Demographic changes in the burden of Diabetes and Cancer

SDC
Januar 2015
http://BendixCarstensen.com/DMCa/EpiDMCa
Version 2.0-r

Compiled Monday 26th January, 2015, 19:15
from: /home/bendix/sdc/DMreg/DemoDMCa/Demo-DMr-Ca.tex

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

Contents

Chapter 1

DM and cancer

1.1 Introduction

The link between diabetes and cancer occurrence is well established, and many population-based studies have demonstrated that the association relates to both cancer incidence and mortality [**?**, **?**, **?**].

In this paper I focus on the general population impact of diabetes and cancer at the population level, and in particular how the changes in incidence and mortality rates influence changes in lifetime risk of diabetes and cancer.

1.2 The broader picture

Studies of diabetes and cancer incidence and mortality have traditionally focused only on pairwise comparison of the thick and thin transition rates in Figure **??**. It is commendable to describe variations between these rates that may give clues to mechanisms underlying the different (typically higher) rates among persons with diabetes compared with those without diabetes. For most of the rates in Figure **??**, however, the major determinant is age, so by only *comparing* the rates (controlling for age), the impact of the aging in the population is lost.

If all transitions shown in figure 1.1 were known as functions of age, it is possible to compute the probability of being in any state at any age, and in particular to compute the lifetime risk of any of the conditons, by simply working out the probability of being in each of the death states at an age where everyone is dead.

We will use nationwide Danish data to estimate all 9 sets of rates shown in Figure **??** by sex, age, calendar time and date of birth (age-period-cohort models). This will eventually enable us to illustrate what fraction of persons in a given age who will eventually contract cancer, depending on whether they suffer from diabetes at the given time. It will also provide the possibility to quantify the fraction of persons in a birth cohort who will end in each of the 5 "death" states.

The point to note here is that we will use a set of age-specific rates for each of the 9 transitions. Given a set of rates it is straight-forward to compute the state-occupancy priobabilities at each age including the risk of dying from each of the 5 states.

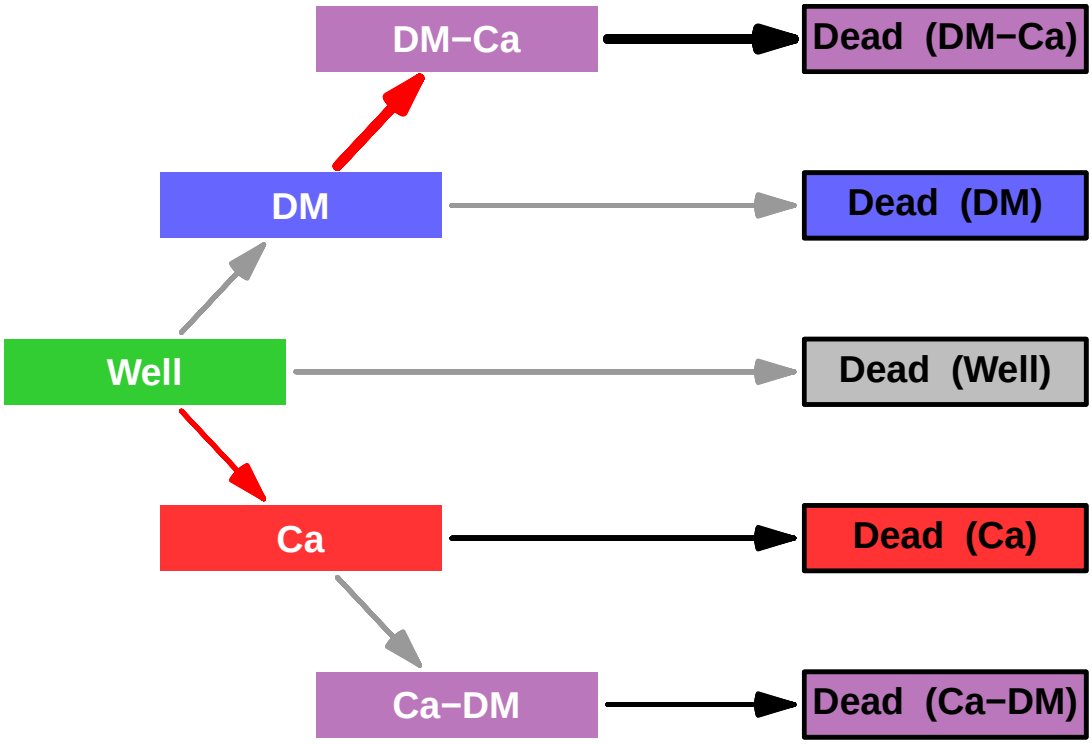

Figure 1.1: *Illustration of the incidence and mortality rates of diabetes and cancer as used to describe the joint population burden of the two diseases.*

1.3 Extensions

It will be possible to use different sets of rates for the calculations; in the paper by Carstensen [1] the cross-sectional rates as of 1 January 2005 were used for the calculations of the lifetime risk of DM and cancer, but it is possible to describe the time-trend by doing the calculations for each year 1995–2012.

Moreover we will explore how a cohort-perspective will modify the results, by trying to define cohort-specific rates. The challenge here is that we only have observations over a 17-year period, and therfore will have to extrapolate cohort-specific rates way outside the obeservation frame.

1.3.1 Counterfactuals

In a scenario like this it will also be possible to quantify the effect of changing the incidence or mortality rates to see how the *relative* size of these influence the lifetime risk time spent in different states. This will give insight into the relative contribution of the incidence rates.

1.3.2 Duration dependence

While it is known that both mortality and cancer incidence depends strongly on diabetes duration, in that it is elevated during the initial period after diagnosis (surveillance bias), the period is for most types of events quite short, so ignoring the duration effects will have only minor influence on the summary measures.

1.4 Methods overview

We merged the Danish National Diabetes Register [2, 3] with the Danish Cancer Register [**?**], and classified all follow up time after 1995 and after any of the two diagnoses by sex, age, calendar time and date of birth in 1-year classes (Lexis triangles). We classified deaths and diagnoses of diabetes and cancer similarly. We also extracted the total population size and number of deaths from the Human Mortality Data Base [**?**]. By subtracting the total number of person-years and deaths in the diabetes and/or cancer population, we obtained the risk time and person-years in the part of the population not diagnosed with any of the two diseases (the "Well" state in Figure **??**).

We then modelled all 9 transition rates shown in Figure **??** using age-period-cohort (APC) models with natural splines [4]. We used separate APC models for all transitions, separately for men and women.

We used the estimated age-specific rates from these models to calculate the burden of disease in a hypothetical population under the scenario of age-specific rates equal to the estimated cross-sectional age-specific rates as of 1 January 1995—2012. The practical calculations were done by multiplying a vector of initial state-distribution (with all persons starting at age 0 in state "Well") successively by the age-specific transition matrices derived from the the rates for every month of age (1220 ages 0–102 years).

A complete account of the data acquisition, rate-estimation and state-probability calculations and graphical displays are available as
http://BendixCarstensen.com/DMCa/EpiDMCa/Demo-DM-Ca.pdf.

Chapter 2

Data base

First we attach the relevant packages and read in a function to acquire data from the Human Mortality Database:

```
> library( foreign )
> library( Epi )

> print( sessionInfo(), l=F )
  R version 3.1.2 (2014-10-31)
  Platform: x86_64-pc-linux-gnu (64-bit)

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

  other attached packages:
  [1] Epi_1.1.67     foreign_0.8-62
```

Referring to figure **??**, all incidence rates and all mortality rates except the transition from the green to the gray box are available from the combined diabetes and cancer register.

The missing mortality rate, namely that among persons without any diagnosis of cancer or diabetes must be derived from the total population mortality by subtracting the mortality among persons with either diabetes or cancer. So we start by acquiring data for the total population mortality.

2.1 Total population follow-up

To this end we first retrieve the total number of deaths from the human mortality database, but we also need data from Statistics Denmark, because deaths in Lexis triangles are only avaiable till 2011, and we have register follow-up to 2012 included.

2.1.1 Mortality data from the Human Mortality Database

In order to fetch mortality from the HMD in 1×1 Lexis triangles we needed to provide a user id and a password, which is hidden in the output here; but they are put in the variables `HMDBusr` and `HMDBpwd`, respectively. We can now get the mortality data for Denmark, and reshape them to our purpose. First we get the deaths in Lexis triangles; note that we also compute the average age and calendar time in the Lexis triangles, since this is going to be used in the modelling:

```
> HMDK <- read.table( "./data/DNK-Deaths-Lexis-HMD.txt",
+                     header=TRUE, skip=2 )[,-6]
> head( HMDK )
  Year Age Cohort  Female    Male
1 1835   0   1835 2158.52 2771.68
2 1835   0   1834 1156.48 1604.32
3 1835   1   1834  502.26  561.56
4 1835   1   1833  363.68  402.14
5 1835   2   1833  293.20  332.44
6 1835   2   1832  288.86  324.86
> str( HMDK )
'data.frame':        39117 obs. of  5 variables:
 $ Year  : int  1835 1835 1835 1835 1835 1835 1835 1835 1835 1835 ...
 $ Age   : Factor w/ 111 levels "0","1","10","100",..: 1 1 2 2 24 24 35 35 46 46 ...
 $ Cohort: Factor w/ 288 levels ".","1725","1726",..: 112 111 111 110 110 109 109 108 108 107 ...
 $ Female: num  2159 1156 502 364 293 ...
 $ Male  : num  2772 1604 562 402 332 ...


> newnames <-  c("P","A","C","F","M")
> cbind( names( HMDK ),   newnames )
               newnames
[1,] "Year"   "P"
[2,] "Age"    "A"
[3,] "Cohort" "C"
[4,] "Female" "F"
[5,] "Male"   "M"
>        names( HMDK ) <- newnames
> HMDK <- transform( HMDK, A = as.numeric(as.character(A)),
+                          C = as.numeric(as.character(C)) )
> HMDK <- subset( HMDK, A < 100 & P > 1994 )
> str( HMDK )
'data.frame':        3400 obs. of  5 variables:
 $ P: int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
 $ A: num  0 0 1 1 2 2 3 3 4 4 ...
 $ C: num  1995 1994 1994 1993 1993 ...
 $ F: num  137 16 8 7 5 3 2 4 2 1 ...
 $ M: num  179 21 13 8 2 7 4 6 5 8 ...
> HMDK$U <- with( HMDK, P-A-C )
> M.dk <- reshape( HMDK, direction = "long",
+                        varying = c("M","F"),
+                        v.names = "D.tot",
+                        timevar = "sex" )#[,-7]
> M.dk <- transform( M.dk, sex = factor( sex, labels=c("M","F") ),
+                          A = A + (1+U)/3,
+                          P = P + (2-U)/3 )[,c("sex","A","P","D.tot")]
> str( M.dk )
'data.frame':        6800 obs. of  4 variables:
 $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A    : num  0.333 0.667 1.333 1.667 2.333 ...
 $ P    : num  1996 1995 1996 1995 1996 ...
 $ D.tot: num  179 21 13 8 2 7 4 6 5 8 ...
> table( round(M.dk$A,1) )

 0.3  0.7  1.3  1.7  2.3  2.7  3.3  3.7  4.3  4.7  5.3  5.7  6.3  6.7  7.3  7.7  8.3  8.7
  34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
 9.3  9.7 10.3 10.7 11.3 11.7 12.3 12.7 13.3 13.7 14.3 14.7 15.3 15.7 16.3 16.7 17.3 17.7
  34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
18.3 18.7 19.3 19.7 20.3 20.7 21.3 21.7 22.3 22.7 23.3 23.7 24.3 24.7 25.3 25.7 26.3 26.7
  34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
27.3 27.7 28.3 28.7 29.3 29.7 30.3 30.7 31.3 31.7 32.3 32.7 33.3 33.7 34.3 34.7 35.3 35.7
  34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
```

```
  36.3 36.7 37.3 37.7 38.3 38.7 39.3 39.7 40.3 40.7 41.3 41.7 42.3 42.7 43.3 43.7 44.3 44.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  45.3 45.7 46.3 46.7 47.3 47.7 48.3 48.7 49.3 49.7 50.3 50.7 51.3 51.7 52.3 52.7 53.3 53.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  54.3 54.7 55.3 55.7 56.3 56.7 57.3 57.7 58.3 58.7 59.3 59.7 60.3 60.7 61.3 61.7 62.3 62.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  63.3 63.7 64.3 64.7 65.3 65.7 66.3 66.7 67.3 67.7 68.3 68.7 69.3 69.7 70.3 70.7 71.3 71.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  72.3 72.7 73.3 73.7 74.3 74.7 75.3 75.7 76.3 76.7 77.3 77.7 78.3 78.7 79.3 79.7 80.3 80.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  81.3 81.7 82.3 82.7 83.3 83.7 84.3 84.7 85.3 85.7 86.3 86.7 87.3 87.7 88.3 88.7 89.3 89.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  90.3 90.7 91.3 91.7 92.3 92.7 93.3 93.7 94.3 94.7 95.3 95.7 96.3 96.7 97.3 97.7 98.3 98.7
    34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34   34
  99.3 99.7
    34   34
```

```
> table( round(M.dk$P,1) )
```

```
  1995.3 1995.7 1996.3 1996.7 1997.3 1997.7 1998.3 1998.7 1999.3 1999.7 2000.3 2000.7
     200    200    200    200    200    200    200    200    200    200    200    200
  2001.3 2001.7 2002.3 2002.7 2003.3 2003.7 2004.3 2004.7 2005.3 2005.7 2006.3 2006.7
     200    200    200    200    200    200    200    200    200    200    200    200
  2007.3 2007.7 2008.3 2008.7 2009.3 2009.7 2010.3 2010.7 2011.3 2011.7
     200    200    200    200    200    200    200    200    200    200
```

```
> range( M.dk$A )
```

```
  [1]  0.3333333 99.6666667
```

```
> range( M.dk$P )
```

```
  [1] 1995.333 2011.667
```

The data frame `M.dk` now have the number of deaths in Lexis triangles between 1995-01-01 and 2011-12-31 in the ages between 0 and 100.

2.1.2 Population data from the Epi package

The total population risk time in Denmark is available from the Epi package in Lexis-triangles in the dataset `Y.dk`

```
> data( Y.dk )
> Y.dk <- subset( Y.dk, P>1994 & P<2012 & A<99 )
> names(Y.dk)[grep("Y",names(Y.dk))] <- "Y.tot"
> Y.dk <- transform( Y.dk, sex = factor( sex, labels=c("M","F") ),
+                          A = A + (1+upper)/3,
+                          P = P + (2-upper)/3 )[,c("sex","A","P","Y.tot")]
```

The data frame `Y.dk` now have the amount of follow-up time in Lexis triangles between 1995-01-01 and 2012-12-31 in the ages between 0 and 99.

We then merge the two dataframe to one; recall that the variable `A` and `P` refer to Lexis triangles, and are coded as the mean age and period in the triangles:

```
> All.dk <- merge( Y.dk, M.dk )
> str( All.dk )
```

```
  'data.frame':        6732 obs. of  5 variables:
   $ sex  : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ Y.tot: num  17026 16470 16434 16066 16198 ...
   $ D.tot: num  137 134 152 132 95 136 138 114 115 110 ...
```

```
> head( All.dk )
```

```
  sex         A        P   Y.tot D.tot
1   F 0.3333333 1995.667 17025.5   137
2   F 0.3333333 1996.667 16469.5   134
3   F 0.3333333 1997.667 16434.0   152
4   F 0.3333333 1998.667 16066.0   132
5   F 0.3333333 1999.667 16198.5    95
6   F 0.3333333 2000.667 16336.5   136
```

We now have all deaths and follow-up time in the total Danish population in the 18-year period 1995-01-01 to 2011-12-31 distributed by Lexis-triangles.

2.2 Follow-up after DM and Cancer

We merged the diabetes register and the cancer register, restricting the cancer register to the first primary tumour in a person, and excluding non-melanoma skin cancers.

Thus the resulting data set has one record per person, and comprises persons that have a diagnosis of cancer or diabetes (incuding person with both diagnoses). Thus we have in this dataset follow-up (and deaths) of patients in the Danish population corresponding to all boxes in figure **??** except the "Well" state.

From the human mortality database we extract the no. of deaths in 1-year Lexis triangles. We also extract the population size, which is used for calculation of person-years in 1-year Lexis triangles. Thus we have deaths and risk time for the total population. We can obtain the figures for the "Well" state by subtraction of risk time and deaths in the patient population from that in the total population.

The patient follow-up is based on the single records of follow-up derived from the merge of the cancer register and the diabetes register.

2.2.1 Follow-up records

First we read the follow-up file from all *patients*, generated by this SAS-program:

The dataset is generated in Lexis-ready-format, so that it can be put into a `Lexis` object anfter a bit of name-grooming and transformaton of the dates to fractions of calendar years:

```
> dc <- read.xport( file="./data/allPTr.xpt" )
> names( dc ) <- gsub( "_", ".", tolower( names(dc) ) )
> str( dc )
  'data.frame':        1031778 obs. of  11 variables:
   $ id   : num  1 2 3 4 5 6 7 8 9 10 ...
   $ diag : num  70 70 NA NA 33 999 91 70 70 NA ...
   $ doca : num  6575 14823 NA NA 17459 ...
   $ sex  : num  2 2 1 2 1 2 1 2 2 2 ...
   $ dobt : num  -11204 -15479 -10166 -14347 -12535 ...
   $ dodd : num  NA 18086 15989 17833 18141 ...
   $ dodm : num  NA NA 13982 10981 NA ...
   $ en.st: Factor w/ 4 levels "Ca","Ca-DM","DM",..: 1 1 3 3 1 1 1 1 1 3 ...
   $ ex.st: Factor w/ 5 levels "Ca","Ca-DM","Dead",..: 1 3 3 3 3 1 1 3 1 4 ...
   $ entry: num  12784 14823 13982 12784 17459 ...
   $ exit : num  18992 18086 15989 17833 18141 ...

> wh <- c( grep( "do", names(dc) ),
+          grep( "ent", names(dc) ),
+          grep( "exi", names(dc) ) )
> names( dc )[wh]

  [1] "doca"  "dobt"  "dodd"  "dodm"  "entry" "exit"
```

```
> dc[,wh] <- dc[,wh]/365.25 + 1960
> dc$sex  <- factor( dc$sex, labels=c("M","F") )
> summary( dc )
```

```
       id              diag            doca        sex            dobt
 Min.   :     1   Min.   : 21.0   Min.   :1943   M:508203   Min.   :1860
 1st Qu.:248382   1st Qu.: 70.0   1st Qu.:1996   F:523575   1st Qu.:1926
 Median :496598   Median : 91.0   Median :2002              Median :1937
 Mean   :496413   Mean   :220.9   Mean   :2000              Mean   :1938
 3rd Qu.:744210   3rd Qu.:241.0   3rd Qu.:2008              3rd Qu.:1948
 Max.   :992753   Max.   :999.0   Max.   :2013              Max.   :2012
                  NA's   :310896  NA's   :311291
      dodd            dodm          en.st          ex.st          entry
 Min.   :1995   Min.   :1942   Ca   :597935   Ca   :231483   Min.   :1995
 1st Qu.:2000   1st Qu.:1996   Ca-DM: 33330   Ca-DM: 42500   1st Qu.:1996
 Median :2004   Median :2002   DM   :357200   Dead :480992   Median :2002
 Mean   :2004   Mean   :2002   DM-Ca: 43313   DM   :220782   Mean   :2002
 3rd Qu.:2009   3rd Qu.:2008                  DM-Ca: 56021   3rd Qu.:2008
 Max.   :2013   Max.   :2012                                 Max.   :2012
 NA's   :485181 NA's   :569694
      exit
 Min.   :1995
 1st Qu.:2004
 Median :2011
 Mean   :2008
 3rd Qu.:2012
 Max.   :2012
```

```
> Ldc <- Lexis( entry = list( age = entry-dobt,
+                             per = entry ),
+                exit = list( per = exit ),
+               entry.status = en.st,
+                exit.status = factor( ex.st,
+                                      levels=c("Well",levels(ex.st)) ),
+                 id = id,
+               data = dc )
```

```
  Incompatible factor levels in entry.status and exit.status:
   both lex.Cst and lex.Xst now have levels:
   Ca Ca-DM DM DM-Ca Well Dead
```

```
> Ldc <- Relevel( Ldc, c(5,3,4,1,2,6) )
> system.time( summary( Ldc ) )
```

```
    user  system elapsed
   0.353   0.000   0.353
```

We illustrate the follow-up among our patients in a figure:

```
> pbox <- boxes( Ldc, boxpos=list(x=c(10,20,50,20,50,80),
+                                 y=c(50,70,90,30,10,50)),
+             scale.Y=1000,
+             show.BE=TRUE, hmult=1.2, wmult=1.1, cex=0.8 )
```

2.3 The analysis data frame

Before we can analyze rates of cancer and diabetes we must include the part of the population that is without any of the two diseases. We have the total amount of person-years and no. of deaths in the data frame `All.dk`. But we must then subtract all risk time and deaths that occur subsequent to either DM or Cancer in order to get the right amount of deaths and PY in the "Well" state.

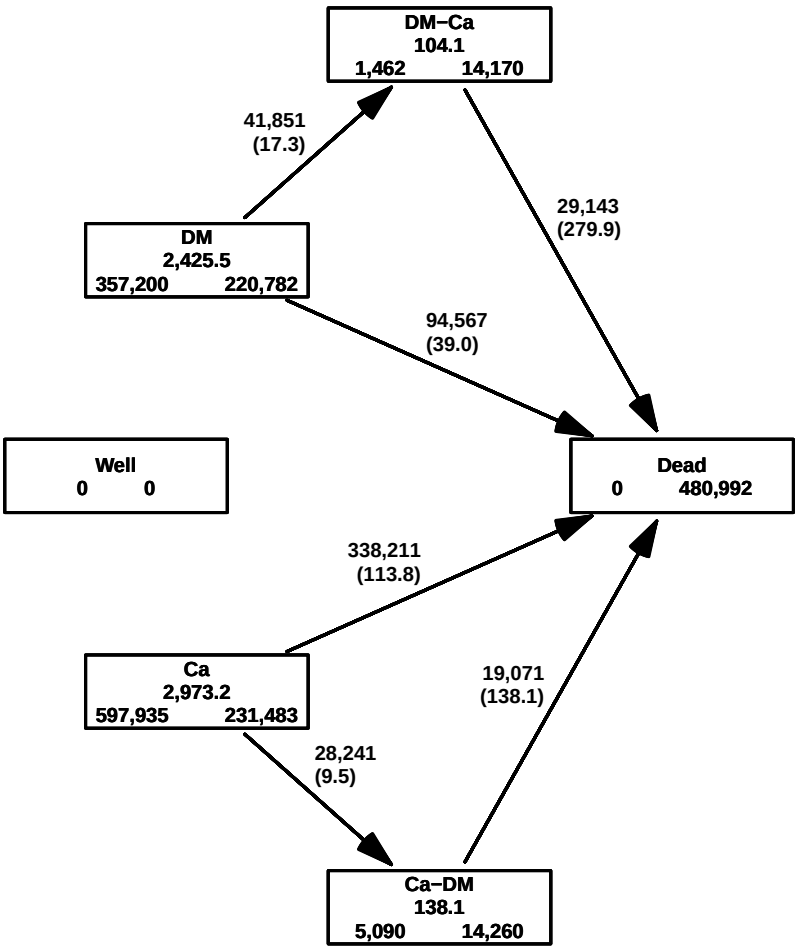

Figure 2.1: *The follow-up of the patients alone. The central number in each box is the amount of follow-up time (in 1000 PY) and the two number at the bottom are the number of persons that enter resp. exit the study in the state. Those entering also include persons that are prevalent cases as of 1.1.1995.*

2.3.1 Patient follow-up

In order to get the risk time among patients we must split the follow-up in the patients by age and calendar time. This is done the classical way, by successively aggreating the risk time and events in tabular form.

