseline (`tfb`):

```
> L0 <- Lexis( entry = list( per = D_Recrui,
+                            age = Age_Recr,
+                            tfb = 0),
+              exit = list( per = pmin( date_death, xdate, na.rm=TRUE ) ),
+       exit.status = factor( !is.na( date_death ),
+                             labels = c("Base","Dead") ),
+              data = mm )
NOTE: entry.status has been set to "Base" for all.
> summary( L0 )
Transitions:
     To
From     Base  Dead  Records:  Events: Risk time:  Persons:
  Base 159353 16541    175894    16541    2118684    175894
```

We see there are 16541 deaths recorded.

Once this is accomplished we can split the follow-up at the recorded events; the events occurring at the dates in `dvar[2:4]` will be called Can, T2D and CHD, respectively:

```
> system.time(
+ L1 <- mcutLexis( L0,
+           timescale = 1,
+                  wh = dvar[2:4],
+          new.states = c("Can","T2D","CHD"),
+    precursor.states = "Base" ) )
```

```
   user  system elapsed
 27.834   0.108  27.943
```

```
> summary( L1 )
```

```
Transitions:
     To
From             Base  Dead   Can   CHD CHD-Can Can-CHD T2D T2D-CHD CHD-T2D T2D-CHD-Can
  Base         140233 12960 15602  6789       0       0 306       0       0           0
  Can               0  2356 12665     0       0     574   0       0       0           0
  CHD               0   900     0  5305     362       0   0       0     224           0
  CHD-Can           0    79     0     0     277       0   0       0       0           0
  Can-CHD           0   189     0     0       0     377   0       0       0           0
  T2D               0     0     0     0       0       0   0     286       0           0
  T2D-CHD           0    27     0     0       0       0   0     240       0          19
  CHD-T2D           0    10     0     0       0       0   0       0     195           0
  T2D-CHD-Can       0     7     0     0       0       0   0       0       0          12
  Can-CHD-T2D       0     1     0     0       0       0   0       0       0           0
  CHD-T2D-Can       0     3     0     0       0       0   0       0       0           0
  CHD-Can-T2D       0     3     0     0       0       0   0       0       0           0
  T2D-Can           0     0     0     0       0       0   0       0       0           0
  T2D-Can-CHD       0     4     0     0       0       0   0       0       0           0
  Can-T2D           0     0     0     0       0       0   0       0       0           0
  Can-T2D-CHD       0     2     0     0       0       0   0       0       0           0
  Sum          140233 16541 28267 12094     639     951 306     526     419          31

Transitions:
     To
From          Can-CHD-T2D CHD-T2D-Can CHD-Can-T2D T2D-Can T2D-Can-CHD Can-T2D Can-T2D-CHD
  Base                  0           0           0       0           0       0           0
  Can                   0           0           0       0           0       9           0
  CHD                   0           0           0       0           0       0           0
  CHD-Can               0           0           6       0           0       0           0
  Can-CHD               8           0           0       0           0       0           0
  T2D                   0           0           0      20           0       0           0
  T2D-CHD               0           0           0       0           0       0           0
  CHD-T2D               0          19           0       0           0       0           0
  T2D-CHD-Can           0           0           0       0           0       0           0
  Can-CHD-T2D           7           0           0       0           0       0           0
  CHD-T2D-Can           0          16           0       0           0       0           0
  CHD-Can-T2D           0           0           3       0           0       0           0
  T2D-Can               0           0           0       0          20       0           0
  T2D-Can-CHD           0           0           0       0          16       0           0
  Can-T2D               0           0           0       0           0       0           9
  Can-T2D-CHD           0           0           0       0           0       0           7
  Sum                  15          35           9      20          36       9          16

Transitions:
     To
From           Records:  Events: Risk time:  Persons:
  Base            175890    35657 2012954.32    175890
  Can              15604     2939   74888.33     15604
  CHD               6791     1486   25226.70      6791
  CHD-Can            362       85    1024.80       362
  Can-CHD            574      197    1357.92       574
  T2D                306      306    1144.91       306
  T2D-CHD            286       46     904.84       286
  CHD-T2D            224       29     922.74       224
  T2D-CHD-Can         19        7      29.88        19
```

```
  Can-CHD-T2D          8        1      26.47         8
  CHD-T2D-Can         19        3      42.13        19
  CHD-Can-T2D          6        3      16.53         6
  T2D-Can             20       20      60.89        20
  T2D-Can-CHD         20        4      39.88        20
  Can-T2D              9        9      23.55         9
  Can-T2D-CHD          9        2      20.19         9
  Sum             200147    40794 2118684.07    175894
```

With this large number of states, the summary is not very helpful.

