Demographic changes in the burden of Diabetes and Cancer in Denmark 1995–2012

Steno Diabetes Center, Clinical Epidemiology
June 2015
http://BendixCarstensen.com/DMCa/EpiDMCa
Version 1.3

Compiled Friday 5th June, 2015, 10:43
from: /home/bendix/sdc/DMreg/DemoDMCa/Demo-DM-Ca

Bendix Carstensen
Senior Statistician

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

Contents

Part I

Demographic trends in Diabetes and Cancer

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 [1, 2, 3].

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 1.1. 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 1.1, 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.

```
library( Epi )
tm <- matrix( NA,10,10)
ast <- c("Well","DM","Ca","DM-Ca","Ca-DM")
dst <- paste( "Dead  (", ast, ")", sep="" )
rownames(tm) <-
colnames(tm) <- c(ast,dst)
tm[1,6] <-
tm[1,2] <-
tm[1,3] <-
tm[2,4] <-
tm[2,7] <-
tm[3,5] <-
tm[3,8] <-
tm[4,9] <-
tm[5,10] <- 1
tm
```

```
              Well DM Ca DM-Ca Ca-DM Dead  (Well) Dead  (DM) Dead  (Ca) Dead  (DM-Ca)
Well            NA  1  1    NA    NA            1          NA          NA             NA
DM              NA NA NA     1    NA           NA           1          NA             NA
Ca              NA NA NA    NA     1           NA          NA           1             NA
DM-Ca           NA NA NA    NA    NA           NA          NA          NA              1
```

```
Ca-DM           NA NA NA    NA    NA           NA         NA         NA           NA
Dead  (Well)    NA NA NA    NA    NA           NA         NA         NA           NA
Dead  (DM)      NA NA NA    NA    NA           NA         NA         NA           NA
Dead  (Ca)      NA NA NA    NA    NA           NA         NA         NA           NA
Dead  (DM-Ca)   NA NA NA    NA    NA           NA         NA         NA           NA
Dead  (Ca-DM)   NA NA NA    NA    NA           NA         NA         NA           NA
               Dead  (Ca-DM)
Well                     NA
DM                       NA
Ca                       NA
DM-Ca                    NA
Ca-DM                     1
Dead  (Well)             NA
Dead  (DM)               NA
Dead  (Ca)               NA
Dead  (DM-Ca)            NA
Dead  (Ca-DM)            NA
```

```
acol <- rep( gray(0.6), 9 )
acol[c(2,4)] <- "red"
acol[c(7:9)] <- "black"
alwd <- rep(3,9)
alwd[c(4,8)] <- 6
bxs <-
boxes.Lexis( tm, boxpos=list( x = c(10,25,25,40,40,rep(87,5)),
                              y = c(50,70,30,90,10,50,70,30,90,10) ),
             wmult=1.1, hmult=2.5, lwd=2, lwd.arr=alwd, col.arr=acol )
```

We then for use in the remainder of the report define a set of colors for the different states — note that the last two colors refeering to "diabetes then cancer" (DM-Ca) and "cancer then diabetes" (Ca-DM) are defined as sligthly different weighted avearages of the colors used for the states "DM" and "Ca":

```
clr <- c("limegreen","#6666FF","#FF3333","#BB77BB","black","gray","white")
clx <- clr[c(1:7,NA,NA)]
clx[8:9] <- rgb( t(col2rgb(clx[c(2,3)]) %*%
                 cbind(c(0.65,0.35),c(0.35,0.65))),
                 max=255 )
names(clx) <- c("Well","DM","Ca","DM+Ca","black","Dead","white","DM-Ca","Ca-DM")
save( clr, clx, file="./data/cols.Rda" )
```

Then we use these to give an easier comprehensible picture.

```
bxs$Boxes$col.border <- c("transparent","black")[rep(1:2,each=5)]
# bxs$Boxes$col.bg     <- clr[c(1,2,3,4,4,6,2,3,4,4)]
  bxs$Boxes$col.bg     <- clx[c(1,2,3,8,9,6,2,3,8,9)]
bxs$Boxes$col.txt    <- c("white","black")[rep(1:2,each=5)]
bxs$Boxes$lwd        <- rep(3,10)
bxs$Boxes$wd         <- bxs$Boxes$wd * 1.1
bxs$Boxes$ht         <- bxs$Boxes$ht * 1.2
boxes.MS( bxs )
```

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 1.1 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.

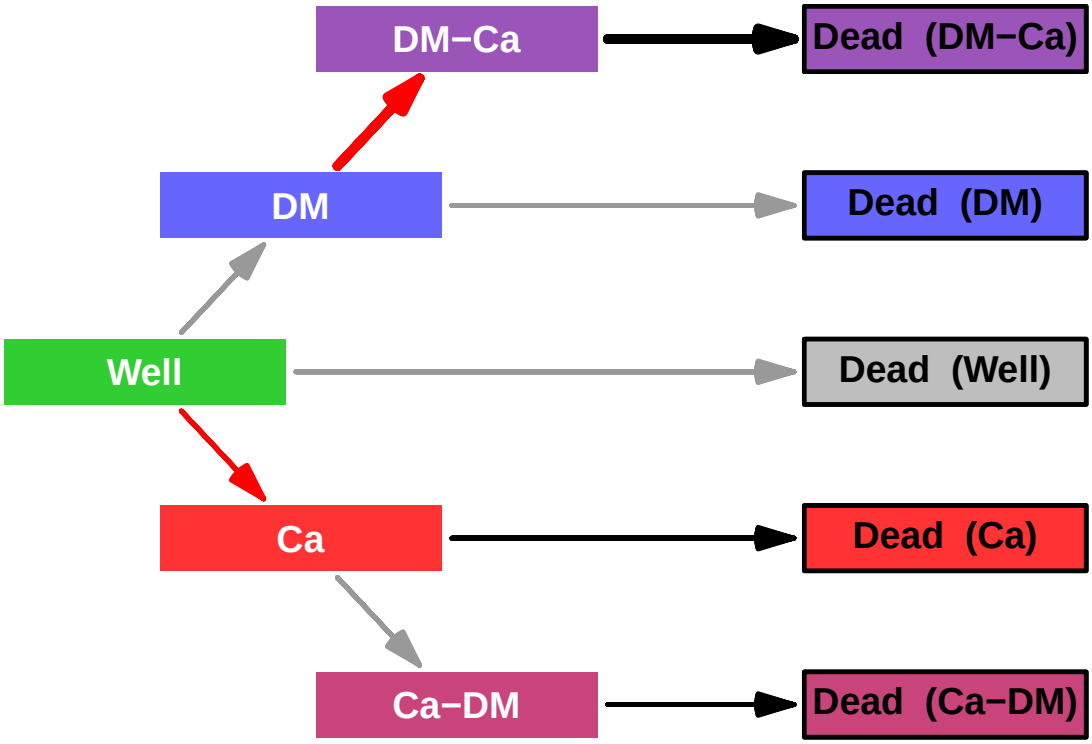

Figure 1.1: *Illustration of the incidence and mortality rates of diabetes and cancer as used to describe the joint population burden of the two diseases.*

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.

1.3 Extensions

It will be possible to use different sets of rates for the calculations; in the paper by Carstensen [4] 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 [5, 6] with the Danish Cancer Register [7], 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 [8]. 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 1.1).

We then modelled all 9 transition rates shown in Figure 1.1 using age-period-cohort (APC) models with natural splines [9]. 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.

Part II

Trends from the Danish NDR & CR

```
> options( width=90,
+ #        prompt=" ", continue=" ",
+          SweaveHooks=list( fig=function()
+          par(mar=c(3,3,1,1),mgp=c(3,1,0)/1.6) ) )
```

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 3.2.0 (2015-04-16)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
[1] Epi_1.1.68     foreign_0.8-63

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7    MASS_7.3-39     parallel_3.2.0  survival_2.38-1 etm_0.6-2
[6] splines_3.2.0   grid_3.2.0      lattice_0.20-29
```

Referring to figure 1.1, 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 99.

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 1.1 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:

```
1                                   "Program: DMCaLex.sas"        21:27 Monday, October 27, 2014

NOTE: Copyright (c) 2002-2008 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.2 (TS2M3)
      Licensed to NOVO NORDISK - BASIC PACKAGE, Site 50800704.
NOTE: This session is executing on the W32_VSPRO  platform.



NOTE: SAS initialization used:
      real time           1.84 seconds
      cpu time            0.26 seconds


NOTE: AUTOEXEC processing beginning; file is c:\stat\sas\autoexec.sas.

----------------------------------------------------------------
C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\sas\DMCaLex.sas
----------------------------------------------------------------
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\sas
NOTE: Libref DATA was successfully assigned as follows:
```

```
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\data

NOTE: AUTOEXEC processing completed.

1          ****************************************************************
2          NOTE: This version of the program takes all patients of either
3                DM or cancer, subdivide their follow-up (using the variables
4                entry, exit and fail) according to their status as being
5                either DM, Ca, DM-Ca or Ca-DM. The coding of the fail
6                variable is: 0: censored, 1: DM, 2: Cancer, 3: Dead
7          ***************************************************************** ;
8
9          * The date from which we trust the inclusion date to be the first ;
10         %let  validdate = '01JAN1995'd ;
11         * Set the entry and exit dates for the entire follow-up endeavour ;
12         %let  truncdate = '01JAN1995'd ;
13         %let   censdate = '31DEC2011'd ;
14         * Just to check it all wemt well ;
15         %put validdate = &validdate.
16              truncdate = &truncdate.
17               censdate = &censdate. ;
validdate = '01JAN1995'd      truncdate = '01JAN1995'd       censdate = '31DEC2011'd
18         * Set the selector of subgroups to analyse ;
19         %let dgrp = 21,22,241,242,243,249,251,26,28,
20                     33,
21                     51,
22                     70,
23                     82,83,84,
24                     91,92,
25                    101,103,
26                    113,
27                    121,
28                    131,132,133,139 ;
29         %let diagselect = diag in (&dgrp.) ;
30         * Variable names for tabulation purposes, note DX and D259 here ;
31         %let dvars = D0 D999
32                     D21 D22 D241 D242 D243 D249 D251 D259 D26 D28
33                     D33
34                     D51g
35                     D70
36                     D82 D83 D84
37                     D91 D92
38                    D101 D103
39                    D113
40                    D121
41                    D131 D132 D133 D139 ;
42
43         * Get the formats and the Lexis macro ;
44         options nosource2 ;
45         %inc "c:\bendix\steno\DM-register\NDR\projects\Cancer\sas\CRG-fmts.sas" ;
NOTE: Format SEX has been output.
NOTE: Format DIAG has been output.

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.05 seconds
      cpu time            0.01 seconds


130        libname DMCA "c:\bendix\steno\DM-register\NDR\projects\Cancer\data" ;
NOTE: Libref DMCA was successfully assigned as follows:
      Engine:        V9
      Physical Name: c:\bendix\steno\DM-register\NDR\projects\Cancer\data
131
132        *-------------------------------------------------------------------;
133        * Preprocessing of the cancer register to first primary tumours only ;
134
135        * First take the cancer registry, remove all non-cancers ;
136        data cancer ;
137          set DMCA.crg2012 ;
138          doca = d_diagnosedato ;
139        * Remove 'not counted as cancer' and non-melanoma skin cancer ;
140          if ( diag in (52,150) ) then delete ;
141        * Recode the leukaemias to one group (139 is a not used value in formats) ;
142          if diag in (134,135,136,137) then diag = 139 ;
143        * Recode the colon cancers to the three separate subsites and the rest ;
144          * 24.1 Ascending colon C18.0, C18.1, C18.2
145          * 24.2 Transverse colon C18.3, C18.4, C18.5
146          * 24.3 Descending and sigmoid colon C18.6, C18.7, C19, C19.9
147          * 24.9 Other colon (unspec. or multiple)
148          * 25.1 Rectum (excl. anus) C20, C209
149          * This means that colorectal cancers are to be taken as the sum of these
150          * 5 groups, but also that the group 24.9 is NOT of interest per se ;
151          if( diag eq 24 )                            then diag = 249 ;
152          if( icdpyrs in ("C180","C181","C182") )       then diag = 241 ;
153          if( icdpyrs in ("C183","C184","C185") )       then diag = 242 ;
154          if( icdpyrs in ("C186","C187","C19","C199") ) then diag = 243 ;
155          if( icdpyrs in ("C20","C209") )               then diag = 251 ;
156          * Finally make a single code for the sites not among those to be anlysed ;
```

```
157         if not ( diag in ( &dgrp. ) ) then diag = 999 ;
158       run ;

NOTE: There were 1929170 observations read from the data set DMCA.CRG2012.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: DATA statement used (Total process time):
      real time           18.03 seconds
      cpu time            2.05 seconds


159
160       * Sort by id and date of diagnosis ;
161       proc sort data = cancer ;
162         by id doCA ;
163       run ;

NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           17.39 seconds
      cpu time            3.54 seconds


164
165       * Sort by id ;
166       proc sort data = DMCA.dmr2012  out = diabetes ;
167         by id ;
168       run ;

NOTE: There were 497232 observations read from the data set DMCA.DMR2012.
NOTE: The data set WORK.DIABETES has 497232 observations and 12 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           14.65 seconds
      cpu time            0.85 seconds


169
170       * Then merge with the diabetes register ;
171       data DMCR;
172         merge cancer diabetes ;
173         by id ;
174         keep id sex diag
175              doBT doDM doCA doDD ;
176       * C_SEX is coded (1/2) in CAreg and (M/K) in DMreg ;
177         sex = ( C_SEX in ("1","M") ) + 2 * ( C_SEX in ("2","K") ) ;
178         if sex in (1,2) ;
179       * Demografic dates collected from CRG and NDR ;
180         doBT = min( D_foddto  , D_fdsdato ) ;
181         doDD = min( D_statdato, D_dodsdto ) ;
182       * Event-dates ;
183         doDM = D_inkldto ;
184         doI  = D_ins ;
185         doCA = D_diagnosedato ;
186       * If date of diabetes or cancer is equal to date of death, remove it ;
187         if doDD gt .z then do;
188            if doDM ge doDD then doDM = . ;
189            if doCA ge doDD then doCA = . ;
190            end ;
191       * If date of diabetes and cancer is the same, diabetes first ;
192         if doDM eq doCA then doDM = doCA - 2 ;
193         if doDM > .z or doCA > .z ;
194       * Only persons alive on 1.1.1995 (or born later) ;
195         if doDD gt '31DEC94'd or doDD le .z ;
196       * Only persons with one or the other disease ;
197         if doDM > .z or doCA > .z ;
198       run ;

NOTE: Missing values were generated as a result of performing an operation on missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      543533 at 181:10   63334 at 192:36
NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: There were 497232 observations read from the data set WORK.DIABETES.
NOTE: The data set WORK.DMCR has 1063649 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           1.83 seconds
      cpu time            1.07 seconds


199
200       * The dataset DMCR now has a record for each person who has either a
201       * a diabetes diagnosis or a cancer diagnosis. Persons with more than
202       * one recorded tumour are represented by a record for each tumour ;
203       * We then construct the records of follow-up in different states ;
204
205       data toLex ;
206         set DMCR ;
207         id = _n_ ;
208         keep id sex diag
209              doBT doCa doDM doDD
```

```
210              entry exit en_st ex_st ;
211          length en_st ex_st $5 ;
212        *** Only Cancer ;
213          if ( doDM le .z ) then do ;
214             entry = max( doCa, &truncdate. ) ;
215             en_st = "Ca" ;
216             exit  = min( doDD, &censdate ) ;
217             if exit eq doDD then ex_st = "Dead" ; else
218                                  ex_st = en_st ;
219             if entry lt exit then output ;
220             end ;
221        *** Only diabetes ;
222          else
223          if ( doCa le .z ) then do ;
224             entry = max( doDM, &truncdate. ) ;
225             en_st = "DM" ;
226             exit  = min( doDD, &censdate ) ;
227             if exit eq doDD then ex_st = "Dead" ; else
228                                  ex_st = en_st ;
229             if entry lt exit then output ;
230             end ;
231        *** DM before Cancer ;
232          else
233          if ( doCa gt doDM ) then do ;
234             * from DM to Ca ;
235             entry = max( doDM, &truncdate. ) ;
236             en_st = "DM" ;
237             exit  = min( doCa, &censdate ) ;
238             if exit eq doCa then ex_st = "DM-Ca" ; else
239                                  ex_st = en_st ;
240             if entry lt exit then output ;
241             * from Ca to end ;
242             entry = max( doCa, &truncdate. ) ;
243             en_st = ex_st ;
244             exit  = min( doDD, &censdate ) ;
245             if exit eq doDD then ex_st = "Dead" ; else
246                                  ex_st = en_st ;
247             if entry lt exit then output ;
248             end ;
249        *** Cancer before DM ;
250          else
251          if ( doCa lt doDM ) then do ;
252             * from Ca to DM ;
253             entry = max( doCa, &truncdate. ) ;
254             en_st = "Ca" ;
255             exit  = min( doDM, &censdate ) ;
256             if exit eq doDM then ex_st = "Ca-DM" ; else
257                                  ex_st = en_st ;
258             if entry lt exit then output ;
259             * from DM to end ;
260             entry = max( doDM, &truncdate. ) ;
261             en_st = ex_st ;
262             exit  = min( doDD, &censdate ) ;
263             if exit eq doDD then ex_st = "Dead" ; else
264                                  ex_st = en_st ;
265             if entry lt exit then output ;
266             end ;
267        run ;

NOTE: There were 1063649 observations read from the data set WORK.DMCR.
NOTE: The data set WORK.TOLEX has 1119678 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.53 seconds
      cpu time            0.45 seconds


268
269        libname allPT xport '../data/allPT.xpt' ;
NOTE: Libref ALLPT was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\data\allPT.xpt
270        proc copy in = work
271                 out = allPT ;
272          select toLex ;
273        run;

NOTE: Copying WORK.TOLEX to ALLPT.TOLEX (memtype=DATA).
NOTE: There were 1119678 observations read from the data set WORK.TOLEX.
NOTE: The data set ALLPT.TOLEX has 1119678 observations and 11 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           29.35 seconds
      cpu time            1.18 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           1:23.93
      cpu time            9.51 seconds
```

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

```
> dc <- read.xport( file="./data/allPT.xpt" )
> names( dc ) <- gsub( "_", ".", tolower( names(dc) ) )
> str( dc )
'data.frame':        1119678 obs. of  11 variables:
 $ id   : num  1 2 3 4 5 6 7 8 9 10 ...
 $ diag : num  70 NA 70 NA NA 33 999 91 70 NA ...
 $ doca : num  6575 NA 14823 NA NA ...
 $ sex  : num  2 2 2 1 2 1 2 1 2 2 ...
 $ dobt : num  -11204 -11164 -15479 -10166 -14347 ...
 $ dodd : num  NA NA 18086 15989 17833 ...
 $ dodm : num  NA 17127 NA 13172 10981 ...
 $ en.st: Factor w/ 4 levels "Ca","Ca-DM","DM",..: 1 3 1 3 3 1 1 1 1 3 ...
 $ ex.st: Factor w/ 5 levels "Ca","Ca-DM","Dead",..: 1 4 3 3 3 3 1 1 3 3 ...
 $ entry: num  12784 17127 14823 13172 12784 ...
 $ exit : num  18992 18992 18086 15989 17833 ...