The aggregated data frame must be classified by the relevant factors, and must allow counting of events of cancer, diabetes and death.

```
> Agg <- data.frame( A=0, P=0, U=0,
+                    Ldc[1,c("sex","lex.Cst")],
+                    Y=0, D.ca=0, D.dm=0, D.dd=0 )[NULL,]
> names( Agg )[5] <- "state"
> str( Agg )
```

```
  'data.frame':        0 obs. of  9 variables:
   $ A    : num
   $ P    : num
   $ U    : num
   $ sex  : Factor w/ 2 levels "M","F":
   $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..:
   $ Y    : num
   $ D.ca : num
   $ D.dm : num
   $ D.dd : num
> n.chunks <- 20
> lm <- round( seq(0,nrow(Ldc),,n.chunks+1) )
> system.time(
+ for( i in 1:n.chunks )
+ {
+ whr <- (lm[i]+1):(lm[i+1])
+ sLx <- splitLexis( Ldc[whr,], 0:120, time.scale="age" )
+ sLx <- splitLexis( sLx,   1990:2020, time.scale="per" )
+ agg <- with( sLx, aggregate( cbind( y = lex.dur,
+                                  d.dm = ( lex.Xst %in% c("DM","Ca-DM") &
+                                           lex.Xst != lex.Cst )*1,
+                                  d.ca = ( lex.Xst %in% c("Ca","DM-Ca") &
+                                           lex.Xst != lex.Cst )*1,
+                                  d.dd = ( lex.Xst %in% c("Dead") )*1 ),
+                              list( A = floor(age),
+                                    P = floor(per),
+                                    U = floor(per)-floor(age)-floor(dobt),
+                                  sex = sex,
+                                state = lex.Cst ),
+                              FUN = sum ) )
+ Agg <- merge( Agg, agg, by=names( Agg )[1:5], all=TRUE )
+ Agg <- transform( Agg, Y = pmax(Y   ,0,na.rm=TRUE) + pmax(y   ,0,na.rm=TRUE),
+                     D.ca = pmax(D.ca,0,na.rm=TRUE) + pmax(d.ca,0,na.rm=TRUE),
+                     D.dm = pmax(D.dm,0,na.rm=TRUE) + pmax(d.dm,0,na.rm=TRUE),
+                     D.dd = pmax(D.dd,0,na.rm=TRUE) + pmax(d.dd,0,na.rm=TRUE) )[,
+                   c("A","P","U","sex","state","Y","D.ca","D.dm","D.dd")]
+ cat( "Merged in chunk", i, "now", nrow(Agg), "rows, at",
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ } )
  Merged in chunk 1 now 19244 rows, at 2015-01-14 15:14:22
  Merged in chunk 2 now 20707 rows, at 2015-01-14 15:15:00
  Merged in chunk 3 now 21475 rows, at 2015-01-14 15:15:38
  Merged in chunk 4 now 22101 rows, at 2015-01-14 15:16:16
  Merged in chunk 5 now 22480 rows, at 2015-01-14 15:16:54
  Merged in chunk 6 now 22903 rows, at 2015-01-14 15:17:32
  Merged in chunk 7 now 23227 rows, at 2015-01-14 15:18:11
  Merged in chunk 8 now 23446 rows, at 2015-01-14 15:18:48
  Merged in chunk 9 now 23633 rows, at 2015-01-14 15:19:26
  Merged in chunk 10 now 23782 rows, at 2015-01-14 15:20:04
  Merged in chunk 11 now 23994 rows, at 2015-01-14 15:20:41
  Merged in chunk 12 now 24225 rows, at 2015-01-14 15:21:19
  Merged in chunk 13 now 24358 rows, at 2015-01-14 15:21:58
  Merged in chunk 14 now 24643 rows, at 2015-01-14 15:22:36
  Merged in chunk 15 now 24792 rows, at 2015-01-14 15:23:15
  Merged in chunk 16 now 24845 rows, at 2015-01-14 15:23:53
  Merged in chunk 17 now 24948 rows, at 2015-01-14 15:24:31
  Merged in chunk 18 now 25110 rows, at 2015-01-14 15:25:08
  Merged in chunk 19 now 25169 rows, at 2015-01-14 15:25:45
  Merged in chunk 20 now 25238 rows, at 2015-01-14 15:26:22
     user  system elapsed
  754.857   3.265 757.747
> Agg <- transform( Agg, A = A + (1+U)/3,
+                        P = P + (2-U)/3 )
> Agg <- subset( Agg, A<99 & A>0 )
> str( Agg )
```

```
  'data.frame':        23532 obs. of  9 variables:
   $ A    : num  0.333 0.333 0.333 0.667 0.667 ...
   $ P    : num  1996 1996 1996 1995 1995 ...
   $ U    : num  0 0 0 1 1 1 1 0 0 0 ...
   $ sex  : Factor w/ 2 levels "M","F": 1 1 2 1 1 2 2 1 2 2 ...
   $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 2 4 4 2 4 2 4 4 2 4 ...
   $ Y    : num  0.805 1.933 1.136 0.862 0.567 ...
   $ D.ca : num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.dm : num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.dd : num  0 3 0 0 0 0 1 0 0 2 ...
> save( Agg, file="./data/Agg-r.Rda" )
```

2.3.2 Non-patient follow-up

Now `Agg` contains all the follow-up and deaths among the patients, but we will need to subtract the person-years ad the deaths from the total population aggregated (`Agg`) across states:

```
> load( file="./data/Agg-r.Rda" )
> Ptt.dk <- with( Agg, aggregate( cbind( Y.ptt = Y,
+                                        D.ptt = D.dd ),
+                                 list( A=A, P=P, U=U, sex=sex ),
+                                 FUN = sum ) )
```

We then merge the patient risk time and deaths with the total population and subtract them to get the risk time and deaths from the well state:

```
> str( All.dk )
  'data.frame':        6732 obs. of  5 variables:
   $ sex  : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ Y.tot: num  17026 16470 16434 16066 16198 ...
   $ D.tot: num  137 134 152 132 95 136 138 114 115 110 ...
> str( Ptt.dk )
  'data.frame':        6732 obs. of  6 variables:
   $ A    : num  0.333 1.333 2.333 3.333 4.333 ...
   $ P    : num  1996 1996 1996 1996 1996 ...
   $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
   $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ Y.ptt: num  2.74 8.82 10.34 11.86 18.38 ...
   $ D.ptt: num  3 1 0 0 0 0 2 1 0 0 ...
> Well <- merge( All.dk, Ptt.dk, all.x=TRUE )
> Well <- transform( Well, Y = Y.tot - pmax(Y.ptt,0,na.rm=TRUE),
+                          D.dd = D.tot - pmax(D.ptt,0,na.rm=TRUE) )
> Well$D.dd <- pmax( Well$D.dd, 0, na.rm=TRUE )
> str( Well )
  'data.frame':        6732 obs. of  10 variables:
   $ sex  : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ Y.tot: num  17026 16470 16434 16066 16198 ...
   $ D.tot: num  137 134 152 132 95 136 138 114 115 110 ...
   $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
   $ Y.ptt: num  1.14 2.57 2.2 2.75 1.39 ...
   $ D.ptt: num  0 2 4 0 0 0 0 0 1 0 ...
   $ Y    : num  17024 16467 16432 16063 16197 ...
   $ D.dd : num  137 132 148 132 95 136 138 114 114 110 ...
```

2.3.3 Incident cases of DM and Cancer

Finally we must tabulate the number of newly diagnosed DM and Cancer cases (incidences) — the transitions from the "Well" state. This is simply a tabulation in `Ldc` of the entry age and date for records with `lex.Cst` equal to either "DM" or "Ca" with an entry date strictly greater than 1995-01-01 (avoiding counting the persons prevalent at 1995):

```
> summary( Ldc, by=factor(Ldc$per>1995.001,labels=c("Prevalent","Incident")) )
  $Prevalent

  Transitions:
       To
  From    Well    DM DM-Ca    Ca Ca-DM   Dead  Records:  Events: Risk time:  Persons:
    DM       0 26951 11792     0     0  41537     80280   53329  813919.17     80280
    DM-Ca    0     0   106     0     0   1356      1462    1356    7204.68      1462
    Ca       0     0     0 43669 10112  86890    140671   97002 1336329.91    140671
    Ca-DM    0     0     0     0   552   4538      5090    4538   33210.81      5090
    Sum      0 26951 11898 43669 10664 134321    227503  156225 2190664.56    227503

  $Incident

  Transitions:
       To
  From    Well     DM DM-Ca     Ca Ca-DM   Dead  Records:  Events: Risk time:  Persons:
    DM       0 193831 30059      0     0  53030    276920    83089 1611614.70    276920
    DM-Ca    0      0 14064      0     0  27787     41851    27787   96914.14     41851
    Ca       0      0     0 187814 18129 251321    457264   269450 1636842.16    457264
    Ca-DM    0      0     0      0 13707  14533     28240    14533  104884.79     28240
    Sum      0 193831 44123 187814 31836 346671    804275   394859 3450255.79    756088
> Inc <- with( subset( Ldc, per>1995.001 ),
+              aggregate( list( D.dm = (lex.Cst=="DM")*1,
+                               D.ca = (lex.Cst=="Ca")*1 ),
+                         list( sex = sex,
+                                 A = floor(age),
+                                 P = floor(per),
+                                 U = floor(per)-floor(age)-floor(dobt) ),
+                         FUN = sum ) )
> Inc <- transform( Inc, A = A + (1+U)/3,
+                        P = P + (2-U)/3 )
> Inc <- subset( Inc, A < 99 & A > 0 )
```

Then we merge in the number of DM cancer diagnoses from the "Well" state:

```
> str( Well )
  'data.frame':        6732 obs. of  10 variables:
   $ sex  : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ Y.tot: num  17026 16470 16434 16066 16198 ...
   $ D.tot: num  137 134 152 132 95 136 138 114 115 110 ...
   $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
   $ Y.ptt: num  1.14 2.57 2.2 2.75 1.39 ...
   $ D.ptt: num  0 2 4 0 0 0 0 0 1 0 ...
   $ Y    : num  17024 16467 16432 16063 16197 ...
   $ D.dd : num  137 132 148 132 95 136 138 114 114 110 ...
> str( Inc )
  'data.frame':        6713 obs. of  6 variables:
   $ sex : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A   : num  0.333 0.333 1.333 1.333 2.333 ...
   $ P   : num  1996 1996 1996 1996 1996 ...
   $ U   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.dm: num  1 0 4 2 5 1 3 1 5 1 ...
   $ D.ca: num  4 3 7 4 3 4 5 2 1 1 ...
```

```
> intersect( names(Well), names(Inc) )

  [1] "sex" "A"   "P"   "U"

> Well <- transform( merge( Well, Inc, all=TRUE ),
+                    D.dm = pmax( D.dm, 0, na.rm=TRUE ),
+                    D.ca = pmax( D.ca, 0, na.rm=TRUE ),
+                   state = factor( "Well",
+                                   levels=levels(Agg$state),
+                                   labels=levels(Agg$state) ) )
> str( Well )

  'data.frame':        6732 obs. of  13 variables:
   $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 ...
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
   $ Y.tot: num  18028 17426 17387 17038 16953 ...
   $ D.tot: num  179 189 172 142 156 188 149 137 136 151 ...
   $ Y.ptt: num  2.738 0.936 1.125 3.743 2.021 ...
   $ D.ptt: num  3 0 0 0 0 0 1 0 1 0 ...
   $ Y    : num  18025 17426 17386 17034 16951 ...
   $ D.dd : num  176 189 172 142 156 188 148 137 135 151 ...
   $ D.dm : num  1 0 1 2 1 1 0 1 1 1 ...
   $ D.ca : num  4 2 1 4 4 1 2 5 2 5 ...
   $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 1 1 1 1 1 1 1 1 1 1 ...

> str( Agg )

  'data.frame':        23532 obs. of  9 variables:
   $ A    : num  0.333 0.333 0.333 0.667 0.667 ...
   $ P    : num  1996 1996 1996 1995 1995 ...
   $ U    : num  0 0 0 1 1 1 1 0 0 0 ...
   $ sex  : Factor w/ 2 levels "M","F": 1 1 2 1 1 2 2 1 2 2 ...
   $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 2 4 4 2 4 2 4 4 2 4 ...
   $ Y    : num  0.805 1.933 1.136 0.862 0.567 ...
   $ D.ca : num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.dm : num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.dd : num  0 3 0 0 0 0 1 0 0 2 ...
```

Finally we can stack the two databases:

```
> dcd <- rbind( Well[,names(Agg)], Agg )
> str( dcd )
  'data.frame':        30264 obs. of  9 variables:
   $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
   $ P    : num  1996 1997 1998 1999 2000 ...
   $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
   $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 1 1 1 1 1 1 1 1 1 1 ...
   $ Y    : num  18025 17426 17386 17034 16951 ...
   $ D.ca : num  4 2 1 4 4 1 2 5 2 5 ...
   $ D.dm : num  1 0 1 2 1 1 0 1 1 1 ...
   $ D.dd : num  176 189 172 142 156 188 148 137 135 151 ...

> save( dcd, file="./data/dcd-r.Rda" )
```

A tabulation of the possible events from various states shows that we have precisely nine entries with events corresponding to the 9 transitions in figure 1.1 and precisely 5 entries with person-years, corresponding to the 5 transient states in the figure.

```
> cbind(
+ xtabs( cbind( D.ca, D.dm, D.dd ) ~ state, data=dcd ), round(
+ xtabs( Y/1000 ~ state, data=dcd ), 1 ) )
```

```
         D.ca   D.dm   D.dd
Well   457106 276847 484737 86078.4
DM      41844      0  94088  2424.6
DM-Ca       0      0  29122   104.1
Ca          0  28225 337066  2968.8
Ca-DM       0      0  19014   138.0
Dead        0      0      0     0.0
```

```
> ftable( xtabs( cbind( D.dm, D.ca, D.dd ) ~ floor(P) + state,
+                data=dcd ),
+         row.vars=c(3,1) )
```

```
                state  Well    DM DM-Ca    Ca Ca-DM  Dead
      floor(P)
D.dm  1995            12307     0     0  1007     0     0
      1996            12458     0     0  1039     0     0
      1997            12057     0     0  1073     0     0
      1998            13235     0     0  1130     0     0
      1999            13752     0     0  1287     0     0
      2000            13770     0     0  1309     0     0
      2001            14157     0     0  1356     0     0
      2002            16581     0     0  1689     0     0
      2003            17964     0     0  1871     0     0
      2004            18304     0     0  1778     0     0
      2005            16323     0     0  1520     0     0
      2006            15919     0     0  1656     0     0
      2007            17047     0     0  1768     0     0
      2008            18081     0     0  1965     0     0
      2009            18901     0     0  2109     0     0
      2010            20141     0     0  2415     0     0
      2011            25850     0     0  3253     0     0
D.ca  1995            21108  1226     0     0     0     0
      1996            23435  1183     0     0     0     0
      1997            23914  1390     0     0     0     0
      1998            24617  1526     0     0     0     0
      1999            24830  1793     0     0     0     0
      2000            24894  1687     0     0     0     0
      2001            25087  1949     0     0     0     0
      2002            25612  1967     0     0     0     0
      2003            25797  2205     0     0     0     0
      2004            26898  2616     0     0     0     0
      2005            27539  2741     0     0     0     0
      2006            28582  3016     0     0     0     0
      2007            29139  3154     0     0     0     0
      2008            30728  3516     0     0     0     0
      2009            32360  3759     0     0     0     0
      2010            31257  3872     0     0     0     0
      2011            31309  4244     0     0     0     0
D.dd  1995            35779  5128   852 20076   895     0
      1996            34640  4814   884 19428   889     0
      1997            33036  4861  1025 19745   912     0
      1998            31737  4892  1122 19519   827     0
      1999            31488  5104  1241 19926   993     0
      2000            30304  5147  1323 19822   967     0
      2001            30342  5244  1411 19860  1051     0
      2002            30209  5523  1461 19845  1104     0
      2003            26981  5635  1678 19849  1191     0
      2004            29620  5552  1754 19060  1106     0
      2005            26480  5755  1920 19126  1158     0
      2006            25979  5874  2074 19871  1167     0
      2007            25626  5977  2165 20023  1250     0
      2008            24575  5818  2250 20079  1297     0
      2009            24216  6304  2512 19942  1282     0
      2010            22705  6265  2667 20571  1430     0
      2011            21020  6195  2783 20324  1495     0
```

```
> print( sessionInfo(), l=F )

  R version 3.1.2 (2014-10-31)
  Platform: x86_64-pc-linux-gnu (64-bit)

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

  other attached packages:
  [1] Epi_1.1.67
```

Chapter 3

Modelling of rates

First we load the data and chek the number of events of different types from different states:

```
> clear()
> load( file="./data/dcd-r.Rda" )
> dcd <- subset( dcd, P<2012 )
> ftable( round(
+         addmargins( xtabs( cbind(D.dm,D.ca,D.dd,PY=Y/1000) ~ sex + state, data=dcd ),
+                     1 ) ),
+         row.vars=1:2 )
                D.dm   D.ca   D.dd     PY
  sex state
  M   Well    153749 221971 235823  42826
      DM           0  23871  49758   1310
      DM-Ca        0      0  16841     54
      Ca       13500      0 164078   1129
      Ca-DM        0      0   9188     57
      Dead         0      0      0      0
  F   Well    123098 235135 248914  43252
      DM           0  17973  44330   1115
      DM-Ca        0      0  12281     50
      Ca       14725      0 172988   1840
      Ca-DM        0      0   9826     81
      Dead         0      0      0      0
  Sum Well    276847 457106 484737  86078
      DM           0  41844  94088   2425
      DM-Ca        0      0  29122    104
      Ca       28225      0 337066   2969
      Ca-DM        0      0  19014    138
      Dead         0      0      0      0
```

From the table we see that we have events for estimating 9 different rates, and also that we have ample data for estimating them. To decide how to distribute knots in modelling of the age-effects, we make histograms of the age-distribution of the events:

```
> par( mfrow=c(5,3), mar=c(3,1,1,1), mgp=c(3,1,0)/1.6 )
> par( mfg=c(1,1) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dm), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="DM | Well" ) )
> par( mfg=c(1,2) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.ca), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Ca | Well" ) )
> par( mfg=c(1,3) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
```

```
+                          ylab="", xlab="Dead | Well" ) )
> par( mfg=c(2,2) ) ; with( subset( dcd, state=="DM" ),
>                          hist( rep(A,D.ca), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="Ca | DM" ) )
> par( mfg=c(2,3) ) ; with( subset( dcd, state=="DM" ),
>                          hist( rep(A,D.dd), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="Dead | DM" ) )
> par( mfg=c(3,3) ) ; with( subset( dcd, state=="DM-Ca" ),
>                          hist( rep(A,D.dd), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="Dead | DM-Ca" ) )
> par( mfg=c(4,1) ) ; with( subset( dcd, state=="Ca" ),
>                          hist( rep(A,D.dm), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="DM | Ca" ) )
> par( mfg=c(4,3) ) ; with( subset( dcd, state=="Ca" ),
>                          hist( rep(A,D.dd), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="Dead | Ca" ) )
> par( mfg=c(5,3) ) ; with( subset( dcd, state=="Ca-DM" ),
>                          hist( rep(A,D.dd), breaks=0:100,
+                          col="black", main="", yaxt="n",
+                          ylab="", xlab="Dead | Ca-DM" ) )
```

3.1 APC-models for the transition rates

We model the 9 different rates by separate age-period-cohort (APC) models. For convenience we wrap the fitting in a function calling `apc.fit`. In the definition of the function we put in the default number of knots for the age-, period- and cohort-effects.

Moreover, for estimates of age-effects we want both a parametrization with a reference period (2010) and a cohort effect as 0 on average, and one with a reference cohort (1935) and a period effect as 0 on average. 2010 is chosen an a conveniently recent date for evaluation of crossectional rates and 1935 as the cohort which is contributing risk time from ages 60 trough 77, an age range where both diabetes and cancer is relatively common.

The models returned by the `apc.fit` function in the element `Model` is a model which is parametrized in a special way, using specially constructed design matrices in the linear predictor, and thus is not suitable for prediction based on the input data frame. Hence we also append a model fitted using a parametrization suitable for prediction, which we shall need later:

```
> tr.apc <-
+ function( event, st, sx, rf.p=2005, rf.c=1935 )
+ {
+ dfr <- subset( dcd, state==st & sex==sx )
+ dfr$D <- dfr[,event]
+ dfr <- dfr[,c("A","P","D","Y")]
+ qnt <- function(x,n) quantile(x,probs=(1:n-0.5)/n)
+ kpos <- list( A = qnt( with(dfr,rep(  A,D)), 10 ),
+               P = qnt( with(dfr,rep(P  ,D)),  5 ),
+               C = qnt( with(dfr,rep(P-A,D)),  7 ) )
+ apc <- apc.fit( dfr, parm = "APC", npar = kpos, ref.p = rf.p, scale = 1000 )
+ acp <- apc.fit( dfr, parm = "ACP", npar = kpos, ref.c = rf.c, scale = 1000,
+                 print.AOV = FALSE)
+ # chop off the cohort effects after 1990 for nicer plots
+ apc$Coh <- apc$Coh[apc$Coh[,"Coh"]<1990,]
+ acp$Coh <- acp$Coh[acp$Coh[,"Coh"]<1990,]
```