In order to illustrate what happened we find a person from each of the three-disease boxes:

```
> maxl <- levels(L1)[nchar( levels(L1) )==11]
> wh <- NULL
> for( l in maxl ) wh <- c( wh, L1[L1$lex.Cst==l,"lex.id"][1] )
> wh
[1]  2752  3080  7709 30443 35909 47277
> L1[L1$lex.id %in% wh,c("lex.id","per","age","tfb","lex.dur","lex.Cst","lex.Xst")]
       lex.id      per      age        tfb     lex.dur     lex.Cst     lex.Xst
199824   2752 1995.027 58.79808  0.0000000  6.21765914        Base         T2D
199825   2752 2001.244 65.01574  6.2176591  1.62354552         T2D     T2D-CHD
199826   2752 2002.868 66.63928  7.8412047  3.71526352     T2D-CHD T2D-CHD-Can
199827   2752 2006.583 70.35455 11.5564682  0.10677618 T2D-CHD-Can        Dead
199900   3080 1995.142 50.79808  0.0000000  4.01916496        Base         Can
199901   3080 1999.161 54.81724  4.0191650  4.77481177         Can     Can-CHD
199902   3080 2003.936 59.59206  8.7939767  0.26557153     Can-CHD Can-CHD-T2D
199903   3080 2004.201 59.85763  9.0595483  3.79876797 Can-CHD-T2D Can-CHD-T2D
199932   7709 1997.504 52.66804  0.0000000  1.75496235        Base         CHD
199933   7709 1999.259 54.42300  1.7549624  5.10061602         CHD     CHD-T2D
199934   7709 2004.360 59.52362  6.8555784  0.72279261     CHD-T2D CHD-T2D-Can
199935   7709 2005.083 60.24641  7.5783710  0.69267625 CHD-T2D-Can        Dead
200008  30443 1993.146 64.15606  0.0000000  0.46269678        Base         CHD
200009  30443 1993.608 64.61876  0.4626968  1.57152635         CHD     CHD-Can
200010  30443 1995.180 66.19028  2.0342231  6.87474333     CHD-Can CHD-Can-T2D
200011  30443 2002.055 73.06503  8.9089665  3.69609856 CHD-Can-T2D        Dead
200032  35909 1993.105 58.46680  0.0000000 11.00068446        Base         T2D
200033  35909 2004.105 69.46748 11.0006845  3.35660507         T2D     T2D-Can
200034  35909 2007.462 72.82409 14.3572895  0.49828884     T2D-Can T2D-Can-CHD
200035  35909 2007.960 73.32238 14.8555784  0.03969884 T2D-Can-CHD T2D-Can-CHD
200112  47277 1993.841 55.67967  0.0000000  0.11772758        Base         Can
200113  47277 1993.959 55.79740  0.1177276  8.82956879         Can     Can-T2D
200114  47277 2002.789 64.62697  8.9472964  4.39425051     Can-T2D Can-T2D-CHD
200115  47277 2007.183 69.02122 13.3415469  0.81724846 Can-T2D-CHD Can-T2D-CHD
```

Here, `lex.dur` is the risk time the individual named in `lex.id` spends in the state `lex.Cst` (Current state). `lex.Xst` (eXit state) is the state the person is in after spending `lex.dur` in the state `lex.Cst`. The time scales `per`, `age` and `tfb` indicates the time *at the beginning* of the interval of follow.up of length `lex.dur`.