> 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:545690   Min.   :1860
 1st Qu.: 266178   1st Qu.: 70.0    1st Qu.:1996     F:573988   1st Qu.:1926
 Median : 531968   Median : 91.0    Median :2002                Median :1937
 Mean   : 531833   Mean   :220.7    Mean   :2000                Mean   :1938
 3rd Qu.: 797324   3rd Qu.:241.0    3rd Qu.:2008                3rd Qu.:1948
 Max.   :1063649   Max.   :999.0    Max.   :2013                Max.   :2012
                   NA's   :381722   NA's   :382182
      dodd             dodm          en.st          ex.st            entry
 Min.   :1995     Min.   :1942     Ca   :587107   Ca   :223463   Min.   :1995
 1st Qu.:2000     1st Qu.:1996     Ca-DM: 40223   Ca-DM: 52892   1st Qu.:1996
 Median :2004     Median :2003     DM   :438780   Dead :495117   Median :2002
 Mean   :2004     Mean   :2002     DM-Ca: 53568   DM   :278550   Mean   :2002
 3rd Qu.:2009     3rd Qu.:2008                    DM-Ca: 69656   3rd Qu.:2008
 Max.   :2013     Max.   :2012                                   Max.   :2012
 NA's   :549842   NA's   :552546
      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
  0.341   0.000   0.341
```

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 )
```

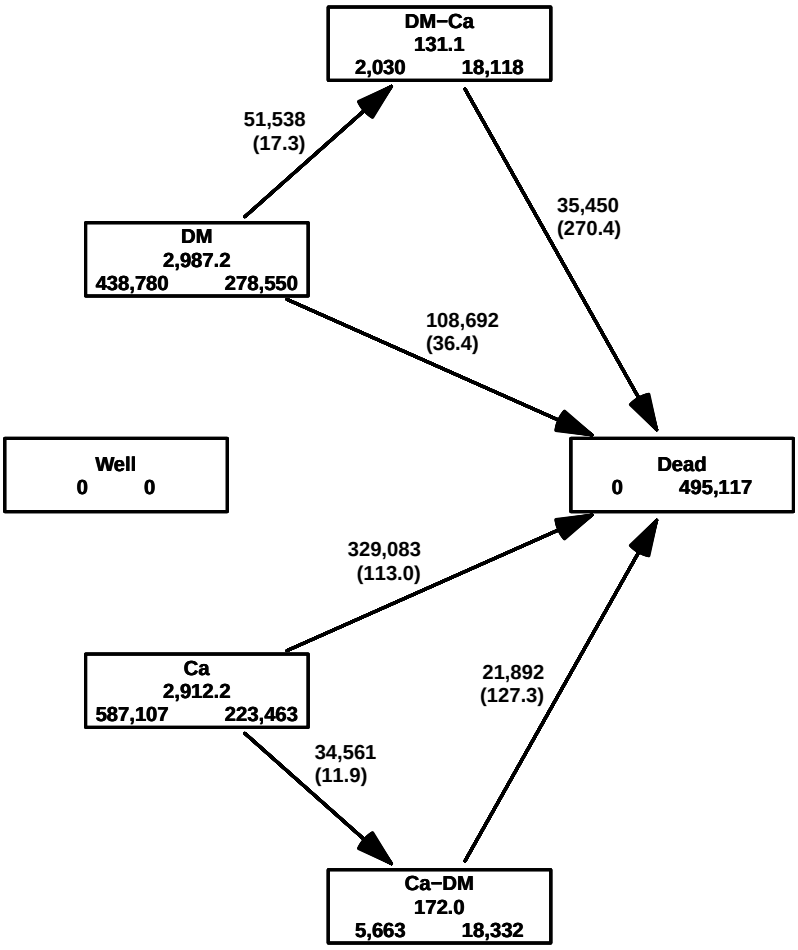

Figure 2.1: *The follow-up of the patients alone. The central number in each box is the amount of follow-up time (in 1000 PY) and the two number at the bottom are the number of persons that enter resp. exit the study in the state. Those entering also include persons that are prevalent cases as of 1.1.1995.*

2.3 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.

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 19543 rows, at 2015-06-03 11:21:44
Merged in chunk 2 now 20978 rows, at 2015-06-03 11:22:14
Merged in chunk 3 now 21763 rows, at 2015-06-03 11:22:43
Merged in chunk 4 now 22312 rows, at 2015-06-03 11:23:13
Merged in chunk 5 now 22738 rows, at 2015-06-03 11:23:41
Merged in chunk 6 now 23145 rows, at 2015-06-03 11:24:11
Merged in chunk 7 now 23474 rows, at 2015-06-03 11:24:39
Merged in chunk 8 now 23664 rows, at 2015-06-03 11:25:08
Merged in chunk 9 now 23826 rows, at 2015-06-03 11:25:37
Merged in chunk 10 now 23956 rows, at 2015-06-03 11:26:05
Merged in chunk 11 now 24197 rows, at 2015-06-03 11:26:34
Merged in chunk 12 now 24432 rows, at 2015-06-03 11:27:02
Merged in chunk 13 now 24564 rows, at 2015-06-03 11:27:30
Merged in chunk 14 now 24830 rows, at 2015-06-03 11:27:58
Merged in chunk 15 now 24982 rows, at 2015-06-03 11:28:27
Merged in chunk 16 now 25031 rows, at 2015-06-03 11:28:56
Merged in chunk 17 now 25121 rows, at 2015-06-03 11:29:25
Merged in chunk 18 now 25273 rows, at 2015-06-03 11:29:52
Merged in chunk 19 now 25342 rows, at 2015-06-03 11:30:20
Merged in chunk 20 now 25421 rows, at 2015-06-03 11:30:48
   user  system elapsed
569.652   4.125 573.546
```

```
> Agg <- transform( Agg, A = A + (1+U)/3,
+                        P = P + (2-U)/3 )
> Agg <- subset( Agg, A<99 & A>0 )
> str( Agg )
```

```
'data.frame':        23670 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.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.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 31597 14239     0     0  47462     93298    61701  956212.10     93298
  DM-Ca    0     0   158     0     0   1872      2030     1872   10429.72      2030
  Ca       0     0     0 42057 12566  84907    139530    97473 1313607.13    139530
  Ca-DM    0     0     0     0   687   4976      5663     4976   38513.44      5663
  Sum      0 31597 14397 42057 13253 139217    240521   166022 2318762.39    240521

$Incident

Transitions:
     To
From    Well     DM DM-Ca     Ca Ca-DM   Dead  Records:  Events: Risk time:  Persons:
  DM       0 246953 37299      0     0  61230    345482    98529  2030992.0    345482
  DM-Ca    0      0 17960      0     0  33578     51538    33578   120674.6     51538
  Ca       0      0     0 181406 21995 244176    447577   266171  1598636.5    447577
  Ca-DM    0      0     0      0 17644  16916     34560    16916   133525.1     34560
  Sum      0 246953 55259 181406 39639 355900    879157   415194  3883828.2    819864
> 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':        6718 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 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':        23670 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':        30402 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.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  447421 345400 470708 85517.0
DM     51529      0 108118  2986.0
DM-Ca      0      0  35419   131.1
Ca         0  34547 327961  2908.0
Ca-DM      0      0  21822   171.9
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           14086     0     0  1158     0     0
     1996           14653     0     0  1179     0     0
     1997           14615     0     0  1256     0     0
     1998           15941     0     0  1375     0     0
     1999           16857     0     0  1542     0     0
     2000           17407     0     0  1589     0     0
     2001           18483     0     0  1708     0     0
     2002           20447     0     0  2025     0     0
     2003           22253     0     0  2163     0     0
     2004           22448     0     0  2140     0     0
     2005           20760     0     0  1956     0     0
     2006           21310     0     0  2235     0     0
     2007           22712     0     0  2304     0     0
     2008           24487     0     0  2646     0     0
     2009           24426     0     0  2704     0     0
     2010           25284     0     0  2985     0     0
     2011           29231     0     0  3582     0     0
D.ca 1995           20917  1444     0     0     0     0
     1996           23203  1388     0     0     0     0
     1997           23653  1651     0     0     0     0
     1998           24334  1809     0     0     0     0
     1999           24493  2161     0     0     0     0
     2000           24554  1996     0     0     0     0
     2001           24631  2405     0     0     0     0
     2002           25184  2395     0     0     0     0
     2003           25335  2670     0     0     0     0
     2004           26334  3177     0     0     0     0
     2005           26916  3364     0     0     0     0
```

```
     2006           27872  3726     0     0     0     0
     2007           28411  3882     0     0     0     0
     2008           29810  4434     0     0     0     0
     2009           31275  4844     0     0     0     0
     2010           30215  4914     0     0     0     0
     2011           30284  5269     0     0     0     0
D.dd 1995           35354  5553  1042 19828   953     0
     1996           34150  5304  1072 19167   962     0
     1997           32551  5346  1230 19476   976     0
     1998           31230  5399  1331 19204   933     0
     1999           30851  5741  1510 19547  1103     0
     2000           29637  5814  1613 19398  1101     0
     2001           29665  5921  1693 19448  1181     0
     2002           29416  6316  1782 19350  1278     0
     2003           26163  6453  2021 19350  1347     0
     2004           28753  6420  2109 18544  1267     0
     2005           25563  6672  2277 18591  1336     0
     2006           25004  6849  2518 19245  1349     0
     2007           24610  6993  2639 19296  1503     0
     2008           23447  6946  2729 19364  1533     0
     2009           23057  7463  3090 19126  1520     0
     2010           21468  7502  3294 19668  1706     0
     2011           19789  7426  3469 19359  1774     0
```

```
> library( Epi )
> library( splines )
> options( width=90,
+ #         prompt=" ", continue=" ",
+          SweaveHooks=list( fig=function()
+          par(mar=c(3,3,1,1),mgp=c(3,1,0)/1.6,bty="n",las=1) ) )
```

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

```
R version 3.2.0 (2015-04-16)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
[1] Epi_1.1.68

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7    MASS_7.3-39     parallel_3.2.0  survival_2.38-1 etm_0.6-2
[6] grid_3.2.0      lattice_0.20-29
```

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/cols.Rda" )
> load( file="./data/dcd.Rda" )
> dcd <- subset( dcd, P<2012 & !is.na(Y) )
> 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   182757 216941 229328  42592
    DM          0  28901  56253   1543
    DM-Ca       0      0  20222     67
    Ca      16000      0 159491   1105
    Ca-DM       0      0  10394     68
    Dead        0      0      0      0
F   Well   162643 230480 241380  42925
    DM          0  22628  51865   1443
    DM-Ca       0      0  15197     64
    Ca      18547      0 168470   1803
    Ca-DM       0      0  11428    104
    Dead        0      0      0      0
Sum Well   345400 447421 470708  85517
    DM          0  51529 108118   2986
    DM-Ca       0      0  35419    131
    Ca      34547      0 327961   2908
    Ca-DM       0      0  21822    172
    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=clx["Well"], border=clx["Well"], 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=clx["Well"], border=clx["Well"], 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=clx["Well"], border=clx["Well"], 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=clx["DM"], border=clx["DM"], 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=clx["DM"], border=clx["DM"], 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=clx["DM-Ca"], border=clx["DM-Ca"], 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=clx["Ca"], border=clx["Ca"], 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=clx["Ca"], border=clx["Ca"], 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=clx["Ca-DM"], border=clx["Ca-DM"], 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,]
```

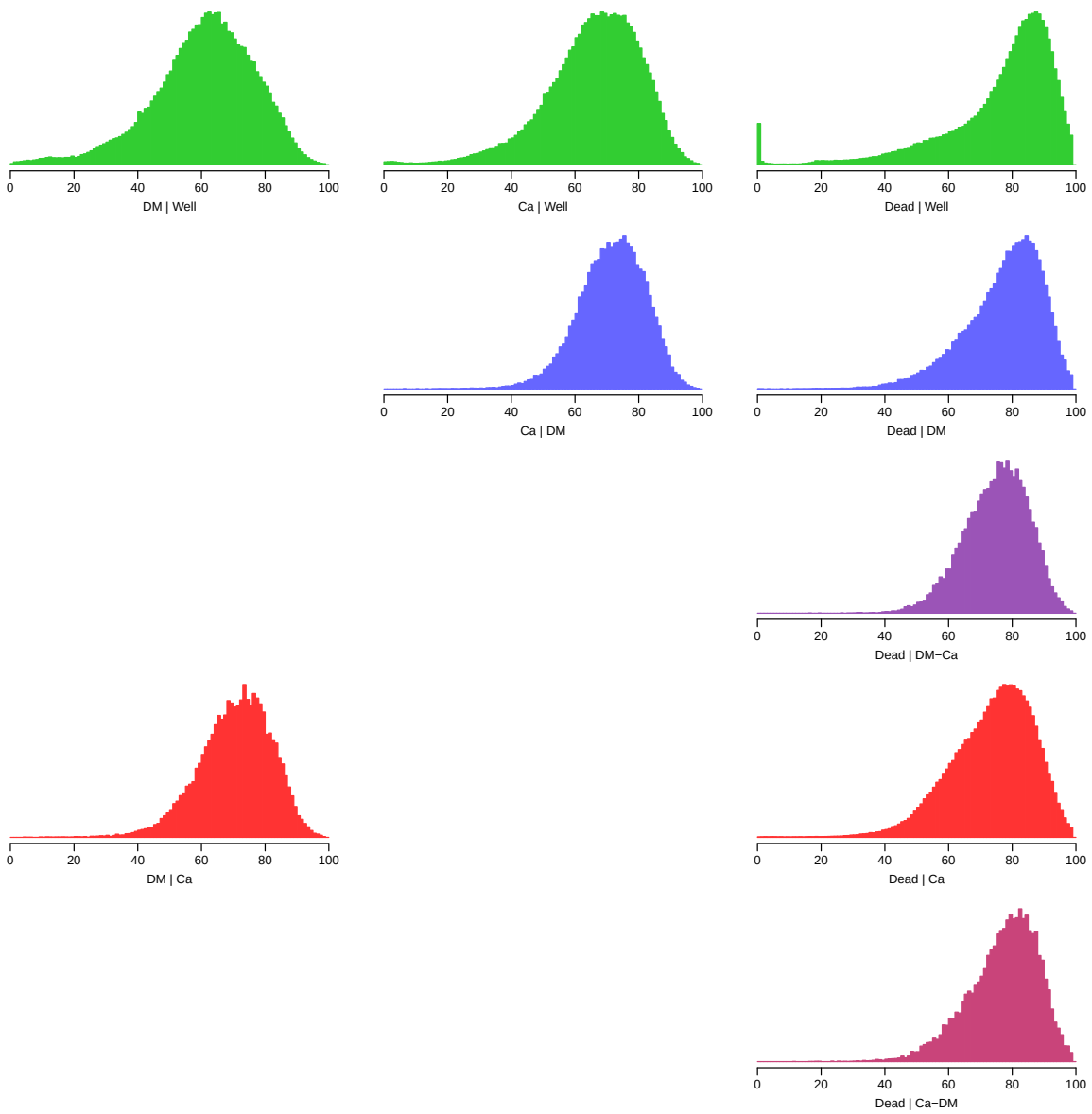

Figure 3.1: *Histograms of the age at event for the 9 possible transitions. Clearly, nothing much is happening in the younger ages, so we shall have age-knots a little closer in the older ages.*

```
+ acp$Coh <- acp$Coh[acp$Coh[,"Coh"]<1990,]
+ Kn <- apc$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    11379.0
Age-drift              3355     5055.7  1   6323.3 < 2.2e-16
Age-Cohort             3350     5012.2  5     43.4 3.051e-08
Age-Period-Cohort      3347     4800.4  3    211.9 < 2.2e-16
Age-Period             3352     4839.1 -5    -38.7 2.737e-07
Age-drift              3355     5055.7 -3   -216.6 < 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     8039.8
Age-drift              3355     5538.5  1  2501.29 < 2.2e-16
Age-Cohort             3350     5369.4  5   169.19 < 2.2e-16
Age-Period-Cohort      3347     5279.6  3    89.73 < 2.2e-16
Age-Period             3352     5446.6 -5  -167.01 < 2.2e-16
Age-drift              3355     5538.5 -3   -91.90 < 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      29864
Age-drift              3355      22588  1   7276.0 < 2.2e-16
Age-Cohort             3350      21196  5   1392.2 < 2.2e-16
Age-Period-Cohort      3347      21123  3     72.5  1.23e-15
Age-Period             3352      22504 -5  -1380.8 < 2.2e-16
Age-drift              3355      22588 -3    -83.9 < 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     2830.0
Age-drift              3350     2645.0  1  184.984  < 2e-16
Age-Cohort             3345     2635.6  5    9.423  0.09334
Age-Period-Cohort      3342     2631.6  3    3.962  0.26558
Age-Period             3347     2641.0 -5   -9.355  0.09572
Age-drift              3350     2645.0 -3   -4.030  0.25821
```

```
> 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     5767.2
Age-drift              3350     3153.6  1  2613.67 < 2.2e-16
Age-Cohort             3345     2947.1  5   206.47 < 2.2e-16
Age-Period-Cohort      3342     2934.5  3    12.54  0.005739
Age-Period             3347     3151.4 -5  -216.82 < 2.2e-16
Age-drift              3350     3153.6 -3    -2.19  0.533872
```

```
> 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     3387.0
Age-drift              3355     3109.7  1  277.312 < 2.2e-16
Age-Cohort             3350     3092.7  5   16.964  0.004569
Age-Period-Cohort      3347     3052.7  3   40.086 1.022e-08
Age-Period             3352     3070.9 -5  -18.205  0.002701
Age-drift              3355     3109.7 -3  -38.845 1.872e-08
> 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    11543.4
Age-drift              3355     5514.3  1   6029.1 < 2.2e-16
Age-Cohort             3350     5003.2  5    511.1 < 2.2e-16
Age-Period-Cohort      3347     4966.1  3     37.1 4.341e-08
Age-Period             3352     5431.1 -5   -465.0 < 2.2e-16
Age-drift              3355     5514.3 -3    -83.2 < 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                    2717     2971.6
Age-drift              2716     2664.3  1  307.265 < 2.2e-16
Age-Cohort             2711     2641.2  5   23.087 0.0003248
Age-Period-Cohort      2708     2631.9  3    9.297 0.0255926
Age-Period             2713     2651.7 -5  -19.754 0.0013897
Age-drift              2716     2664.3 -3  -12.630 0.0055082
> 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                    2346     3979.2
Age-drift              2345     2473.5  1  1505.76  < 2e-16
Age-Cohort             2340     2462.7  5    10.73  0.05695
Age-Period-Cohort      2337     2452.1  3    10.67  0.01365
Age-Period             2342     2462.5 -5   -10.46  0.06316
Age-drift              2345     2473.5 -3   -10.94  0.01205
> # 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    11153.5
Age-drift              3355     5455.3  1   5698.2 < 2.2e-16
Age-Cohort             3350     5315.5  5    139.8 < 2.2e-16
Age-Period-Cohort      3347     5026.2  3    289.2 < 2.2e-16
Age-Period             3352     5170.8 -5   -144.6 < 2.2e-16
Age-drift              3355     5455.3 -3   -284.5 < 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     6888.3
Age-drift              3355     5345.0  1  1543.35 < 2.2e-16
Age-Cohort             3350     5211.2  5   133.81 < 2.2e-16
Age-Period-Cohort      3347     5098.1  3   113.06 < 2.2e-16
Age-Period             3352     5239.6 -5  -141.48 < 2.2e-16
Age-drift              3355     5345.0 -3  -105.39 < 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      28042
Age-drift              3355      24159  1   3882.8 < 2.2e-16
Age-Cohort             3350      22625  5   1533.9 < 2.2e-16
Age-Period-Cohort      3347      22588  3     37.3 3.981e-08
Age-Period             3352      24100 -5  -1512.5 < 2.2e-16
Age-drift              3355      24159 -3    -58.7 1.117e-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     2919.3
Age-drift              3346     2752.0  1  167.298 < 2.2e-16
Age-Cohort             3341     2740.7  5   11.279  0.046120
Age-Period-Cohort      3338     2725.6  3   15.175  0.001673
Age-Period             3343     2736.7 -5  -11.101  0.049411
Age-drift              3346     2752.0 -3  -15.352  0.001539
```

```
> 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     5117.3
Age-drift              3346     3197.1  1  1920.20 < 2.2e-16
Age-Cohort             3341     2934.3  5   262.88 < 2.2e-16
Age-Period-Cohort      3338     2886.6  3    47.68 2.489e-10
Age-Period             3343     3177.5 -5  -290.95 < 2.2e-16
Age-drift              3346     3197.1 -3   -19.61 0.0002042
```

```
> 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     3619.0
Age-drift              3355     3016.8  1   602.22 < 2.2e-16
Age-Cohort             3350     3008.0  5     8.83   0.11591
Age-Period-Cohort      3347     2942.0  3    65.98 3.102e-14
Age-Period             3352     2954.7 -5   -12.74   0.02595
Age-drift              3355     3016.8 -3   -62.07 2.123e-13
```

```
> 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     7408.5
Age-drift               3355     4715.1  1  2693.39  < 2e-16
Age-Cohort              3350     4348.8  5   366.32  < 2e-16
Age-Period-Cohort       3347     4345.1  3     3.64  0.30289
Age-Period              3352     4705.7 -5  -360.60  < 2e-16
Age-drift               3355     4715.1 -3    -9.36  0.02485
```

```
> 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                    2635     2881.6
Age-drift              2634     2614.5  1  267.126   <2e-16
Age-Cohort             2629     2612.3  5    2.228   0.8168
Age-Period-Cohort      2626     2607.6  3    4.648   0.1995
Age-Period             2631     2609.5 -5   -1.833   0.8718
Age-drift              2634     2614.5 -3   -5.043   0.1687
```

```
> 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                    2482     3131.0
Age-drift              2481     2524.0  1   607.02 < 2.2e-16
Age-Cohort             2476     2500.9  5    23.10 0.0003232
Age-Period-Cohort      2473     2499.6  3     1.22 0.7481820
Age-Period             2478     2522.6 -5   -22.99 0.0003388
Age-drift              2481     2524.0 -3    -1.33 0.7226354
```