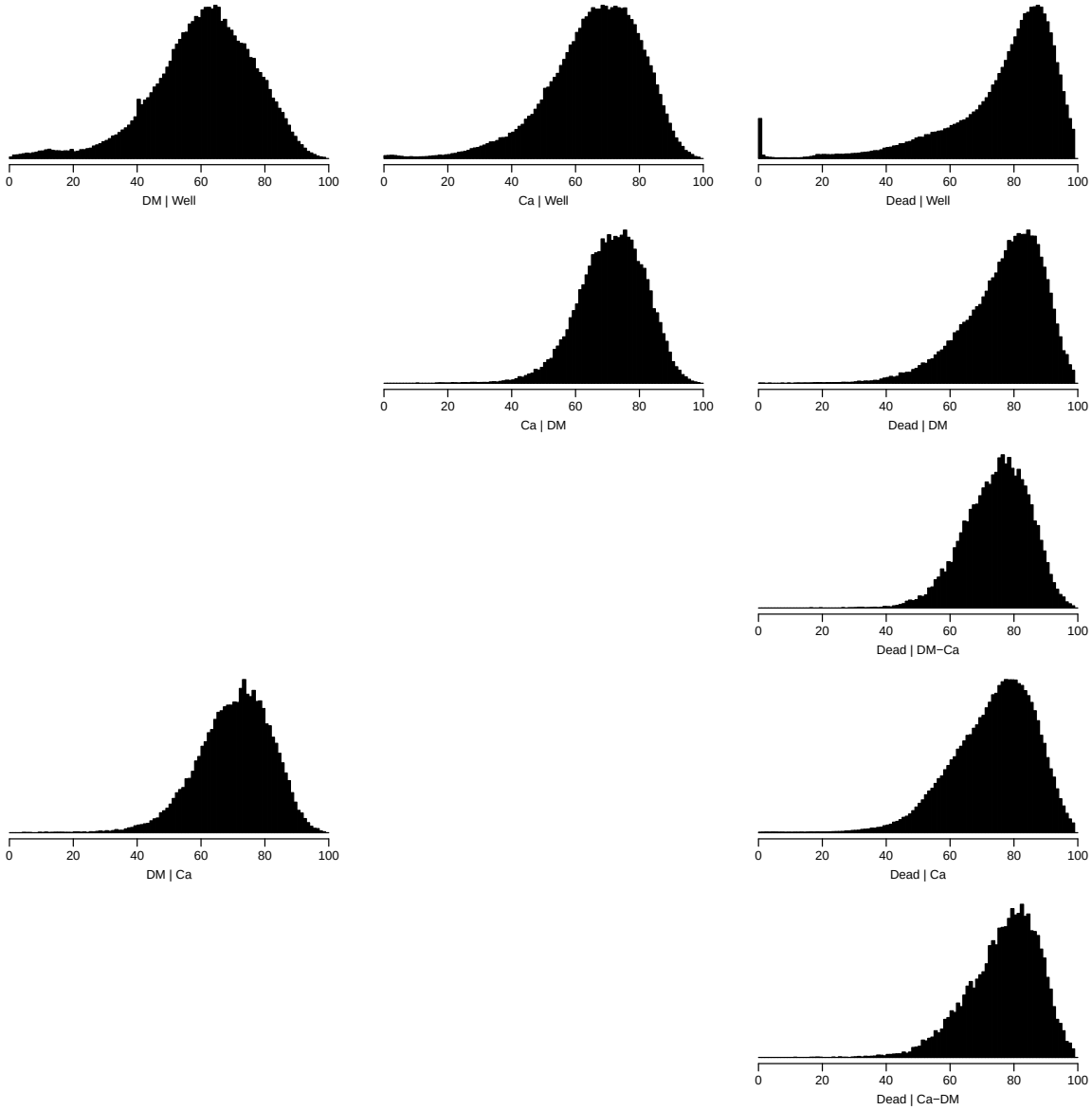

Figure 3.1: *Histograms of the age at event for the 9 possible transitions. Clearly, nothing much is happening in the younger ages, so we shall have age-knots a little closer in the older ages.*

```
+ Kn <- apc$Knots
+ c( list( apc=apc, acp=acp ),
+    list( model = glm( D ~ Ns(  A,knots=Kn$Age) +
+                           Ns(P  ,knots=Kn$Per) +
+                           Ns(P-A,knots=Kn$Coh)[,-1], # avoid singularity
+                       offset = log(Y),
+                       family = poisson,
+                         data = dfr ) ) )
+ }
> # Men
> M.w2dm  <- tr.apc( "D.dm", "Well" , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model
```

```
                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3356     9952.0
  Age-drift              3355     5479.0  1   4473.0 < 2.2e-16
  Age-Cohort             3350     5433.7  5     45.3 1.258e-08
  Age-Period-Cohort      3347     4820.4  3    613.3 < 2.2e-16
  Age-Period             3352     4861.8 -5    -41.3 8.069e-08
  Age-drift              3355     5479.0 -3   -617.3 < 2.2e-16
> M.w2ca  <- tr.apc( "D.ca", "Well" , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3356     8087.7
  Age-drift              3355     5536.6  1  2551.11 < 2.2e-16
  Age-Cohort             3350     5375.3  5   161.34 < 2.2e-16
  Age-Period-Cohort      3347     5288.3  3    86.99 < 2.2e-16
  Age-Period             3352     5447.6 -5  -159.36 < 2.2e-16
  Age-drift              3355     5536.6 -3   -88.97 < 2.2e-16
> M.w2dd  <- tr.apc( "D.dd", "Well" , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3356      29868
  Age-drift              3355      22566  1   7302.2 < 2.2e-16
  Age-Cohort             3350      21173  5   1393.4 < 2.2e-16
  Age-Period-Cohort      3347      21100  3     72.6 1.201e-15
  Age-Period             3352      22483 -5  -1382.7 < 2.2e-16
  Age-drift              3355      22566 -3    -83.3 < 2.2e-16
> M.dm2ca <- tr.apc( "D.ca", "DM"   , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                    3351     2738.4
  Age-drift              3350     2568.6  1  169.775  < 2e-16
  Age-Cohort             3345     2555.9  5   12.667  0.02671
  Age-Period-Cohort      3342     2552.1  3    3.874  0.27545
  Age-Period             3347     2564.8 -5  -12.760  0.02573
  Age-drift              3350     2568.6 -3   -3.780  0.28621
> M.dm2dd <- tr.apc( "D.dd", "DM"   , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3351     5267.6
  Age-drift              3350     3134.2  1  2133.33 < 2.2e-16
  Age-Cohort             3345     2970.8  5   163.49 < 2.2e-16
  Age-Period-Cohort      3342     2958.2  3    12.53  0.005782
  Age-Period             3347     3131.8 -5  -173.62 < 2.2e-16
  Age-drift              3350     3134.2 -3    -2.40  0.493520
> M.ca2dm <- tr.apc( "D.dm", "Ca"   , "M" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
```

```
Age                    3356     3326.3
Age-drift              3355     3147.1  1  179.147 < 2.2e-16
Age-Cohort             3350     3112.8  5   34.336 2.041e-06
Age-Period-Cohort      3347     3022.5  3   90.265 < 2.2e-16
Age-Period             3352     3056.0 -5  -33.466 3.041e-06
Age-drift              3355     3147.1 -3  -91.136 < 2.2e-16
```

```
> M.ca2dd <- tr.apc( "D.dd", "Ca"   , "M" )
```

```
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   3356    11760.5
  Age-drift             3355     5551.9  1   6208.5 < 2.2e-16
  Age-Cohort            3350     5008.4  5    543.5 < 2.2e-16
  Age-Period-Cohort     3347     4968.6  3     39.8 1.186e-08
  Age-Period            3352     5464.3 -5   -495.7 < 2.2e-16
  Age-drift             3355     5551.9 -3    -87.6 < 2.2e-16
```

```
> M.cd2dd <- tr.apc( "D.dd", "Ca-DM", "M" )
```

```
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   2686     2763.3
  Age-drift             2685     2553.9  1  209.399 < 2.2e-16
  Age-Cohort            2680     2534.0  5   19.895  0.001308
  Age-Period-Cohort     2677     2526.6  3    7.392  0.060398
  Age-Period            2682     2544.3 -5  -17.688  0.003364
  Age-drift             2685     2553.9 -3   -9.599  0.022298
```

```
> M.dc2dd <- tr.apc( "D.dd", "DM-Ca", "M" )
```

```
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   2314     3815.9
  Age-drift             2313     2489.4  1  1326.52 < 2.2e-16
  Age-Cohort            2308     2484.0  5     5.38 0.3718256
  Age-Period-Cohort     2305     2465.1  3    18.93 0.0002827
  Age-Period            2310     2471.8 -5    -6.75 0.2399052
  Age-drift             2313     2489.4 -3   -17.55 0.0005434
```

```
> # Women
> F.w2dm  <- tr.apc( "D.dm", "Well" , "F" )
```

```
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   3356     9759.4
  Age-drift             3355     6631.7  1  3127.65 < 2.2e-16
  Age-Cohort            3350     6456.8  5   174.87 < 2.2e-16
  Age-Period-Cohort     3347     5529.9  3   926.97 < 2.2e-16
  Age-Period            3352     5689.2 -5  -159.35 < 2.2e-16
  Age-drift             3355     6631.7 -3  -942.49 < 2.2e-16
```

```
> F.w2ca  <- tr.apc( "D.ca", "Well" , "F" )
```

```
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
```

```
  Age                    3356     6949.3
  Age-drift              3355     5380.8  1  1568.57 < 2.2e-16
  Age-Cohort             3350     5248.7  5   132.02 < 2.2e-16
  Age-Period-Cohort      3347     5129.2  3   119.51 < 2.2e-16
  Age-Period             3352     5267.7 -5  -138.51 < 2.2e-16
  Age-drift              3355     5380.8 -3  -113.02 < 2.2e-16
> F.w2dd  <- tr.apc( "D.dd", "Well" , "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   3356      27952
  Age-drift             3355      24136  1   3816.1 < 2.2e-16
  Age-Cohort            3350      22638  5   1498.4 < 2.2e-16
  Age-Period-Cohort     3347      22602  3     35.5 9.391e-08
  Age-Period            3352      24080 -5  -1478.3 < 2.2e-16
  Age-drift             3355      24136 -3    -55.7 4.827e-12
> F.dm2ca <- tr.apc( "D.ca", "DM"   , "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                   3347     2886.8
  Age-drift             3346     2708.2  1  178.530  < 2e-16
  Age-Cohort            3341     2699.8  5    8.436  0.13377
  Age-Period-Cohort     3338     2693.5  3    6.338  0.09626
  Age-Period            3343     2701.8 -5   -8.365  0.13722
  Age-drift             3346     2708.2 -3   -6.410  0.09330
> F.dm2dd <- tr.apc( "D.dd", "DM"   , "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   3347     4635.8
  Age-drift             3346     3119.7  1  1516.10 < 2.2e-16
  Age-Cohort            3341     2921.4  5   198.36 < 2.2e-16
  Age-Period-Cohort     3338     2871.9  3    49.46 1.040e-10
  Age-Period            3343     3097.1 -5  -225.16 < 2.2e-16
  Age-drift             3346     3119.7 -3   -22.66 4.751e-05
> F.ca2dm <- tr.apc( "D.dm", "Ca"   , "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   3356     3414.0
  Age-drift             3355     3073.7  1   340.36 < 2.2e-16
  Age-Cohort            3350     3051.0  5    22.67 0.0003903
  Age-Period-Cohort     3347     2881.8  3   169.22 < 2.2e-16
  Age-Period            3352     2904.5 -5   -22.76 0.0003755
  Age-drift             3355     3073.7 -3  -169.13 < 2.2e-16
> F.ca2dd <- tr.apc( "D.dd", "Ca"   , "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                   3356     7455.6
```

```
  Age-drift              3355     4707.2  1  2748.44  < 2e-16
  Age-Cohort             3350     4329.0  5   378.21  < 2e-16
  Age-Period-Cohort      3347     4325.0  3     3.99  0.26293
  Age-Period             3352     4697.0 -5  -372.05  < 2e-16
  Age-drift              3355     4707.2 -3   -10.15  0.01737
> F.cd2dd <- tr.apc( "D.dd", "Ca-DM", "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                   2617     2783.5
  Age-drift             2616     2617.8  1  165.659   <2e-16
  Age-Cohort            2611     2616.4  5    1.396   0.9248
  Age-Period-Cohort     2608     2612.9  3    3.479   0.3235
  Age-Period            2613     2614.5 -5   -1.598   0.9015
  Age-drift             2616     2617.8 -3   -3.277   0.3508
> F.dc2dd <- tr.apc( "D.dd", "DM-Ca", "F" )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                   2425     3022.9
  Age-drift             2424     2513.3  1   509.61 < 2.2e-16
  Age-Cohort            2419     2496.3  5    16.96  0.004575
  Age-Period-Cohort     2416     2491.8  3     4.47  0.214701
  Age-Period            2421     2509.9 -5   -18.04  0.002895
  Age-drift             2424     2513.3 -3    -3.39  0.334993
```

Having fitted all 18 APC-models we can graph the estimated rates as well as the cohort- and period effects from the two different parametrizations:

```
> apc.fr <- function( rl, rt, rf=1, ...){
+ apc.frame( a.lab = seq(10,90,20),
+             a.tic = seq(15,95,5),
+            cp.lab = seq(1900,2015,20),
+            cp.tic = seq(1900,2015,5),
+             r.lab = rl,
+             r.tic = rt,
+            rr.ref = rf,
+             a.txt = "",
+            cp.txt = "",
+             r.txt = "",
+            rr.txt = "",
+          ref.line = TRUE,
+               gap = 10, ... )
+ }
> inc.fr <- function(...){
+ apc.fr( rl = c(c(5)/100,c(1,2,5)/10,c(1,2,5),c(1,2,5)*10,100),
+         rt = c(2:9/100,1:9/10,1:9,1:7*10), ... )
+ }
> mort.fr <- function(...){
+ apc.fr( rl = c(c(2,5)/10,c(1,2,5),c(1,2,5)*10,c(1,2,5)*100),
+         rt = c(2:9/10,1:9,1:9*10,1:9*100,1000), rf=10, ... )
+ }
> par( mfcol=c(2,2), mar=c(0,0,0.5,1), oma=c(4,4,2,3),
+      mgp=c(3,1,0)/1.6, las=1, bty="n" )
> inc.fr(sides=2,col.grid=gray(0.9))
> lines( M.dm2ca$apc, col="blue", lend=1, lwd=4 )
> lines( M.w2ca$apc , col="forestgreen", lend=1, lwd=4 )
> lines( M.w2dm$apc , col="forestgreen", lty="21", lend=1, lwd=4 )
> lines( M.ca2dm$apc, col="red", lty="21", lend=1, lwd=4 )
```

```
> mort.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.w2dd$apc , col="forestgreen", lend=1, lwd=4 )
> lines( M.dm2dd$apc, col="blue", lend=1, lwd=4 )
> lines( M.ca2dd$apc, col="red", lend=1, lwd=4 )
> lines( M.cd2dd$apc, col="red", lty="12", lend=1, lwd=4 )
> lines( M.dc2dd$apc, col="red", lty="21", lend=1, lwd=4 )
> inc.fr(sides=4,col.grid=gray(0.9))
> lines( F.dm2ca$apc, col="blue", lend=1, lwd=4 )
> lines( F.w2ca$apc, col="forestgreen", lend=1, lwd=4 )
> lines( F.w2dm$apc, col="forestgreen", lty="21", lend=1, lwd=4 )
> lines( F.ca2dm$apc, col="red", lty="21", lend=1, lwd=4 )
> mort.fr(sides=c(1,4),col.grid=gray(0.9))
> lines( F.w2dd$apc, col="forestgreen", lend=1, lwd=4 )
> lines( F.dm2dd$apc, col="blue", lend=1, lwd=4 )
> lines( F.ca2dd$apc, col="red", lend=1, lwd=4 )
> lines( F.cd2dd$apc, col="red", lty="12", lend=1, lwd=4 )
> lines( F.dc2dd$apc, col="red", lty="21", lend=1, lwd=4 )
> mtext( "Men"  , at=0.25, side=3, outer=TRUE, line=0.5 )
> mtext( "Women", at=0.75, side=3, outer=TRUE, line=0.5 )
> mtext( "Incidence rates per 1000 PY", at=0.75, side=2, outer=TRUE, line=2.5, las=0 )
> mtext( "Mortality rates per 1000 PY", at=0.25, side=2, outer=TRUE, line=2.5, las=0 )
> mtext( "Rate ratio", at=0.5, side=4, outer=TRUE, line=1.5, las=0 )


> par( mfcol=c(2,2), mar=c(0,0,0.5,1), oma=c(4,4,2,3),
+      mgp=c(3,1,0)/1.6, las=1 )
> inc.fr(sides=2,col.grid=gray(0.9))
> lines( M.dm2ca$acp, col="blue", lend=1, lwd=4 )
> lines( M.w2ca$acp , col="forestgreen", lend=1, lwd=4 )
> lines( M.w2dm$acp , col="forestgreen", lty="21", lend=1, lwd=4 )
> lines( M.ca2dm$acp, col="red", lty="21", lend=1, lwd=4 )
> mort.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.w2dd$acp , col="forestgreen", lend=1, lwd=4 )
> lines( M.dm2dd$acp, col="blue", lend=1, lwd=4 )
> lines( M.ca2dd$acp, col="red", lend=1, lwd=4 )
> lines( M.cd2dd$acp, col="red", lty="12", lend=1, lwd=4 )
> lines( M.dc2dd$acp, col="red", lty="21", lend=1, lwd=4 )
> inc.fr(sides=4,col.grid=gray(0.9))
> lines( F.dm2ca$acp, col="blue", lend=1, lwd=4 )
> lines( F.w2ca$acp, col="forestgreen", lend=1, lwd=4 )
> lines( F.w2dm$acp, col="forestgreen", lty="21", lend=1, lwd=4 )
> lines( F.ca2dm$acp, col="red", lty="21", lend=1, lwd=4 )
> mort.fr(sides=c(1,4),col.grid=gray(0.9))
> lines( F.w2dd$acp, col="forestgreen", lend=1, lwd=4 )
> lines( F.dm2dd$acp, col="blue", lend=1, lwd=4 )
> lines( F.ca2dd$acp, col="red", lend=1, lwd=4 )
> lines( F.cd2dd$acp, col="red", lty="12", lend=1, lwd=4 )
> lines( F.dc2dd$acp, col="red", lty="21", lend=1, lwd=4 )
> mtext( "Men"  , at=0.25, side=3, outer=TRUE, line=0.5 )
> mtext( "Women", at=0.75, side=3, outer=TRUE, line=0.5 )
> mtext( "Incidence rates per 1000 PY", at=0.75, side=2, outer=TRUE, line=2.5, las=0 )
> mtext( "Mortality rates per 1000 PY", at=0.25, side=2, outer=TRUE, line=2.5, las=0 )
> mtext( "Rate ratio", at=0.5, side=4, outer=TRUE, line=1.5, las=0 )
```

It is clear from the figures 3.2 and 3.3 that only mortality rates exhibit non-linearity by date of birth, and in particular that incidence rates are increasing with time and mortality rates are decreasing with time.

Finally we save the fitted APC-models for further use:

```
> save( M.w2dm,M.w2ca,M.w2dd,M.dm2ca,M.dm2dd,M.ca2dm,M.ca2dd,M.cd2dd,M.dc2dd,
+       F.w2dm,F.w2ca,F.w2dd,F.dm2ca,F.dm2dd,F.ca2dm,F.ca2dd,F.cd2dd,F.dc2dd,
+       file = "./data/APC-r.Rda" )
```

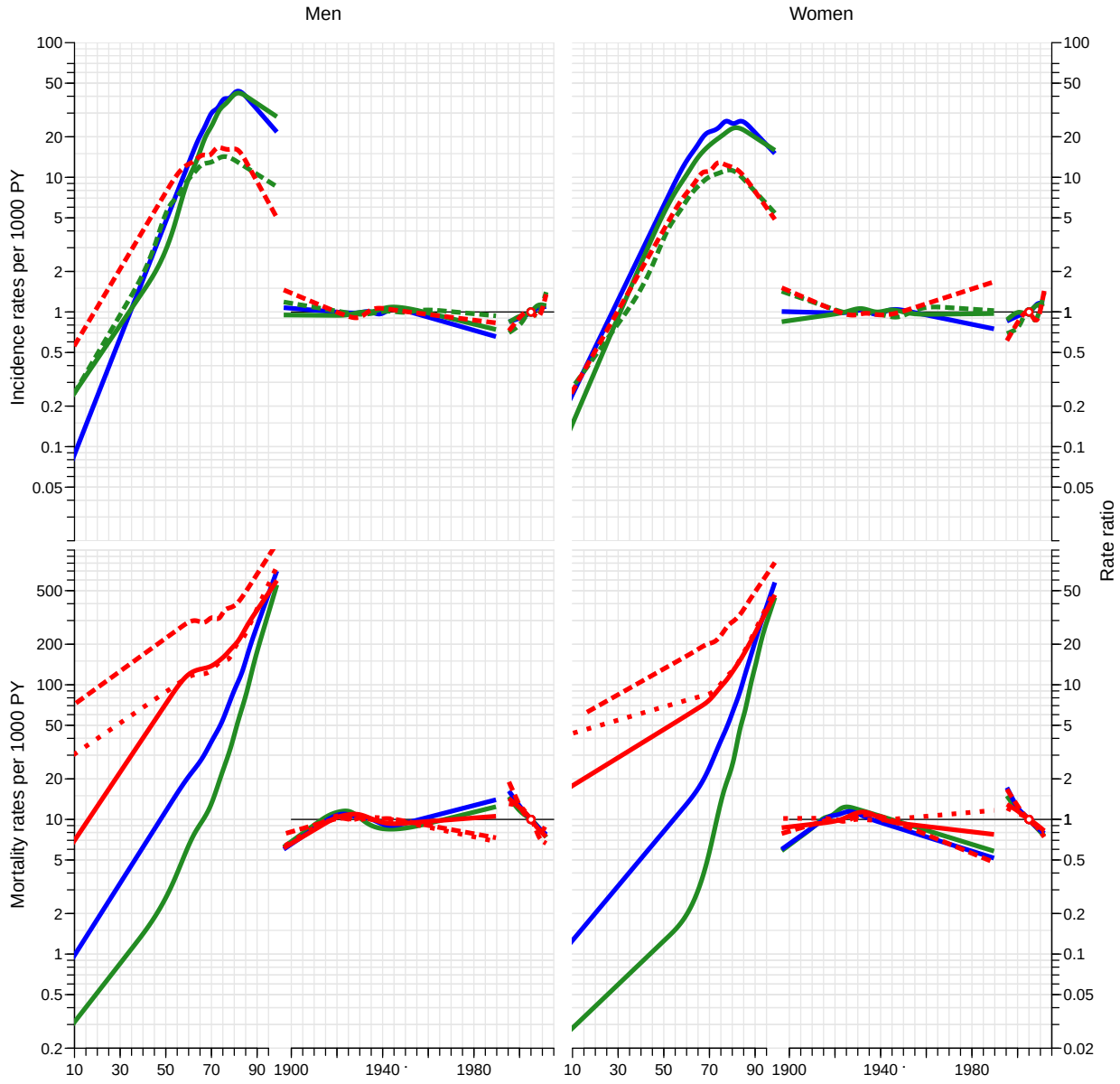

Figure 3.2: *Parameters from the fitted APC-models for rates using the period effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes incidence rates.
Bottom panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates from states with cancer; full lines: mortality rates from states "Well", "DM" and "Ca"; broken lines: mortality in cancer patients with pre-existing diabetes; dotted lines: mortality in cancer patients with post-cancer diabetes.

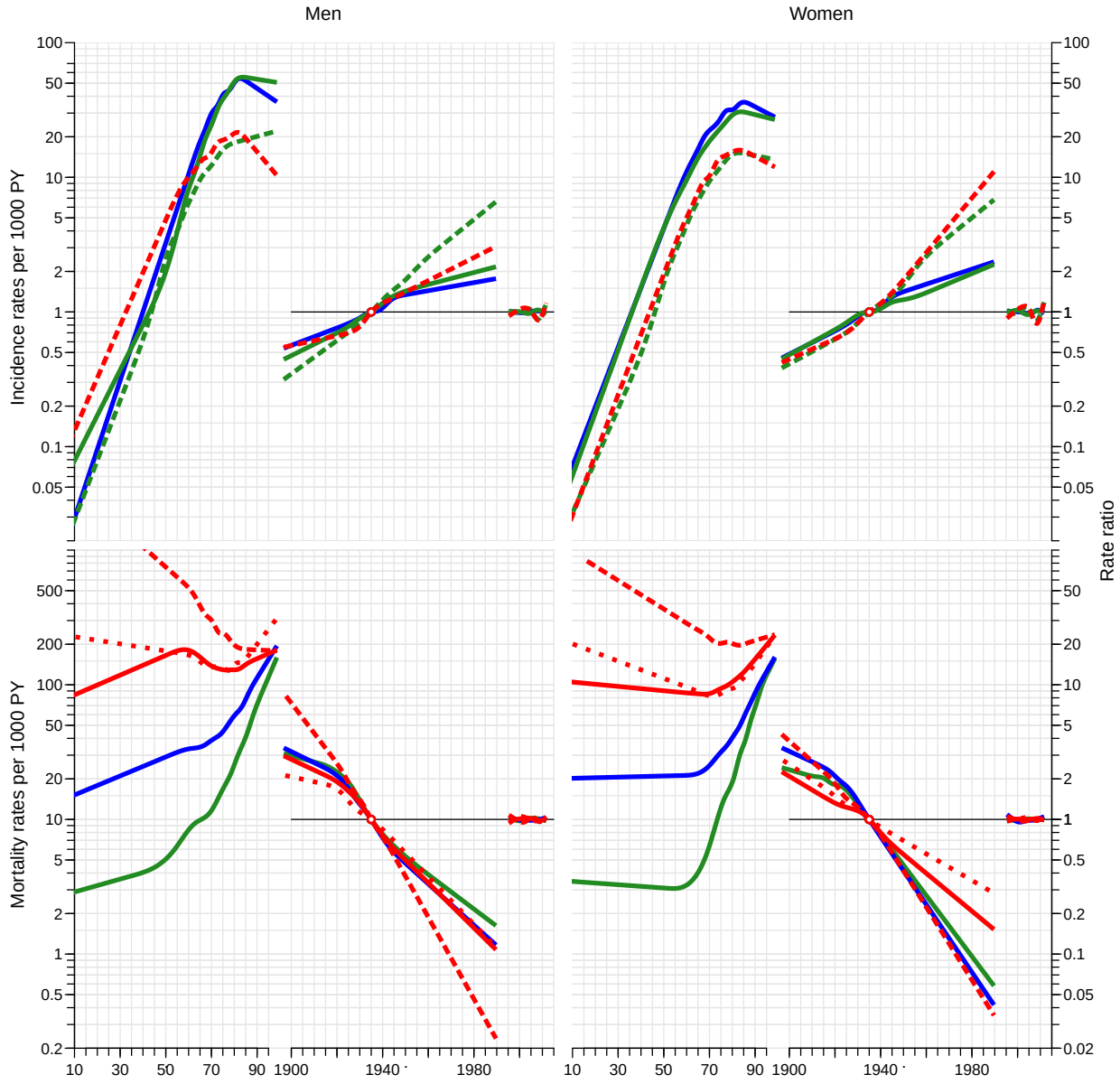

Figure 3.3: *Parameters from the fitted APC-models for rates using the cohort effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes incidence rates.
Bottom panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates from states with cancer; full lines: mortality rates from states "Well", "DM" and "Ca"; broken lines: mortality in cancer patients with pre-existing diabetes; dotted lines: mortality in cancer patients with post-cancer diabetes.