We can illustrate the number of transitions and the rates using this Lexis object. To this end it is convenient to reorder the factor levels so that setting up the position of the state boxes becomes easier. We also omit drawing of the mortality rates by setting the color for these transitions to transparent. The fidgeting with the sequence in which the arrows are referred to relies on the (unfortunately yet undocumented) fact that arrows are ordered lexicographically by (`lex.Cst`,`lex.Xst`):

```
> levels( L1 )
 [1] "Base"         "Dead"         "Can"          "CHD"          "CHD-Can"      "Can-CHD"
 [7] "T2D"          "T2D-CHD"      "CHD-T2D"      "T2D-CHD-Can" "Can-CHD-T2D" "CHD-T2D-Can"
[13] "CHD-Can-T2D" "T2D-Can"      "T2D-Can-CHD" "Can-T2D"      "Can-T2D-CHD"
```

```
> perm <- c(1,3,4,7,6,16,5,9,14,8,11,17,13,12,15,10,2)
> levels( L1 )[perm]
 [1] "Base"         "Can"          "CHD"          "T2D"          "Can-CHD"      "Can-T2D"
 [7] "CHD-Can"      "CHD-T2D"      "T2D-Can"      "T2D-CHD"      "Can-CHD-T2D" "Can-T2D-CHD"
[13] "CHD-Can-T2D" "CHD-T2D-Can" "T2D-Can-CHD" "T2D-CHD-Can" "Dead"
```

```
> L1 <- Relevel( L1, perm )
> bp <- list( x=c(seq(10,90,,4)[rep(1:4,c(1,3,6,6))],10),
+             y=c(50,seq(20,80,,3),
+                    seq(10,90,,6),
+                    seq(10,90,,6),10) )
> bb <- boxes( L1, boxpos=bp, scale.R=1000, wmult=0.35, hmult=0.7, show.BE=TRUE )
> LM <- !is.na(bb$Tmat)
> ( toD <- cumsum(apply(LM,1,sum))[LM[,17]][-1] )
        Can          CHD     Can-CHD     CHD-Can     CHD-T2D     T2D-CHD Can-CHD-T2D
          7           10          14          17          19          22          23
Can-T2D-CHD CHD-Can-T2D CHD-T2D-Can T2D-Can-CHD T2D-CHD-Can
         24          25          26          27          28
```

```
> bb$Arrows$col.arr[toD] <- "transparent"
> bb$Arrows$col.txt[toD] <- "transparent"
> bb$Arrows$col.txt.arr[toD] <- "transparent"
> class(bb)
[1] "MS"
```

```
> boxes( bb )
```

Finally, we save the dataset for further analysis:

```
> save( L1, file="../data/L1.Rda" )
```

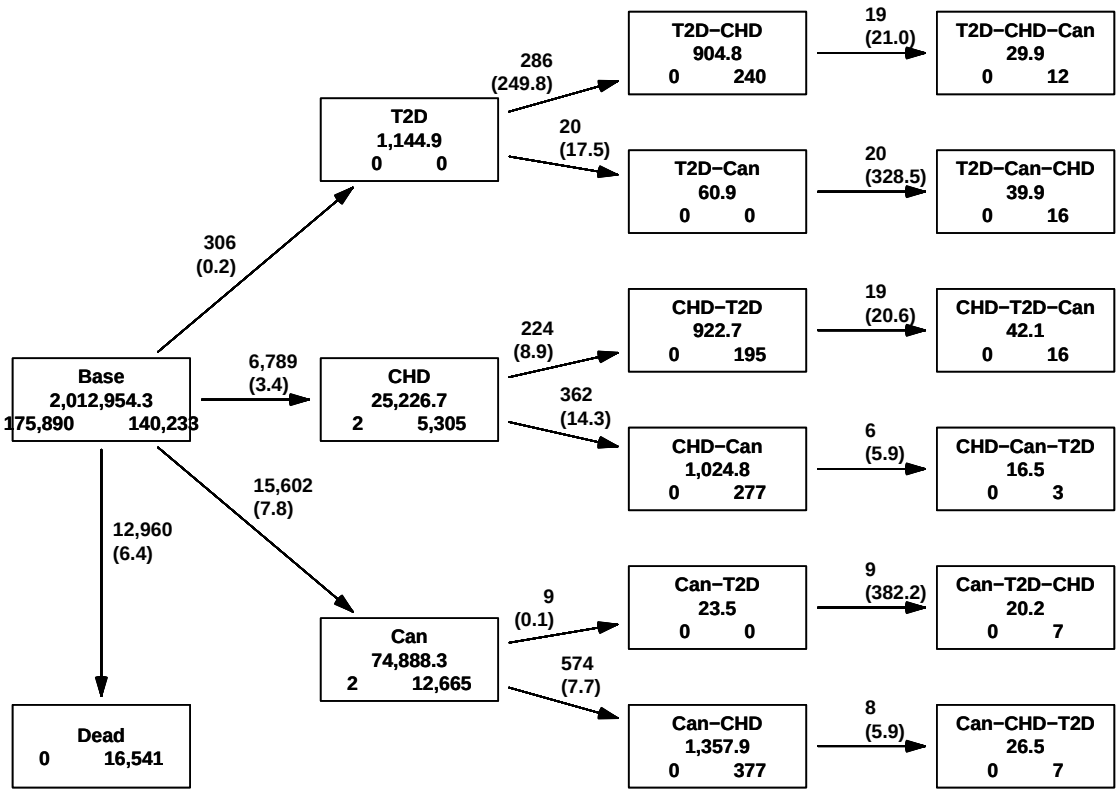

Figure 2.1: *Transitions present in the dataset, numbers on the arrows are the number of transitions and rates per 1000 PY. The numbers in the boxes are the number of person-years, and the number of persons starting, resp. ending their follow-up in each state. Obviously the number to T2D events in this dataset is coded wrongly (or read wrongly).* `./read-L1-box`

Chapter 3

Analysis of rates

First we load the created Lexis object

```
> library(    Epi )
> library( popEpi )
> library( splines )
> library( splines2 )
> clear()
> load( "../data/L1.Rda" )
> lls()
  name mode class            dim              size(Kb)
1 L1   list Lexis data.frame 200147 141      165,627.4
> sum( L1$lex.dur )
[1] 2118684
```

So we have in total a bit more than 2 million PY.