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(1,0,0.5,1), oma=c(3,4,2,3),
+      mgp=c(3,1,0)/1.6, las=1, bty="n" )
> inc.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.dm2ca$apc, col=clx["DM"]    , lend=1, lwd=4 )
> lines( M.w2ca$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( M.w2dm$apc , col=clx["Well"] , lend=1, lwd=4, lty="21" )
```

```
> lines( M.ca2dm$apc, col=clx["Ca"]   , lend=1, lwd=4, lty="21" )
> text(rep(101,2), c((rev(M.ca2dm$apc$Age[,2])[1]+
+                     rev( M.w2dm$apc$Age[,2])[1])/2,
+                    (rev(M.dm2ca$apc$Age[,2])[1]+
+                     rev( M.w2ca$apc$Age[,2])[1])/2), c("DM inc.","Ca inc."), adj=0 )
> text( c(145,208), c(0.25,0.25), c("Cohort","Period") )
> mort.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.w2dd$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( M.dm2dd$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( M.ca2dd$apc, col=clx["Ca"]   , lend=1, lwd=4 )
> lines( M.cd2dd$apc, col=clx["Ca-DM"], lend=1, lwd=4 )
> lines( M.dc2dd$apc, col=clx["DM-Ca"], lend=1, lwd=4 )
> text( c(145,208), c(2.5,2.5), c("Cohort","Period") )
> mtext( "Age", at=55, side=1, line=2 )
> mtext( "Calendar time", at=165, side=1, line=2 )
> inc.fr(sides=1,col.grid=gray(0.9))
> lines( F.dm2ca$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( F.w2ca$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( F.w2dm$apc , col=clx["Well"] , lend=1, lwd=4, lty="21" )
> lines( F.ca2dm$apc, col=clx["Ca"]   , lend=1, lwd=4, lty="21" )
> text(rep(101,2), c((rev(F.ca2dm$apc$Age[,2])[1]+
+                     rev( F.w2dm$apc$Age[,2])[1])/2,
+                    (rev(F.dm2ca$apc$Age[,2])[1]+
+                     rev( F.w2ca$apc$Age[,2])[1])/2), c("DM inc.","Ca inc."), adj=0 )
> text( c(145,208), c(0.25,0.25), c("Cohort","Period") )
> mort.fr(sides=1,col.grid=gray(0.9))
> lines( F.w2dd$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( F.dm2dd$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( F.ca2dd$apc, col=clx["Ca"]   , lend=1, lwd=4 )
> lines( F.cd2dd$apc, col=clx["Ca-DM"], lend=1, lwd=4 )
> lines( F.dc2dd$apc, col=clx["DM-Ca"], lend=1, lwd=4 )
> text( c(145,208), c(2.5,2.5), c("Cohort","Period") )
> 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 )
> mtext( "Age", at=55, side=1, line=2 )
> mtext( "Calendar time", at=165, side=1, line=2 )
```

```
> 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 7.2 and 7.3 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.Rda" )
```

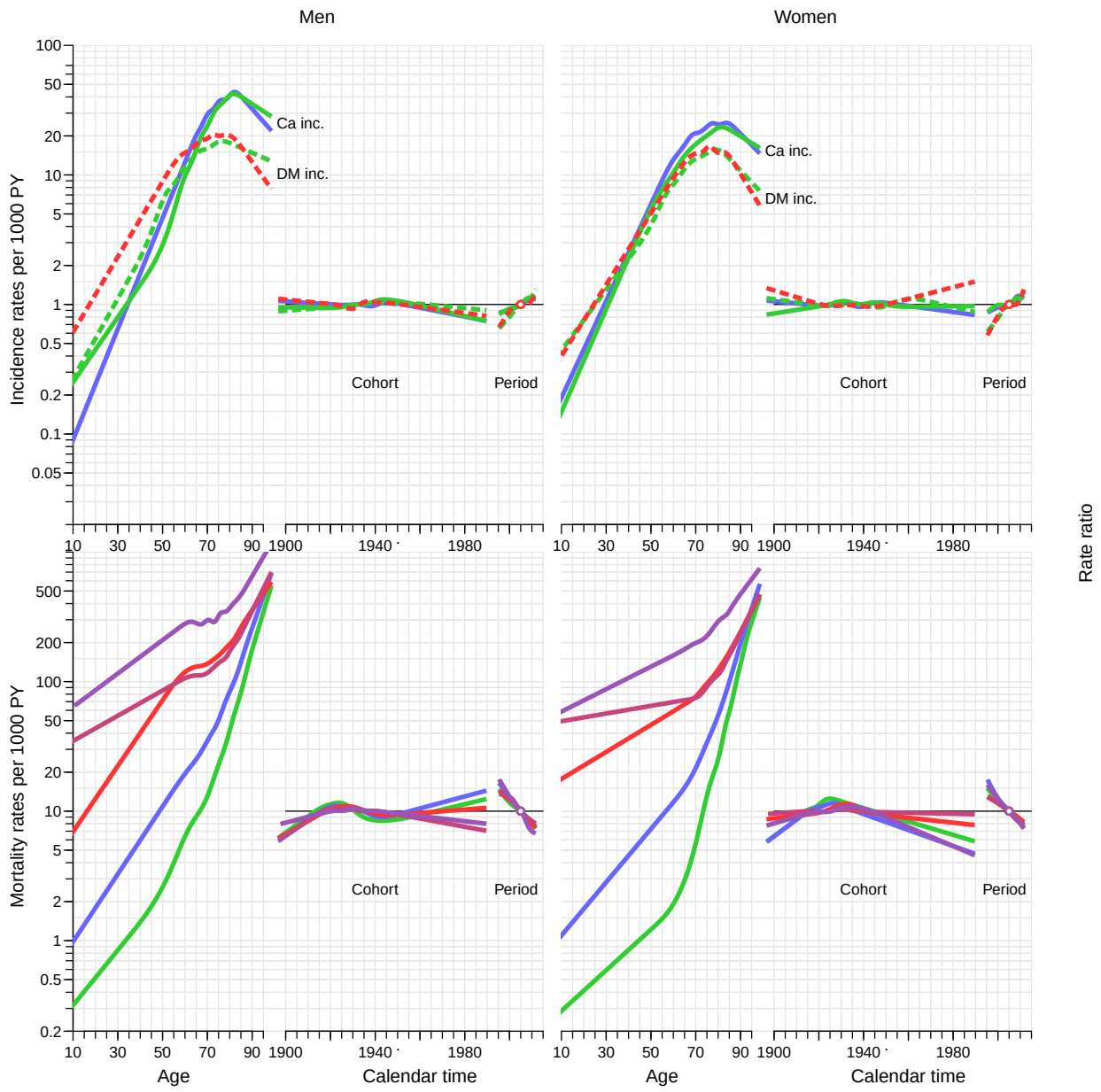

Figure 3.2: *Parameters from the fitted APC-models for rates using the period effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes 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.

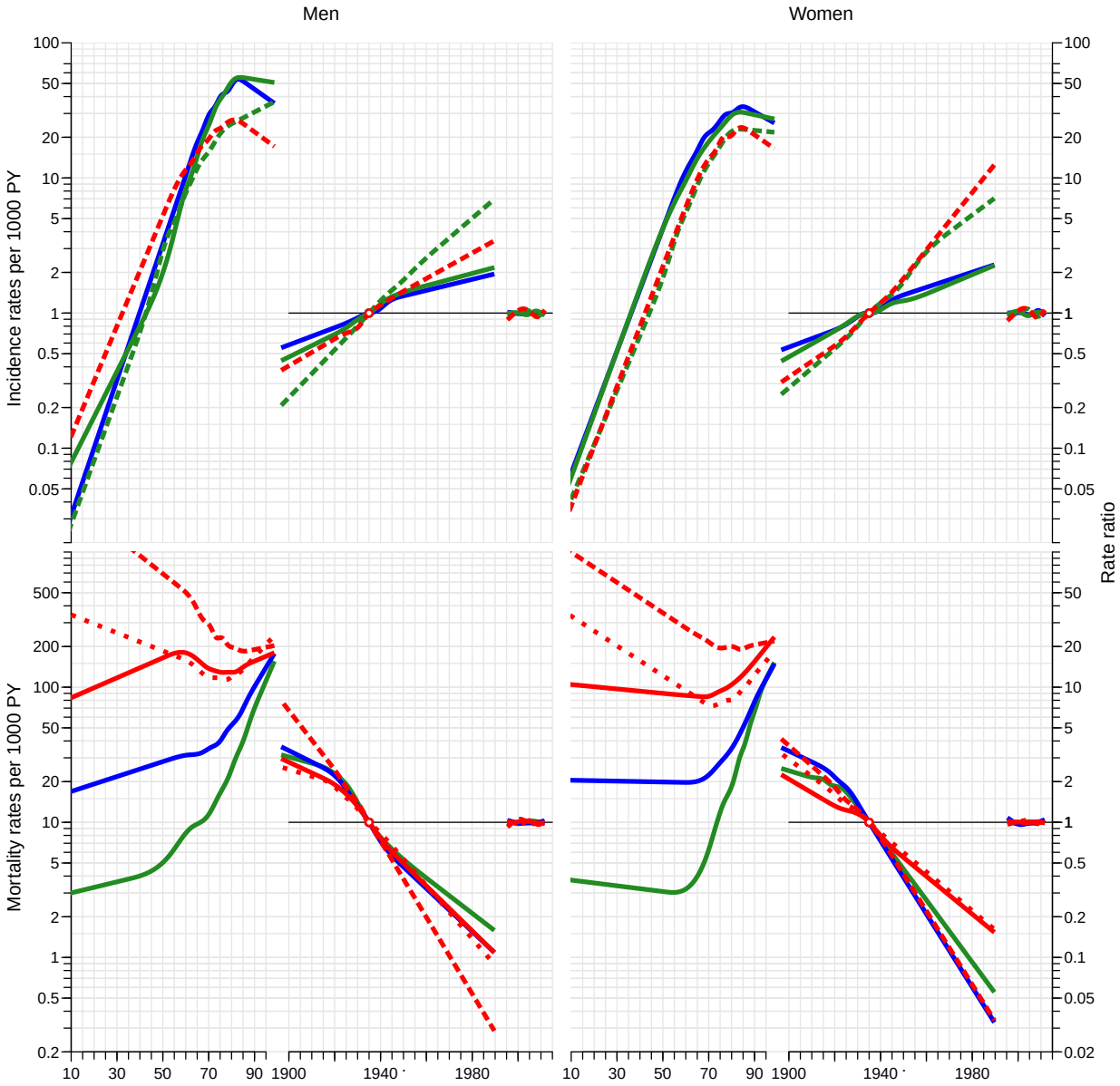

Figure 3.3: *Parameters from the fitted APC-models for rates using the cohort effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes 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           3.8  3.7  3.9   3.9  3.8  4.0
Ca to DM          2.7  2.4  3.1   3.9  3.6  4.3
W to Ca           2.0  1.9  2.1   1.6  1.5  1.7
DM to Ca          1.8  1.5  2.0   1.9  1.6  2.2
W to Dth         -3.9 -4.0 -3.8  -4.0 -4.1 -3.9
DM to Dth        -4.6 -4.8 -4.5  -4.7 -4.9 -4.5
Ca to Dth        -4.1 -4.2 -4.0  -2.7 -2.8 -2.6
DMCa to Dth      -5.9 -6.2 -5.6  -4.5 -4.8 -4.1
CaDM to Dth      -3.7 -4.1 -3.3  -3.2 -3.6 -2.8
```

```
> 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.

```
> library( Epi )
> library( splines )
> clear()
> options( width=130,
+ #         prompt=" ", continue=" ",
+          SweaveHooks=list( fig=function()
```

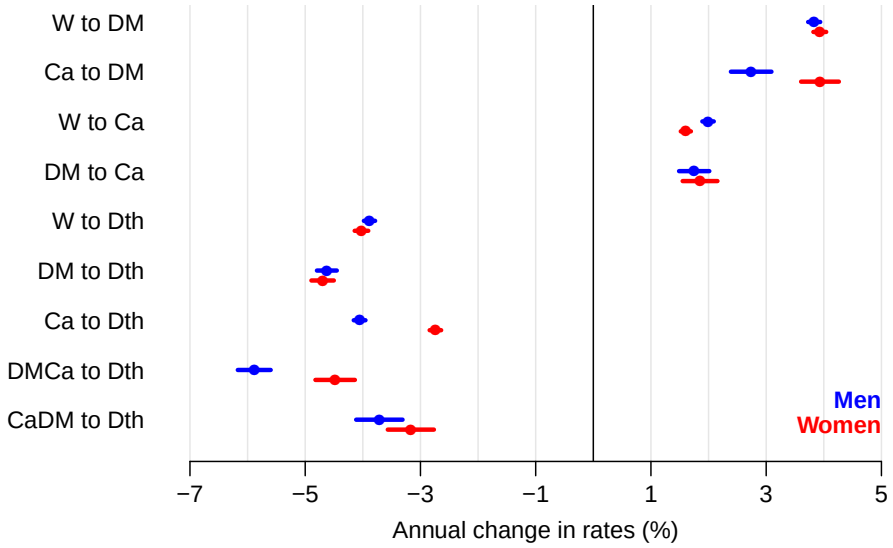

Figure 3.4: *Annual changes in the 9 incidence and mortality rates considered. Blue: men, red: women.*