3.1.1 Rate drift

From the apc objects we can extract the annual drift:

```
> Drift <- NArray( list( type = c("W to DM","W to Ca","W to Dth",
+                                 "DM to Ca","DM to Dth",
+                                 "Ca to DM","Ca to Dth",
+                                 "DMCa to Dth", "CaDM to Dth"),
+                       sex = levels( dcd$sex ),
+                       res = c("Drift","lo","up") ) )
> str( Drift )
   logi [1:9, 1:2, 1:3] NA NA NA NA NA NA ...
   - attr(*, "dimnames")=List of 3
    ..$ type: chr [1:9] "W to DM" "W to Ca" "W to Dth" "DM to Ca" ...
    ..$ sex : chr [1:2] "M" "F"
    ..$ res : chr [1:3] "Drift" "lo" "up"

> Drift["W to DM"    ,"M",] <- M.w2dm$apc$Drift[1,]
> Drift["W to Ca"    ,"M",] <- M.w2ca$apc$Drift[1,]
> Drift["W to Dth"   ,"M",] <- M.w2dd$apc$Drift[1,]
> Drift["DM to Ca"   ,"M",] <- M.dm2ca$apc$Drift[1,]
> Drift["DM to Dth"  ,"M",] <- M.dm2dd$apc$Drift[1,]
> Drift["Ca to DM"   ,"M",] <- M.ca2dm$apc$Drift[1,]
> Drift["Ca to Dth"  ,"M",] <- M.ca2dd$apc$Drift[1,]
> Drift["DMCa to Dth","M",] <- M.dc2dd$apc$Drift[1,]
> Drift["CaDM to Dth","M",] <- M.cd2dd$apc$Drift[1,]
> Drift["W to DM"    ,"F",] <- F.w2dm$apc$Drift[1,]
> Drift["W to Ca"    ,"F",] <- F.w2ca$apc$Drift[1,]
> Drift["W to Dth"   ,"F",] <- F.w2dd$apc$Drift[1,]
> Drift["DM to Ca"   ,"F",] <- F.dm2ca$apc$Drift[1,]
> Drift["DM to Dth"  ,"F",] <- F.dm2dd$apc$Drift[1,]
> Drift["Ca to DM"   ,"F",] <- F.ca2dm$apc$Drift[1,]
> Drift["Ca to Dth"  ,"F",] <- F.ca2dd$apc$Drift[1,]
> Drift["DMCa to Dth","F",] <- F.dc2dd$apc$Drift[1,]
> Drift["CaDM to Dth","F",] <- F.cd2dd$apc$Drift[1,]
> round( ftable( (Drift[c(1,6,2,4,3,5,7:9),,]-1)*100, row.vars=1 ), 1 )
             sex     M               F
             res Drift   lo   up Drift   lo   up
type
W to DM          3.6  3.5  3.7   3.5  3.4  3.6
Ca to DM         2.5  2.1  2.9   3.5  3.1  3.8
W to Ca          2.0  1.9  2.1   1.6  1.5  1.7
DM to Ca         1.8  1.5  2.1   2.1  1.8  2.4
W to Dth        -3.8 -3.9 -3.7  -3.9 -4.0 -3.8
DM to Dth       -4.4 -4.6 -4.3  -4.5 -4.7 -4.3
Ca to Dth       -4.1 -4.2 -4.0  -2.7 -2.8 -2.6
DMCa to Dth     -6.1 -6.4 -5.8  -4.5 -4.9 -4.2
CaDM to Dth     -3.2 -3.6 -2.8  -2.6 -3.0 -2.1

> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( (Drift[c(1,6,2,4,3,5,7:9),1,]-1)*100, lwd=3, col="blue", y=9:1+0.1,
+          xlab="Annual change in rates (%)", vref=0, xtic=seq(-7,5,2), grid=-7:5 )
> linesEst( (Drift[c(1,6,2,4,3,5,7:9),2,]-1)*100, lwd=3, col="red", y=9:1-0.1 )
> text( c(5,5), 3:2/2, c("Men","Women"), col=c("blue","red"), font=2, adj=1 )
```

This brief overview shows that the incidence of DM is increasing about 4% per year, of cancer 2% per year, largely independent of preexisting diabetes/cancer. And very broadly speaking the mortality rates are decreasing by some 3–5% per year.

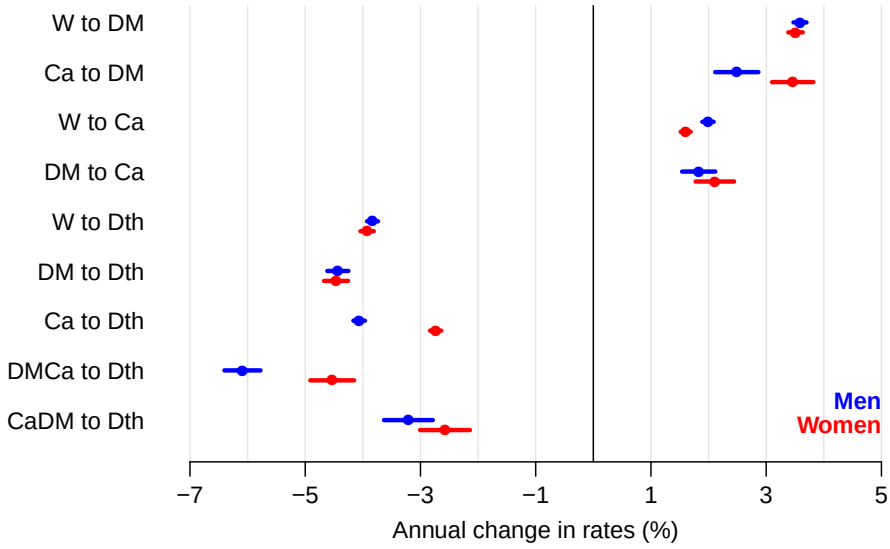

Figure 3.4: *Annual changes in the 9 incidence and mortality rates considered. Blue: men, red: women.*

Chapter 4

State probabilities

If we want to compute the fraction of persons in a given state at a given time, that is in any of the other possible states at a later time we must know the *transition matrices* between states for any pair of times. However, it suffices to know the transition matrices for a sequence if tightly spaced times since the matrices for more distantly spaced times can be constructed by multiplication of the matrices for the intervals between.

In the following we shall compute the state occupancy probabilities under different scenarios used to construct the age-specific transition rates.

4.1 Scenarios

Since we have restricted ourselves to a scenery where we have only one time scale, namely age, we can do the calculations in closed form by setting up the transition probability matrix for small age intervals (of length `int` years).

To illustrate the time-trends in risk of DM and cancer, we will use the *predicted* cross-sectional rates from the APC-models as of 1 January 1995,. . . ,2012.

For a longitudinal counterpart of this we would ideally want predicted rates from the models for the birth cohorts, say, 1920,1922,. . . ,1950. These are however only observed in ages 75–92, 73–90,. . . ,45–62, so this would require predictions many decades outside the observed age-span as we will need rates in ages from 0 to 100 (or more). Instead we use rates predicted for these cohorts for the calendar time span 1990–2017, that is only extrapolation 5 years outside the observed range. For the ages not covered in this period we use the cross-sectional rates for the dates 1990, resp 2017.

For a start we define two arrays to hold the predicted rates in these two scenarios; we shall use these for:

- plotting the predicted rates together with the corresponding age-effecst from the APC-models
- defining matrices of transition probabilities

4.2 Transition matrices

Hence we first set up the arrays to hold the transition rates at intervals of 1 month; we compute the rates at the midpoint of each age interval:

```
> int <- 1/12
> a.pt <- seq(int,102,int) - int/2
> ( states <- c( levels( dcd$state )[-6],
+                c("D-W","D-DM","D-Ca","D-DC","D-CD") ) )
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"
> pnam <-
+ cnam <- list( from = states,
+                 to = states,
+               age = a.pt,
+               per = 1995:2012,
+               sex = c("M","F") )
> names(cnam)[4] <- "coh"
> cnam[["coh"]] <- seq(1920,1950,2)
> pnam[-3]
  $from
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

  $to
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

  $per
   [1] 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

  $sex
  [1] "M" "F"
> cnam[-3]
  $from
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

  $to
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

  $coh
   [1] 1920 1922 1924 1926 1928 1930 1932 1934 1936 1938 1940 1942 1944 1946 1948 1950

  $sex
  [1] "M" "F"
> PR <- ZArray( pnam )
> CR <- ZArray( cnam )
```

Then we can fill in the age-specific rates that will later be used in the calculations of state occupancy probabilities; note that we are using the quantity `int` for `Y` in the prediction frame, that way we get the incidence rate per this length of time, or more specifically, the cumulative incidence over an interval of this length, for an interval (centered) at the age `a.pt`, assuming constant rate over the interval. Which seems reasonable for intervals of length 1 month.

Thus, we now compute the transition rates, or rather the cumulative transition rates for an interval of length 1 month, both for cross-secions at 1 january 1995,...,2012 (`PR`: Period Rates) and for the birth cohorts 1920, 1922, ..., 1950 (`CR`: Cohort Rates).

```
> system.time(
+ for( yy in dimnames(PR)[[4]] )
+ {
+ nd <- data.frame( A=a.pt, P=as.numeric(yy), Y=int )
+
+ PR["Well" ,"DM"   ,,yy,"M"] <- ci.pred( M.w2dm$model , newdata=nd )[,1]
+ PR["Well" ,"Ca"   ,,yy,"M"] <- ci.pred( M.w2ca$model , newdata=nd )[,1]
+ PR["Well" ,"D-W"  ,,yy,"M"] <- ci.pred( M.w2dd$model , newdata=nd )[,1]
+ PR["DM"   ,"DM-Ca",,yy,"M"] <- ci.pred( M.dm2ca$model, newdata=nd )[,1]
+ PR["DM"   ,"D-DM" ,,yy,"M"] <- ci.pred( M.dm2dd$model, newdata=nd )[,1]
```

```
+ PR["Ca"   ,"Ca-DM",,yy,"M"] <- ci.pred( M.ca2dm$model, newdata=nd )[,1]
+ PR["Ca"   ,"D-Ca" ,,yy,"M"] <- ci.pred( M.ca2dd$model, newdata=nd )[,1]
+ PR["DM-Ca","D-DC" ,,yy,"M"] <- ci.pred( M.dc2dd$model, newdata=nd )[,1]
+ PR["Ca-DM","D-CD" ,,yy,"M"] <- ci.pred( M.cd2dd$model, newdata=nd )[,1]
+
+ PR["Well" ,"DM"   ,,yy,"F"] <- ci.pred( F.w2dm$model , newdata=nd )[,1]
+ PR["Well" ,"Ca"   ,,yy,"F"] <- ci.pred( F.w2ca$model , newdata=nd )[,1]
+ PR["Well" ,"D-W"  ,,yy,"F"] <- ci.pred( F.w2dd$model , newdata=nd )[,1]
+ PR["DM"   ,"DM-Ca",,yy,"F"] <- ci.pred( F.dm2ca$model, newdata=nd )[,1]
+ PR["DM"   ,"D-DM" ,,yy,"F"] <- ci.pred( F.dm2dd$model, newdata=nd )[,1]
+ PR["Ca"   ,"Ca-DM",,yy,"F"] <- ci.pred( F.ca2dm$model, newdata=nd )[,1]
+ PR["Ca"   ,"D-Ca" ,,yy,"F"] <- ci.pred( F.ca2dd$model, newdata=nd )[,1]
+ PR["DM-Ca","D-DC" ,,yy,"F"] <- ci.pred( F.dc2dd$model, newdata=nd )[,1]
+ PR["Ca-DM","D-CD" ,,yy,"F"] <- ci.pred( F.cd2dd$model, newdata=nd )[,1]
+ }
+ )
   user  system elapsed
  2.957   0.052   3.009
> for( bb in dimnames(CR)[[4]] )
+ {
+ nd <- data.frame( A=a.pt, P=as.numeric(bb)+a.pt, Y=int )
+
+ CR["Well" ,"DM"   ,,bb,"M"] <- ci.pred( M.w2dm$model , newdata=nd )[,1]
+ CR["Well" ,"Ca"   ,,bb,"M"] <- ci.pred( M.w2ca$model , newdata=nd )[,1]
+ CR["Well" ,"D-W"  ,,bb,"M"] <- ci.pred( M.w2dd$model , newdata=nd )[,1]
+ CR["DM"   ,"DM-Ca",,bb,"M"] <- ci.pred( M.dm2ca$model, newdata=nd )[,1]
+ CR["DM"   ,"D-DM" ,,bb,"M"] <- ci.pred( M.dm2dd$model, newdata=nd )[,1]
+ CR["Ca"   ,"Ca-DM",,bb,"M"] <- ci.pred( M.ca2dm$model, newdata=nd )[,1]
+ CR["Ca"   ,"D-Ca" ,,bb,"M"] <- ci.pred( M.ca2dd$model, newdata=nd )[,1]
+ CR["DM-Ca","D-DC" ,,bb,"M"] <- ci.pred( M.dc2dd$model, newdata=nd )[,1]
+ CR["Ca-DM","D-CD" ,,bb,"M"] <- ci.pred( M.cd2dd$model, newdata=nd )[,1]
+
+ CR["Well" ,"DM"   ,,bb,"F"] <- ci.pred( F.w2dm$model , newdata=nd )[,1]
+ CR["Well" ,"Ca"   ,,bb,"F"] <- ci.pred( F.w2ca$model , newdata=nd )[,1]
+ CR["Well" ,"D-W"  ,,bb,"F"] <- ci.pred( F.w2dd$model , newdata=nd )[,1]
+ CR["DM"   ,"DM-Ca",,bb,"F"] <- ci.pred( F.dm2ca$model, newdata=nd )[,1]
+ CR["DM"   ,"D-DM" ,,bb,"F"] <- ci.pred( F.dm2dd$model, newdata=nd )[,1]
+ CR["Ca"   ,"Ca-DM",,bb,"F"] <- ci.pred( F.ca2dm$model, newdata=nd )[,1]
+ CR["Ca"   ,"D-Ca" ,,bb,"F"] <- ci.pred( F.ca2dd$model, newdata=nd )[,1]
+ CR["DM-Ca","D-DC" ,,bb,"F"] <- ci.pred( F.dc2dd$model, newdata=nd )[,1]
+ CR["Ca-DM","D-CD" ,,bb,"F"] <- ci.pred( F.cd2dd$model, newdata=nd )[,1]
+ }


> save( PR, CR, file="rates-r.Rda" )
```

4.2.1 Estimated rates

We can now plot the estimated transition rates, that is the incidence or mortality rates for the successive periods/cohorts. To this end we need a couple of functions to simplify the task; first a function that returns coordinates a specified proportion from the llh corner:

```
> cnr <-
+ function( xf, yf )
+ {
+ # A function that gives the coordinates of the
+ # point (xf,yf) from ll corner in the current plot.
+ # if xf or yf are > 1 they are considered percentages
+ #
+ cn <- par()$usr
+ xf <- ifelse( xf>1, xf/100, xf )
+ yf <- ifelse( yf>1, yf/100, yf )
```

```
+ xx <- ( 1 - xf ) * cn[1] + xf * cn[2]
+ yy <- ( 1 - yf ) * cn[3] + yf * cn[4]
+ if ( par()$xlog ) xx <- 10^xx
+ if ( par()$ylog ) yy <- 10^yy
+ list( x=xx, y=yy )
+ }
```

Then a function to plot the rates from state `f` to state `t`:

```
> pl1 <-
+ function( M, f, t, sx, yf, parm )
+ {
+ plot( NA, xlim=c(10,100), ylim=yf*c(1,10000), log="y",
+           xlab="", ylab="", xaxt="n", yaxt="n" )
+ abline( v=1:10*10, h=outer(1:9,10^(-3:5),"*"), col=gray(0.9) )
+ text( cnr(0.05,0.95), paste(f,"to",t), adj=c(0,1) )
+ matlines( a.pt, M[f,t,,,sx]*10^4,
+           type="l", lty=1, lwd=1, col=if(sx=="M") "blue" else "red" )
+ lines( parm[,1], parm[,2], lwd=2 )
+ }
```

And finally a function to plot the 4 incidence rates and 5 mortality rates

```
> pl9 <-
+ function( M, sx, mod )
+ {
+ par( mfrow=c(2,5), mar=rep(0,4), mgp=c(3,1,0)/1.6, oma=c(4,4,1,1),las=1,bty="n")
+ pl1(M,"Well", "Ca",sx,0.02,get(paste(sx,".w2ca",sep=""))[[mod]][["Age"]])
+ axis( side=2, at=outer(c(1,2,5),10^(-2:2),"*")[2:14],
+     labels=paste(outer(c(1,2,5),10^(-2:2),"*")[2:14]) )
+ pl1(M,"DM","DM-Ca",sx,0.02,get(paste(sx,".dm2ca",sep=""))[[mod]][["Age"]])
+ pl1(M,"Well", "DM",sx,0.02,get(paste(sx,".w2dm",sep=""))[[mod]][["Age"]])
+ pl1(M,"Ca","Ca-DM",sx,0.02,get(paste(sx,".ca2dm",sep=""))[[mod]][["Age"]])
+ par( mfg=c(2,1) )
+ pl1(M,"Well" , "D-W",sx,0.2,get(paste(sx,".w2dd",sep=""))[[mod]][["Age"]]) ; axis( side=1 )
+ axis( side=2, at=outer(c(1,2,5),10^(-1:3),"*")[2:14],
+     labels=paste(outer(c(1,2,5),10^(-1:3),"*")[2:14]) )
+ pl1(M,"DM"   ,"D-DM",sx,0.2,get(paste(sx,".dm2dd",sep=""))[[mod]][["Age"]]) ; axis( side=1 )
+ pl1(M,"Ca"   ,"D-Ca",sx,0.2,get(paste(sx,".ca2dd",sep=""))[[mod]][["Age"]]) ; axis( side=1 )
+ pl1(M,"DM-Ca","D-DC",sx,0.2,get(paste(sx,".dc2dd",sep=""))[[mod]][["Age"]]) ; axis( side=1 )
+ pl1(M,"Ca-DM","D-CD",sx,0.2,get(paste(sx,".cd2dd",sep=""))[[mod]][["Age"]]) ; axis( side=1 )
+ mtext( "Age (years)", side=1, line=2.5, cex=0.8, outer=TRUE )
+ mtext( "Incidence rates (per 1000 PY)", side=2, line=2.5, cex=0.8, outer=TRUE, at=0.75, las=0 )
+ mtext( "Mortality rates (per 1000 PY)", side=2, line=2.5, cex=0.8, outer=TRUE, at=0.25, las=0 )
+ }
```

```
> pl9( PR, "M", "apc" )
```

```
> pl9( PR, "F", "apc" )
```

```
> pl9( CR, "M", "acp" )
```

```
> pl9( CR, "F", "acp" )
```

Inspection of the predicted incidence and mortality rates in Figures 4.2 and 4.4 clearly shows that the construction of "cohort" rates by using the estimated cross-sectional rates at 1990 and 2017 in conjunction with the cohort rates for the years between is not an attractive feature; the mortality rates are hardly credible as shown, which we will bear in mind when reporting results from these.

Thus, we shall in the first place use the period rates for calculation of state occupancy probabilities.

4.3 Transition probabilities

Now we have the transition rates corresponding to 1 month in the array `PR`, but we need to fill in the diagonals to get a proper transition matrix for every combination of age, period and sex. To this end we need a function that does this properly; note that the entries in `PR` are cumulative rates corresponding to a period of length 1 month (well, formally `int`). Thus if cumulative transition rates *from* a given state are, say, Λ_1, Λ_2 Λ_3, then the diagonal element in the row must be $\exp\big(-(\Lambda_1+\Lambda_2+\Lambda_3)\big)$ and the off-diagonal elements in the row should be $\Big(1-\exp\big(-(\Lambda_1+\Lambda_2+\Lambda_3)\big)\Big)\times\Lambda_i/(\Lambda_1+\Lambda_2+\Lambda_3), i=1,2,3$, that is the cumulative rates[1] multiplied by $\Big(1-\exp\big(-(\Lambda_1+\Lambda_2+\Lambda_3)\big)\Big)/(\Lambda_1+\Lambda_2+\Lambda_3)$. We wrap this calculation in a small function:

```
> ci2pr <-
+ function( M )
+ {
+ sm  <- apply( M, 1, sum )
+ res <- sweep( M, 1, (1-exp(-sm))/sm, "*" )
+ # Rows corresponding to absorbing states have sum 0 so the above
+ # returns NA, which must then be converted to 0 before the diagonal is
+ # filled with the survival probabilities
+ res[is.na(res)] <- 0
+ diag( res ) <- exp( -sm )
+ res
+ }
```

First we check that the function does the right thing:

```
> print.table( round(        PR[,,800,1,1]  *10^4 ), zero.print="." )
```

```
        to
from     Well  DM DM-Ca  Ca Ca-DM D-W D-DM D-Ca D-DC D-CD
  Well      .   7     .  13     .  14    .    .    .    .
  DM        .   .    16   .     .   .   43    .    .    .
  DM-Ca     .   .     .   .     .   .    .    .  477    .
  Ca        .   .     .   .     8   .    .  175    .    .
  Ca-DM     .   .     .   .     .   .    .    .    .  130
  D-W       .   .     .   .     .   .    .    .    .    .
  D-DM      .   .     .   .     .   .    .    .    .    .
  D-Ca      .   .     .   .     .   .    .    .    .    .
  D-DC      .   .     .   .     .   .    .    .    .    .
  D-CD      .   .     .   .     .   .    .    .    .    .
```

```
> print.table( round( addmargins(
+                     ci2pr( PR[,,800,1,1] )*10^4, margin=2 ) ),
+                                                zero.print="." )
```

```
        to
from     Well    DM DM-Ca    Ca Ca-DM   D-W  D-DM  D-Ca  D-DC  D-CD   Sum
  Well   9967     7     .    13     .    14     .     .     .     . 10000
  DM        .  9941    16     .     .     .    43     .     .     . 10000
  DM-Ca     .     .  9534     .     .     .     .     .   466     . 10000
  Ca        .     .     .  9819     8     .     .   173     .     . 10000
  Ca-DM     .     .     .     .  9871     .     .     .     .   129 10000
  D-W       .     .     .     .     . 10000     .     .     .     . 10000
  D-DM      .     .     .     .     .     . 10000     .     .     . 10000
  D-Ca      .     .     .     .     .     .     . 10000     .     . 10000
  D-DC      .     .     .     .     .     .     .     . 10000     . 10000
  D-CD      .     .     .     .     .     .     .     .     . 10000 10000
```

[1] Formally we should use the instantaneous rates in the fraction, but since our intervals are small this difference is immaterial

We can then convert the matrices of cumulative transition intensities to matrices of transition probabilities:

```
> PRp <- apply( PR, 3:5, ci2pr )
```

Note that apply does not recognize the dim attribute of what the FUN argument returns, so we fix it and check:

```
> dim( PRp )
```

```
       age  per  sex
   100 1224   18    2
```

```
> dim( PRp ) <- c(10,10,dim(PRp)[-1])
> dimnames( PRp ) <- dimnames( PR )
> print.table( round( PRp[,,800,1,1]*10^4 ), zero.print="." )
```

```
        to
from      Well    DM DM-Ca    Ca Ca-DM   D-W  D-DM  D-Ca  D-DC  D-CD
  Well    9967     7     .    13     .    14     .     .     .     .
  DM         .  9941    16     .     .     .    43     .     .     .
  DM-Ca      .     .  9534     .     .     .     .     .   466     .
  Ca         .     .     .  9819     8     .     .   173     .     .
  Ca-DM      .     .     .     .  9871     .     .     .     .   129
  D-W        .     .     .     .     . 10000     .     .     .     .
  D-DM       .     .     .     .     .     . 10000     .     .     .
  D-Ca       .     .     .     .     .     .     . 10000     .     .
  D-DC       .     .     .     .     .     .     .     . 10000     .
  D-CD       .     .     .     .     .     .     .     .     . 10000
```

```
> names( dimnames( PRp ) )
```

```
[1] "from" "to"   "age"  "per"  "sex"
```

So now in `PRp` we have the matrices of transition probabilities based on the cross-sectional rates for ages from 0 to 102 years, at 1995,. . . ,2012, separately for the two sexes.

4.4 State occupancy and lifetime risk

The just printed matrix is the transition matrix (multiplied by 10,000) from age 799 to 800 months (approx 68 years), so in order to get the state distribution at age 800 months, we just multiply the state distribution at age 799 months (as a row vector) with the transition matrix. This must of course be looped over ages from 0 and upward, as well as over all the other dimensions of PR.