3.1 Splitting the follow-up

However with the `L1` dataset, we will have to resort to Cox-models that will not provide the baseline, so in order to be able to use parametric models for the baseline, we split the dataset in 1-year classes of follow-up, using the `splitMulti` function from the `popEpi` package that is vastly more efficient than `splitLexis`, and has the facility by default to exclude follow-up outside the range of the breaks specified:

```
> names( LR <- L1[,c(1:7,9,26)] )
[1] "per"     "age"     "tfb"     "lex.dur" "lex.Cst" "lex.Xst" "lex.id"  "sex"
[9] "Bmi_Adj"
> system.time( M1 <- splitMulti( LR, age=0:100 ) )
   user  system elapsed
  6.861   1.814  13.162
> summary( M1 )
Transitions:
     To
From             Base    Can   CHD  T2D Can-CHD Can-T2D CHD-Can CHD-T2D T2D-Can T2D-CHD
  Base         2151602  15602  6788  306       0       0       0       0       0       0
  Can                0  87393     0    0     574       9       0       0       0       0
  CHD                0      0 30438    0       0       0     362     224       0       0
```

T2D	0	0	0	1148	0	0	0	0	20	286
Can-CHD	0	0	0	0	1725	0	0	0	0	0
Can-T2D	0	0	0	0	0	23	0	0	0	0
CHD-Can	0	0	0	0	0	0	1294	0	0	0
CHD-T2D	0	0	0	0	0	0	0	1125	0	0
T2D-Can	0	0	0	0	0	0	0	0	61	0
T2D-CHD	0	0	0	0	0	0	0	0	0	1143
Can-CHD-T2D	0	0	0	0	0	0	0	0	0	0
Can-T2D-CHD	0	0	0	0	0	0	0	0	0	0
CHD-Can-T2D	0	0	0	0	0	0	0	0	0	0
CHD-T2D-Can	0	0	0	0	0	0	0	0	0	0
T2D-Can-CHD	0	0	0	0	0	0	0	0	0	0
T2D-CHD-Can	0	0	0	0	0	0	0	0	0	0
Sum	2151602	102995	37226	1454	2299	32	1656	1349	81	1429

Transitions:

From \ To	Can-CHD-T2D	Can-T2D-CHD	CHD-Can-T2D	CHD-T2D-Can	T2D-Can-CHD	T2D-CHD-Can
Base	0	0	0	0	0	0
Can	0	0	0	0	0	0
CHD	0	0	0	0	0	0
T2D	0	0	0	0	0	0
Can-CHD	8	0	0	0	0	0
Can-T2D	0	9	0	0	0	0
CHD-Can	0	0	6	0	0	0
CHD-T2D	0	0	0	19	0	0
T2D-Can	0	0	0	0	20	0
T2D-CHD	0	0	0	0	0	19
Can-CHD-T2D	33	0	0	0	0	0
Can-T2D-CHD	0	25	0	0	0	0
CHD-Can-T2D	0	0	18	0	0	0
CHD-T2D-Can	0	0	0	56	0	0
T2D-Can-CHD	0	0	0	0	53	0
T2D-CHD-Can	0	0	0	0	0	41
Sum	41	34	24	75	73	60

Transitions:

From \ To	Dead	Records:	Events:	Risk time:	Persons:
Base	12960	2187258	35656	2012927.18	175890
Can	2356	90332	2939	74885.99	15604
CHD	900	31924	1486	25214.05	6790
T2D	0	1454	306	1144.91	306
Can-CHD	189	1922	197	1357.92	574
Can-T2D	0	32	9	23.55	9
CHD-Can	79	1379	85	1024.80	362
CHD-T2D	10	1154	29	922.74	224
T2D-Can	0	81	20	60.89	20
T2D-CHD	27	1189	46	904.84	286
Can-CHD-T2D	1	34	1	26.47	8
Can-T2D-CHD	2	27	2	20.19	9
CHD-Can-T2D	3	21	3	16.53	6
CHD-T2D-Can	3	59	3	42.13	19
T2D-Can-CHD	4	57	4	39.88	20
T2D-CHD-Can	7	48	7	29.88	19
Sum	16541	2316971	40793	2118641.94	175894