```
+          par(mar=c(3,3,1,1),mgp=c(3,1,0)/1.6,bty="n",las=1) ) )
> load( file="./data/APC.Rda" )
> load( file="./data/dcd.Rda" )
> load( file="./data/cols.Rda" )
```

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 of 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$model, newdata=nd )[,1]
+ }
+ )
   user  system elapsed
  2.521   0.032   2.553
> for( bb in dimnames(CR)[[4]] )
+ {
+ nd <- data.frame( A=a.pt, P=as.numeric(bb)+a.pt, Y=int )
+
+ CR["Well" ,"DM"   ,,bb,"M"] <- ci.pred( M.w2dm$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="./data/rates.Rda" )
> load(         file="./data/rates.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 estimated age-specific 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 8.2 and 8.4 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.

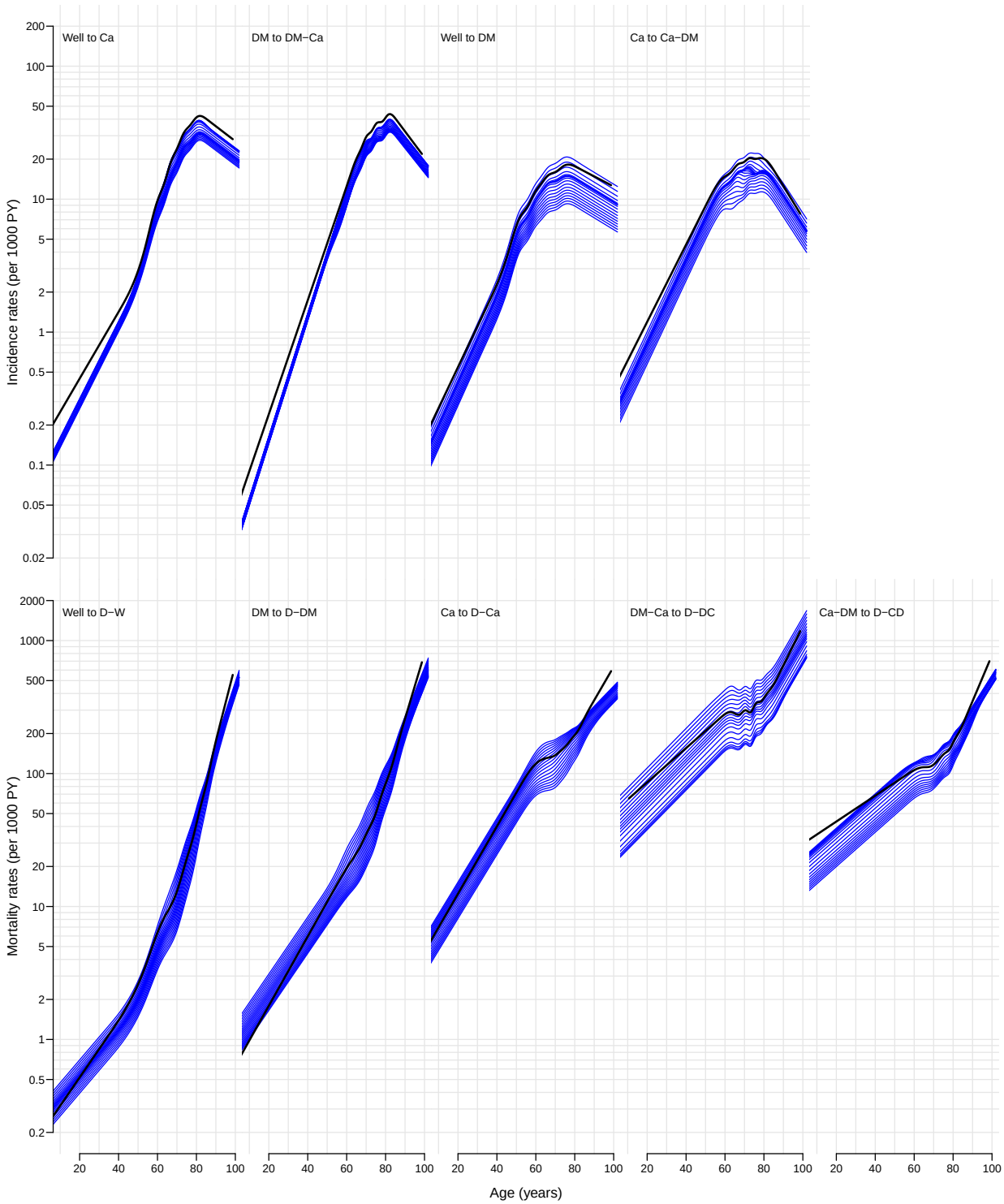

Figure 4.1: *Cross-sectional rates 1995–2012 for men, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

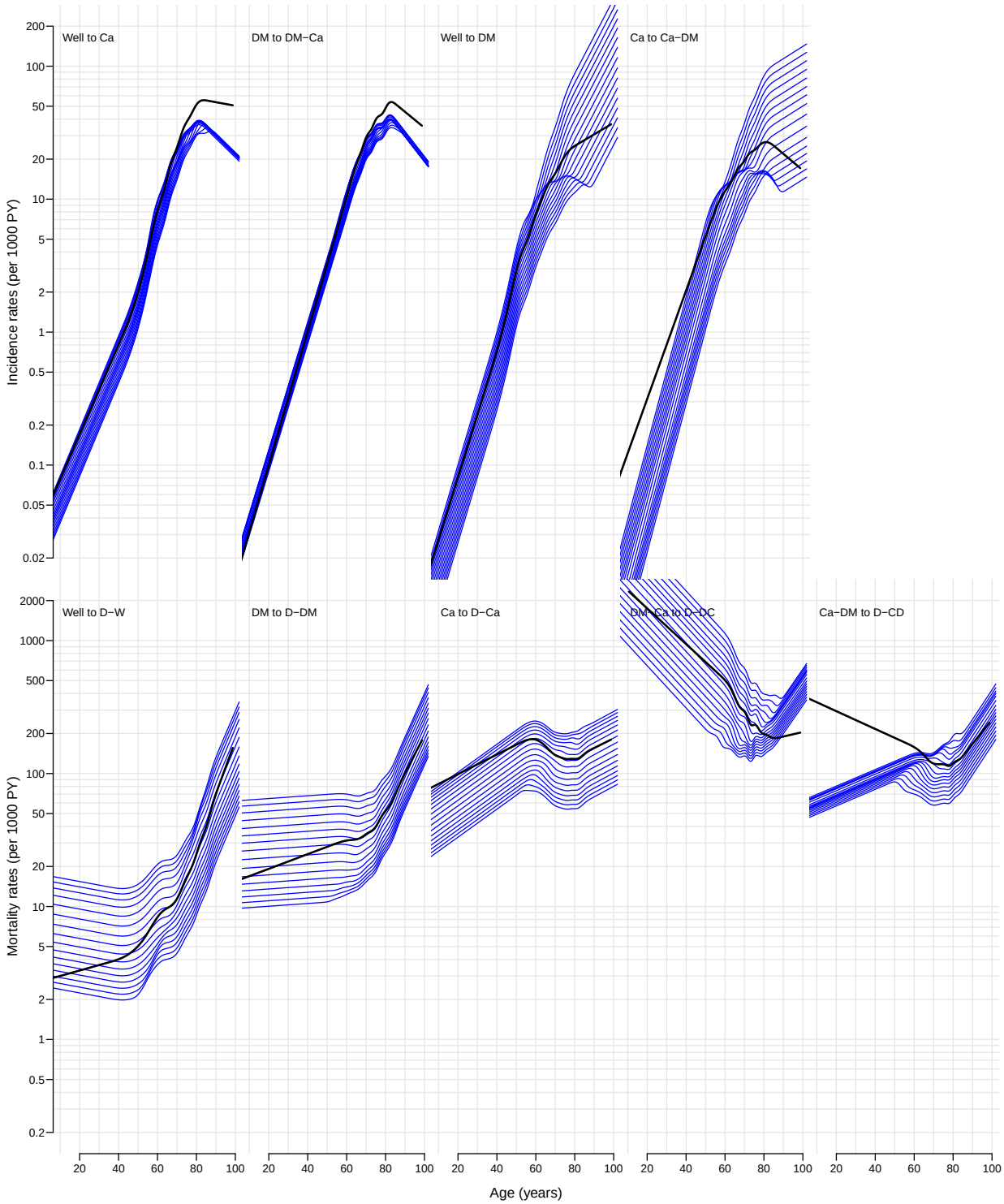

Figure 4.2: *Longitudinal rates 1930–1970 for men, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

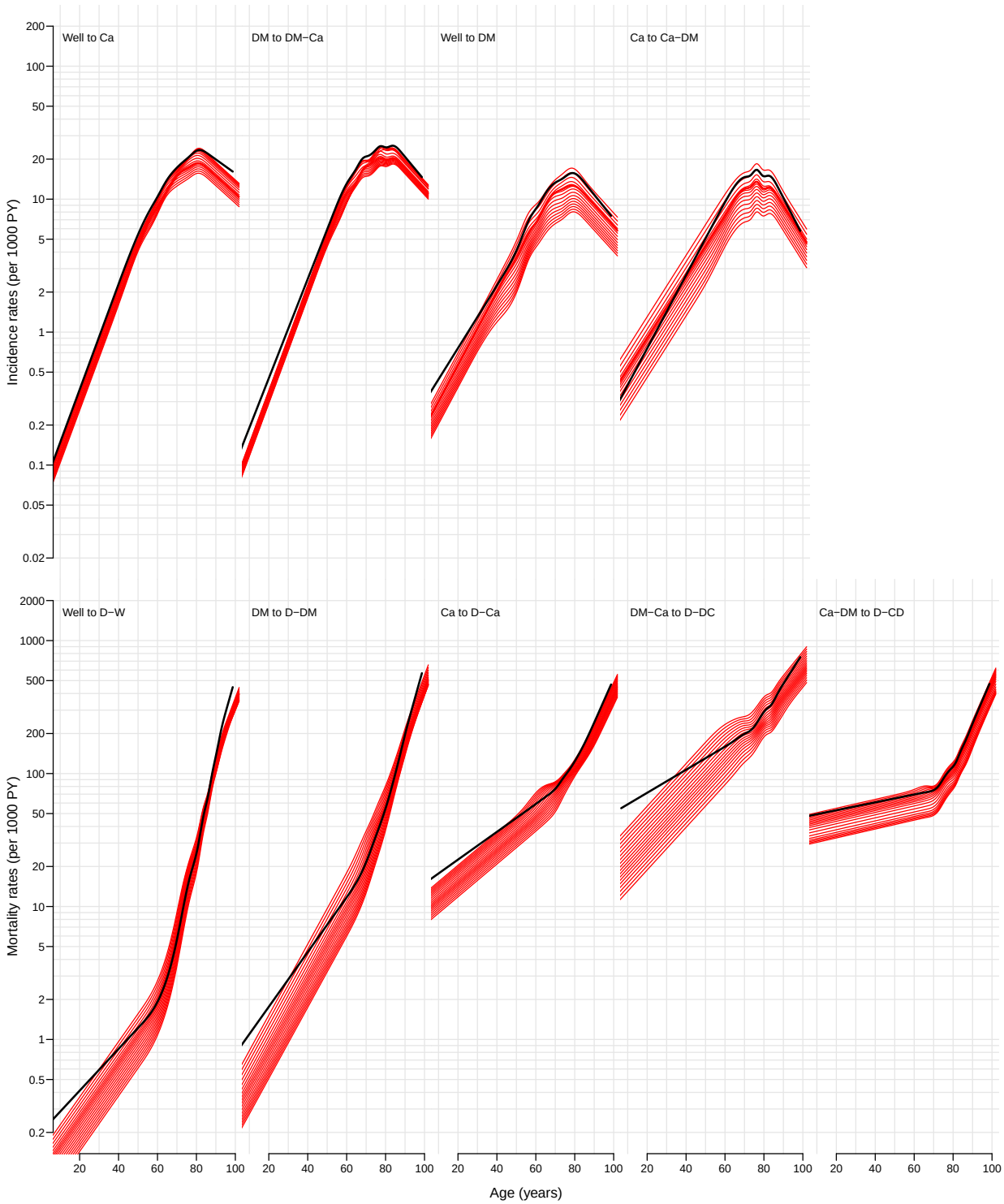

Figure 4.3: *Cross-sectional rates 1995–2012 for women, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

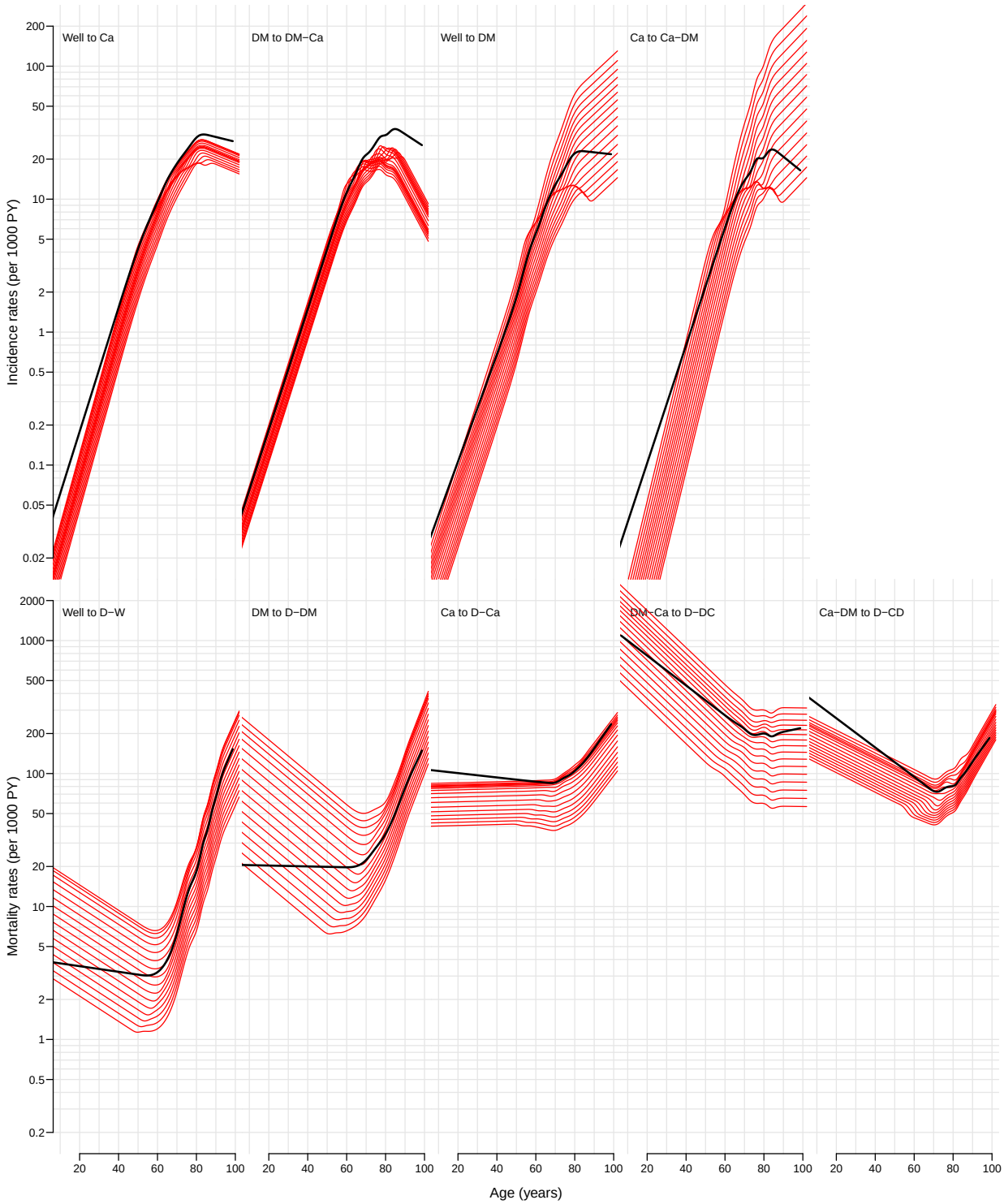

Figure 4.4: *Longitudinal rates 1930–1970 for women, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

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     .   8     .  13     .  14    .    .    .    .
  DM       .   .    16   .     .   .   41    .    .    .
  DM-Ca    .   .     .   .     .   .    .    .  427    .
  Ca       .   .     .   .     9   .    .  174    .    .
  Ca-DM    .   .     .   .     .   .    .    .    .  135
  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   9966     8     .    13     .    14     .     .     .     . 10000
  DM        .  9943    16     .     .     .    41     .     .     . 10000
  DM-Ca     .     .  9582     .     .     .     .     .   418     . 10000
  Ca        .     .     .  9819     9     .     .   172     .     . 10000
  Ca-DM     .     .     .     .  9866     .     .     .     .   134 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   9966     8     .    13     .    14     .     .     .     .
  DM        .  9943    16     .     .     .    41     .     .     .
  DM-Ca     .     .  9582     .     .     .     .     .   418     .
  Ca        .     .     .  9819     9     .     .   172     .     .
  Ca-DM     .     .     .     .  9866     .     .     .     .   134
  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"
> system.time(
+ 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]
+    } )
   user  system elapsed
  0.383   0.000   0.384
```

```
> str( PV )
 num [1:10, 1:1224, 1:18, 1:2] 1.00 6.28e-06 0.00 6.94e-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       51.0 21.2 33.6   5.8 49.4 19.8 35.8   5.3
1996       49.1 22.3 34.9   6.3 47.6 20.9 37.1   5.9
1997       47.3 23.5 36.2   7.0 45.8 21.9 38.4   6.5
1998       45.4 24.7 37.5   7.6 44.0 23.1 39.7   7.1
1999       43.6 25.9 38.7   8.3 42.4 24.2 40.9   7.9
2000       42.0 27.2 39.8   9.1 41.0 25.5 41.7   8.6
2001       40.5 28.6 40.7   9.8 40.0 26.8 42.1   9.3
2002       39.2 30.0 41.4  10.7 39.0 28.4 42.3  10.1
2003       37.8 31.5 42.2  11.5 38.0 30.0 42.4  10.8
2004       36.4 32.7 43.2  12.4 37.0 31.3 42.7  11.4
2005       35.0 33.5 44.6  13.2 35.9 31.9 43.6  11.9
2006       33.6 33.7 46.4  13.8 34.9 31.9 45.1  12.3
2007       32.2 34.0 48.3  14.5 33.8 31.7 46.7  12.8
2008       30.8 34.6 49.8  15.3 32.7 32.2 48.2  13.6
2009       29.3 36.1 50.8  16.4 31.3 33.5 49.3  14.6
2010       27.7 38.5 51.4  17.7 29.7 35.4 49.9  15.6
2011       26.1 41.3 51.8  19.3 28.1 37.8 50.2  16.8
2012       24.4 44.2 52.1  20.9 26.5 40.4 50.5  18.1
```

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

```
> 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 )
```

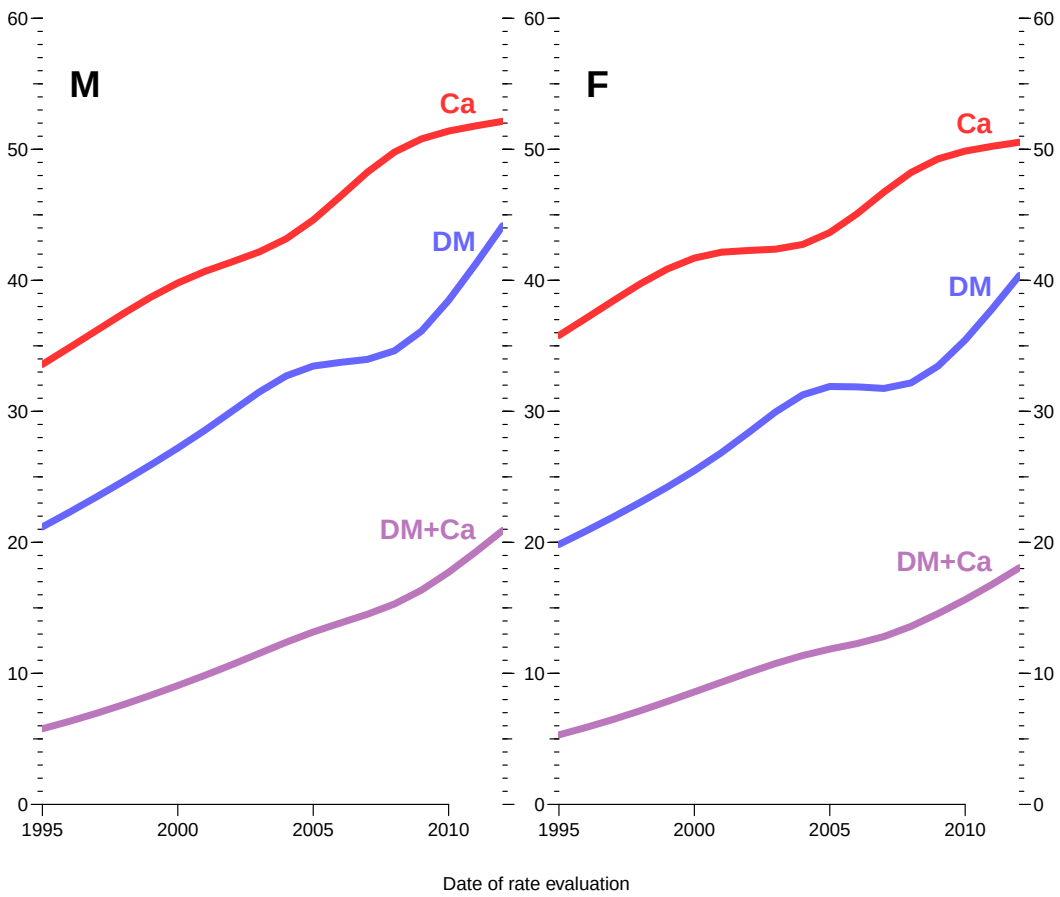

Figure 4.5: *Lifetime risk of DM (blue), cancer(red) and both diseases (purple) by year of rate calculation.*

For the corresponding cumulative plots we also define an array of cumulative lifetime probabilities over the states in the order: DM / DM+Ca / Ca / Well. 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
     state        D-DM  D-DC  D-CD  D-Ca   D-W        D-DM  D-DC  D-CD  D-Ca   D-W
per
1995         0.0  15.4  19.8  21.2  49.0 100.0   0.0  14.5  18.1  19.8  50.3 100.0
1996         0.0  16.0  20.7  22.3  50.8 100.0   0.0  15.0  18.9  20.9  52.1 100.0
1997         0.0  16.5  21.7  23.5  52.7 100.0   0.0  15.5  19.8  21.9  53.9 100.0
1998         0.0  17.1  22.7  24.7  54.5 100.0   0.0  15.9  20.7  23.1  55.6 100.0
1999         0.0  17.6  23.8  25.9  56.3 100.0   0.0  16.4  21.6  24.2  57.2 100.0
2000         0.0  18.1  24.8  27.2  57.9 100.0   0.0  16.9  22.6  25.5  58.6 100.0
2001         0.0  18.7  26.0  28.6  59.4 100.0   0.0  17.5  23.7  26.8  59.7 100.0
2002         0.0  19.3  27.3  30.0  60.7 100.0   0.0  18.3  25.1  28.4  60.6 100.0
2003         0.0  19.9  28.6  31.5  62.1 100.0   0.0  19.2  26.5  30.0  61.6 100.0
2004         0.0  20.3  29.6  32.7  63.5 100.0   0.0  19.9  27.6  31.3  62.6 100.0
2005         0.0  20.3  30.2  33.5  64.9 100.0   0.0  20.0  28.1  31.9  63.7 100.0
2006         0.0  19.9  30.2  33.7  66.3 100.0   0.0  19.6  27.9  31.9  64.7 100.0
2007         0.0  19.5  30.2  34.0  67.7 100.0   0.0  18.9  27.6  31.7  65.7 100.0
2008         0.0  19.3  30.5  34.6  69.1 100.0   0.0  18.6  27.8  32.2  66.8 100.0
2009         0.0  19.8  31.8  36.1  70.6 100.0   0.0  18.9  28.8  33.5  68.2 100.0
2010         0.0  20.8  33.8  38.5  72.1 100.0   0.0  19.8  30.4  35.4  69.7 100.0
2011         0.0  22.0  36.2  41.3  73.8 100.0   0.0  21.0  32.2  37.8  71.3 100.0
2012         0.0  23.3  38.7  44.2  75.4 100.0   0.0  22.3  34.2  40.4  72.9 100.0
```

In order to plot the cooresponding 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 )
```

We do the same thing, cunulating in a different order: Ca / DM+Ca / DM / Neither:

```
> LcP <- apply( PV[c("D-DM","D-DC","D-CD","D-Ca","D-W")[c(4:1,5)],nA,,], 2:3, cumsum )
> LcP <- LcP[c(1,1:5),,]
> LcP[1,,] <- 0
```