We start by setting up the state vector, which is classified as the transition matrix, bar the first dimension:

```
> PV <- PR[1,,,,]*0
> names( dimnames(PV) )[1] <- "state"
> for( sc in dimnames(PRp)[["per"]] )
+ for( sx in dimnames(PRp)[["sex"]] )
+    {
+    # Initialize to all well at age 0:
+    PV[,1,sc,sx] <- c(1,rep(0,9))
+    # Compute distribution at endpoint of each age-interval
+    for( ag in 1:dim(PRp)[3] ) PV[,ag,sc,sx] <- PV[ ,max(ag-1,1),sc,sx] %*%
+                                                PRp[,    ag     ,sc,sx]
+    }
> str( PV )
```

```
 num [1:10, 1:1224, 1:18, 1:2] 1.00 6.85e-06 0.00 6.93e-06 0.00 ...
 - attr(*, "dimnames")=List of 4
  ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
  ..$ age  : chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex  : chr [1:2] "M" "F"
```

The array `PV` now contains the probability of being in a given state as a function of age. Thus the sum over the first dimension should be 1 for any combination of the remaining 3 classifiers:

```
> summary( apply( PV, 2:4, sum ) )
     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
        1       1       1       1       1       1
```

4.4.1 Timetrend in lifetime risks

First we compute the the lifetime cumulative probability of DM, Cancer and both as a function of calendar time. The entry correponding to the latest age will give the life-time risk of each of the conditions, so it is simple to compute the lifetime risk of DM, Ca and both:

```
> nA <- dim(PV)[2]
> pp <- as.numeric( dimnames(PV)[["per"]] )
> LrP <- PV[c(1,2,4,3),nA,,]*0
> dimnames(LrP)[[1]][4] <- "DM+Ca"
> dimnames(LrP)[[1]]
  [1] "Well"  "DM"    "Ca"    "DM+Ca"
> dimnames(PV)[[1]]
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"
> LrP["Well" ,,] <- PV["D-W" ,nA,,]
> LrP["DM"   ,,] <- PV["D-DM",nA,,] + PV["D-DC",nA,,] + PV["D-CD",nA,,]
> LrP["Ca"   ,,] <- PV["D-Ca",nA,,] + PV["D-DC",nA,,] + PV["D-CD",nA,,]
> LrP["DM+Ca",,] <-                   PV["D-DC",nA,,] + PV["D-CD",nA,,]
> ftable( round( LrP*100, 1 ), col.vars=c(3,1) )
      sex      M                    F
      state Well   DM   Ca DM+Ca Well   DM   Ca DM+Ca
per
1995       52.1 19.3 33.6   5.1 51.0 17.5 35.8   4.5
1996       50.6 20.0 34.9   5.5 49.6 17.8 37.2   4.8
1997       49.0 20.7 36.2   6.0 48.2 18.2 38.5   5.2
1998       47.4 21.5 37.5   6.4 46.9 18.6 39.9   5.7
1999       45.9 22.3 38.8   7.0 45.6 19.1 41.0   6.1
2000       44.4 23.1 39.9   7.5 44.6 19.8 41.9   6.6
2001       43.1 24.2 40.8   8.1 43.7 20.8 42.3   7.2
2002       41.7 25.6 41.5   8.9 42.8 22.3 42.4   7.8
2003       40.3 27.0 42.3   9.6 41.7 23.9 42.4   8.5
2004       39.0 28.1 43.2  10.4 40.7 25.2 42.8   9.0
2005       37.8 28.5 44.7  11.0 39.9 25.2 43.7   9.3
2006       36.6 28.1 46.5  11.3 39.5 24.0 45.2   9.2
2007       35.4 27.6 48.4  11.6 38.9 22.8 46.9   9.2
2008       34.1 27.9 50.0  12.1 37.9 22.7 48.4   9.5
2009       32.5 29.5 50.9  13.0 36.2 24.2 49.4  10.4
2010       30.6 32.6 51.4  14.7 34.1 27.3 49.9  11.9
2011       28.4 36.5 51.7  16.7 31.7 31.3 50.3  14.0
2012       26.2 40.8 51.9  19.0 29.1 35.9 50.7  16.4
```

We can now plot the secular trends in the life-time risk of the two diseases:

```
> clr <- c("limegreen","#6666FF","#FF3333","#BB77BB","black","gray","white")
> par( mfrow=c(1,2), mar=c(2,1,1,1.5), oma=c(2,2,0,2), mgp=c(3,1,0)/1.6, las=1 )
> for( sx in dimnames(LrP)[["sex"]] )
+    {
+ plot( NA, xlim=range(pp), ylim=c(0,60),
+           xlab="", ylab="", xaxs="i", yaxs="i", yaxt="n" )
+ axis( side=4, lwd=0, lwd.ticks=1, labels=(sx=="F") )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ axis( side=2, lwd=0, lwd.ticks=1 )
+ axis( side=2, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=2, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.4 )
+ axis( side=2, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ matlines( pp, t(LrP[2:4,,sx])*100,
+           lty=1, lwd=6, col=clr[2:4] )
+ text( 1996, 55, sx, font=2, cex=2, adj=0 )
+ text( rep(2011,3), LrP[2:4,"2011",sx]*100+1, dimnames(LrP)[[1]][2:4],
+       col=clr[2:4], font=2, cex=1.5, adj=c(1,0) )
+    }
> mtext( "Date of rate evaluation", side=1, line=1, outer=T )
```

For the corresponding cumulative plots we also define an array of cumulative lifetime probabilities over the states in a sensible order order. For convenience of programming we add a 0 layer:

```
> LcP <- apply( PV[c("D-DM","D-DC","D-CD","D-Ca","D-W"),nA,,], 2:3, cumsum )
> LcP <- LcP[c(1,1:5),,]
> LcP[1,,] <- 0
> LcP[6,,] <- 1 # Compensate the rounding errors
> dimnames( LcP )[[1]][1] <- ""
> ftable( round( LcP*100, 1 ), col.vars=c(3,1) )
```

sex	M						F					
		D-DM	D-DC	D-CD	D-Ca	D-W		D-DM	D-DC	D-CD	D-Ca	D-W
per												
1995	0.0	14.2	18.1	19.3	47.8	100.0	0.0	13.0	16.0	17.5	48.8	100.0
1996	0.0	14.5	18.6	20.0	49.4	100.0	0.0	13.0	16.2	17.8	50.1	100.0
1997	0.0	14.8	19.2	20.7	51.0	100.0	0.0	13.0	16.4	18.2	51.5	100.0
1998	0.0	15.0	19.8	21.5	52.5	100.0	0.0	12.9	16.6	18.6	52.8	100.0
1999	0.0	15.3	20.4	22.3	54.1	100.0	0.0	13.0	16.8	19.1	54.0	100.0
2000	0.0	15.6	21.1	23.1	55.5	100.0	0.0	13.1	17.3	19.8	55.0	100.0
2001	0.0	16.1	22.0	24.2	56.9	100.0	0.0	13.6	18.1	20.8	55.9	100.0
2002	0.0	16.7	23.2	25.6	58.2	100.0	0.0	14.5	19.5	22.3	56.8	100.0
2003	0.0	17.3	24.5	27.0	59.6	100.0	0.0	15.5	21.0	23.9	57.9	100.0
2004	0.0	17.7	25.5	28.1	60.9	100.0	0.0	16.1	22.1	25.2	58.9	100.0
2005	0.0	17.5	25.7	28.5	62.1	100.0	0.0	15.9	22.0	25.2	59.6	100.0
2006	0.0	16.8	25.1	28.1	63.3	100.0	0.0	14.8	20.8	24.0	60.0	100.0
2007	0.0	16.1	24.5	27.6	64.5	100.0	0.0	13.6	19.6	22.8	60.5	100.0
2008	0.0	15.8	24.6	27.9	65.8	100.0	0.0	13.2	19.4	22.7	61.6	100.0
2009	0.0	16.5	25.9	29.5	67.4	100.0	0.0	13.8	20.7	24.2	63.2	100.0
2010	0.0	17.9	28.6	32.6	69.3	100.0	0.0	15.3	23.3	27.3	65.3	100.0
2011	0.0	19.8	31.9	36.5	71.4	100.0	0.0	17.3	26.5	31.3	67.6	100.0
2012	0.0	21.8	35.6	40.8	73.6	100.0	0.0	19.5	30.1	35.9	70.2	100.0

In order to plot the stacked cumulative probabilities we use the `polygon` trick, and in order to visualize the joint occurrence of diabetes and cancer we define blue for DM, red for cancer and purple for both:

```
> pp <- as.numeric( dimnames(LcP)[["per"]] )
> par( mfrow=c(1,2), mar=c(2,1,1,1.5), oma=c(2,2,0,2), mgp=c(3,1,0)/1.6, las=1 )
> for( sx in dimnames(LcP)[["sex"]] )
+    {
+ plot( NA, xlim=range(pp), ylim=c(0,100),
+           xlab="", ylab="", xaxs="i", yaxs="i" )
```

```
+ axis( side=4, lwd=0, lwd.ticks=1, labels=(sx=="F") )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.3 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ polygon( c(pp,rev(pp)), c(LcP[1,,sx],rev(LcP[2,,sx]))*100,
+          col = clr[2], border="transparent")
+ polygon( c(pp,rev(pp)), c(LcP[2,,sx],rev(LcP[4,,sx]))*100,
+          col = clr[4], border="transparent")
+ polygon( c(pp,rev(pp)), c(LcP[4,,sx],rev(LcP[5,,sx]))*100,
+          col = clr[3], border="transparent")
+ polygon( c(pp,rev(pp)), c(LcP[5,,sx],rev(LcP[6,,sx]))*100,
+          col = clr[6], border="transparent")
+ lines( pp, LcP[3,,sx]*100, col=clr[7] )
+ text( 1996, 95, sx, font=2, col=clr[7], cex=2 )
+ text( rep(2011.5,4), 60*LcP[(1:5)[-3],dim(LcP)[2],sx] +
+                      40*LcP[(2:6)[-3],dim(LcP)[2],sx],
+       c("DM","DM+Ca","Ca","Neither"),
+       font=2, adj=1, cex=1.5, col="white" )
+    }
> mtext( "Date of rate evaluation", side=1, line=1, outer=T )
> mtext( "Lifetime risk", side=2, line=1, outer=T, las=0 )
```

From the figures 4.10 and 4.6 we see that there is a dramaic increase in the life-time risk of both diabetes and cancer, but also that the the main driver is the increasing risk of both diseases, the lifetine risk og having a cancer without concomitant diabetes and vice-versa has not changed dramatically over the 18-year observation period.

4.5 States by age

We have the distribution of the persons in the different states under various scenarios, and also want to plot the resulting distribution of the states as function of age; for each of the 4 combinations of scenario and sex we can plot the probabilities of being in each of the 10 states. However we must put them in a sensible order to make a meaningful plot, with the transient states first, the states with DM and cancer between the diabetes state and the cancer state:

```
> perm <- c(2,3,5,4,1,6,8,10,9,7)
> round( t(PV[perm,600+1:5,1,1])*100, 1 )
                   state
  age                DM DM-Ca Ca-DM  Ca Well D-W D-Ca D-CD D-DC D-DM
    50.0416666666667 3.3     0     0 1.6 87.0 5.9  1.5    0  0.1  0.5
    50.125           3.3     0     0 1.6 87.0 5.9  1.6    0  0.1  0.5
    50.2083333333333 3.3     0     0 1.6 86.9 5.9  1.6    0  0.1  0.5
    50.2916666666667 3.4     0     0 1.6 86.8 5.9  1.6    0  0.1  0.5
    50.375           3.4     0     0 1.7 86.8 6.0  1.6    0  0.1  0.5

> cPV <- apply( PV[perm,,,], 2:4, cumsum )
> str( PV )

   num [1:10, 1:1224, 1:18, 1:2] 1.00 6.85e-06 0.00 6.93e-06 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
    ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex  : chr [1:2] "M" "F"

> cPV <- cPV[c(1,1:10),,,]
> cPV[ 1,,,] <- 0
> cPV[11,,,] <- 1
> dimnames( cPV )[[1]][1] <- ""
> str( cPV )
```

```
   num [1:11, 1:1224, 1:18, 1:2] 0.00 6.85e-06 6.85e-06 6.85e-06 1.38e-05 ...
   - attr(*, "dimnames")=List of 4
    ..$    : chr [1:11] "" "DM" "DM-Ca" "Ca-DM" ...
    ..$ age: chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
    ..$ per: chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex: chr [1:2] "M" "F"
> round( t(cPV[,600+1:5,1,1])*100, 1 )
  age                  DM DM-Ca Ca-DM  Ca Well  D-W D-Ca D-CD D-DC D-DM
    50.0416666666667 0 3.3   3.3   3.4 5.0 92.0 97.9 99.5 99.5 99.5  100
    50.125           0 3.3   3.4   3.4 5.0 92.0 97.9 99.5 99.5 99.5  100
    50.2083333333333 0 3.3   3.4   3.4 5.1 92.0 97.9 99.4 99.5 99.5  100
    50.2916666666667 0 3.4   3.4   3.4 5.1 91.9 97.9 99.4 99.5 99.5  100
    50.375           0 3.4   3.4   3.5 5.1 91.9 97.8 99.4 99.5 99.5  100
```

In order to plot the state occupancy probabilities by age we again use the `polygon` trick, and the same colors as before.

```
> sx <- 1
> sc <- 1
> aa <- as.numeric( dimnames(CR)[["age"]] )
> nul <- aa * 0
> crpl <- function( sc )
+    {
+ par( mfrow=c(1,2), mar=c(2,2,1,3), oma=c(2,2,0,0), mgp=c(3,1,0)/1.6, las=1 )
+ for( sx in dimnames(cPV)[["sex"]] )
+    {
+ plot( NA, xlim=c(50,100), ylim=c(0,100),
+           xlab="", ylab="", xaxs="i", yaxs="i" )
+ axis( side=4, lwd=0, lwd.ticks=1 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.3 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ polygon( c(aa,rev(aa)), c(cPV[1,,sc,sx],
+                           rev(cPV[2,,sc,sx]))*100,
+          col = clr[2], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[2,,sc,sx],
+                           rev(cPV[4,,sc,sx]))*100,
+          col = clr[4], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[4,,sc,sx],
+                           rev(cPV[5,,sc,sx]))*100,
+          col = clr[3], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[5,,sc,sx],
+                           rev(cPV[6,,sc,sx]))*100,
+          col = clr[1], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[6,,sc,sx],
+                           rev(cPV[7,,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[7,,sc,sx],
+                           rev(cPV[8,,sc,sx]))*100,
+          col = clr[3], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[8,,sc,sx],
+                           rev(cPV[10,,sc,sx]))*100,
+          col = clr[4], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[10,,sc,sx],
+                           rev(cPV[11,,sc,sx]))*100,
+          col = clr[2], border="transparent")
+ matlines( aa, 100*t(cPV[c(3,6,9),,sc,sx]),
+           lty=1, col=c("white","black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+ text( 55, 70, sx, font=2, cex=1.5, col="white" )
+ mtext( "Age (years)", side=1, outer=TRUE )
+    }
+ mtext( "Fraction of persons (%)", side=2, outer=TRUE, las=0 )
+ text( 98, 98, sc, adj=c(1,1), col="white", font=2, cex=1.5 )
+    }
> crpl( sc="2010" )
```

```
> pdf( "crisk-film.pdf", width=11, height=8 )
> for( sc in dimnames(cPV)[[3]] ) crpl( sc )
> dev.off()
  pdf
    2
```

4.5.1 Cumulative risk by age

We also want to see the cumulative risks of getting DM, cancer and both before a given age, so we make graphs of this for men and women:

```
> dimnames(PV)[[1]]
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"
> dmlev <- c(2,3,5,7,9,10)
> calev <- c(3:5,8:10)
> dclev <- intersect(dmlev,calev)
> dimnames(PV)[[1]][dmlev]
  [1] "DM"    "DM-Ca" "Ca-DM" "D-DM"  "D-DC"  "D-CD"
> dimnames(PV)[[1]][calev]
  [1] "DM-Ca" "Ca"    "Ca-DM" "D-Ca"  "D-DC"  "D-CD"
> dimnames(PV)[[1]][dclev]
  [1] "DM-Ca" "Ca-DM" "D-DC"  "D-CD"
> par( mfrow=c(1,2), mar=c(2,2,1,3), oma=c(2,2,0,0), mgp=c(3,1,0)/1.6, las=1 )
> for( sc in "2010" ) # dimnames(cPV)[[3]][1] )
+ for( sx in dimnames(cPV)[[4]] )
+    {
+ plot( NA, xlim=c(50,100), ylim=c(0,60),
+           xlab="", ylab="", xaxs="i", yaxs="i" )
+ axis( side=4, lwd=0, lwd.ticks=1 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.3 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ text( 55, 55, sx, cex=1.5, font=2 )
+ matlines( aa, zz <- cbind( apply( PV[dmlev,,sc,sx]*100, 2, sum ),
+                            apply( PV[calev,,sc,sx]*100, 2, sum ),
+                            apply( PV[dclev,,sc,sx]*100, 2, sum ) ),
+           col=clr[2:4], lty=1, lwd=5 )
+ text( rep(99,3), zz[99/int,]+2, c("DM","Ca","DM+Ca"),
+       col=clr[2:4], adj=c(1,0),cex=1.5, font=2 )
+ mtext( "Age (years)", side=1, outer=TRUE )
+ mtext( "Fraction of persons (%)", side=2, outer=TRUE, las=0 )
+     }
```

4.5.2 Conditional state probabilities

We can of course also make the same exercise *conditional* being alive at age 50, 60 etc, but as is seen from figure **??** the ultimate distribution of the fraction of persons that get the two diseases is not dramatically changed by condtioning on survival to ages 50, 60 or 70.

We set up the machinery in parallel for the three conditioning ages

```
> DM50 <- DM60 <- DM70 <-
+ PV50 <- PV60 <- PV70 <- PV*0
> dimnames( PV )[[2]][50/int]
  [1] "49.9583333333333"
```

```
> dimnames( PV )[[1]]
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"
> for( sc in dimnames(PR)[["per"]] )
+ for( sx in dimnames(PR)[["sex"]] )
+    {
+    # Initialize to all being well at age 50, 60, 70
+    PV50[,50/int,sc,sx] <-
+    PV60[,60/int,sc,sx] <-
+    PV70[,70/int,sc,sx] <- c(1,rep(0,9))
+    # Initialize to all being DM at age 50, 60, 70
+    DM50[,50/int,sc,sx] <-
+    DM60[,60/int,sc,sx] <-
+    DM70[,70/int,sc,sx] <- c(0,1,rep(0,8))
+    for( ag in (50/int+1):dim(PV)[2] )
+       {
+                    PV50[,ag,sc,sx] <- PV50[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+    if( ag>60/int ) PV60[,ag,sc,sx] <- PV60[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+    if( ag>70/int ) PV70[,ag,sc,sx] <- PV70[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+                    DM50[,ag,sc,sx] <- DM50[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+    if( ag>60/int ) DM60[,ag,sc,sx] <- DM60[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+    if( ag>70/int ) DM70[,ag,sc,sx] <- DM70[,ag-1,sc,sx] %*% PRp[,,ag,sc,sx]
+       }
+    }
```

4.5.3 Plotting the lifecourse

We can now plot the comparison between the life-long outloook of a person with and without diabetes, conditioning on status at ages 50, 60 and 70. To this end we define a function that will plot the stacked state occupancies for a given array, for a given year (`sc`) and given sex (`sx`), omitting a part of the age-scale (`rm`):

```
> CRpl <-
+ function( PV, sc, sx, rm, sepcol="white" )
+ {
+ CR <- apply( PV[perm,,,], 2:4, cumsum )
+ plot( NA, xlim=c(50,100), ylim=c(0,100),
+ xlab="", ylab="", xaxs="i", yaxs="i" )
+ axis( side=4, lwd=0, lwd.ticks=1 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.3 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[1,-rm,sc,sx],rev(nul[-rm]))*100,
+          col = clr[2], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[1,-rm,sc,sx],
+          rev(CR[3,-rm,sc,sx]))*100,
+          col = clr[4], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[3,-rm,sc,sx],
+          rev(CR[4,-rm,sc,sx]))*100,
+          col = clr[3], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[4,-rm,sc,sx],
+          rev(CR[5,-rm,sc,sx]))*100,
+          col = clr[1], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[5,-rm,sc,sx],
+          rev(CR[6,-rm,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[6,-rm,sc,sx],
+          rev(CR[7,-rm,sc,sx]))*100,
+          col = clr[3], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[7,-rm,sc,sx],
+          rev(CR[9,-rm,sc,sx]))*100,
+          col = clr[4], border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[ 9,-rm,sc,sx],
```

```
+          rev(CR[10,-rm,sc,sx]))*100,
+          col = clr[2], border="transparent")
+ matlines( aa[-rm], 100*t(CR[c(2,5,8),-rm,sc,sx]),
+ lty=1, col=c(sepcol,"black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+ }
```

With this plotting function defined we can plot the different lay-outs

```
> par( mfcol=c(3,4), mar=c(2,2,1,3), oma=c(2,2,2,0), mgp=c(3,1,0)/1.6, las=1 )
> for( sc in "2012" ) # dimnames(PV50)[[3]] )
+ for( sx in dimnames(PV50)[[4]] )
+ {
+ CRpl( PV50, sc, sx, 1:500 )
+ CRpl( PV60, sc, sx, 1:600 )
+ CRpl( PV70, sc, sx, 1:700 )
+ CRpl( DM50, sc, sx, 1:500, "transparent" )
+ CRpl( DM60, sc, sx, 1:600, "transparent" )
+ CRpl( DM70, sc, sx, 1:700, "transparent" )
+ }
> mtext( "Age (years)", side=1, outer=TRUE )
> mtext( "Fraction of persons (%)", side=2, outer=TRUE, las=0 )
> mtext( "Men, no DM" , side=3, outer=TRUE, las=0, at=1/8 )
> mtext( "Men, DM" , side=3, outer=TRUE, las=0, at=3/8 )
> mtext( "Women, no DM", side=3, outer=TRUE, las=0, at=5/8 )
> mtext( "Women, DM" , side=3, outer=TRUE, las=0, at=7/8 )
```

4.5.4 Lifetime risk

For further comparisons we extract the state distribution at age 102 years, corresponding to the lifetime risk:

```
> library( abind )
> LRp <- abind( PV[,dim(PV)[2],,],
+             PV50[,dim(PV)[2],,],
+             PV60[,dim(PV)[2],,],
+             PV70[,dim(PV)[2],,],
+             DM50[,dim(PV)[2],,],
+             DM60[,dim(PV)[2],,],
+             DM70[,dim(PV)[2],,], along=4 )
> str( PV )
   num [1:10, 1:1224, 1:18, 1:2] 1.00 6.85e-06 0.00 6.93e-06 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
    ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex  : chr [1:2] "M" "F"

> dimnames(LRp)[4] <- list( cond=c("0","W-50", "W-60", "W-70",
+                                  "DM-50","DM-60","DM-70") )
> str( LRp )
   num [1:10, 1:18, 1:2, 1:7] 2.27e-04 1.26e-05 2.03e-07 9.51e-05 3.58e-06 ...
   - attr(*, "dimnames")=List of 4
    ..$ : chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ : chr [1:2] "M" "F"
    ..$ : chr [1:7] "0" "W-50" "W-60" "W-70" ...

> round( ftable( LRp, row.vars=c(3,2,4) )*100, 1 )
                Well   DM DM-Ca   Ca Ca-DM  D-W D-DM D-Ca D-DC D-CD

  M 1995 0       0.0  0.0   0.0  0.0   0.0 52.1 14.2 28.5  3.8  1.3
         W-50    0.0  0.0   0.0  0.0   0.0 53.1 12.9 29.2  3.4  1.3
         W-60    0.0  0.0   0.0  0.0   0.0 56.2 10.8 29.0  2.7  1.2
```

```
      W-70    0.0  0.0   0.0  0.0   0.0 62.8  8.1 26.4  1.6  1.1
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 76.9  0.0 23.1  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 76.3  0.0 23.7  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 78.5  0.0 21.5  0.0
1996  0       0.0  0.0   0.0  0.0   0.0 50.6 14.5 29.4  4.1  1.4
      W-50    0.0  0.0   0.0  0.0   0.0 51.6 13.2 30.1  3.7  1.4
      W-60    0.0  0.0   0.0  0.0   0.0 54.7 11.1 30.0  2.9  1.4
      W-70    0.1  0.0   0.0  0.0   0.0 61.6  8.2 27.2  1.7  1.1
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 75.7  0.0 24.3  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 75.2  0.0 24.8  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 77.7  0.0 22.3  0.0
1997  0       0.0  0.0   0.0  0.0   0.0 49.0 14.8 30.2  4.4  1.6
      W-50    0.0  0.0   0.0  0.0   0.0 50.0 13.4 31.0  3.9  1.6
      W-60    0.0  0.0   0.0  0.0   0.0 53.1 11.3 31.0  3.1  1.5
      W-70    0.1  0.0   0.0  0.0   0.0 60.4  8.4 28.0  1.8  1.2
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 74.5  0.0 25.5  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 74.2  0.0 25.8  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 76.9  0.0 23.1  0.0
1998  0       0.0  0.0   0.0  0.0   0.0 47.4 15.0 31.1  4.7  1.7
      W-50    0.0  0.0   0.0  0.0   0.0 48.4 13.7 31.9  4.2  1.7
      W-60    0.0  0.0   0.0  0.0   0.0 51.6 11.5 31.9  3.3  1.6
      W-70    0.1  0.0   0.0  0.0   0.0 59.2  8.6 28.9  1.9  1.3
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 73.3  0.0 26.7  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 73.0  0.0 27.0  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 76.0  0.0 24.0  0.0
1999  0       0.0  0.0   0.0  0.0   0.0 45.9 15.3 31.8  5.1  1.9
      W-50    0.0  0.0   0.0  0.0   0.0 46.9 13.9 32.7  4.5  1.9
      W-60    0.1  0.0   0.0  0.0   0.0 50.1 11.8 32.8  3.5  1.8
      W-70    0.1  0.0   0.0  0.0   0.0 57.9  8.8 29.7  2.0  1.4
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 72.1  0.0 27.9  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 71.9  0.0 28.1  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 75.1  0.0 24.8  0.0
2000  0       0.0  0.0   0.0  0.0   0.0 44.4 15.6 32.4  5.5  2.1
      W-50    0.0  0.0   0.0  0.0   0.0 45.4 14.2 33.4  4.8  2.1
      W-60    0.1  0.0   0.0  0.0   0.0 48.7 12.1 33.5  3.8  1.9
      W-70    0.1  0.0   0.0  0.0   0.0 56.7  9.0 30.4  2.2  1.6
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 70.8  0.0 29.2  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 70.8  0.0 29.2  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 74.2  0.0 25.8  0.0
2001  0       0.0  0.0   0.0  0.0   0.0 43.1 16.1 32.6  5.9  2.2
      W-50    0.1  0.0   0.0  0.0   0.0 44.1 14.7 33.7  5.2  2.2
      W-60    0.1  0.0   0.0  0.0   0.0 47.4 12.5 33.9  4.1  2.1
      W-70    0.1  0.0   0.0  0.0   0.0 55.6  9.3 31.0  2.3  1.7
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 69.6  0.0 30.4  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 69.6  0.0 30.4  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 73.3  0.0 26.7  0.0
2002  0       0.0  0.0   0.0  0.0   0.0 41.7 16.7 32.7  6.5  2.4
      W-50    0.1  0.0   0.0  0.0   0.0 42.7 15.3 33.8  5.8  2.4
      W-60    0.1  0.0   0.0  0.0   0.0 46.0 13.0 34.1  4.5  2.2
      W-70    0.1  0.0   0.0  0.0   0.0 54.5  9.8 31.3  2.6  1.7
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 68.3  0.0 31.7  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 68.3  0.0 31.7  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 72.2  0.0 27.7  0.0
2003  0       0.0  0.0   0.0  0.0   0.0 40.3 17.3 32.6  7.1  2.5
      W-50    0.1  0.0   0.0  0.0   0.0 41.3 15.9 33.9  6.3  2.5
      W-60    0.1  0.0   0.0  0.0   0.0 44.7 13.6 34.3  4.9  2.4
      W-70    0.1  0.0   0.0  0.0   0.0 53.3 10.2 31.7  2.8  1.8
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 67.0  0.0 33.0  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 67.0  0.0 33.0  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 71.1  0.0 28.9  0.0
2004  0       0.0  0.0   0.0  0.0   0.0 39.0 17.7 32.9  7.8  2.6
      W-50    0.0  0.0   0.0  0.0   0.0 39.9 16.2 34.2  6.9  2.7
      W-60    0.1  0.0   0.0  0.0   0.0 43.3 14.0 34.8  5.3  2.5
      W-70    0.1  0.0   0.0  0.0   0.0 52.1 10.5 32.2  3.0  1.9
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 65.5  0.0 34.5  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 65.5  0.0 34.5  0.0
```