To illustrate what happened by the splitting we list the records from one of the previously listed persons:

```
> LR[LR$lex.id==2752,]
            per      age       tfb   lex.dur     lex.Cst     lex.Xst lex.id    sex
199824 1995.027 58.79808  0.000000 6.2176591        Base         T2D   2752 Female
199825 2001.244 65.01574  6.217659 1.6235455         T2D     T2D-CHD   2752 Female
199826 2002.868 66.63928  7.841205 3.7152635     T2D-CHD T2D-CHD-Can   2752 Female
199827 2006.583 70.35455 11.556468 0.1067762 T2D-CHD-Can        Dead   2752 Female
       Bmi_Adj
199824   26.15
199825   26.15
199826   26.15
199827   26.15
> M1[M1$lex.id==2752,]
    lex.id      per      age       tfb    lex.dur     lex.Cst     lex.Xst    sex Bmi_Adj
 1:   2752 1995.027 58.79808  0.000000 0.20192000        Base        Base Female   26.15
 2:   2752 1995.229 59.00000  0.201920 1.00000000        Base        Base Female   26.15
 3:   2752 1996.229 60.00000  1.201920 1.00000000        Base        Base Female   26.15
 4:   2752 1997.229 61.00000  2.201920 1.00000000        Base        Base Female   26.15
 5:   2752 1998.229 62.00000  3.201920 1.00000000        Base        Base Female   26.15
 6:   2752 1999.229 63.00000  4.201920 1.00000000        Base        Base Female   26.15
 7:   2752 2000.229 64.00000  5.201920 1.00000000        Base        Base Female   26.15
 8:   2752 2001.229 65.00000  6.201920 0.01573914        Base         T2D Female   26.15
 9:   2752 2001.244 65.01574  6.217659 0.98426086         T2D         T2D Female   26.15
10:   2752 2002.229 66.00000  7.201920 0.63928465         T2D     T2D-CHD Female   26.15
11:   2752 2002.868 66.63928  7.841205 0.36071535     T2D-CHD     T2D-CHD Female   26.15
12:   2752 2003.229 67.00000  8.201920 1.00000000     T2D-CHD     T2D-CHD Female   26.15
13:   2752 2004.229 68.00000  9.201920 1.00000000     T2D-CHD     T2D-CHD Female   26.15
14:   2752 2005.229 69.00000 10.201920 1.00000000     T2D-CHD     T2D-CHD Female   26.15
15:   2752 2006.229 70.00000 11.201920 0.35454817     T2D-CHD T2D-CHD-Can Female   26.15
16:   2752 2006.583 70.35455 11.556468 0.10677618 T2D-CHD-Can        Dead Female   26.15
```

Thus each person's follow-up is now represented as a number of adjacent time interval (of length at most one year), and the time scale variables indicate the value of the time scale at the left endpoint of these, while the variable `lex.dur` holds the lengths of the intervals.

3.2 Analysis of cancer occurrence

If we want to analyze rates of cancer occurrence, say, we only use the subset of data (the intervals) where persons are at risk of cancer, that is states that do *not* have cancer in them, and conversely we shall use the states that *has* cancer in them as outcomes. So we need to separate the states in two groups:

```
> levels( L1 )
 [1] "Base"         "Can"          "CHD"          "T2D"          "Can-CHD"      "Can-T2D"
 [7] "CHD-Can"      "CHD-T2D"      "T2D-Can"      "T2D-CHD"      "Can-CHD-T2D"  "Can-T2D-CHD"
[13] "CHD-Can-T2D"  "CHD-T2D-Can"  "T2D-Can-CHD"  "T2D-CHD-Can"  "Dead"
> isC <- grep( "Can", levels(L1) )
> ( hasCan <- levels( L1 )[ isC] )
 [1] "Can"          "Can-CHD"      "Can-T2D"      "CHD-Can"      "T2D-Can"      "Can-CHD-T2D"
 [7] "Can-T2D-CHD"  "CHD-Can-T2D"  "CHD-T2D-Can"  "T2D-Can-CHD"  "T2D-CHD-Can"
> (  noCan <- levels( L1 )[-isC] )
```

```
[1] "Base"    "CHD"     "T2D"     "CHD-T2D" "T2D-CHD" "Dead"
```

This illustrates how simple character manipulation tools can be used to select states of interest if the state names have been chosen carefully. A simple replacement of `"Can"` by `"CHD"` will render the analysis of CHD rates instead.