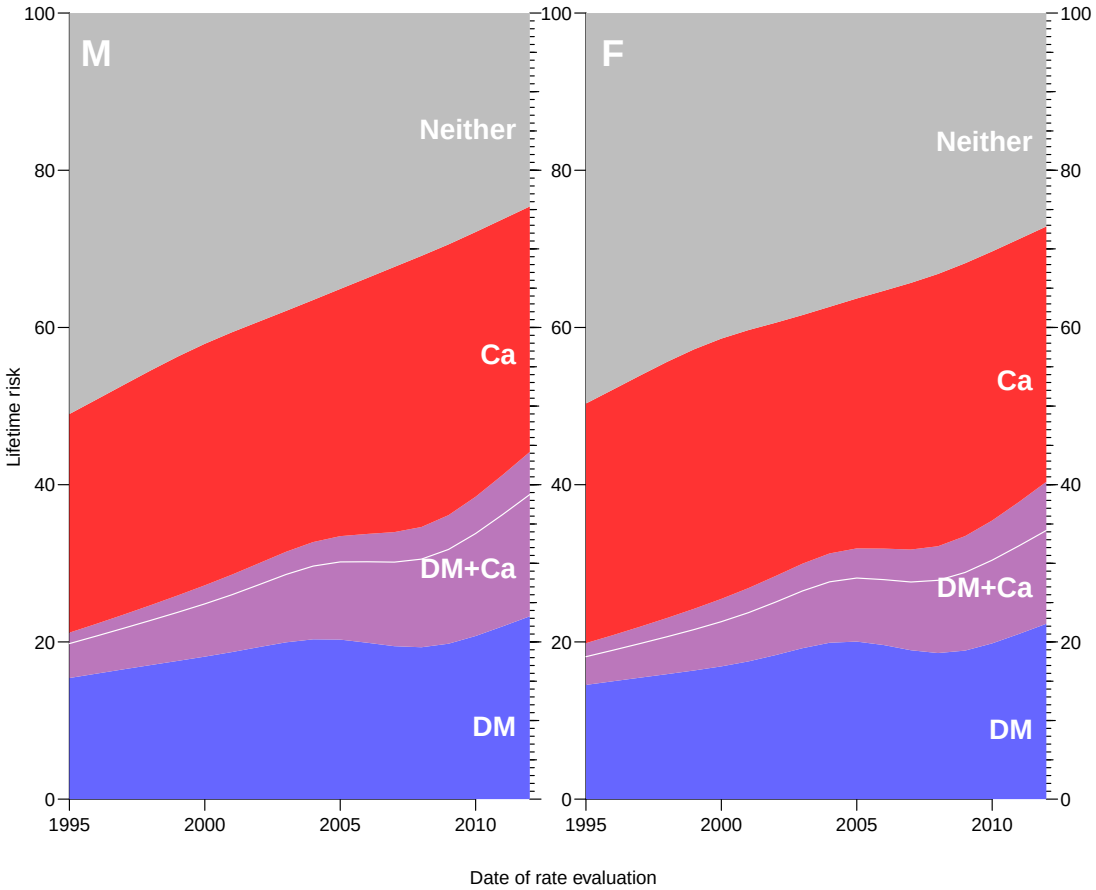

Figure 4.6: *Lifetime risk of DM and cancer by year of rate calculation.*

```
> LcP[6,,] <- 1 # Compensate the rounding errors
> dimnames( LcP )[[1]][1] <- ""
> ftable( round( LcP*100, 1 ), col.vars=c(3,1) )
     sex        M                                   F
     state         D-Ca  D-CD  D-DC  D-DM   D-W         D-Ca  D-CD  D-DC  D-DM   D-W
per
1995          0.0  27.8  29.2  33.6  49.0 100.0    0.0  30.5  32.2  35.8  50.3 100.0
1996          0.0  28.5  30.1  34.9  50.8 100.0    0.0  31.2  33.2  37.1  52.1 100.0
1997          0.0  29.2  31.0  36.2  52.7 100.0    0.0  32.0  34.1  38.4  53.9 100.0
1998          0.0  29.9  31.8  37.5  54.5 100.0    0.0  32.6  35.0  39.7  55.6 100.0
1999          0.0  30.4  32.5  38.7  56.3 100.0    0.0  33.0  35.7  40.9  57.2 100.0
2000          0.0  30.7  33.1  39.8  57.9 100.0    0.0  33.1  36.0  41.7  58.6 100.0
2001          0.0  30.8  33.4  40.7  59.4 100.0    0.0  32.8  35.9  42.1  59.7 100.0
2002          0.0  30.7  33.5  41.4  60.7 100.0    0.0  32.2  35.5  42.3  60.6 100.0
2003          0.0  30.6  33.5  42.2  62.1 100.0    0.0  31.6  35.1  42.4  61.6 100.0
2004          0.0  30.8  33.9  43.2  63.5 100.0    0.0  31.4  35.0  42.7  62.6 100.0
2005          0.0  31.4  34.7  44.6  64.9 100.0    0.0  31.8  35.6  43.6  63.7 100.0
2006          0.0  32.6  36.1  46.4  66.3 100.0    0.0  32.8  36.7  45.1  64.7 100.0
2007          0.0  33.7  37.6  48.3  67.7 100.0    0.0  33.9  38.1  46.7  65.7 100.0
2008          0.0  34.5  38.6  49.8  69.1 100.0    0.0  34.6  39.0  48.2  66.8 100.0
2009          0.0  34.4  38.8  50.8  70.6 100.0    0.0  34.7  39.3  49.3  68.2 100.0
2010          0.0  33.7  38.4  51.4  72.1 100.0    0.0  34.2  39.3  49.9  69.7 100.0
2011          0.0  32.5  37.6  51.8  73.8 100.0    0.0  33.4  39.0  50.2  71.3 100.0
2012          0.0  31.2  36.6  52.1  75.4 100.0    0.0  32.5  38.7  50.5  72.9 100.0

> 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")[c(3:1,4)],
+       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 )
```

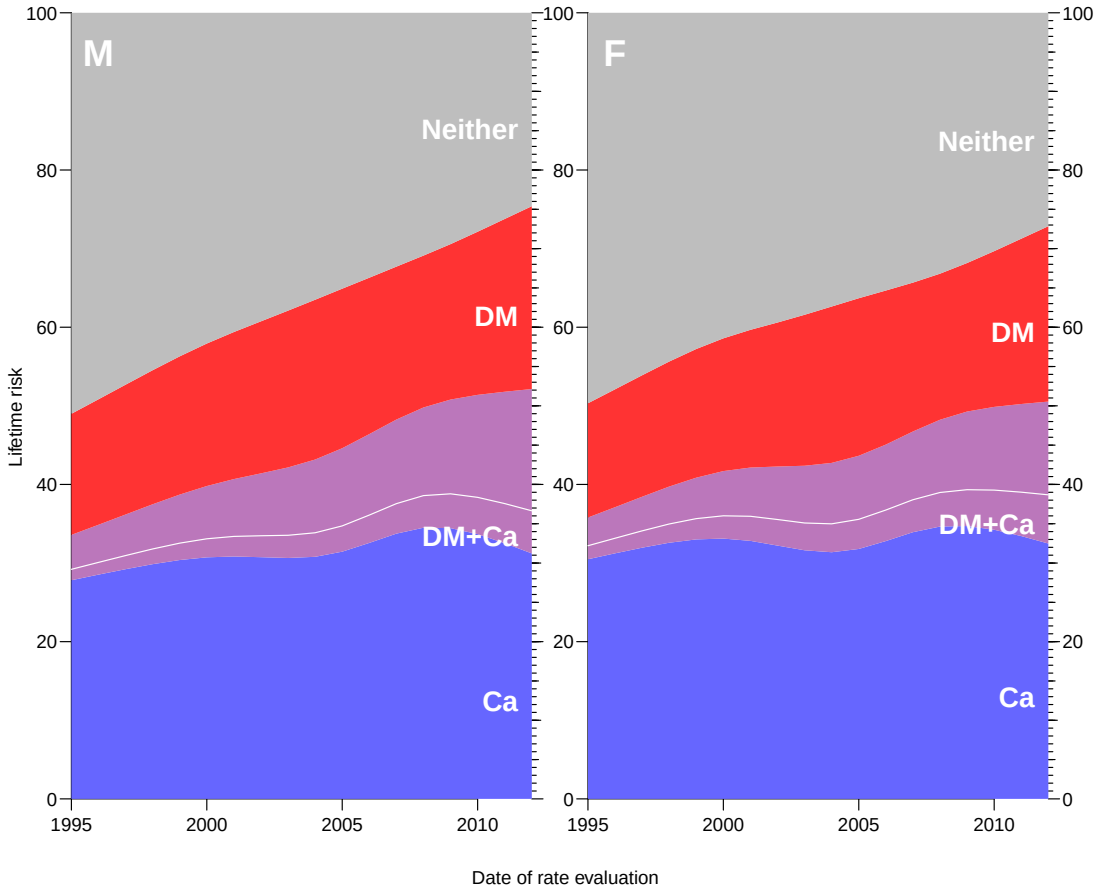

Figure 4.7: *Lifetime risk of DM and cancer by year of rate calculation.*

From the figures 8.5 and 8.6 we see that there is a dramatic 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 3.6     0     0 1.6 86.8 5.8  1.5    0  0.1  0.5
  50.125           3.6     0     0 1.6 86.7 5.9  1.5    0  0.1  0.5
  50.2083333333333 3.6     0     0 1.6 86.6 5.9  1.6    0  0.1  0.5
  50.2916666666667 3.7     0     0 1.6 86.6 5.9  1.6    0  0.1  0.5
  50.375           3.7     0     0 1.6 86.5 5.9  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 6.28e-06 0.00 6.94e-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 6.28e-06 6.28e-06 6.28e-06 1.32e-05 ...
 - attr(*, "dimnames")=List of 4
  ..$ state: 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 )
                  state
age                   DM DM-Ca Ca-DM  Ca Well  D-W D-Ca D-CD D-DC D-DM
  50.0416666666667 0 3.6   3.6   3.7 5.3 92.1 97.9 99.4 99.5 99.5  100
  50.125           0 3.6   3.6   3.7 5.3 92.0 97.9 99.4 99.5 99.5  100
  50.2083333333333 0 3.6   3.7   3.7 5.3 92.0 97.9 99.4 99.5 99.5  100
  50.2916666666667 0 3.7   3.7   3.7 5.4 91.9 97.8 99.4 99.5 99.5  100
  50.375           0 3.7   3.7   3.8 5.4 91.9 97.8 99.4 99.5 99.5  100


> crapl <- function( sc, aa, sx="M" ) # sc is the year of rate evaluation,
+                                     # aa the age
+    {
+ an <- aa*12
+ plot( NA, xlim=c(50,100), ylim=c(0,100),
+           xlab="Age", ylab="Probability (%)", xaxs="i", yaxs="i" )
+ csq <- clx[c("DM","DM-Ca","Ca-DM","Ca","Well","Dead","Ca","Ca-DM","DM-Ca","DM")]
+ if( aa>95 ) csq[1:5] <- "transparent"
+ for( i in 1:10)
+ rect( aa-1, cPV[i  ,an,sc,sx]*100,
+       aa+1, cPV[i+1,an,sc,sx]*100,
+       col = csq[i], border="transparent")
+ segments( aa-1, cPV[6,an,sc,sx]*100,
```

```
+           aa+1, cPV[6,an,sc,sx]*100, lwd=2 )
+ pm <- ( aa<80 ) - ( aa>=80 )
+ text( rep(aa+pm*4,10), seq(5,95,,10),
+       c("DM","DM-Ca","Ca-DM","Ca","Well","Dead(W)","D(Ca)","D(Ca-DM)","D(DM-Ca)","D(DM)"),
+       col=csq, cex=1.1, font=2, adj=(1-pm)/2 )
+ segments( rep(aa+pm*1.0,10), (cPV[1:10,an,sc,sx]+cPV[1:10+1,an,sc,sx])/2*100,
+           rep(aa+pm*3.8), seq(5,95,,10), col=csq, lwd=2 )
+ }
> for( a in seq(55,100,5) )
+    {
+ pdf( paste("./graph/demo-crh-",a,".pdf",sep=""),
+      height=5, width=6 )
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
+ crapl( sc="2010", aa=a )
+ dev.off()
+    }
```

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

```
> aa <- as.numeric( dimnames(CR)[["age"]] )
> nul <- aa * 0
> crpl <- function( sc ) # sc is the year of rate evaluation
+    {
+ 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 = clx["DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[2,,sc,sx],
+                        rev(cPV[3,,sc,sx]))*100,
+          col = clx["DM-Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[3,,sc,sx],
+                        rev(cPV[4,,sc,sx]))*100,
+          col = clx["Ca-DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[4,,sc,sx],
+                        rev(cPV[5,,sc,sx]))*100,
+          col = clx["Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[5,,sc,sx],
+                        rev(cPV[6,,sc,sx]))*100,
+          col = clx["Well"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[6,,sc,sx],
+                        rev(cPV[7,,sc,sx]))*100,
+          col = clx["Dead"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[7,,sc,sx],
+                        rev(cPV[8,,sc,sx]))*100,
+          col = clx["Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[8,,sc,sx],
+                        rev(cPV[9,,sc,sx]))*100,
+          col = clx["Ca-DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[9,,sc,sx],
+                        rev(cPV[10,,sc,sx]))*100,
+          col = clx["DM-Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[10,,sc,sx],
+                        rev(cPV[11,,sc,sx]))*100,
+          col = clx["DM"], border="transparent")
+ matlines( aa, 100*cPV[6,,sc,sx],
+           lty=1, col="black", lwd=3, 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( "demo-film.pdf", width=11, height=8 )
> for( sc in dimnames(cPV)[[3]] ) crpl( sc )
> dev.off()
```

```
pdf
  2
```

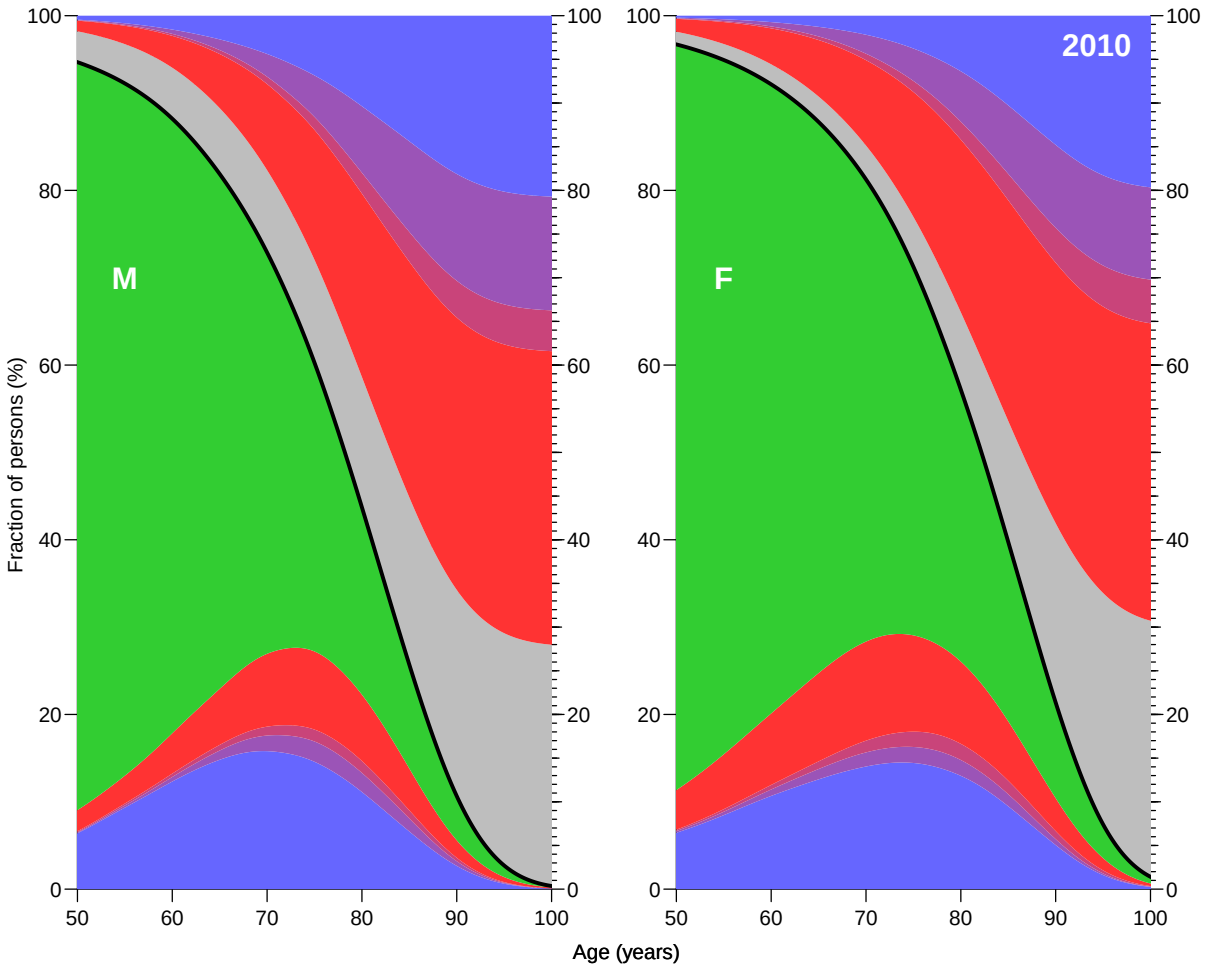

Figure 4.8: *Occupation probabilities for the various states in figure 1 at various ages, assuming all start in "Well" at age 0. Based on cross-sectional rates from 2010.*
The thick black line is the overall survival curve, with "Dead" states are above and "Alive" below the line. The blue states are 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.

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 these 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* on 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] <-
```

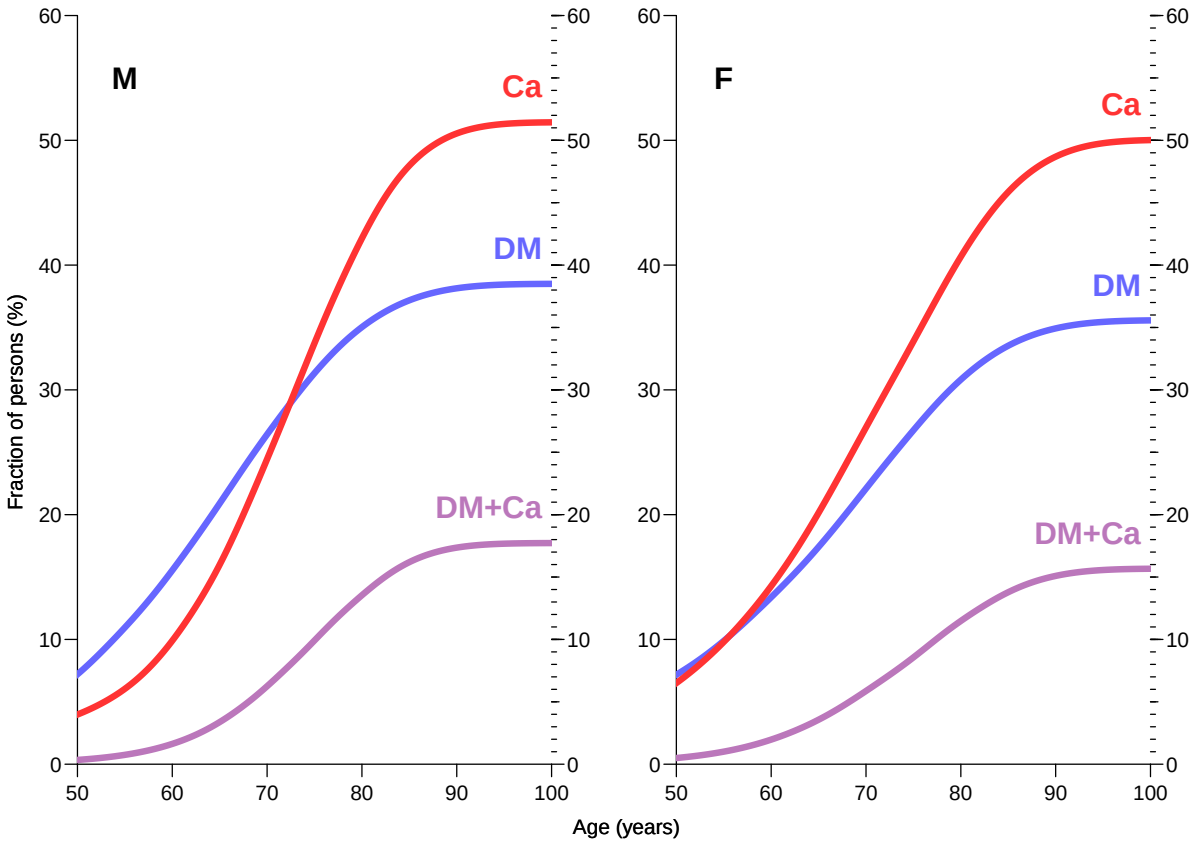

Figure 4.9: *Cumulative risk of acquiring DM, cancer or both before a given age, using rates as of 2010.*

```
+     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 make the same plot as above, calssified by sex, conditioning age (50, 60, 70) and state conditioned on (DM/no DM), in total 12 combinations:

```
> 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:

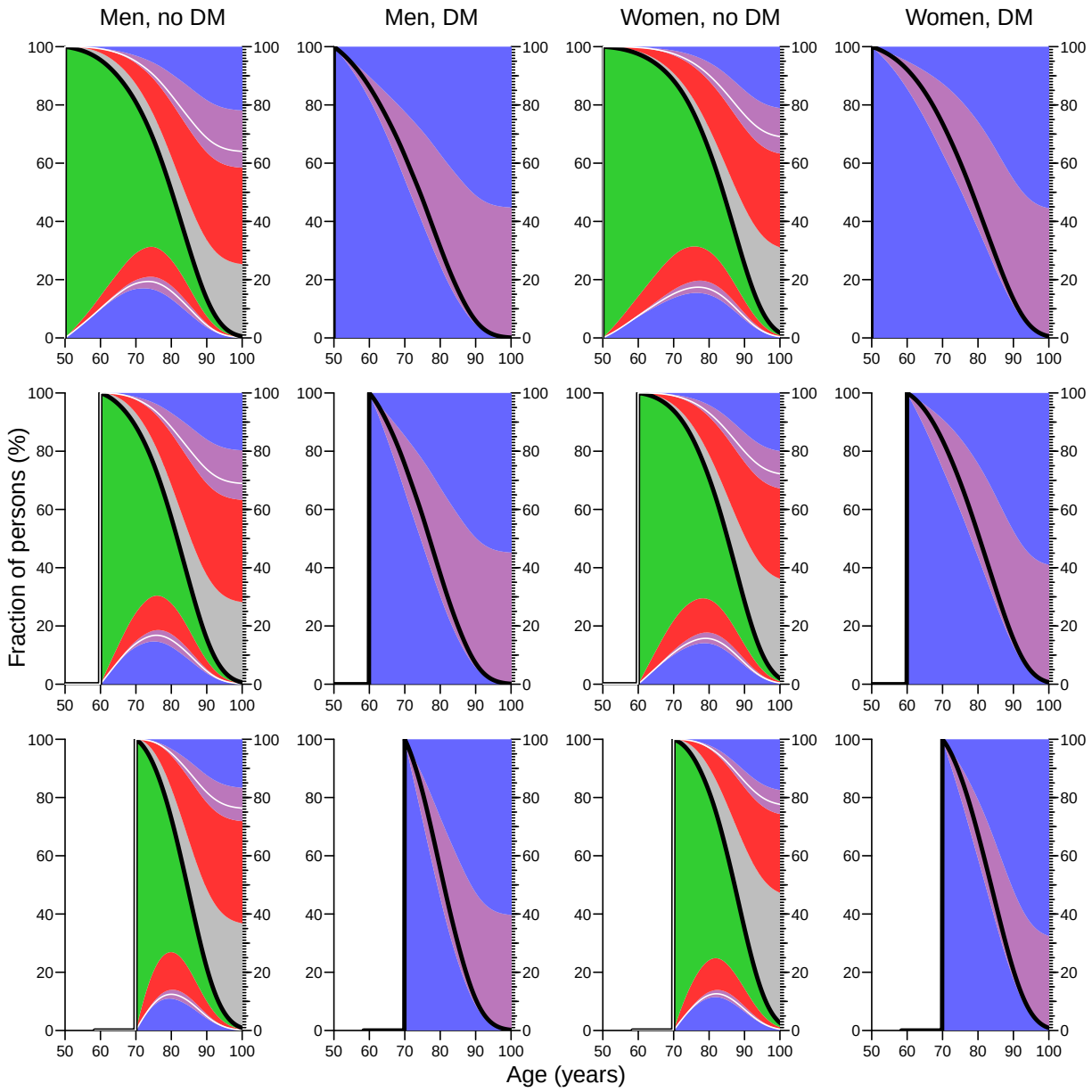

Figure 4.10: *Plots of state occupancies conditional on being either well or diabetic at different ages. Based on cross-sectional rates as of 2012-01-01.*

```
> 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 6.28e-06 0.00 6.94e-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"
> str( LRp )
 num [1:10, 1:18, 1:2, 1:7] 2.21e-04 1.54e-05 2.55e-07 9.41e-05 5.16e-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"
  ..$ : NULL
> 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] 2.21e-04 1.54e-05 2.55e-07 9.41e-05 5.16e-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 51.0 15.4 27.8  4.4  1.4
       W-50    0.0  0.0   0.0  0.0   0.0 52.0 14.1 28.5  4.0  1.4
       W-60    0.0  0.0   0.0  0.0   0.0 55.2 11.8 28.5  3.1  1.4
       W-70    0.0  0.0   0.0  0.0   0.0 62.2  8.7 26.1  1.8  1.1
       DM-50   0.0  0.0   0.0  0.0   0.0  0.0 75.9  0.0 24.1  0.0
       DM-60   0.0  0.0   0.0  0.0   0.0  0.0 75.3  0.0 24.7  0.0
       DM-70   0.0  0.0   0.0  0.0   0.0  0.0 77.6  0.0 22.4  0.0
  1996 0       0.0  0.0   0.0  0.0   0.0 49.1 16.0 28.5  4.8  1.6
       W-50    0.0  0.0   0.0  0.0   0.0 50.1 14.6 29.3  4.3  1.6
       W-60    0.0  0.0   0.0  0.0   0.0 53.4 12.3 29.3  3.4  1.5
       W-70    0.0  0.0   0.0  0.0   0.0 60.8  9.1 26.8  2.0  1.3
       DM-50   0.0  0.0   0.0  0.0   0.0  0.0 74.7  0.0 25.3  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 76.8  0.0 23.2  0.0
  1997 0       0.0  0.0   0.0  0.0   0.0 47.3 16.5 29.2  5.2  1.7
       W-50    0.0  0.0   0.0  0.0   0.0 48.3 15.2 30.0  4.7  1.7
       W-60    0.0  0.0   0.0  0.0   0.0 51.6 12.8 30.1  3.7  1.7
       W-70    0.1  0.0   0.0  0.0   0.0 59.4  9.5 27.5  2.2  1.4
       DM-50   0.0  0.0   0.0  0.0   0.0  0.0 73.4  0.0 26.6  0.0
       DM-60   0.0  0.0   0.0  0.0   0.0  0.0 73.1  0.0 26.9  0.0
       DM-70   0.0  0.0   0.0  0.0   0.0  0.0 75.9  0.0 24.1  0.0
  1998 0       0.0  0.0   0.0  0.0   0.0 45.4 17.1 29.9  5.7  1.9
       W-50    0.0  0.0   0.0  0.0   0.0 46.4 15.7 30.7  5.1  2.0
       W-60    0.0  0.0   0.0  0.0   0.0 49.8 13.3 30.9  4.0  1.9
       W-70    0.1  0.0   0.0  0.0   0.0 57.9  9.9 28.3  2.3  1.5
       DM-50   0.0  0.0   0.0  0.0   0.0  0.0 72.2  0.0 27.8  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 75.0  0.0 25.0  0.0
  1999 0       0.0  0.0   0.0  0.0   0.0 43.6 17.6 30.4  6.2  2.1
       W-50    0.0  0.0   0.0  0.0   0.0 44.6 16.2 31.4  5.6  2.2
       W-60    0.1  0.0   0.0  0.0   0.0 48.1 13.9 31.5  4.4  2.1
       W-70    0.1  0.0   0.0  0.0   0.0 56.4 10.3 29.0  2.5  1.7
       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 70.8  0.0 29.2  0.0
       DM-70   0.0  0.0   0.0  0.0   0.0  0.0 74.1  0.0 25.9  0.0
  2000 0       0.0  0.0   0.0  0.0   0.0 42.0 18.1 30.7  6.7  2.4
       W-50    0.0  0.0   0.0  0.0   0.0 43.0 16.7 31.8  6.0  2.4
       W-60    0.1  0.0   0.0  0.0   0.0 46.4 14.4 32.1  4.8  2.3
       W-70    0.1  0.0   0.0  0.0   0.0 55.0 10.8 29.5  2.8  1.8
       DM-50   0.0  0.0   0.0  0.0   0.0  0.0 69.7  0.0 30.3  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 73.2  0.0 26.8  0.0
  2001 0       0.0  0.0   0.0  0.0   0.0 40.5 18.7 30.8  7.3  2.6
```

	W-50	0.0	0.0	0.0	0.0	0.0	41.5	17.3	32.0	6.6	2.6
	W-60	0.1	0.0	0.0	0.0	0.0	45.0	15.0	32.3	5.2	2.5
	W-70	0.1	0.0	0.0	0.0	0.0	53.7	11.3	29.9	3.0	2.0
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	68.4	0.0	31.6	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	68.4	0.0	31.6	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	72.2	0.0	27.8	0.0
2002	0	0.0	0.0	0.0	0.0	0.0	39.2	19.3	30.7	7.9	2.7
	W-50	0.0	0.0	0.0	0.0	0.0	40.1	17.9	31.9	7.1	2.8
	W-60	0.1	0.0	0.0	0.0	0.0	43.6	15.6	32.4	5.7	2.7
	W-70	0.1	0.0	0.0	0.0	0.0	52.5	11.8	30.2	3.3	2.1
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	67.2	0.0	32.8	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	67.2	0.0	32.8	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	71.2	0.0	28.8	0.0
2003	0	0.0	0.0	0.0	0.0	0.0	37.8	19.9	30.6	8.6	2.9
	W-50	0.0	0.0	0.0	0.0	0.0	38.7	18.5	31.9	7.8	2.9
	W-60	0.1	0.0	0.0	0.0	0.0	42.2	16.2	32.6	6.2	2.8
	W-70	0.1	0.0	0.0	0.0	0.0	51.2	12.4	30.4	3.6	2.2
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	65.9	0.0	34.1	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	65.9	0.0	34.1	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	70.1	0.0	29.9	0.0
2004	0	0.0	0.0	0.0	0.0	0.0	36.4	20.3	30.8	9.3	3.1
	W-50	0.0	0.0	0.0	0.0	0.0	37.3	18.9	32.2	8.4	3.1
	W-60	0.1	0.0	0.0	0.0	0.0	40.7	16.6	32.9	6.7	3.0
	W-70	0.1	0.0	0.0	0.0	0.0	49.9	12.8	30.9	3.9	2.3
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	64.5	0.0	35.5	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	64.5	0.0	35.5	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	69.0	0.0	31.0	0.0
2005	0	0.0	0.0	0.0	0.0	0.0	35.0	20.3	31.4	9.9	3.3
	W-50	0.0	0.0	0.0	0.0	0.0	35.8	18.9	32.9	8.9	3.4
	W-60	0.1	0.0	0.0	0.0	0.0	39.1	16.6	33.8	7.1	3.2
	W-70	0.1	0.0	0.0	0.0	0.0	48.3	13.0	31.9	4.2	2.5
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	63.1	0.0	36.9	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	63.0	0.0	37.0	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	67.7	0.0	32.3	0.0
2006	0	0.0	0.0	0.0	0.0	0.0	33.6	19.9	32.6	10.3	3.5
	W-50	0.0	0.0	0.0	0.0	0.0	34.3	18.6	34.1	9.3	3.6
	W-60	0.1	0.0	0.0	0.0	0.0	37.5	16.4	35.1	7.4	3.5
	W-70	0.1	0.0	0.0	0.1	0.0	46.6	12.9	33.3	4.4	2.7
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	61.5	0.0	38.4	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	61.4	0.0	38.6	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	66.3	0.0	33.6	0.0
2007	0	0.0	0.0	0.0	0.0	0.0	32.2	19.5	33.7	10.7	3.8
	W-50	0.0	0.0	0.0	0.0	0.0	32.8	18.2	35.4	9.7	3.9
	W-60	0.1	0.0	0.0	0.0	0.0	35.8	16.1	36.5	7.7	3.8
	W-70	0.1	0.0	0.0	0.1	0.0	44.8	12.8	34.7	4.6	2.9
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	60.0	0.0	39.9	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	59.9	0.0	40.1	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	65.0	0.0	35.0	0.0
2008	0	0.0	0.0	0.0	0.0	0.0	30.8	19.3	34.5	11.2	4.1
	W-50	0.0	0.0	0.0	0.0	0.0	31.3	18.1	36.2	10.1	4.2
	W-60	0.1	0.0	0.0	0.1	0.0	34.3	16.0	37.4	8.1	4.1
	W-70	0.1	0.0	0.0	0.1	0.0	43.1	12.9	35.8	4.8	3.1
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	58.7	0.0	41.3	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	58.4	0.0	41.5	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	63.8	0.0	36.2	0.0
2009	0	0.0	0.0	0.0	0.0	0.0	29.3	19.8	34.4	12.0	4.4
	W-50	0.1	0.0	0.0	0.1	0.0	29.8	18.5	36.2	10.8	4.5
	W-60	0.1	0.0	0.0	0.1	0.0	32.7	16.5	37.6	8.7	4.4
	W-70	0.1	0.0	0.0	0.1	0.0	41.5	13.4	36.3	5.2	3.4
	DM-50	0.0	0.0	0.0	0.0	0.0	0.0	57.6	0.0	42.4	0.0
	DM-60	0.0	0.0	0.0	0.0	0.0	0.0	57.3	0.0	42.7	0.0
	DM-70	0.0	0.0	0.0	0.0	0.0	0.0	62.8	0.0	37.2	0.0
2010	0	0.1	0.0	0.0	0.0	0.0	27.7	20.8	33.7	13.0	4.7
	W-50	0.1	0.0	0.0	0.1	0.0	28.2	19.5	35.5	11.8	4.8
	W-60	0.1	0.0	0.0	0.1	0.0	31.1	17.4	37.1	9.5	4.8
	W-70	0.1	0.0	0.0	0.1	0.0	39.8	14.3	36.2	5.8	3.7

		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	56.7	0.0	43.3	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	56.3	0.0	43.6	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	61.9	0.0	38.0	0.0
	2011	0	0.1	0.0	0.0	0.1	0.0	26.1	22.0	32.5	14.2	5.0
		W-50	0.1	0.0	0.0	0.1	0.0	26.6	20.7	34.4	12.9	5.2
		W-60	0.1	0.0	0.0	0.1	0.0	29.4	18.5	36.2	10.4	5.2
		W-70	0.1	0.0	0.0	0.1	0.0	38.0	15.5	35.8	6.4	4.1
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	55.9	0.0	44.0	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	55.6	0.0	44.4	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	61.2	0.0	38.8	0.0
	2012	0	0.1	0.0	0.0	0.1	0.0	24.4	23.3	31.2	15.5	5.4
		W-50	0.1	0.0	0.0	0.1	0.0	25.0	21.9	33.2	14.1	5.6
		W-60	0.1	0.0	0.0	0.1	0.0	27.8	19.8	35.2	11.4	5.7
		W-70	0.1	0.0	0.0	0.1	0.0	36.2	16.7	35.2	7.0	4.5
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	55.3	0.0	44.7	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	54.8	0.0	45.1	0.0
		DM-70	0.0	0.1	0.0	0.0	0.0	0.0	60.4	0.0	39.5	0.0
F	1995	0	0.2	0.0	0.0	0.0	0.0	49.4	14.5	30.5	3.6	1.7
		W-50	0.2	0.0	0.0	0.0	0.0	52.4	13.3	29.4	2.9	1.7
		W-60	0.3	0.0	0.0	0.0	0.0	57.1	12.1	26.6	2.4	1.5
		W-70	0.3	0.0	0.0	0.0	0.0	66.8	9.6	20.8	1.4	1.0
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	74.3	0.0	25.7	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	76.2	0.0	23.8	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	81.4	0.0	18.6	0.0
	1996	0	0.2	0.0	0.0	0.0	0.0	47.6	15.0	31.2	3.9	1.9
		W-50	0.3	0.0	0.0	0.0	0.0	50.7	13.7	30.2	3.2	1.9
		W-60	0.3	0.0	0.0	0.0	0.0	55.5	12.6	27.4	2.6	1.6
		W-70	0.4	0.0	0.0	0.0	0.0	65.4	10.0	21.5	1.6	1.1
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	72.9	0.0	27.1	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	75.0	0.0	25.0	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	80.5	0.0	19.4	0.0
	1997	0	0.2	0.0	0.0	0.0	0.0	45.8	15.5	32.0	4.3	2.1
		W-50	0.3	0.0	0.0	0.0	0.0	48.9	14.1	30.9	3.5	2.1
		W-60	0.3	0.0	0.0	0.0	0.0	53.8	13.0	28.2	2.8	1.8
		W-70	0.4	0.0	0.0	0.1	0.0	63.9	10.4	22.3	1.7	1.2
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	71.5	0.0	28.5	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	73.7	0.0	26.2	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	79.6	0.0	20.4	0.0
	1998	0	0.3	0.0	0.0	0.0	0.0	44.0	15.9	32.6	4.8	2.4
		W-50	0.3	0.0	0.0	0.0	0.0	47.2	14.6	31.6	3.9	2.3
		W-60	0.4	0.0	0.0	0.0	0.0	52.1	13.5	28.9	3.1	2.0
		W-70	0.5	0.0	0.0	0.1	0.0	62.4	10.8	23.0	1.9	1.3
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	70.0	0.0	30.0	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	72.5	0.0	27.5	0.0
		DM-70	0.0	0.0	0.0	0.0	0.0	0.0	78.6	0.0	21.3	0.0
	1999	0	0.3	0.0	0.0	0.0	0.0	42.4	16.4	33.0	5.2	2.6
		W-50	0.3	0.0	0.0	0.0	0.0	45.7	15.0	32.1	4.2	2.6
		W-60	0.4	0.0	0.0	0.1	0.0	50.6	13.9	29.4	3.4	2.2
		W-70	0.5	0.0	0.0	0.1	0.0	61.0	11.2	23.6	2.0	1.5
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	68.6	0.0	31.4	0.0
		DM-60	0.0	0.0	0.0	0.0	0.0	0.0	71.3	0.0	28.7	0.0
		DM-70	0.0	0.1	0.0	0.0	0.0	0.0	77.6	0.0	22.3	0.0
	2000	0	0.3	0.0	0.0	0.0	0.0	41.0	16.9	33.1	5.7	2.9
		W-50	0.3	0.0	0.0	0.1	0.0	44.3	15.5	32.3	4.6	2.8
		W-60	0.4	0.0	0.0	0.1	0.0	49.3	14.5	29.6	3.7	2.4
		W-70	0.5	0.0	0.0	0.1	0.0	59.8	11.7	24.0	2.2	1.6
		DM-50	0.0	0.0	0.0	0.0	0.0	0.0	67.2	0.0	32.7	0.0
		DM-60	0.0	0.1	0.0	0.0	0.0	0.0	70.1	0.0	29.9	0.0
		DM-70	0.0	0.1	0.0	0.0	0.0	0.0	76.7	0.0	23.2	0.0
	2001	0	0.3	0.0	0.0	0.0	0.0	40.0	17.5	32.8	6.2	3.1
		W-50	0.3	0.0	0.0	0.1	0.0	43.3	16.1	32.1	5.0	3.0
		W-60	0.4	0.0	0.0	0.1	0.0	48.3	15.1	29.5	4.1	2.6
		W-70	0.5	0.1	0.0	0.1	0.0	58.9	12.2	24.1	2.4	1.8
		DM-50	0.0	0.1	0.0	0.0	0.0	0.0	66.1	0.0	33.9	0.0
		DM-60	0.0	0.1	0.0	0.0	0.0	0.0	69.1	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