	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	69.9	0.0	30.1	0.0
2005	0	0.0	0.0	0.0	0.0	0.0	37.8	17.5	33.7	8.2	2.8
	W-50	0.0	0.0	0.0	0.0	0.0	38.7	16.0	35.2	7.2	2.8
	W-60	0.1	0.0	0.0	0.0	0.0	41.9	13.8	35.8	5.6	2.7
	W-70	0.1	0.0	0.0	0.0	0.0	50.8	10.5	33.3	3.2	2.0
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	64.0	0.0	36.0	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	63.9	0.0	36.1	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	68.5	0.0	31.4	0.0
2006	0	0.0	0.0	0.0	0.0	0.0	36.6	16.8	35.2	8.3	3.0
	W-50	0.0	0.0	0.0	0.0	0.0	37.4	15.4	36.7	7.4	3.0
	W-60	0.1	0.0	0.0	0.0	0.0	40.5	13.3	37.5	5.7	2.9
	W-70	0.1	0.0	0.0	0.1	0.0	49.4	10.1	34.9	3.3	2.1
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	62.5	0.0	37.5	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	62.3	0.0	37.6	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	67.2	0.0	32.8	0.0
2007	0	0.0	0.0	0.0	0.0	0.0	35.4	16.1	36.8	8.4	3.1
	W-50	0.1	0.0	0.0	0.0	0.0	36.1	14.7	38.4	7.5	3.2
	W-60	0.1	0.0	0.0	0.1	0.0	39.1	12.7	39.2	5.8	3.0
	W-70	0.1	0.0	0.0	0.1	0.0	47.8	9.8	36.7	3.3	2.2
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	61.1	0.0	38.9	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	60.9	0.0	39.1	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	66.0	0.0	34.0	0.0
2008	0	0.1	0.0	0.0	0.0	0.0	34.1	15.8	37.9	8.7	3.3
	W-50	0.1	0.0	0.0	0.1	0.0	34.7	14.5	39.6	7.7	3.4
	W-60	0.1	0.0	0.0	0.1	0.0	37.6	12.5	40.5	6.0	3.2
	W-70	0.1	0.0	0.0	0.1	0.0	46.3	9.7	38.0	3.4	2.4
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	59.9	0.0	40.1	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	59.6	0.0	40.4	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	64.9	0.0	35.1	0.0
2009	0	0.1	0.0	0.0	0.1	0.0	32.5	16.5	37.9	9.5	3.6
	W-50	0.1	0.0	0.0	0.1	0.0	33.1	15.1	39.6	8.4	3.6
	W-60	0.1	0.0	0.0	0.1	0.0	36.0	13.1	40.8	6.5	3.5
	W-70	0.1	0.0	0.0	0.1	0.0	44.6	10.2	38.6	3.8	2.6
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	58.9	0.0	41.1	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	58.5	0.0	41.4	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	64.0	0.0	36.0	0.0
2010	0	0.1	0.0	0.0	0.1	0.0	30.6	17.9	36.7	10.7	4.0
	W-50	0.1	0.0	0.0	0.1	0.0	31.2	16.5	38.6	9.5	4.1
	W-60	0.1	0.0	0.0	0.1	0.0	34.1	14.3	40.0	7.4	4.0
	W-70	0.1	0.0	0.0	0.1	0.0	42.8	11.4	38.4	4.3	3.0
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	58.1	0.0	41.9	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	57.6	0.0	42.4	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	63.2	0.0	36.7	0.0
2011	0	0.1	0.0	0.0	0.1	0.0	28.4	19.8	34.9	12.2	4.5
	W-50	0.1	0.0	0.0	0.1	0.0	29.1	18.2	36.9	10.9	4.7
	W-60	0.1	0.0	0.0	0.1	0.0	32.1	16.0	38.7	8.5	4.6
	W-70	0.1	0.0	0.0	0.1	0.0	40.8	12.8	37.7	5.0	3.5
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	57.3	0.0	42.6	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	56.8	0.0	43.2	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	62.5	0.0	37.5	0.0
2012	0	0.1	0.0	0.0	0.1	0.0	26.2	21.8	32.8	13.9	5.2
	W-50	0.1	0.0	0.0	0.1	0.0	26.9	20.2	34.9	12.4	5.3
	W-60	0.1	0.0	0.0	0.1	0.0	29.9	17.7	37.0	9.8	5.3
	W-70	0.2	0.0	0.0	0.1	0.0	38.6	14.4	36.7	5.8	4.1
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	56.6	0.0	43.4	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	56.0	0.0	43.9	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	61.7	0.0	38.3	0.0
F 1995	0	0.2	0.0	0.0	0.0	0.0	51.0	13.0	31.3	3.0	1.5
	W-50	0.2	0.0	0.0	0.0	0.0	53.6	12.1	30.0	2.5	1.4
	W-60	0.3	0.0	0.0	0.0	0.0	58.2	11.0	27.2	2.0	1.3
	W-70	0.3	0.0	0.0	0.0	0.0	67.6	8.8	21.1	1.2	0.8
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	75.2	0.0	24.8	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	77.0	0.0	23.0	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	82.1	0.0	17.9	0.0
1996	0	0.2	0.0	0.0	0.0	0.0	49.6	13.0	32.3	3.2	1.7
	W-50	0.3	0.0	0.0	0.0	0.0	52.3	12.1	31.0	2.7	1.6

```
      W-60    0.3  0.0   0.0  0.0   0.0 57.0 11.0 28.1  2.1  1.4
      W-70    0.4  0.0   0.0  0.0   0.0 66.5  8.8 22.0  1.3  0.9
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 73.8  0.0 26.2  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 75.8  0.0 24.2  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 81.3  0.0 18.7  0.0
1997  0       0.3  0.0   0.0  0.0   0.0 48.2 13.0 33.3  3.4  1.8
      W-50    0.3  0.0   0.0  0.0   0.0 51.0 12.0 32.0  2.8  1.8
      W-60    0.3  0.0   0.0  0.0   0.0 55.7 11.0 29.1  2.3  1.5
      W-70    0.4  0.0   0.0  0.1   0.0 65.4  8.9 22.8  1.4  1.0
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 72.4  0.0 27.6  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 74.5  0.0 25.4  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 80.3  0.0 19.7  0.0
1998  0       0.3  0.0   0.0  0.0   0.0 46.9 12.9 34.2  3.6  2.0
      W-50    0.3  0.0   0.0  0.0   0.0 49.7 12.0 32.9  3.0  2.0
      W-60    0.4  0.0   0.0  0.1   0.0 54.5 11.0 30.0  2.4  1.7
      W-70    0.5  0.0   0.0  0.1   0.0 64.3  8.9 23.7  1.4  1.1
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 70.9  0.0 29.1  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 73.3  0.0 26.7  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 79.3  0.0 20.6  0.0
1999  0       0.3  0.0   0.0  0.0   0.0 45.6 13.0 34.9  3.9  2.2
      W-50    0.3  0.0   0.0  0.1   0.0 48.6 12.0 33.6  3.2  2.1
      W-60    0.4  0.0   0.0  0.1   0.0 53.3 11.1 30.7  2.6  1.9
      W-70    0.5  0.0   0.0  0.1   0.0 63.3  8.9 24.4  1.5  1.2
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 69.5  0.0 30.5  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 72.1  0.0 27.9  0.0
      DM-70   0.0  0.0   0.0  0.0   0.0  0.0 78.4  0.0 21.6  0.0
2000  0       0.3  0.0   0.0  0.1   0.0 44.6 13.1 35.3  4.2  2.4
      W-50    0.4  0.0   0.0  0.1   0.0 47.6 12.2 34.0  3.4  2.3
      W-60    0.4  0.0   0.0  0.1   0.0 52.4 11.3 31.1  2.7  2.0
      W-70    0.5  0.0   0.0  0.1   0.0 62.4  9.1 24.9  1.6  1.4
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 68.2  0.0 31.8  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 70.9  0.0 29.0  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 77.4  0.0 22.5  0.0
2001  0       0.3  0.0   0.0  0.1   0.0 43.7 13.6 35.1  4.5  2.6
      W-50    0.4  0.0   0.0  0.1   0.0 46.7 12.6 34.0  3.7  2.5
      W-60    0.4  0.0   0.0  0.1   0.0 51.6 11.7 31.1  3.0  2.2
      W-70    0.5  0.0   0.0  0.1   0.0 61.7  9.4 25.1  1.7  1.4
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 67.0  0.0 32.9  0.0
      DM-60   0.0  0.0   0.0  0.0   0.0  0.0 69.9  0.0 30.0  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 76.6  0.0 23.4  0.0
2002  0       0.3  0.0   0.0  0.1   0.0 42.8 14.5 34.6  5.0  2.8
      W-50    0.4  0.0   0.0  0.1   0.0 45.9 13.4 33.5  4.1  2.7
      W-60    0.4  0.0   0.0  0.1   0.0 50.8 12.4 30.7  3.3  2.3
      W-70    0.5  0.0   0.0  0.1   0.0 60.9 10.0 24.9  1.9  1.5
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 66.0  0.0 33.9  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 69.0  0.0 30.9  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 75.8  0.0 24.1  0.0
2003  0       0.3  0.0   0.0  0.1   0.0 41.7 15.5 33.9  5.6  2.9
      W-50    0.3  0.0   0.0  0.1   0.0 44.9 14.3 33.0  4.6  2.8
      W-60    0.4  0.0   0.0  0.1   0.0 49.8 13.3 30.3  3.6  2.4
      W-70    0.5  0.0   0.0  0.1   0.0 60.1 10.7 24.8  2.1  1.6
      DM-50   0.0  0.0   0.0  0.0   0.0  0.0 65.2  0.0 34.8  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 68.2  0.0 31.7  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 75.2  0.0 24.7  0.0
2004  0       0.3  0.0   0.0  0.1   0.0 40.7 16.1 33.7  6.0  3.0
      W-50    0.4  0.0   0.0  0.1   0.0 43.9 15.0 32.8  4.9  2.9
      W-60    0.4  0.0   0.0  0.1   0.0 48.9 13.9 30.3  3.9  2.5
      W-70    0.5  0.1   0.0  0.1   0.0 59.2 11.3 24.9  2.3  1.6
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 64.3  0.0 35.6  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 67.4  0.0 32.5  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 74.5  0.0 25.4  0.0
2005  0       0.3  0.0   0.0  0.1   0.0 39.9 15.9 34.4  6.2  3.1
      W-50    0.4  0.0   0.0  0.1   0.0 43.2 14.8 33.5  5.1  3.0
      W-60    0.4  0.0   0.0  0.1   0.0 48.2 13.7 31.1  4.0  2.5
      W-70    0.6  0.1   0.0  0.1   0.0 58.5 11.1 25.6  2.4  1.7
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 63.3  0.0 36.7  0.0
```

```
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 66.4  0.0 33.5  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 73.7  0.0 26.2  0.0
2006  0       0.4  0.0   0.0  0.1   0.0 39.5 14.8 36.0  6.0  3.2
      W-50    0.4  0.0   0.0  0.1   0.0 42.8 13.7 35.0  5.0  3.0
      W-60    0.5  0.1   0.0  0.1   0.0 47.7 12.7 32.5  3.9  2.5
      W-70    0.6  0.1   0.0  0.2   0.0 57.9 10.4 26.9  2.3  1.7
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 61.9  0.0 38.0  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 65.1  0.0 34.8  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 72.6  0.0 27.2  0.0
2007  0       0.4  0.0   0.0  0.1   0.0 38.9 13.6 37.7  6.0  3.2
      W-50    0.4  0.0   0.0  0.1   0.0 42.1 12.7 36.7  4.9  3.0
      W-60    0.5  0.1   0.0  0.1   0.0 47.1 11.7 34.2  3.8  2.5
      W-70    0.6  0.1   0.0  0.2   0.0 57.2  9.7 28.2  2.3  1.7
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 60.1  0.0 39.8  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 63.4  0.0 36.5  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 71.2  0.0 28.6  0.0
2008  0       0.4  0.0   0.0  0.1   0.0 37.9 13.2 38.9  6.2  3.2
      W-50    0.4  0.0   0.0  0.1   0.0 41.1 12.3 37.8  5.1  3.0
      W-60    0.5  0.1   0.0  0.2   0.0 46.0 11.4 35.3  4.0  2.6
      W-70    0.7  0.1   0.0  0.2   0.0 56.2  9.5 29.3  2.4  1.7
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 58.1  0.0 41.8  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 61.5  0.0 38.4  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 69.7  0.0 30.1  0.0
2009  0       0.4  0.1   0.0  0.1   0.0 36.2 13.8 38.9  6.9  3.5
      W-50    0.5  0.1   0.0  0.2   0.0 39.5 12.9 38.0  5.7  3.2
      W-60    0.5  0.1   0.0  0.2   0.0 44.4 12.0 35.6  4.5  2.7
      W-70    0.7  0.1   0.0  0.2   0.0 54.7 10.1 29.7  2.7  1.8
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 56.7  0.0 43.2  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 60.1  0.0 39.8  0.0
      DM-70   0.0  0.1   0.0  0.0   0.0  0.0 68.6  0.0 31.2  0.0
2010  0       0.4  0.1   0.0  0.2   0.0 34.1 15.3 38.0  7.9  4.0
      W-50    0.5  0.1   0.0  0.2   0.0 37.4 14.4 37.2  6.6  3.7
      W-60    0.5  0.1   0.0  0.2   0.0 42.4 13.3 35.2  5.1  3.2
      W-70    0.7  0.1   0.0  0.3   0.0 52.8 11.3 29.6  3.1  2.1
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 56.0  0.0 43.9  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 59.4  0.0 40.5  0.0
      DM-70   0.0  0.2   0.0  0.0   0.0  0.0 68.0  0.0 31.8  0.0
2011  0       0.4  0.1   0.0  0.2   0.0 31.7 17.3 36.4  9.2  4.8
      W-50    0.5  0.1   0.0  0.2   0.0 35.0 16.3 35.8  7.6  4.5
      W-60    0.5  0.1   0.0  0.2   0.0 40.1 15.2 34.1  5.9  3.8
      W-70    0.7  0.1   0.0  0.3   0.0 50.7 13.0 29.1  3.6  2.5
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 55.6  0.0 44.3  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 58.9  0.0 40.9  0.0
      DM-70   0.0  0.2   0.0  0.0   0.0  0.0 67.7  0.0 32.1  0.0
2012  0       0.4  0.1   0.0  0.2   0.0 29.1 19.5 34.3 10.6  5.8
      W-50    0.4  0.1   0.0  0.2   0.0 32.5 18.5 34.0  8.8  5.4
      W-60    0.5  0.1   0.0  0.2   0.0 37.6 17.2 32.8  6.9  4.6
      W-70    0.7  0.1   0.0  0.3   0.0 48.3 14.9 28.4  4.2  3.0
      DM-50   0.0  0.1   0.0  0.0   0.0  0.0 55.2  0.0 44.7  0.0
      DM-60   0.0  0.1   0.0  0.0   0.0  0.0 58.5  0.0 41.3  0.0
      DM-70   0.0  0.2   0.0  0.0   0.0  0.0 67.4  0.0 32.4  0.0
```

We can also show the fraction of a birth cohort that will eventually face a cancer diagnosis, resp. diabetes diagnosis, and both, conditional on being in a particular state at a particular age:

```
> data.frame(1:10,st=dimnames(LRp)[[1]])
   X1.10    st
1      1  Well
2      2    DM
3      3 DM-Ca
4      4    Ca
5      5 Ca-DM
6      6   D-W
7      7  D-DM
```

```
  8      8  D-Ca
  9      9  D-DC
  10    10  D-CD
> LRsum <- abind( apply(LRp[c(7,9,10),,,],2:4,sum),
+                 apply(LRp[c(8,9,10),,,],2:4,sum),
+                 apply(LRp[c(  9,10),,,],2:4,sum),
+                 along=4 )
> dimnames( LRsum )[4] <- list( c("DM","Ca","DM+CA") )
> round( ftable( LRsum[,,c(1,2,5,3,6,4,7),], row.vars=c(4,2,1), col.vars=c(3) )*100, 1 )
                    0  W-50 DM-50  W-60 DM-60  W-70 DM-70

  DM    M 1995   19.3  17.6 100.0  14.8 100.0  10.7 100.0
            1996   20.0  18.3 100.0  15.3 100.0  11.1 100.0
            1997   20.7  18.9 100.0  15.8 100.0  11.5 100.0
            1998   21.5  19.6 100.0  16.4 100.0  11.8 100.0
            1999   22.3  20.3 100.0  17.1 100.0  12.2 100.0
            2000   23.1  21.1 100.0  17.8 100.0  12.7 100.0
            2001   24.2  22.1 100.0  18.7 100.0  13.3 100.0
            2002   25.6  23.4 100.0  19.7 100.0  14.1 100.0
            2003   27.0  24.7 100.0  20.9 100.0  14.9 100.0
            2004   28.1  25.7 100.0  21.8 100.0  15.5 100.0
            2005   28.5  26.1 100.0  22.2 100.0  15.7 100.0
            2006   28.1  25.8 100.0  21.9 100.0  15.5 100.0
            2007   27.6  25.4 100.0  21.5 100.0  15.3 100.0
            2008   27.9  25.6 100.0  21.8 100.0  15.5 100.0
            2009   29.5  27.1 100.0  23.1 100.0  16.6 100.0
            2010   32.6  30.0 100.0  25.7 100.0  18.6 100.0
            2011   36.5  33.8 100.0  29.1 100.0  21.3 100.0
            2012   40.8  37.9 100.0  32.8 100.0  24.3 100.0
          F 1995   17.5  16.0 100.0  14.3 100.0  10.9 100.0
            1996   17.8  16.3 100.0  14.5 100.0  11.1 100.0
            1997   18.2  16.6 100.0  14.8 100.0  11.2 100.0
            1998   18.6  17.0 100.0  15.1 100.0  11.4 100.0
            1999   19.1  17.4 100.0  15.5 100.0  11.7 100.0
            2000   19.8  18.0 100.0  16.0 100.0  12.0  99.9
            2001   20.8  18.9 100.0  16.8  99.9  12.6  99.9
            2002   22.3  20.2 100.0  18.0  99.9  13.5  99.9
            2003   23.9  21.7  99.9  19.3  99.9  14.5  99.9
            2004   25.2  22.8  99.9  20.2  99.9  15.2  99.9
            2005   25.2  22.8  99.9  20.2  99.9  15.2  99.9
            2006   24.0  21.7  99.9  19.1  99.9  14.4  99.9
            2007   22.8  20.6  99.9  18.1  99.9  13.6  99.9
            2008   22.7  20.4  99.9  17.9  99.9  13.6  99.9
            2009   24.2  21.9  99.9  19.2  99.9  14.5  99.8
            2010   27.3  24.7  99.9  21.6  99.9  16.5  99.8
            2011   31.3  28.4  99.9  24.9  99.9  19.1  99.8
            2012   35.9  32.7  99.9  28.7  99.8  22.1  99.8
  Ca    M 1995   33.6  33.9  23.1  33.0  23.7  29.1  21.5
            1996   34.9  35.2  24.3  34.2  24.8  30.1  22.3
            1997   36.2  36.5  25.5  35.5  25.8  31.1  23.1
            1998   37.5  37.9  26.7  36.8  27.0  32.2  24.0
            1999   38.8  39.1  27.9  38.0  28.1  33.2  24.8
            2000   39.9  40.3  29.2  39.2  29.2  34.2  25.8
            2001   40.8  41.2  30.4  40.1  30.4  34.9  26.7
            2002   41.5  42.0  31.7  40.8  31.7  35.6  27.7
            2003   42.3  42.7  33.0  41.6  33.0  36.3  28.9
            2004   43.2  43.8  34.5  42.7  34.5  37.2  30.1
            2005   44.7  45.2  36.0  44.2  36.1  38.5  31.4
            2006   46.5  47.1  37.5  46.1  37.6  40.3  32.8
            2007   48.4  49.1  38.9  48.1  39.1  42.2  34.0
            2008   50.0  50.7  40.1  49.7  40.4  43.8  35.1
            2009   50.9  51.6  41.1  50.8  41.4  44.9  36.0
            2010   51.4  52.2  41.9  51.4  42.4  45.6  36.7
            2011   51.7  52.5  42.6  51.8  43.2  46.1  37.5
            2012   51.9  52.7  43.4  52.1  43.9  46.6  38.3
          F 1995   35.8  34.0  24.8  30.5  23.0  23.2  17.9
```

```
          1996   37.2  35.3  26.2  31.7  24.2  24.2  18.7
          1997   38.5  36.6  27.6  32.9  25.4  25.2  19.7
          1998   39.9  37.9  29.1  34.1  26.7  26.3  20.6
          1999   41.0  39.0  30.5  35.1  27.9  27.2  21.6
          2000   41.9  39.8  31.8  35.9  29.0  27.9  22.5
          2001   42.3  40.2  32.9  36.2  30.0  28.3  23.4
          2002   42.4  40.3  33.9  36.3  30.9  28.4  24.1
          2003   42.4  40.3  34.8  36.4  31.7  28.5  24.7
          2004   42.8  40.6  35.6  36.7  32.5  28.9  25.4
          2005   43.7  41.5  36.7  37.6  33.5  29.7  26.2
          2006   45.2  43.0  38.0  39.0  34.8  30.9  27.2
          2007   46.9  44.6  39.8  40.5  36.5  32.2  28.6
          2008   48.4  46.0  41.8  41.9  38.4  33.4  30.1
          2009   49.4  46.9  43.2  42.9  39.8  34.2  31.2
          2010   49.9  47.5  43.9  43.5  40.5  34.8  31.8
          2011   50.3  47.9  44.3  43.9  40.9  35.2  32.1
          2012   50.7  48.3  44.7  44.3  41.3  35.7  32.4
DM+CA M   1995    5.1   4.7  23.1   3.9  23.7   2.7  21.5
          1996    5.5   5.1  24.3   4.2  24.8   2.9  22.3
          1997    6.0   5.5  25.5   4.5  25.8   3.0  23.1
          1998    6.4   5.9  26.7   4.9  27.0   3.3  24.0
          1999    7.0   6.4  27.9   5.3  28.1   3.5  24.8
          2000    7.5   6.9  29.2   5.7  29.2   3.7  25.8
          2001    8.1   7.5  30.4   6.2  30.4   4.0  26.7
          2002    8.9   8.1  31.7   6.7  31.7   4.3  27.7
          2003    9.6   8.8  33.0   7.3  33.0   4.6  28.9
          2004   10.4   9.5  34.5   7.9  34.5   5.0  30.1
          2005   11.0  10.1  36.0   8.3  36.1   5.2  31.4
          2006   11.3  10.4  37.5   8.6  37.6   5.4  32.8
          2007   11.6  10.6  38.9   8.8  39.1   5.6  34.0
          2008   12.1  11.1  40.1   9.2  40.4   5.8  35.1
          2009   13.0  12.0  41.1  10.0  41.4   6.3  36.0
          2010   14.7  13.6  41.9  11.3  42.4   7.2  36.7
          2011   16.7  15.5  42.6  13.1  43.2   8.5  37.5
          2012   19.0  17.8  43.4  15.1  43.9   9.9  38.3
        F 1995    4.5   4.0  24.8   3.3  23.0   2.1  17.9
          1996    4.8   4.3  26.2   3.5  24.2   2.2  18.7
          1997    5.2   4.6  27.6   3.8  25.4   2.4  19.7
          1998    5.7   5.0  29.1   4.1  26.7   2.6  20.6
          1999    6.1   5.4  30.5   4.4  27.9   2.8  21.6
          2000    6.6   5.8  31.8   4.8  29.0   3.0  22.5
          2001    7.2   6.2  32.9   5.1  30.0   3.2  23.4
          2002    7.8   6.8  33.9   5.6  30.9   3.5  24.1
          2003    8.5   7.3  34.8   6.0  31.7   3.7  24.7
          2004    9.0   7.8  35.6   6.4  32.5   3.9  25.4
          2005    9.3   8.0  36.7   6.5  33.5   4.0  26.2
          2006    9.2   7.9  38.0   6.4  34.8   4.0  27.2
          2007    9.2   7.9  39.8   6.4  36.5   4.0  28.6
          2008    9.5   8.1  41.8   6.6  38.4   4.1  30.1
          2009   10.4   8.9  43.2   7.2  39.8   4.5  31.2
          2010   11.9  10.3  43.9   8.3  40.5   5.2  31.8
          2011   14.0  12.1  44.3   9.7  40.9   6.1  32.1
          2012   16.4  14.2  44.7  11.5  41.3   7.3  32.4
```

```
> clr <- c("limegreen","#6666FF","#FF3333","#BB77BB","black","gray","white")
> par( mfrow=c(1,2), mar=c(2,1,1,1.5), oma=c(2,2,0,2), mgp=c(3,1,0)/1.6, las=1 )
> for( sx in dimnames(LRsum)[[2]] )
+    {
+ plot( NA, xlim=range(pp), ylim=c(0,60),
+           xlab="", ylab="", xaxs="i", yaxs="i", yaxt="n" )
+ axis( side=4, lwd=0, lwd.ticks=1, labels=(sx=="F") )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.4 )
+ axis( side=4, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ axis( side=2, lwd=0, lwd.ticks=1 )
```

```
+ axis( side=2, lwd=0, lwd.ticks=1, at=seq(10,90,10), labels=F, tcl=-0.4 )
+ axis( side=2, lwd=0, lwd.ticks=1, at=seq( 5,95, 5), labels=F, tcl=-0.4 )
+ axis( side=2, lwd=0, lwd.ticks=1, at=1:99, labels=F, tcl=-0.2 )
+ matlines( pp, LRsum[,sx,"0",]*100,
+           lty=1, lwd=6, col=clr[2:4] )
+ text( 1996, 55, sx, font=2, cex=2, adj=0 )
+ text( rep(2011,3), LRsum["2011",sx,"0",]*100+1, dimnames(LRsum)[[4]],
+       col=clr[2:4], font=2, cex=1.5, adj=c(1,0) )
+    }
> mtext( "Date of rate evaluation", side=1, line=1, outer=T )
```

4.5.5 Time spent with disease

The array `PW` contains the probability of being in a given state at a given time:

```
> str( PV )
   num [1:10, 1:1224, 1:18, 1:2] 1.00 6.85e-06 0.00 6.93e-06 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
    ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex  : chr [1:2] "M" "F"
```

The first 5 states are the "alive" states, so the sum of the probabilities of being in these is the survival function. From that we can compute the expected (residual) life time from any age by integration the (conditional) survival function to the end.