We will look at the rates of exiting to states with cancer, but only for persons (or rather follow-up time) at risk of cancer. This means using the subset where `lex.Cst` is in `noCan` and using as events those with `lex.Xst` in `hasCan`. Since these two sets of states are disjoint and no transitions occur from the `hasCan` states to any of the `noCan` states, we can have a maximum of one transition per person counted.

To get an overview of the transitions to cancer we just make a summary of the `M1` object with all cancer states grouped:

```
> summary( Relevel( M1, list(hasCan=hasCan), print=FALSE, first=FALSE ) )
Transitions:
     To
From          Base   CHD  T2D CHD-T2D T2D-CHD  Dead hasCan  Records:  Events: Risk time:
  Base    2151602  6788  306       0       0 12960  15602   2187258    35656 2012927.18
  CHD           0 30438    0     224       0   900    362     31924     1486   25214.05
  T2D           0     0 1148       0     286     0     20      1454      306    1144.91
  CHD-T2D       0     0    0    1125       0    10     19      1154       29     922.74
  T2D-CHD       0     0    0       0    1143    27     19      1189       46     904.84
  hasCan        0     0    0       0       0  2644  91348     93992     2644   77528.22
  Sum     2151602 37226 1454    1349    1429 16541 107370   2316971    40167 2118641.94

Transitions:
     To
From       Persons:
  Base       175890
  CHD          6790
  T2D           306
  CHD-T2D       224
  T2D-CHD       286
  hasCan      16024
  Sum        175894
```

In order to compare cancer rates between different states of multimorbidity we just include `lex.Cst` as a covariate in the model.

We can now fit a model with natural splines, where we arbitrarily choose knots on the age-scale:

```
> a.kn <- seq(50,90,10)
> mca <- glm( (lex.Xst %in% hasCan) ~ Ns(age,knots=a.kn) + lex.Cst + Bmi_Adj,
+             family = poisson,
+             offset = log( lex.dur ),
+               data = subset( M1, lex.Cst %in% noCan ) )
> round( ci.exp( mca ), 3 )
                       exp(Est.)  2.5% 97.5%
(Intercept)                0.005 0.004 0.005
Ns(age, knots = a.kn)1     4.194 3.902 4.508
Ns(age, knots = a.kn)2     3.653 3.251 4.105
Ns(age, knots = a.kn)3     7.679 6.474 9.107
Ns(age, knots = a.kn)4     3.146 2.422 4.087
lex.CstCHD                 1.204 1.084 1.338
lex.CstT2D                 1.620 1.045 2.513
```

```
lex.CstCHD-T2D              1.584 1.010 2.486
lex.CstT2D-CHD              1.640 1.045 2.573
Bmi_Adj                     0.996 0.993 1.000
```

So we see that hazard ratios of cancer increase by co-morbidity status:

```
> hr <- ci.exp( mca, subset="Cst" )
> rownames( hr ) <- gsub( "lex.Cst", "", rownames(hr) )
> round( hr, 3 )
        exp(Est.)  2.5% 97.5%
CHD         1.204 1.084 1.338
T2D         1.620 1.045 2.513
CHD-T2D     1.584 1.010 2.486
T2D-CHD     1.640 1.045 2.573
> plotEst( hr, lwd=5, cex=2, xlog=TRUE, grid=c(1,1.5,2,3), vref=1,
+          xlab="Cancer HR versus baseline" )
```

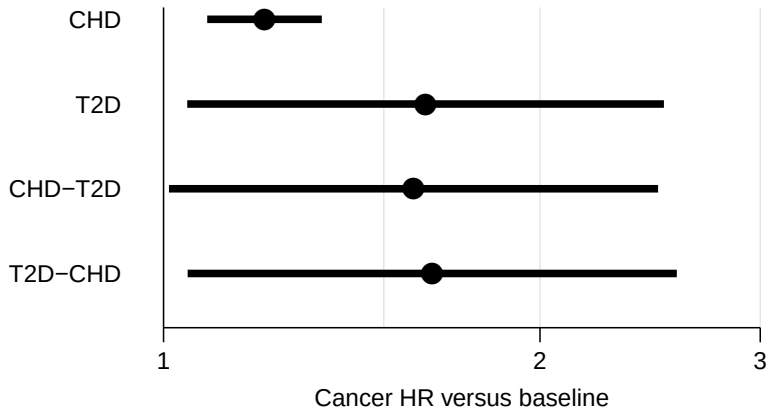

Figure 3.1: *Hazard ratios of cancer relative to baseline associated with different levels of multimorbidity.* `./ana-HR`