2002	0	0.3	0.0	0.0	0.1	0.0	39.0	18.3	32.2	6.7	3.3
	W-50	0.3	0.0	0.0	0.1	0.0	42.4	16.9	31.6	5.5	3.2
	W-60	0.4	0.1	0.0	0.1	0.0	47.4	15.8	29.1	4.4	2.8
	W-70	0.5	0.1	0.0	0.1	0.0	58.1	12.9	23.9	2.6	1.9
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	65.1	0.0	34.9	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	75.1	0.0	24.8	0.0
2003	0	0.3	0.1	0.0	0.1	0.0	38.0	19.2	31.6	7.3	3.5
	W-50	0.3	0.1	0.0	0.1	0.0	41.5	17.8	31.0	5.9	3.4
	W-60	0.4	0.1	0.0	0.1	0.0	46.5	16.7	28.7	4.8	2.9
	W-70	0.5	0.1	0.0	0.1	0.0	57.2	13.6	23.7	2.8	1.9
	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	67.5	0.0	32.4	0.0
	DM-70	0.0	0.1	0.0	0.0	0.0	0.0	74.5	0.0	25.4	0.0
2004	0	0.3	0.1	0.0	0.1	0.0	37.0	19.9	31.4	7.8	3.6
	W-50	0.3	0.1	0.0	0.1	0.0	40.5	18.5	30.9	6.3	3.5
	W-60	0.4	0.1	0.0	0.1	0.0	45.5	17.3	28.6	5.1	3.0
	W-70	0.5	0.1	0.0	0.1	0.0	56.2	14.2	23.8	3.0	2.0
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	63.6	0.0	36.3	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	66.9	0.0	33.0	0.0
	DM-70	0.0	0.1	0.0	0.0	0.0	0.0	73.9	0.0	25.9	0.0
2005	0	0.3	0.1	0.0	0.1	0.0	35.9	20.0	31.8	8.1	3.8
	W-50	0.3	0.1	0.0	0.1	0.0	39.4	18.6	31.3	6.6	3.6
	W-60	0.4	0.1	0.0	0.1	0.0	44.4	17.5	29.2	5.3	3.1
	W-70	0.5	0.1	0.0	0.1	0.0	55.1	14.4	24.4	3.2	2.1
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	62.7	0.0	37.2	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	66.1	0.0	33.8	0.0
	DM-70	0.0	0.1	0.0	0.0	0.0	0.0	73.3	0.0	26.6	0.0
2006	0	0.3	0.1	0.0	0.1	0.0	34.9	19.6	32.8	8.3	3.9
	W-50	0.3	0.1	0.0	0.1	0.0	38.3	18.3	32.3	6.8	3.8
	W-60	0.4	0.1	0.0	0.1	0.0	43.3	17.2	30.2	5.4	3.2
	W-70	0.5	0.1	0.0	0.1	0.0	54.1	14.3	25.4	3.3	2.2
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	61.4	0.0	38.5	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	64.8	0.0	35.1	0.0
	DM-70	0.0	0.2	0.0	0.0	0.0	0.0	72.3	0.0	27.6	0.0
2007	0	0.3	0.1	0.0	0.1	0.0	33.8	18.9	33.9	8.7	4.1
	W-50	0.4	0.1	0.0	0.1	0.0	37.3	17.7	33.4	7.1	3.9
	W-60	0.4	0.1	0.0	0.1	0.0	42.2	16.7	31.4	5.7	3.4
	W-70	0.6	0.1	0.0	0.2	0.0	53.0	14.0	26.4	3.5	2.3
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	59.4	0.0	40.5	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	62.9	0.0	37.0	0.0
	DM-70	0.0	0.2	0.0	0.0	0.0	0.0	70.7	0.0	29.1	0.0
2008	0	0.3	0.1	0.0	0.1	0.0	32.7	18.6	34.6	9.3	4.3
	W-50	0.4	0.1	0.0	0.1	0.0	36.1	17.5	34.2	7.6	4.1
	W-60	0.4	0.1	0.0	0.1	0.0	41.0	16.5	32.2	6.1	3.6
	W-70	0.6	0.1	0.0	0.2	0.0	51.7	14.0	27.2	3.8	2.4
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	57.2	0.0	42.7	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	60.8	0.0	39.1	0.0
	DM-70	0.0	0.2	0.0	0.0	0.0	0.0	69.0	0.0	30.8	0.0
2009	0	0.3	0.1	0.0	0.1	0.0	31.3	18.9	34.7	9.9	4.6
	W-50	0.4	0.1	0.0	0.1	0.0	34.7	17.8	34.3	8.2	4.4
	W-60	0.5	0.1	0.0	0.2	0.0	39.6	16.8	32.5	6.5	3.8
	W-70	0.6	0.1	0.0	0.2	0.0	50.3	14.5	27.6	4.1	2.5
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	55.8	0.0	44.1	0.0
	DM-60	0.0	0.1	0.0	0.0	0.0	0.0	59.4	0.0	40.4	0.0
	DM-70	0.0	0.2	0.0	0.0	0.0	0.0	67.9	0.0	31.9	0.0
2010	0	0.3	0.1	0.0	0.1	0.0	29.7	19.8	34.2	10.6	5.0
	W-50	0.4	0.1	0.0	0.2	0.0	33.1	18.8	33.9	8.7	4.8
	W-60	0.5	0.1	0.0	0.2	0.0	38.0	17.7	32.3	7.0	4.1
	W-70	0.6	0.1	0.0	0.2	0.0	48.8	15.4	27.7	4.4	2.7
	DM-50	0.0	0.1	0.0	0.0	0.0	0.0	55.4	0.0	44.4	0.0
	DM-60	0.0	0.2	0.0	0.0	0.0	0.0	59.0	0.0	40.8	0.0
	DM-70	0.0	0.2	0.0	0.0	0.0	0.0	67.7	0.0	32.1	0.0
2011	0	0.3	0.1	0.0	0.1	0.0	28.1	21.0	33.4	11.2	5.6
	W-50	0.4	0.1	0.0	0.2	0.0	31.5	20.0	33.2	9.3	5.3
	W-60	0.5	0.1	0.0	0.2	0.0	36.4	18.9	31.9	7.4	4.6

```
       W-70    0.6  0.1   0.0  0.3   0.0 47.1 16.5 27.6  4.7  3.1
       DM-50   0.0  0.2   0.0  0.0   0.0  0.0 55.6  0.0 44.2  0.0
       DM-60   0.0  0.2   0.0  0.0   0.0  0.0 59.2  0.0 40.6  0.0
       DM-70   0.0  0.2   0.0  0.0   0.0  0.0 67.8  0.0 31.9  0.0
  2012 0       0.3  0.1   0.0  0.2   0.0 26.5 22.3 32.5 11.9  6.2
       W-50    0.4  0.1   0.0  0.2   0.0 29.8 21.3 32.4  9.8  5.9
       W-60    0.5  0.1   0.0  0.2   0.0 34.7 20.2 31.3  7.8  5.1
       W-70    0.7  0.2   0.0  0.3   0.0 45.4 17.7 27.4  4.9  3.4
       DM-50   0.0  0.2   0.0  0.0   0.0  0.0 55.8  0.0 44.0  0.0
       DM-60   0.0  0.2   0.0  0.0   0.0  0.0 59.3  0.0 40.4  0.0
       DM-70   0.0  0.3   0.1  0.0   0.0  0.0 68.0  0.0 31.7  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   21.2  19.5 100.0  16.3 100.0  11.7 100.0
        1996   22.3  20.5 100.0  17.2 100.0  12.3 100.0
        1997   23.5  21.6 100.0  18.2 100.0  13.0 100.0
        1998   24.7  22.8 100.0  19.2 100.0  13.8 100.0
        1999   25.9  23.9 100.0  20.3 100.0  14.6 100.0
        2000   27.2  25.2 100.0  21.4 100.0  15.4 100.0
        2001   28.6  26.5 100.0  22.6 100.0  16.3 100.0
        2002   30.0  27.9 100.0  23.9 100.0  17.2 100.0
        2003   31.5  29.3 100.0  25.2 100.0  18.2 100.0
        2004   32.7  30.5 100.0  26.3 100.0  19.0 100.0
        2005   33.5  31.2 100.0  27.0 100.0  19.6 100.0
        2006   33.7  31.5 100.0  27.3 100.0  19.9 100.0
        2007   34.0  31.7 100.0  27.6 100.0  20.2 100.0
        2008   34.6  32.4 100.0  28.2 100.0  20.8 100.0
        2009   36.1  33.8 100.0  29.5 100.0  22.0 100.0
        2010   38.5  36.1 100.0  31.6 100.0  23.8 100.0
        2011   41.3  38.8 100.0  34.1 100.0  25.9  99.9
        2012   44.2  41.7 100.0  36.8 100.0  28.2  99.9
      F 1995   19.8  17.9 100.0  15.9 100.0  12.0 100.0
        1996   20.9  18.8 100.0  16.8 100.0  12.7 100.0
        1997   21.9  19.8 100.0  17.7 100.0  13.3 100.0
        1998   23.1  20.8 100.0  18.6 100.0  14.0  99.9
        1999   24.2  21.8 100.0  19.6 100.0  14.8  99.9
        2000   25.5  23.0 100.0  20.6  99.9  15.5  99.9
        2001   26.8  24.2  99.9  21.8  99.9  16.4  99.9
        2002   28.4  25.6  99.9  23.0  99.9  17.4  99.9
        2003   30.0  27.0  99.9  24.3  99.9  18.4  99.9
        2004   31.3  28.2  99.9  25.4  99.9  19.3  99.9
```

		2005	31.9	28.8	99.9	25.9	99.9	19.7	99.8
		2006	31.9	28.8	99.9	25.9	99.9	19.8	99.8
		2007	31.7	28.7	99.9	25.7	99.9	19.8	99.8
		2008	32.2	29.1	99.9	26.1	99.8	20.1	99.8
		2009	33.5	30.4	99.9	27.2	99.8	21.1	99.8
		2010	35.4	32.3	99.8	28.8	99.8	22.5	99.7
		2011	37.8	34.6	99.8	30.9	99.8	24.2	99.7
		2012	40.4	37.0	99.8	33.1	99.8	26.0	99.7
Ca	M	1995	33.6	33.9	24.1	33.0	24.7	29.1	22.4
		1996	34.9	35.2	25.3	34.2	25.8	30.1	23.2
		1997	36.2	36.5	26.6	35.5	26.9	31.1	24.1
		1998	37.5	37.8	27.8	36.8	28.1	32.2	25.0
		1999	38.7	39.1	29.1	38.0	29.2	33.2	25.9
		2000	39.8	40.2	30.3	39.1	30.4	34.1	26.8
		2001	40.7	41.1	31.6	40.0	31.6	34.9	27.8
		2002	41.4	41.9	32.8	40.8	32.8	35.6	28.8
		2003	42.2	42.6	34.1	41.6	34.1	36.2	29.9
		2004	43.2	43.7	35.5	42.6	35.5	37.2	31.0
		2005	44.6	45.2	36.9	44.1	37.0	38.6	32.3
		2006	46.4	47.0	38.4	46.0	38.6	40.3	33.6
		2007	48.3	48.9	39.9	48.0	40.1	42.2	35.0
		2008	49.8	50.5	41.3	49.6	41.5	43.8	36.2
		2009	50.8	51.5	42.4	50.7	42.7	44.9	37.2
		2010	51.4	52.1	43.3	51.4	43.6	45.6	38.0
		2011	51.8	52.5	44.0	51.8	44.4	46.2	38.8
		2012	52.1	52.9	44.7	52.2	45.1	46.7	39.5
	F	1995	35.8	34.0	25.7	30.5	23.8	23.2	18.6
		1996	37.1	35.3	27.1	31.6	25.0	24.2	19.4
		1997	38.4	36.6	28.5	32.8	26.2	25.2	20.4
		1998	39.7	37.8	30.0	34.0	27.5	26.2	21.3
		1999	40.9	38.9	31.4	35.0	28.7	27.2	22.3
		2000	41.7	39.7	32.7	35.8	29.9	27.9	23.2
		2001	42.1	40.1	33.9	36.2	30.9	28.2	24.1
		2002	42.3	40.2	34.9	36.3	31.7	28.4	24.8
		2003	42.4	40.3	35.6	36.4	32.4	28.5	25.4
		2004	42.7	40.6	36.3	36.7	33.0	28.9	25.9
		2005	43.6	41.5	37.2	37.6	33.8	29.7	26.6
		2006	45.1	42.9	38.5	38.9	35.1	30.9	27.6
		2007	46.7	44.5	40.5	40.4	37.0	32.2	29.1
		2008	48.2	45.9	42.7	41.8	39.1	33.4	30.8
		2009	49.3	46.9	44.1	42.8	40.4	34.2	31.9
		2010	49.9	47.4	44.4	43.4	40.8	34.8	32.1
		2011	50.2	47.8	44.2	43.8	40.6	35.3	31.9
		2012	50.5	48.2	44.0	44.2	40.4	35.7	31.7
DM+CA	M	1995	5.8	5.4	24.1	4.5	24.7	3.0	22.4
		1996	6.3	5.9	25.3	4.9	25.8	3.3	23.2
		1997	7.0	6.5	26.6	5.4	26.9	3.6	24.1
		1998	7.6	7.1	27.8	5.9	28.1	3.9	25.0
		1999	8.3	7.7	29.1	6.5	29.2	4.2	25.9
		2000	9.1	8.4	30.3	7.0	30.4	4.6	26.8
		2001	9.8	9.2	31.6	7.7	31.6	5.0	27.8
		2002	10.7	9.9	32.8	8.3	32.8	5.4	28.8
		2003	11.5	10.7	34.1	9.0	34.1	5.8	29.9
		2004	12.4	11.5	35.5	9.7	35.5	6.2	31.0
		2005	13.2	12.3	36.9	10.3	37.0	6.7	32.3
		2006	13.8	12.9	38.4	10.9	38.6	7.1	33.6
		2007	14.5	13.6	39.9	11.5	40.1	7.5	35.0
		2008	15.3	14.3	41.3	12.2	41.5	7.9	36.2
		2009	16.4	15.3	42.4	13.1	42.7	8.6	37.2
		2010	17.7	16.6	43.3	14.2	43.6	9.4	38.0
		2011	19.3	18.1	44.0	15.6	44.4	10.4	38.8
		2012	20.9	19.7	44.7	17.0	45.1	11.5	39.5
	F	1995	5.3	4.6	25.7	3.8	23.8	2.4	18.6
		1996	5.9	5.1	27.1	4.2	25.0	2.6	19.4
		1997	6.5	5.6	28.5	4.7	26.2	2.9	20.4
		1998	7.1	6.2	30.0	5.1	27.5	3.2	21.3

```
1999    7.9   6.8  31.4   5.6  28.7   3.5  22.3
2000    8.6   7.4  32.7   6.2  29.9   3.9  23.2
2001    9.3   8.0  33.9   6.7  30.9   4.2  24.1
2002   10.1   8.7  34.9   7.2  31.7   4.5  24.8
2003   10.8   9.3  35.6   7.7  32.4   4.8  25.4
2004   11.4   9.8  36.3   8.1  33.0   5.1  25.9
2005   11.9  10.2  37.2   8.4  33.8   5.3  26.6
2006   12.3  10.5  38.5   8.7  35.1   5.5  27.6
2007   12.8  11.0  40.5   9.1  37.0   5.7  29.1
2008   13.6  11.7  42.7   9.6  39.1   6.1  30.8
2009   14.6  12.6  44.1  10.3  40.4   6.6  31.9
2010   15.6  13.5  44.4  11.1  40.8   7.1  32.1
2011   16.8  14.6  44.2  12.0  40.6   7.7  31.9
2012   18.1  15.7  44.0  12.9  40.4   8.4  31.7
```

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 6.28e-06 0.00 6.94e-06 0.00 ...
 - attr(*, "dimnames")=List of 4
  ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
  ..$ age  : chr [1:1224] "0.041666666666667" "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 6.28e-06 0.00 6.94e-06 0.00 ...
 - attr(*, "dimnames")=List of 4
  ..$ state: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
  ..$ age  : chr [1:1224] "0.041666666666667" "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] 67.6705 2.5455 0.0779 1.4552 0.0712 ...
 - 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( PY )[-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" "200

$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" "200

$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 well as the percentage of total life spent with diabetes, as a function of the date at which we evaluated rates:

```
> par( mfrow=c(1,2), mar=c(1,1,1,1), mgp=c(3,1,0)/1.6, oma=c(2,2,0,1) )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,],
+                                                100*PY["anyDM",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,2),each=2), lwd=4, col=c("blue","red"),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "Diabetes", adj=0 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyCa",,],
+                                                100*PY["anyCa",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,2),each=2), lwd=4, col=c("blue","red"),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "Cancer", adj=0 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> mtext("Date of rate evaluation", side=1, line=1, cex=1.0, outer=TRUE )
> mtext("Years / % of life spent with disease", side=2, line=1, cex=1.0,
+       outer=TRUE, las=0 )
```

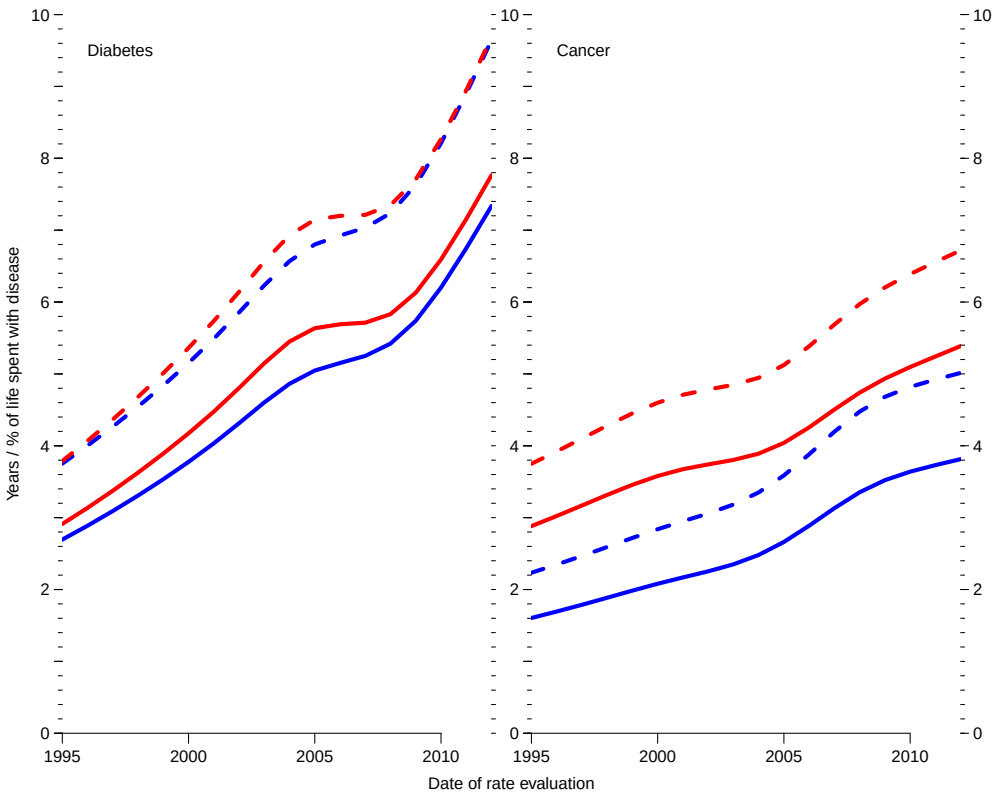

Figure 4.11: *Years (full lines) and percent of life (broken lines) spent with disease (daiebets or cancer); red: women, blue: men.*

```
> par( mfrow=c(1,2), mar=c(1,1,1,1), mgp=c(3,1,0)/1.6, oma=c(2,2,0,1) )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,"M"], PY["anyDM",,"M"],
+                                                100*PY["anyDM",,"M"]/PY["All",,"M"],
+                                                    PY["anyCa",,"M"], PY["anyCa",,"M"],
+                                                100*PY["anyCa",,"M"]/PY["All",,"M"] ),
+          type="l", lty=c("F1","11F1","11"), lwd=6, col=rep(clr[c(2,3)],each=3),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "M", adj=0, font=2, cex=2 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,"F"], PY["anyDM",,"F"],
+                                                100*PY["anyDM",,"F"]/PY["All",,"F"],
+                                                    PY["anyCa",,"F"], PY["anyCa",,"F"],
+                                                100*PY["anyCa",,"F"]/PY["All",,"F"] ),
+          type="l", lty=c("F1","11F1","11"), lwd=6, col=rep(clr[c(2,3)],each=3),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "F", adj=0, font=2, cex=2 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> mtext("Date of rate evaluation", side=1, line=1, cex=1.0, outer=TRUE )
> mtext("Years / % of life spent with disease", side=2, line=1, cex=1.0,
+       outer=TRUE, las=0 )
```