For each of the separate states in which persons are alive, we can based on simple integration compute:

- expected years spent in each state — the sum of which is the expected (residual) lifetime
- fraction of life spent in the state
- average age during the state — or more generally, population distribution of the ages in which persons are in the state

We shall compute these measures based on the derived probabilities in the array `PV`; a trivial operation using `apply`; we use 9 levels of the states, although the dead states does not make any sense, but this is just to use the slots for summaries:

```
> str( PV )
   num [1:10, 1:1224, 1:18, 1:2] 1.00 6.85e-06 0.00 6.93e-06 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1224] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
    ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex  : chr [1:2] "M" "F"
> aa <- as.numeric( dimnames(PV)[[2]] )
> var( diff(aa) )
  [1] 6.010333e-27
> PY <- apply( PV[1:9,,,], c(1,3,4), sum ) * mean( diff(aa) )
> str( PY )
   num [1:9, 1:18, 1:2] 67.935 2.2865 0.0597 1.4815 0.0636 ...
   - attr(*, "dimnames")=List of 3
    ..$ state: chr [1:9] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ per  : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex  : chr [1:2] "M" "F"
```

```
> dimnames( PY )[[1]]
 [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"
> dimnames( PY )[[1]][6:9] <- c("anyDM","anyCa","DM+Ca","All")
> PY["All"  ,,] <- apply( PY[1:5,,], 2:3, sum )
> PY["anyDM" ,,] <- apply( PY[c("DM","DM-Ca","Ca-DM","Ca")[1:3],,], 2:3, sum )
> PY["anyCa" ,,] <- apply( PY[c("DM","DM-Ca","Ca-DM","Ca")[2:4],,], 2:3, sum )
> PY["DM+Ca",,] <- apply( PY[c("DM","DM-Ca","Ca-DM","Ca")[2:3],,], 2:3, sum )
> dimnames( PV )[-2]
 $state
  [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

 $per
  [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2

 $sex
 [1] "M" "F"
> dimnames( PY )
 $state
 [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "anyDM" "anyCa" "DM+Ca" "All"

 $per
  [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2

 $sex
 [1] "M" "F"
```

The array `PY` now contains the expected number of years spent in each state, and so we can plot the expected time spent with diabetes, as a function of the date at which we evaluated rates:

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,],
+                                                100*PY["anyDM",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,3),each=2), lwd=4, col=c("blue","red"),
+          xlab="Date of rate evaluation", ylim=c(0,10), yaxs="i",
+          ylab="Years / % of life spent with DM" )
> text( 1995, 9.5, "Diabetes", adj=0 )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyCa",,],
+                                                100*PY["anyCa",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,3),each=2), lwd=4, col=c("blue","red"),
+          xlab="Date of rate evaluation", ylim=c(0,10), yaxs="i",
+          ylab="Years / % of life spent with Cancer" )
> text( 1995, 9.5, "Cancer", adj=0 )
```

4.5.5.1 Diseased ages of life

A little more tricky is measures of the ages in which the time with diabetes/cancer is spent. But the probabilities in `PV` gives the *distributions* of the ages in each state; we just need to normalize these to proper probability distributions and find the relevant quantiles.

Thus we devise a function that does this for an age-slice of `PV`, which is a vector of probabilities for each of 1224 ages (0–102 years in steps of 1 month):

```
> aPV <- as.numeric( dimnames(PV)[[2]] ) + 1/24
> pct <- c(10,25,50,75,90)/100
> aqnt <-
+ function( pp )
+ {
+ pp <- cumsum( pp / sum(pp) )
+ approx( pp, aPV, xout=pct )$y
+ }
> aqnt( PV["DM",,1,1] )
```

```
  [1] 41.73232 53.68013 63.88541 72.43918 79.13117
```

Thus we compute quantiles of age spent in states Well, DM (regardless of cancer status), Cancer and both, so we set up an array (`AD`, `A`ge `D`iseased) to hold these:

```
> AD <- ZArray( c( list( pct = pct*100,
+                        dis = c("Well","DM","Ca","DM+Ca") ),
+                  dimnames(PY)[-1] ) )
> dimnames( AD )
  $pct
  [1] "10" "25" "50" "75" "90"

  $dis
  [1] "Well"  "DM"    "Ca"    "DM+Ca"

  $per
   [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2

  $sex
  [1] "M" "F"

> dimnames( PV )[-2]
  $state
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"

  $per
   [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2

  $sex
  [1] "M" "F"

> AD["Well" ,,] <- apply( PV["Well",,,],
+                         2:3, aqnt )
> AD["DM"   ,,] <- apply( apply( PV[c("DM","DM-Ca","Ca-DM"),,,],
+                                2:4, sum ),
+                         2:3, aqnt )
> AD["Ca"   ,,] <- apply( apply( PV[c("Ca","DM-Ca","Ca-DM"),,,],
+                                2:4, sum ),
+                         2:3, aqnt )
> AD["DM+Ca",,] <- apply( apply( PV[c("DM-Ca","Ca-DM"),,,],
+                                2:4, sum ),
+                         2:3, aqnt )
> str( AD )
   num [1:5, 1:4, 1:18, 1:2] 6.81 17.09 34.6 53.35 67.27 ...
   - attr(*, "dimnames")=List of 4
    ..$ pct: chr [1:5] "10" "25" "50" "75" ...
    ..$ dis: chr [1:4] "Well" "DM" "Ca" "DM+Ca"
    ..$ per: chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex: chr [1:2] "M" "F"
```

We can the plot the distribution of the ages with DM:

```
> matplot( as.numeric(dimnames(AD)[[3]]), t(rbind(AD["DM",,"M"],
+                                                 AD["DM",,"F"])),
+          type="l", col=rep(c("blue","red"),each=5),
+          lwd=c(1,3,5,3,1), lty=1,
+          ylab="Age with diabetes (10,25,50,75,90 percentiles)",
+          xlab="Date of rates used")
```

We can augment the plot with an indication of the expected *length* of time spent, arbitrarily allocated around the median age spent with disease:

```
> par( mfrow=c(1,2) )
> matplot( pp<-as.numeric(dimnames(AD)[[3]]), t(AD[,"DM",,"M"]),
+          type="l", col="blue", lwd=c(1,3,5,3,1), lty=1,
+          ylab="Age with diabetes (10,25,50,75,90 percentiles)",
+          xlab="Date of rates used", ylim=c(0,100))
> polygon( c(pp,rev(pp)), c( AD["50","DM",,"M"]+PY["anyDM",,"M"]/2,
+                          rev(AD["50","DM",,"M"]-PY["anyDM",,"M"]/2)),
+          col="#0000FF44", border="transparent" )
> matplot( pp<-as.numeric(dimnames(AD)[[3]]), t(AD[,"DM",,"F"]),
+          type="l", col="red", lwd=c(1,3,5,3,1), lty=1,
+          ylab="Age with diabetes (10,25,50,75,90 percentiles)",
+          xlab="Date of rates used", ylim=c(0,100))
> polygon( c(pp,rev(pp)), c( AD["50","DM",,"F"]+PY["anyDM",,"F"]/2,
+                          rev(AD["50","DM",,"F"]-PY["anyDM",,"F"]/2)),
+          col="#FF000044", border="transparent" )
```

The comparison in figure 4.13 is somewhat misleading, because the percentiles of ages in which diabetes are spent are *conditional* om having had diabetes, whereas the expected length spent with the diasease is an average over all persons.

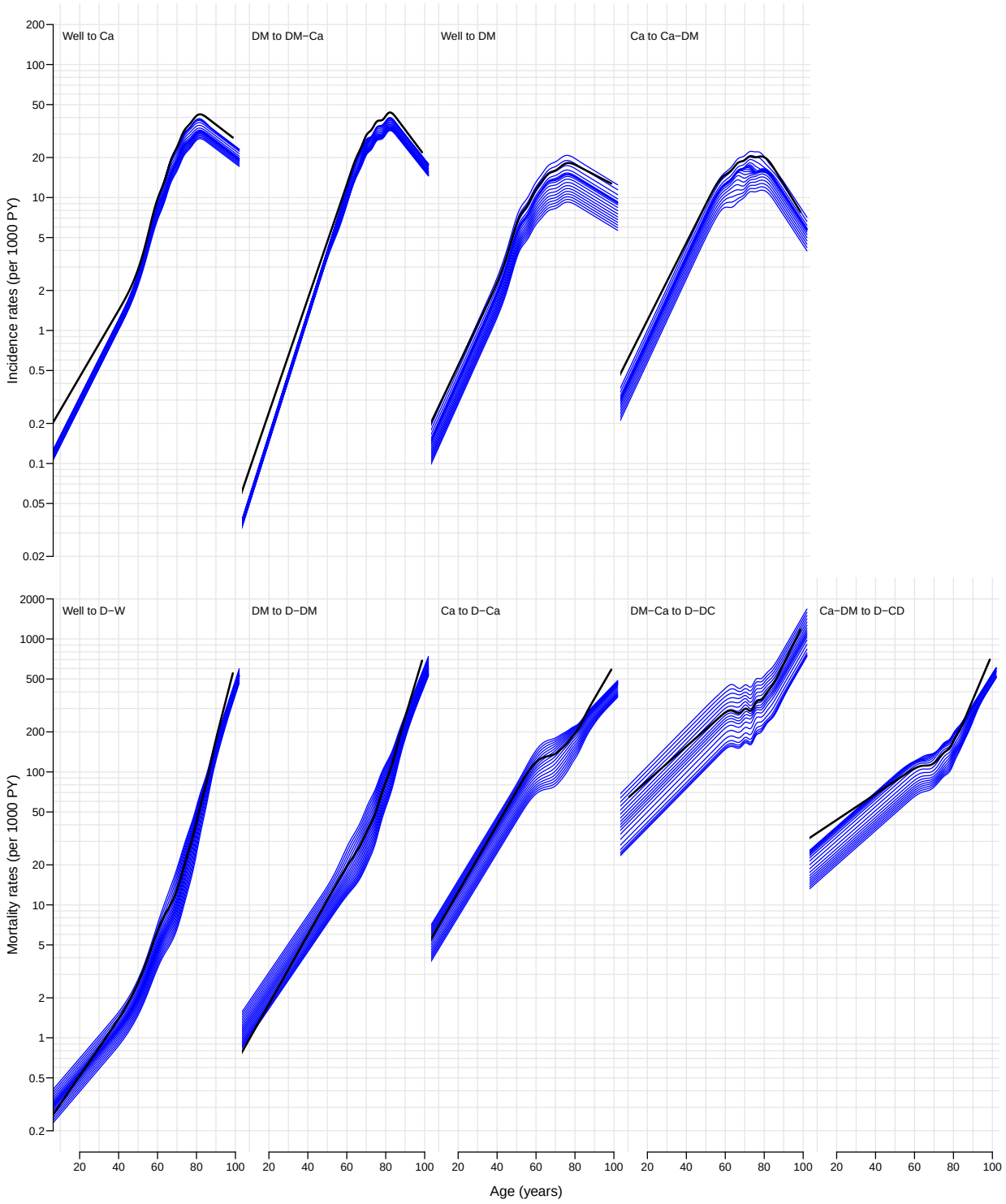

Figure 4.1: *Cross-sectional rates 1995–2012 for men, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

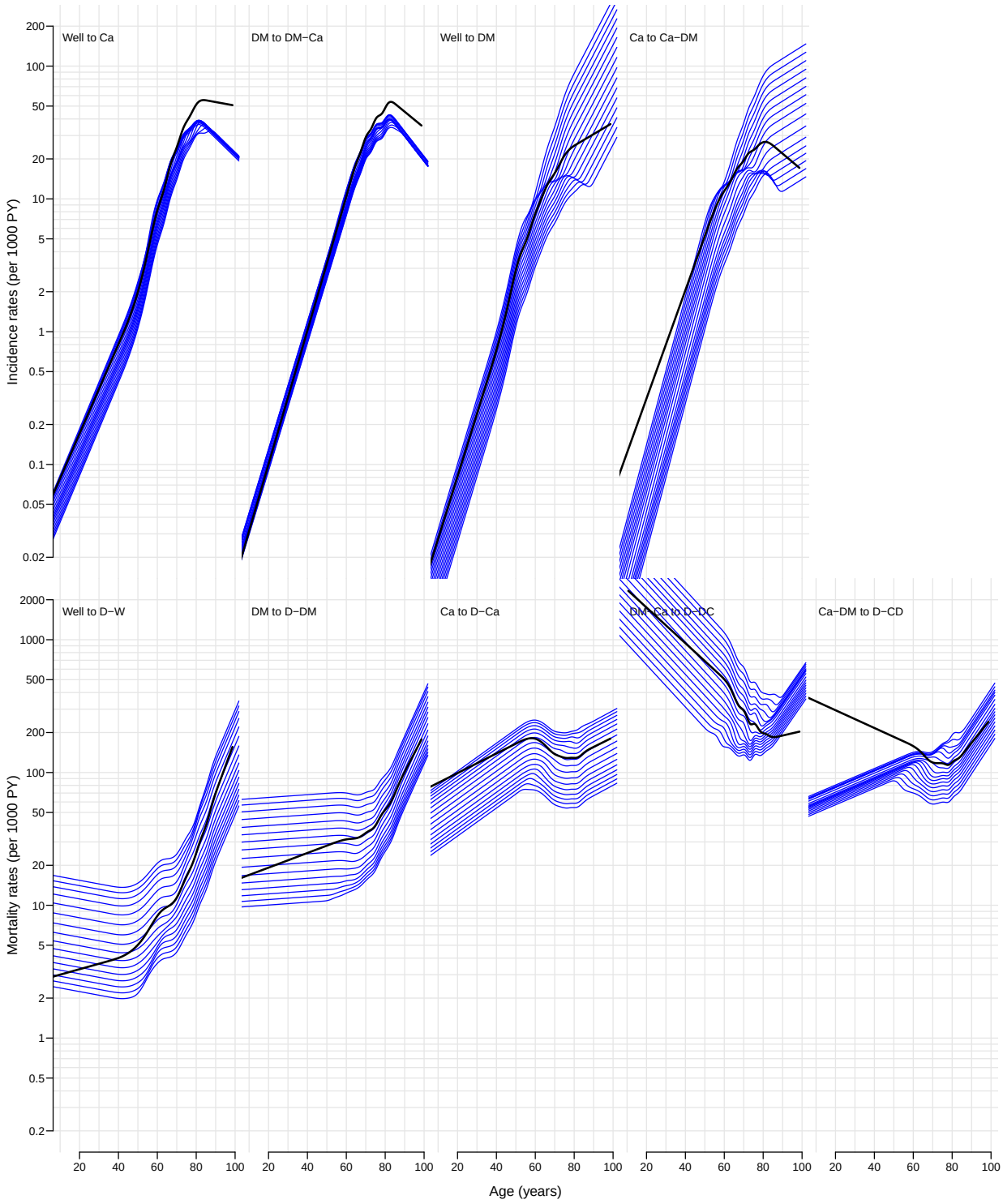

Figure 4.2: *Longitudinal rates 1930–1970 for men, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

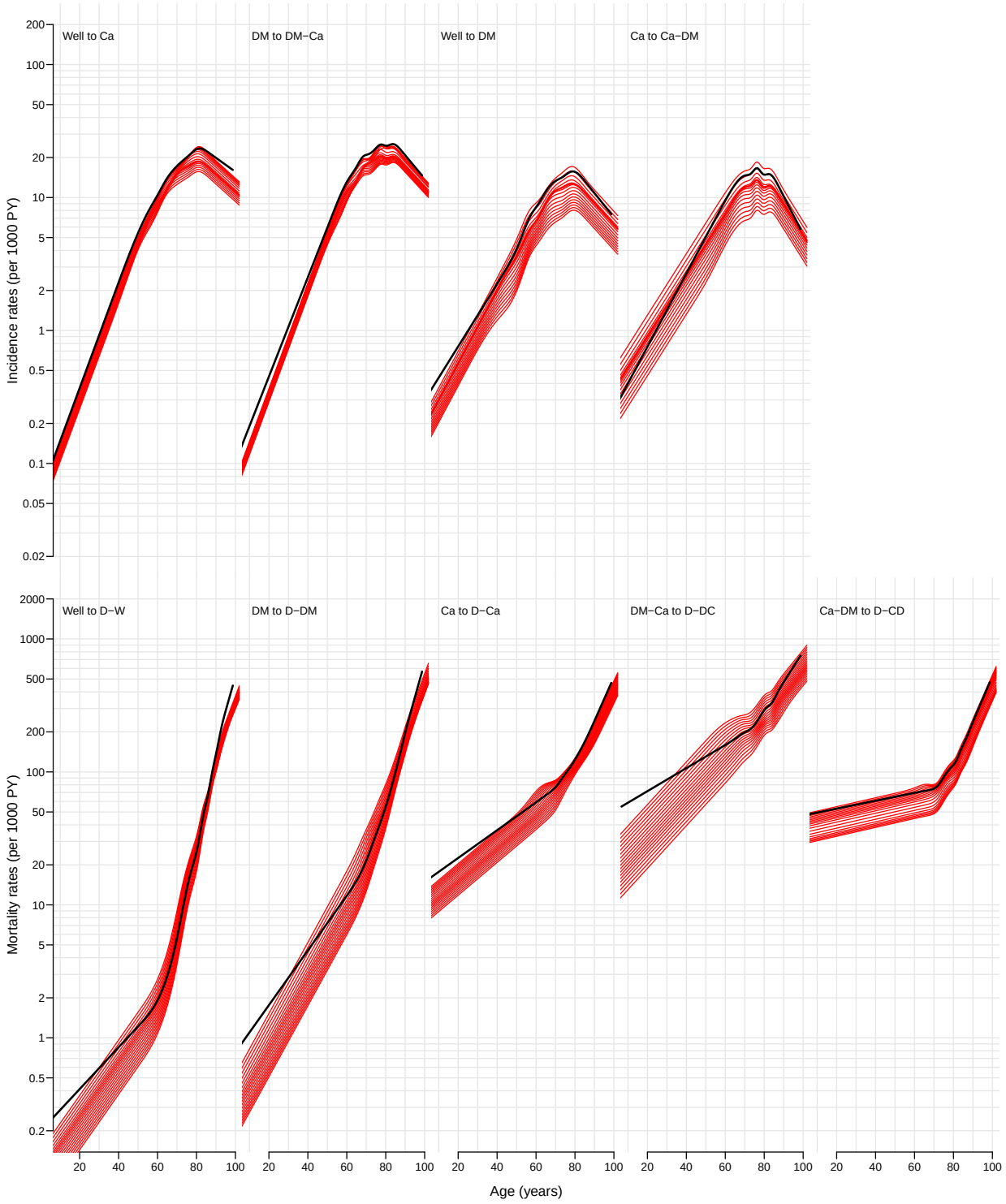

Figure 4.3: *Cross-sectional rates 1995–2012 for women, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

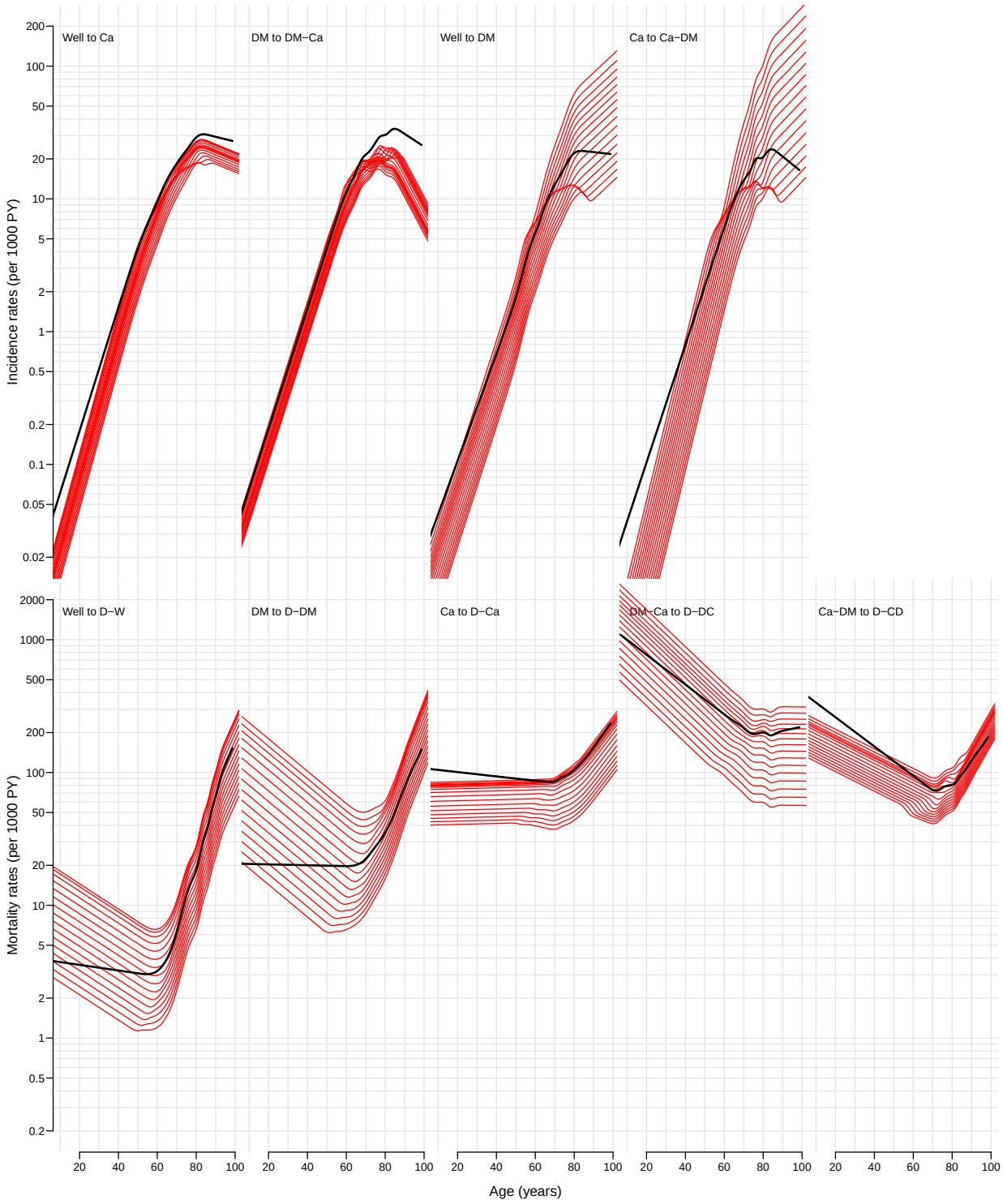

Figure 4.4: *Longitudinal rates 1930–1970 for women, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

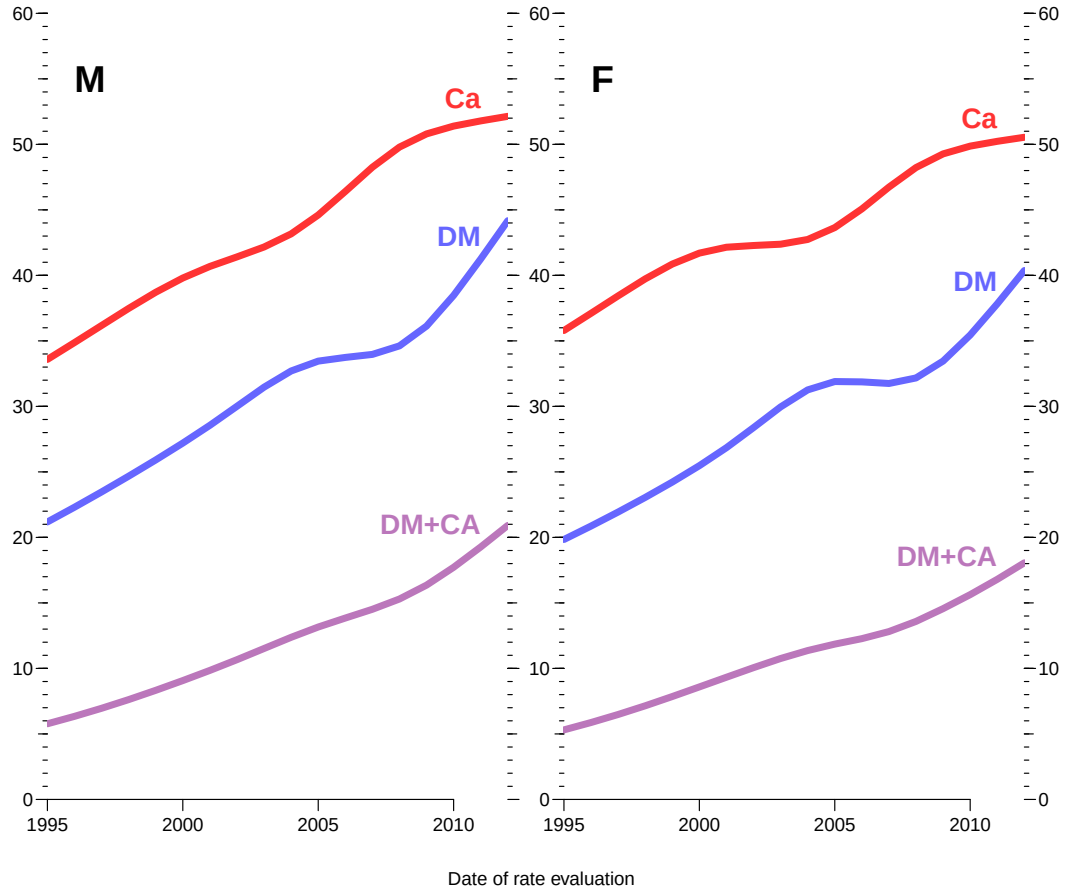

Figure 4.5: *Lifetime risk of DM (blue), cancer(red) and both diseases (purple) by year of rate calculation.*

Figure 4.6: *Lifetime risk of DM and cancer by year of rate calculation.*

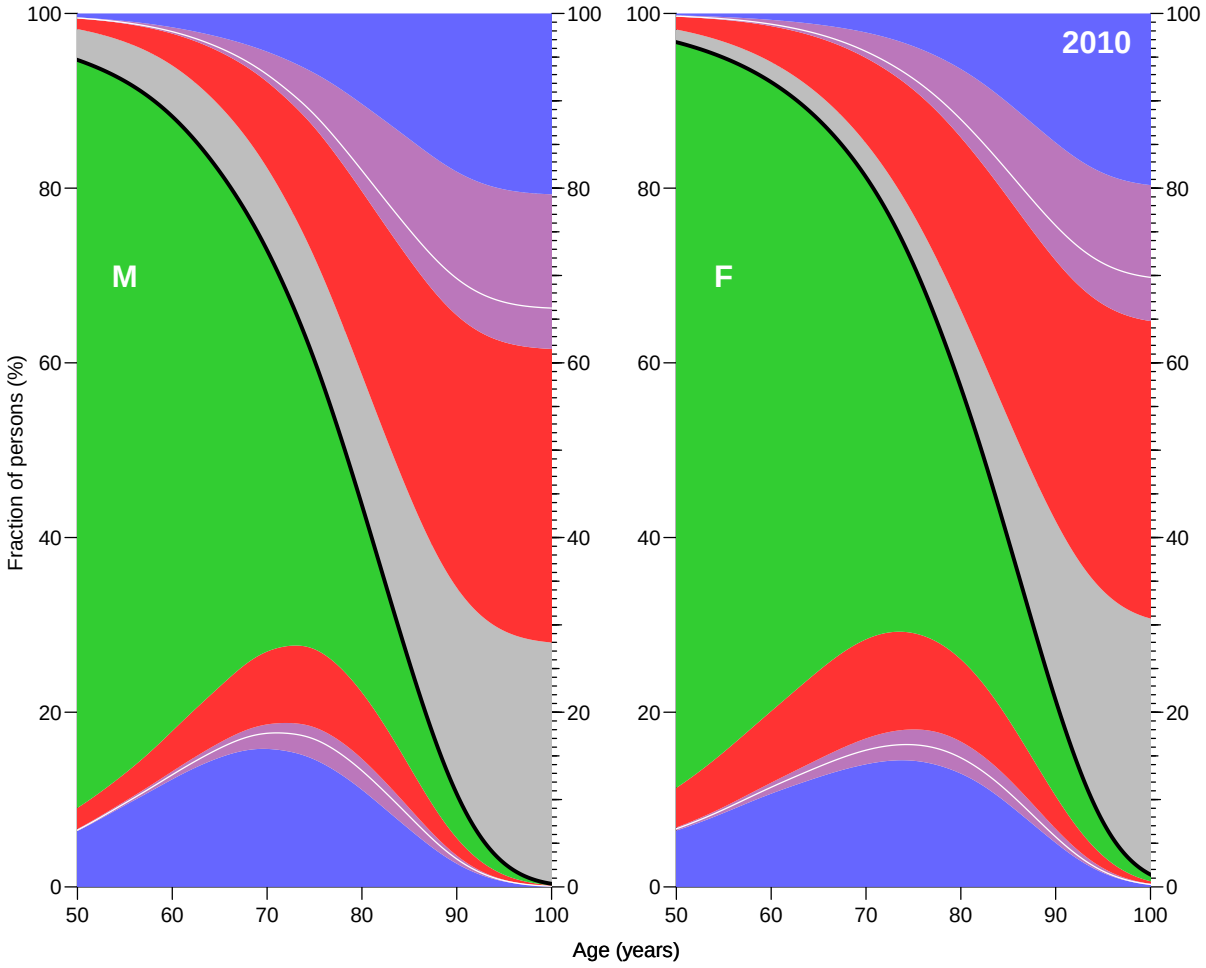

Figure 4.7: *Occupation probabilities for the various states in figure 1 at various ages, assuming all start in "Well" at age 0. Based on cross-sectional rates from 2010.*
The thick black line is the overall survival curve, with "Dead" states are above and "Alive" below the line. The blue states are persons with a diagnosis of diabetes, the red states are persons with a cancer diagnosis, and the purple areas are persons with both diagnoses. The white lines a separate those that have a DM diagnosis first (adjacent to the DM area) from those with a cancer diagnosis first (adjacent to the cancer area). The green and gray areas are those who do not have any of two diseases.

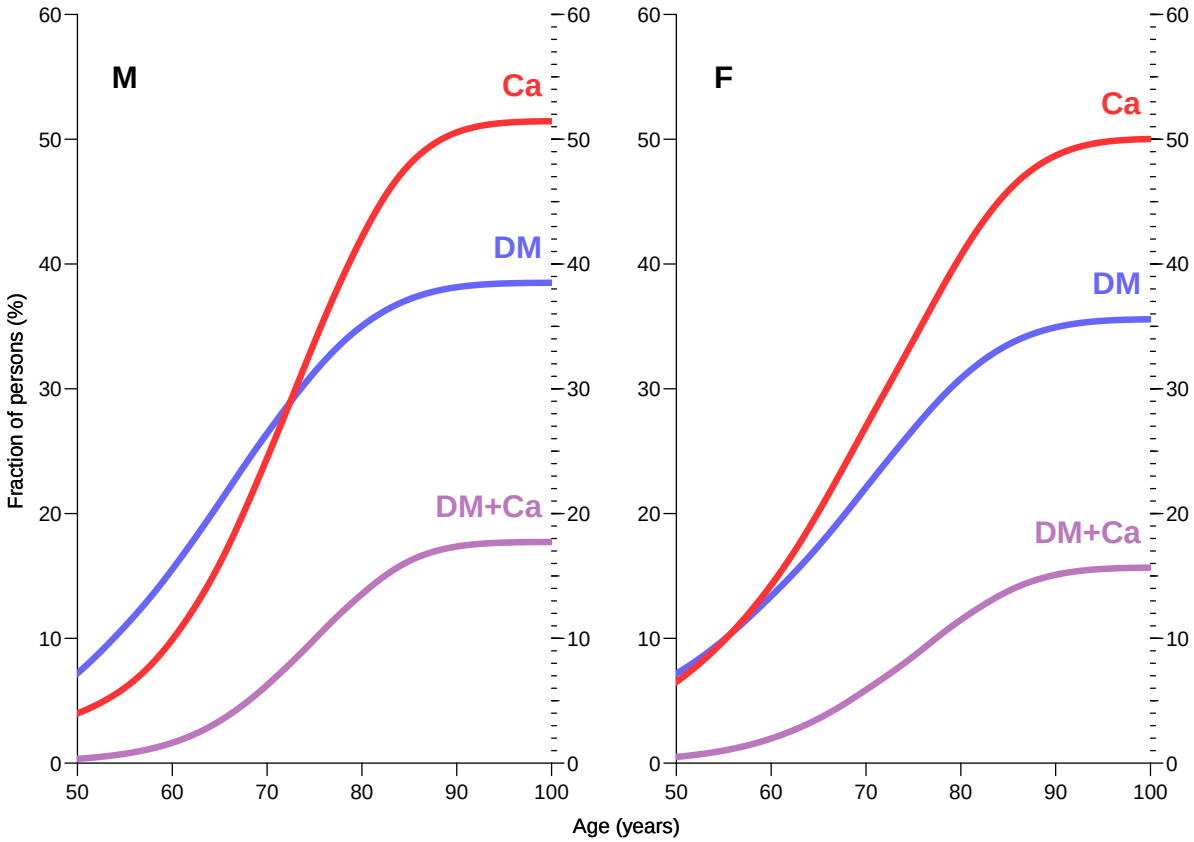

Figure 4.8: *Cumulative risk of acquiring DM, cancer or both before a given age, using rates as of 2010.*

Figure 4.9: *Plots of state occupancies conditional on being either well or diabetic at different ages.*

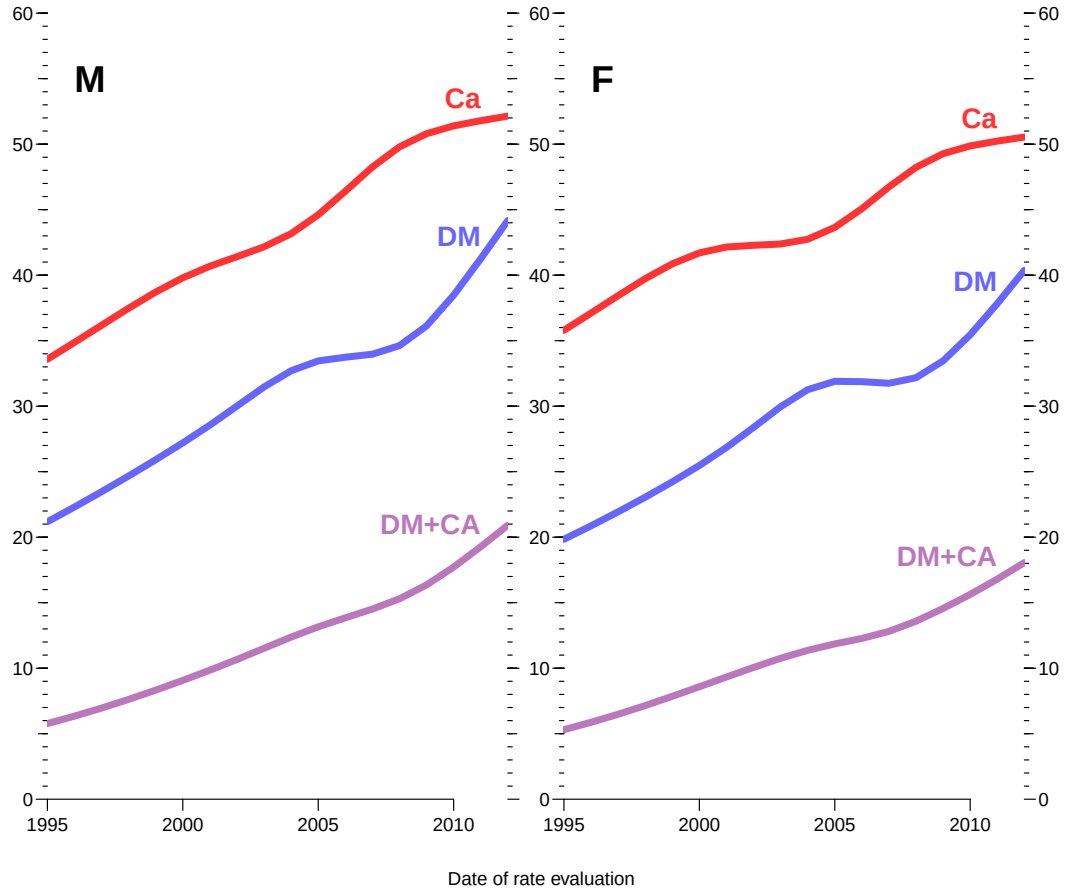

Figure 4.10: *Lifetime risk of DM (blue), cancer(red) and both diseases (purple) by year of rate calculation.*

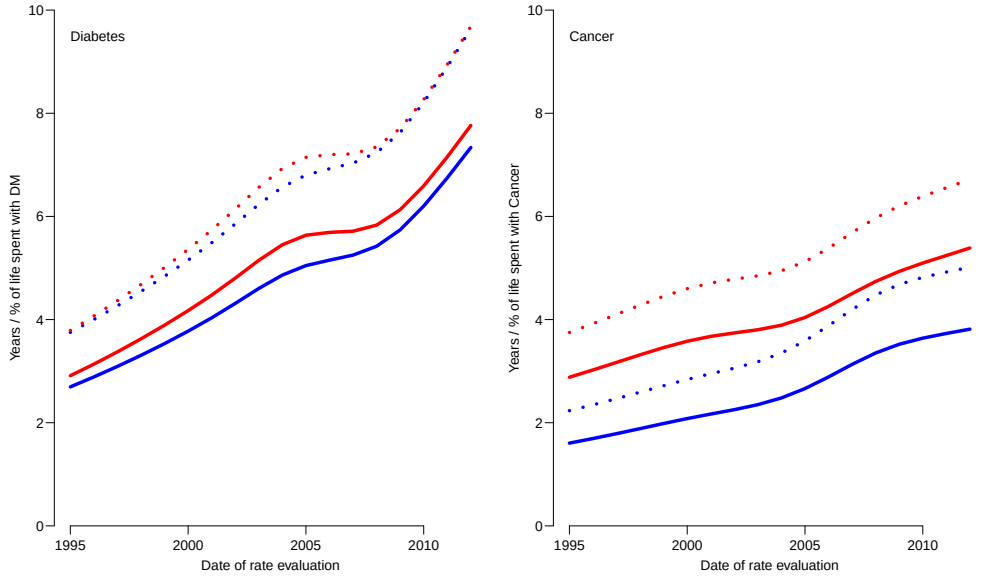

Figure 4.11: *Years (full lines) and percent of life (broken lines) spent with diabetes; red: women, blue: men.*

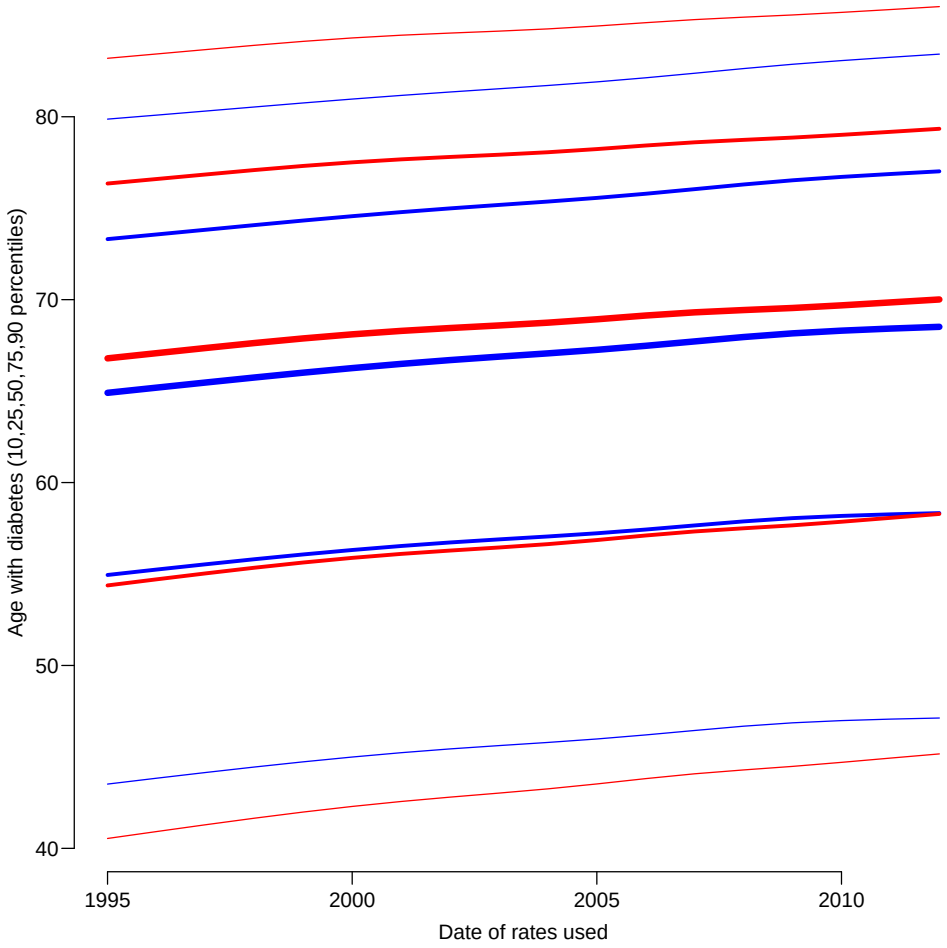

Figure 4.12: *Percentiles of ages spent with diabetes for men (blue) and women (red).*

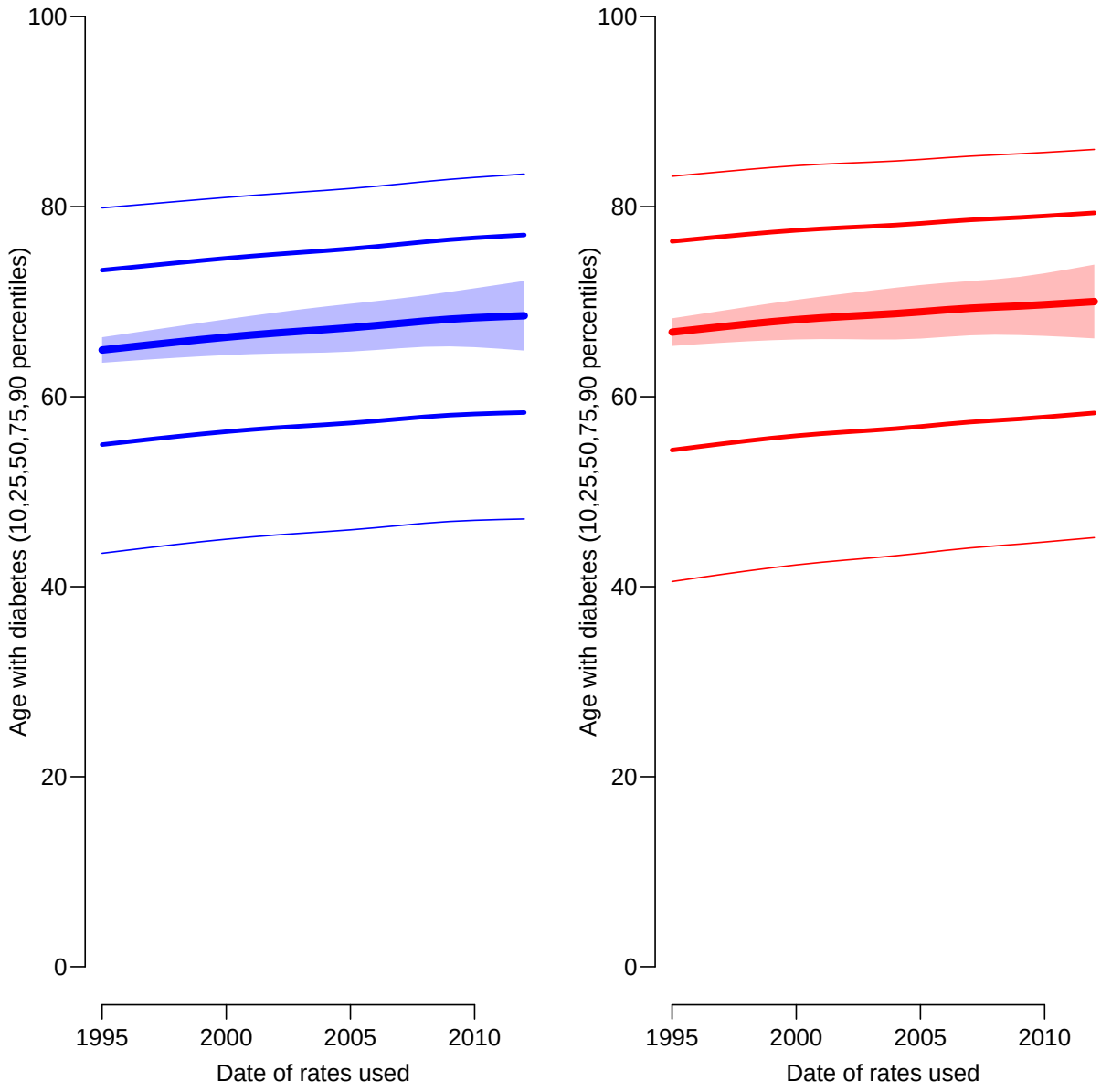

Figure 4.13: *Percentiles of ages spent with diabetes for men (blue) and women (red). The shaded area indicate the expected* length *of time spent with diabetes.*

Bibliography

[1] B. Carstensen, M. E. Jørgensen, and S. Friis. The epidemiology of diabetes and cancer. *Curr. Diab. Rep.*, 14(10):535, Oct 2014.

[2] B. Carstensen, Christensen J.K., Marcussen M.M., and Borch-Johnsen K. The National Diabetes Register. *Scandinavian Journal of Public Health*, 39(7 suppl):58–61, 2011.

[3] B. Carstensen, J.K. Kristensen, P. Ottosen, and K. Borch-Johnsen. The Danish National Diabetes Register: Trends in incidence, prevalence and mortality. *Diabetologia*, 51:2187–2196, 2008.

[4] B Carstensen. Age-Period-Cohort models for the Lexis diagram (author's reply). *Statistics in Medicine*, 27:1561–1564, 2007.