We can do a simple Wald-test of equality of the 4 HRs:

```
> ( c4 <- rbind( c(1,-1,0,0),c(0,1,-1,0),c(0,0,1,-1) ) )
     [,1] [,2] [,3] [,4]
[1,]    1   -1    0    0
[2,]    0    1   -1    0
[3,]    0    0    1   -1
> Wald( mca, subset="Cst", ctr.mat=c4 )
    Chisq      d.f.         P
4.2946798 3.0000000 0.2313519
```

The hypothesis tested here is that the four different kinds of morbidity exercise *the same* effect on cancer rates. A Wald test of all 4 HRs being 1 (that is no effect of *any* morbidity) would be:

```
> Wald( mca, subset="Cst", ctr.mat=diag(4) )
       Chisq         d.f.            P
2.480956e+01 4.000000e+00 5.494542e-05
```

which, not surprisingly is strongly significant.

So there is indeed no tangible difference between the morbidity states relative to the baseline state, but the cancer incidence rates in the morbid states are different from the baseline rates.

However, this is a proportional hazards model where we assume that the hazard of cancer is proportional between morbidity states. We can test the proportionality assumption by adding a simple interaction (2 parameter spline):

```
> # mSpline(60:70,knots=65,degree=2,Boundary.Knots=c(40,75))
> # Ns(60:70,knots=c(40,65,75))
> mci <- update( mca, . ~ . + # lex.Cst:mSpline(age,knots=65,degree=2,Boundary.Knots=c(40,7
+                             lex.Cst:Ns(age,knots=c(40,65,75)) )
> anova( mca, mci, test="Chisq" )
Analysis of Deviance Table

Model 1: (lex.Xst %in% hasCan) ~ Ns(age, knots = a.kn) + lex.Cst + Bmi_Adj
Model 2: (lex.Xst %in% hasCan) ~ Ns(age, knots = a.kn) + lex.Cst + Bmi_Adj +
    lex.Cst:Ns(age, knots = c(40, 65, 75))
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1   2203143     181438
2   2203134     181410  9    27.21 0.001291
```

So there are indeed non-proportional hazards. We can illustrate these by plotting the hazards for persons at BMI 25, say, from the two models:

```
> nd <- data.frame( age = 45:90,
+                   lex.dur = 1000,
+                   Bmi_Adj = 25 )
> prr <- NULL
> for( ll in noCan[1:5] )
+    {
+ prr <- cbind( prr, predict( mca, cbind( nd, lex.Cst=ll ) ),
+                     predict( mci, cbind( nd, lex.Cst=ll ) ) )
+    }
> clr <- c("black","red","blue","purple","violet")
> matplot( nd$age, prr,
+          type="l", lty=c("23","solid"),
+          lend="butt", col=rep(clr,each=2), lwd=5,
+          log="y", ylab="Cancer incidence per 1000 PY at BMI 25",
+          xlab="Age (years)", ylim=c(1,4) )
> text( 70, 1.1^(2:6), noCan[1:5], col=clr, adj=0 )
```

We see there is no effect of BMI, but here might well be an interaction between BMI and the time since baseline because the predictive value of the baseline BMI possibly diminishes by time — note that this is a different type of non-proportional hazards, not testable in the usual way in a Cox-model.

```
> mbi <- update( mca, . ~ . + Bmi_Adj:I(tfb/10) )
> round( ci.exp( mbi, pval=TRUE ), 3 )
                       exp(Est.)  2.5% 97.5%     P
(Intercept)                0.005 0.004 0.005 0.000
Ns(age, knots = a.kn)1     3.998 3.715 4.303 0.000
Ns(age, knots = a.kn)2     3.375 2.998 3.800 0.000
Ns(age, knots = a.kn)3     7.227 6.093 8.572 0.000
Ns(age, knots = a.kn)4     3.037 2.342 3.940 0.000
lex.CstCHD                 1.177 1.059 1.308 0.002
```

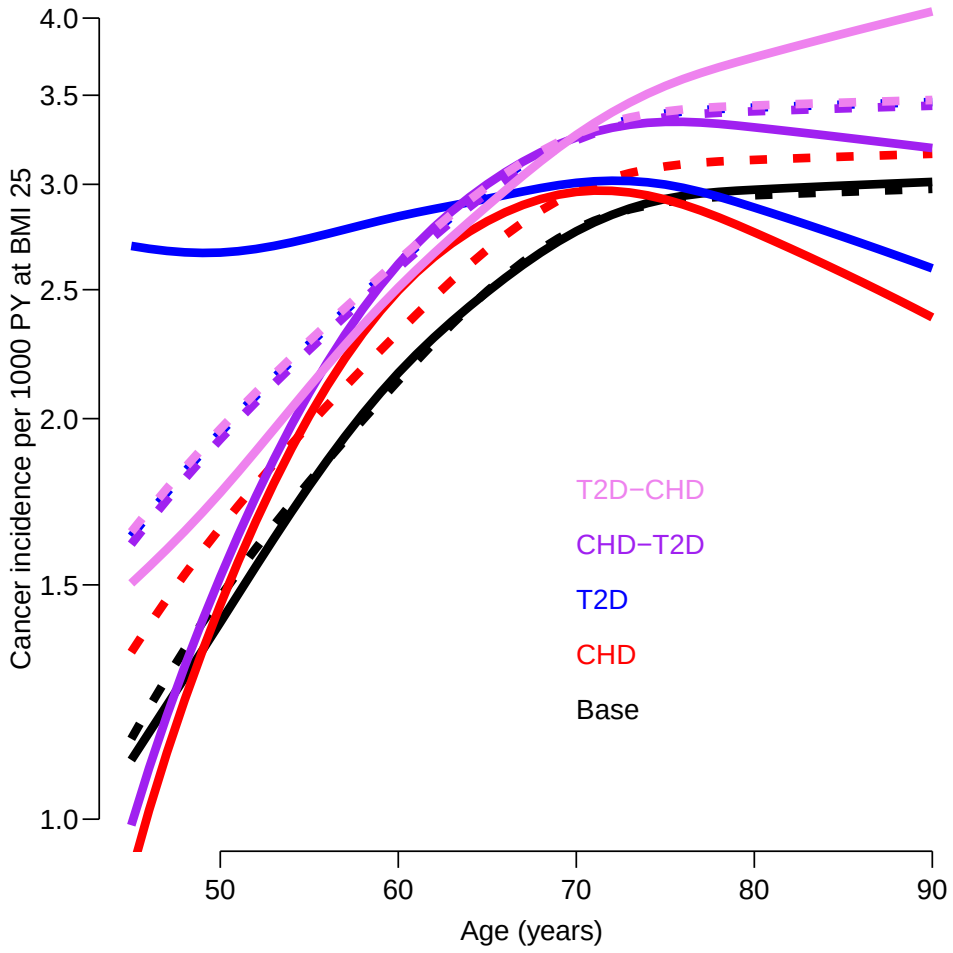

Figure 3.2: *Predicted cancer incidence rates in 5 groups of multimorbidity status, solid lines are from the interaction model, broken lines from the proportional hazards model, and thus parallel. The two groups with the largest number of events (cancer occurrences) are the Base with 15,602 events and the CHD group with 362 events, the remaining three states have only some 20 events each.* `./ana-can-int`

```
lex.CstT2D                  1.611 1.039 2.499 0.033
lex.CstCHD-T2D              1.521 0.969 2.386 0.068
lex.CstT2D-CHD              1.567 0.999 2.459 0.051
Bmi_Adj                     0.993 0.989 0.997 0.000
Bmi_Adj:I(tfb/10)           1.006 1.004 1.008 0.000
```

```
> round( ci.exp( mbi, pval=TRUE,Exp=FALSE ), 3 )
```

```
                      Estimate   2.5%  97.5%     P
(Intercept)             -5.325 -5.426 -5.225 0.000
Ns(age, knots = a.kn)1   1.386  1.312  1.459 0.000
Ns(age, knots = a.kn)2   1.216  1.098  1.335 0.000
Ns(age, knots = a.kn)3   1.978  1.807  2.148 0.000
Ns(age, knots = a.kn)4   1.111  0.851  1.371 0.000
lex.CstCHD               0.163  0.057  0.269 0.002
lex.CstT2D               0.477  0.038  0.916 0.033
lex.CstCHD-T2D           0.419 -0.031  0.870 0.068
lex.CstT2D-CHD           0.449 -0.001  0.900 0.051
```

```
Bmi_Adj                  -0.007 -0.011 -0.003 0.000
Bmi_Adj:I(tfb/10)         0.006  0.004  0.008 0.000
```

Thus there is a formally strong interaction; at time 0 after baseline the HR is 0.993 per kg/m^2, and at 10 years later the effect has all but vanished, and after that it starts to be positive — at least in the model fitted here. The conclusion is that there is a significant effect of BMI at start and a significant effect-modification by time since baseline, but both effects are tiny and of no practical importance.

References

[1] S. Iacobelli and B. Carstensen. Multiple time scales in multi-state models. *Stat Med*, 32(30):5315–5327, Dec 2013.