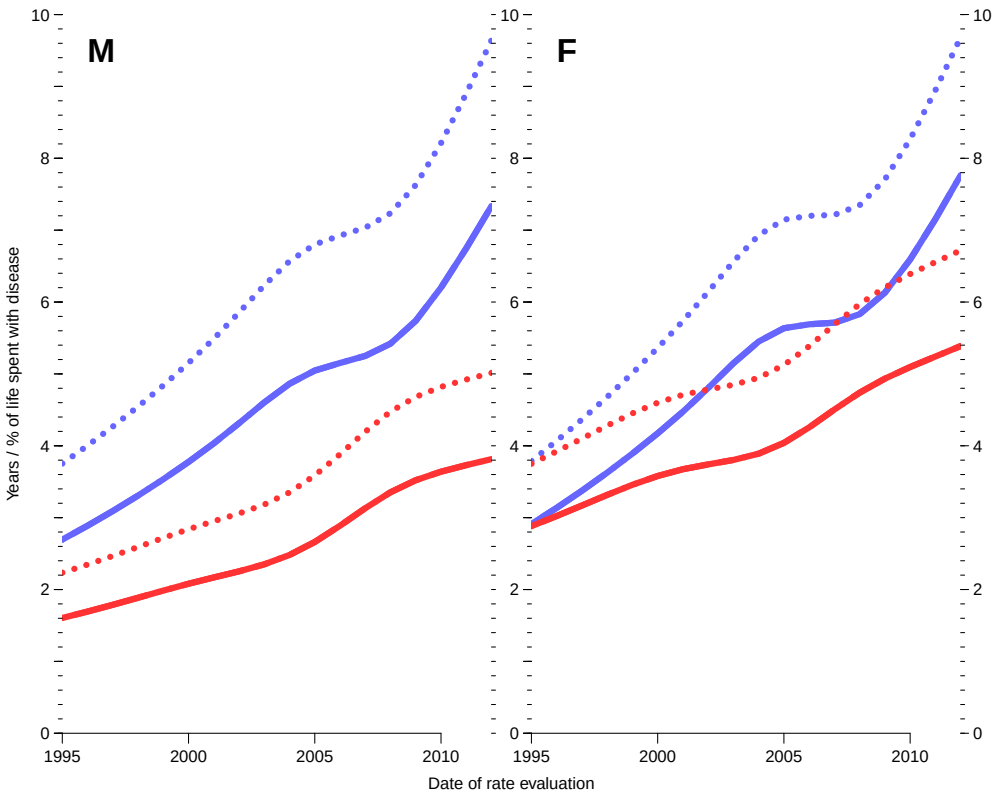

Figure 4.12: *Years (full lines) and percent of life (broken lines) spent with disease; diabetes (blue) or cancer (red).*

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. The probabilities in `PV` gives the *distributions* of persons in states in each age. Since this refers to the distribution of *all* initial persons in the population, normalization of the age-specific occupancy probabilities of, say, the state "DM" to an age-distribution will represent the distribution of time alive spent in the state. This of course is meaningless for the corresponding death states.

Thus we devise a function that does this for a 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] 42.93041 54.42347 64.37286 72.79126 79.42252
```

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`, `Age Diseased`) 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" "200

$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" "200

$sex
[1] "M" "F"
> AD[,"Well" ,,] <- apply( PV["Well",,,],
+                          2:3, aqnt )
> AD[,"DM"   ,,] <- apply( apply( PV[c("DM","DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> AD[,"Ca"   ,,] <- apply( apply( PV[c("Ca","DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> AD[,"DM+Ca",,] <- apply( apply( PV[c("DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> str( AD )
 num [1:5, 1:4, 1:18, 1:2] 6.78 17.02 34.46 53.15 67.08 ...
 - attr(*, "dimnames")=List of 4
  ..$ 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 then show the distribution of the ages with DM, and augment the plot with an indication of the expected *length* of time spent diseased, arbitrarily allocated around the median age spent with disease:

```
> par( mfrow=c(1,2), mar=c(2,2,1,1) )
> 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="",
+          xlab="Date of rates used", ylim=c(0,100), yaxs="i")
> 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), yaxs="i")
> 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" )
> mtext( "Age with diabetes (10,25,50,75,90 percentiles)", side=2,
+        outer=TRUE, line=0 )
```

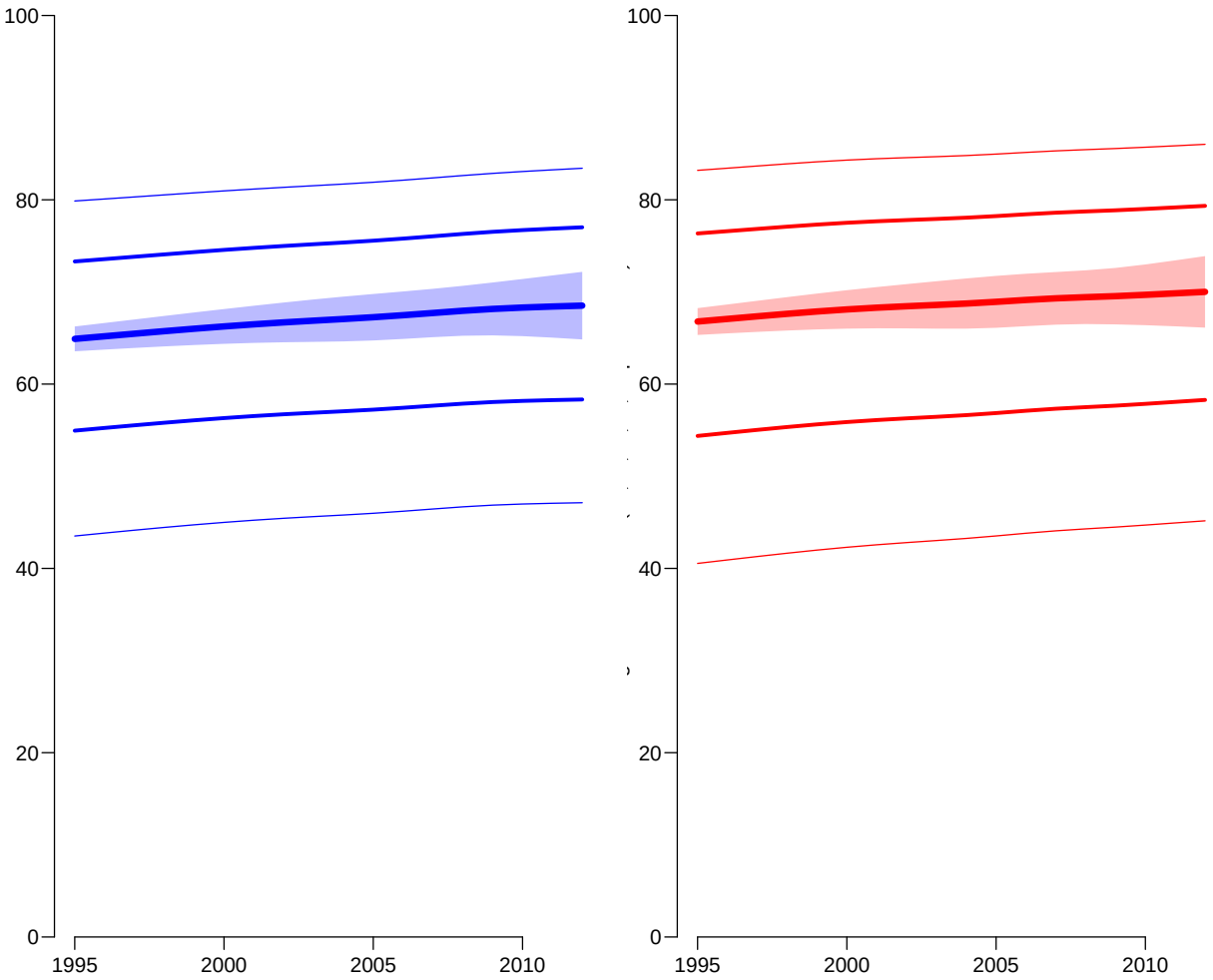

Figure 4.13: *Percentiles of ages in which persons suffer from diabetes for men (blue) and women (red). The* height *of shaded area indicate the expected* length *of time spent with diabetes.*

The comparison in figure 8.13 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.

Bibliography

[1] H. O. Adami, J. McLaughlin, A. Ekbom, C. Berne, D. Silverman, D. Hacker, and I. Persson. Cancer risk in patients with diabetes mellitus. *Cancer Causes Control*, 2:307–314, Sep 1991.

[2] L. Wideroff, G. Gridley, L. Mellemkjær, W. H. Chow, M. Linet, S. Keehn, K. Borch-Johnsen, and J. H. Olsen. Cancer incidence in a population-based cohort of patients hospitalized with diabetes mellitus in Denmark. *J. Natl. Cancer Inst.*, 89:1360–1365, Sep 1997.

[3] S. S. Coughlin, E. E. Calle, L. R. Teras, J. Petrelli, and M. J. Thun. Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. *Am. J. Epidemiol.*, 159:1160–1167, Jun 2004.

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

[5] 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.

[6] 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.

[7] M. L. Gjerstorff. The Danish Cancer Registry. *Scand J Public Health*, 39(7 Suppl):42–45, July 2011.

[8] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). available at www.mortality.org or www.humanmortality.de (data downloaded 28 november 2014). Technical report.

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

[10] A. Green, C. Sortsø, P. B. Jensen, and M. Emneus. Validation of the Danish National Diabetes Register. *Clin Epidemiol*, 7:5–15, 2015.

Part III

Using a more restrictive definition of DM

Chapter 5

Removing blood glucose criteria

It has been questioned whether the originally defined criteria for classification persons and having diabetes [5] are all valid; in particular it has been shown [10], that the criteria based on usage of blood-glucose testing (note, *usage*, not value of test) might be including persons that does have diabetes.

The fraction of persons in the NDR that *only* meet this criterion is non-negligible, and this part of the report therefor repeats all analyses and graphs using a modified definition of diabetes where persons are *not* included on the basis of these criteria. The modification is straight-forward; resulting in some persons not being included in the register at all, others included at a later date, and no change for others.

These effects differ by sex and age and date at inclusion, which is the reason to redo the entire analysis based on the revised definition of diabetes from the register.

Chapter 6

Data base (modified def.)

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 3.2.0 (2015-04-16)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
[1] Epi_1.1.68     foreign_0.8-63

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7    MASS_7.3-39     parallel_3.2.0  survival_2.38-1 etm_0.6-2
[6] splines_3.2.0   grid_3.2.0      lattice_0.20-29
```

Referring to figure 1.1 with transitions between states, 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.

6.1 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.

6.1.1 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.

6.1.2 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.

6.2 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 1.1 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.

6.2.1 Follow-up records

First we read the follow-up file from all *patients*, generated by a SAS-program virtually identical to that :

```
1                                   "Program: DMrCaLex.sas"    09:22 Wednesday, January 14, 2015

NOTE: Copyright (c) 2002-2008 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.2 (TS2M3)
      Licensed to NOVO NORDISK - BASIC PACKAGE, Site 50800704.
NOTE: This session is executing on the W32_VSPRO  platform.



NOTE: SAS initialization used:
      real time           2.60 seconds
      cpu time            0.57 seconds


NOTE: AUTOEXEC processing beginning; file is c:\stat\sas\autoexec.sas.

----------------------------------------------------------------
C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\sas\DMrCaLex.sas
----------------------------------------------------------------
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\sas
NOTE: Libref DATA was successfully assigned as follows:
      Engine:        V9
```

```
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\data

NOTE: AUTOEXEC processing completed.

1           ***************************************************************
2          NOTE: This version of the program takes all patients of either
3                DM or cancer, subdivide their follow-up (using the variables
4                entry, exit and fail) according to their status as being
5                either DM, Ca, DM-Ca or Ca-DM. The coding of the fail
6                variable is: 0: censored, 1: DM, 2: Cancer, 3: Dead
7          ************************************************************** ;
8
9          * The date from which we trust the inclusion date to be the first ;
10         %let  validdate = '01JAN1995'd ;
11         * Set the entry and exit dates for the entire follow-up endeavour ;
12         %let  truncdate = '01JAN1995'd ;
13         %let   censdate = '31DEC2011'd ;
14         * Just to check it all wemt well ;
15         %put validdate = &validdate.
16              truncdate = &truncdate.
17               censdate = &censdate. ;
validdate = '01JAN1995'd      truncdate = '01JAN1995'd       censdate = '31DEC2011'd
18         * Set the selector of subgroups to analyse ;
19         %let dgrp = 21,22,241,242,243,249,251,26,28,
20                     33,
21                     51,
22                     70,
23                     82,83,84,
24                     91,92,
25                    101,103,
26                    113,
27                    121,
28                    131,132,133,139 ;
29         %let diagselect = diag in (&dgrp.) ;
30         * Variable names for tabulation purposes, note DX and D259 here ;
31         %let dvars = D0 D999
32                      D21 D22 D241 D242 D243 D249 D251 D259 D26 D28
33                      D33
34                      D51g
35                      D70
36                      D82 D83 D84
37                      D91 D92
38                     D101 D103
39                     D113
40                     D121
41                     D131 D132 D133 D139 ;
42
43         * Get the formats and the Lexis macro ;
44         options nosource2 ;
45         %inc "c:\bendix\steno\DM-register\NDR\projects\Cancer\sas\CRG-fmts.sas" ;
NOTE: Format SEX has been output.
NOTE: Format DIAG has been output.

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.04 seconds
      cpu time            0.03 seconds


130        libname DMCA "c:\bendix\steno\DM-register\NDR\projects\Cancer\data" ;
NOTE: Libref DMCA was successfully assigned as follows:
      Engine:        V9
      Physical Name: c:\bendix\steno\DM-register\NDR\projects\Cancer\data
131
132        *-------------------------------------------------------------------;
133        * Preprocessing of the cancer register to first primary tumours only ;
134
135        * First take the cancer registry, remove all non-cancers ;
136        data cancer ;
137          set DMCA.crg2012 ;
138          doca = d_diagnosedato ;
139        * Remove 'not counted as cancer' and non-melanoma skin cancer ;
140          if ( diag in (52,150) ) then delete ;
141        * Recode the leukaemias to one group (139 is a not used value in formats) ;
142          if diag in (134,135,136,137) then diag = 139 ;
143        * Recode the colon cancers to the three separate subsites and the rest ;
144          * 24.1 Ascending colon C18.0, C18.1, C18.2
145          * 24.2 Transverse colon C18.3, C18.4, C18.5
146          * 24.3 Descending and sigmoid colon C18.6, C18.7, C19, C19.9
147          * 24.9 Other colon (unspec. or multiple)
148          * 25.1 Rectum (excl. anus) C20, C209
149          * This means that colorectal cancers are to be taken as the sum of these
150          * 5 groups, but also that the group 24.9 is NOT of interest per se ;
151          if( diag eq 24 )                          then diag = 249 ;
152          if( icdpyrs in ("C180","C181","C182") )       then diag = 241 ;
153          if( icdpyrs in ("C183","C184","C185") )       then diag = 242 ;
154          if( icdpyrs in ("C186","C187","C19","C199") ) then diag = 243 ;
155          if( icdpyrs in ("C20","C209") )               then diag = 251 ;
156          * Finally make a single code for the sites not among those to be anlysed ;
157          if not ( diag in ( &dgrp. ) ) then diag = 999 ;
```

```
158        run ;

NOTE: There were 1929170 observations read from the data set DMCA.CRG2012.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: DATA statement used (Total process time):
      real time           9.68 seconds
      cpu time            2.24 seconds


159
160        * Sort by id and date of diagnosis ;
161        proc sort data = cancer ;
162          by id doCA ;
163        run ;

NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           20.67 seconds
      cpu time            3.96 seconds


164
165        * Sort by id ;
166        proc sort data = DMCA.dmr2012  out = diabetes ;
167          by id ;
168        run ;

NOTE: There were 497232 observations read from the data set DMCA.DMR2012.
NOTE: The data set WORK.DIABETES has 497232 observations and 12 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           4.75 seconds
      cpu time            1.23 seconds


169
170        * Then merge with the diabetes register ;
171        data DMCR;
172          merge cancer diabetes ;
173          by id ;
174          keep id sex diag
175               doBT doDM doCA doDD ;
176        * C_SEX is coded (1/2) in CAreg and (M/K) in DMreg ;
177          sex = ( C_SEX in ("1","M") ) + 2 * ( C_SEX in ("2","K") ) ;
178          if sex in (1,2) ;
179        * Demografic dates collected from CRG and NDR ;
180          doBT = min( D_foddto  , D_fdsdato ) ;
181          doDD = min( D_statdato, D_dodsdto ) ;
182        * Event-dates ;
183          * Note the revised date of dm
184          * doDM = D_inkldto ;
185          doDM = min( D_fodt, D_ins, D_oad, D_lpr ) ;
186          doI  = D_ins ;
187          doCA = D_diagnosedato ;
188        * If date of diabetes or cancer is equal to date of death, remove it ;
189          if doDD gt .z then do;
190             if doDM ge doDD then doDM = . ;
191             if doCA ge doDD then doCA = . ;
192             end ;
193        * If date of diabetes and cancer is the same, diabetes first ;
194          if doDM eq doCA then doDM = doCA - 2 ;
195          if doDM > .z or doCA > .z ;
196        * Only persons alive on 1.1.1995 (or born later) ;
197          if doDD gt '31DEC94'd or doDD le .z ;
198        * Only persons with one or the other disease ;
199          if doDM > .z or doCA > .z ;
200        run ;

NOTE: Missing values were generated as a result of performing an operation on missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      543533 at 181:10    1383110 at 185:10   136323 at 194:36
NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: There were 497232 observations read from the data set WORK.DIABETES.
NOTE: The data set WORK.DMCR has 992753 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           1.98 seconds
      cpu time            1.26 seconds


201
202        * The dataset DMCR now has a record for each person who has either a
203        * a diabetes diagnosis or a cancer diagnosis. Persons with more than
204        * one recorded tumour are represented by a record for each tumour ;
205        * We then construct the records of follow-up in different states ;
206
207        data toLex ;
208          set DMCR ;
209          id = _n_ ;
210          keep id sex diag
```

```
211               doBT doCa doDM doDD
212               entry exit en_st ex_st ;
213          length en_st ex_st $5 ;
214        *** Only Cancer ;
215          if ( doDM le .z ) then do ;
216             entry = max( doCa, &truncdate. ) ;
217             en_st = "Ca" ;
218             exit  = min( doDD, &censdate ) ;
219             if exit eq doDD then ex_st = "Dead" ; else
220                                  ex_st = en_st ;
221             if entry lt exit then output ;
222             end ;
223        *** Only diabetes ;
224          else
225          if ( doCa le .z ) then do ;
226             entry = max( doDM, &truncdate. ) ;
227             en_st = "DM" ;
228             exit  = min( doDD, &censdate ) ;
229             if exit eq doDD then ex_st = "Dead" ; else
230                                  ex_st = en_st ;
231             if entry lt exit then output ;
232             end ;
233        *** DM before Cancer ;
234          else
235          if ( doCa gt doDM ) then do ;
236             * from DM to Ca ;
237             entry = max( doDM, &truncdate. ) ;
238             en_st = "DM" ;
239             exit  = min( doCa, &censdate ) ;
240             if exit eq doCa then ex_st = "DM-Ca" ; else
241                                  ex_st = en_st ;
242             if entry lt exit then output ;
243             * from Ca to end ;
244             entry = max( doCa, &truncdate. ) ;
245             en_st = ex_st ;
246             exit  = min( doDD, &censdate ) ;
247             if exit eq doDD then ex_st = "Dead" ; else
248                                  ex_st = en_st ;
249             if entry lt exit then output ;
250             end ;
251        *** Cancer before DM ;
252          else
253          if ( doCa lt doDM ) then do ;
254             * from Ca to DM ;
255             entry = max( doCa, &truncdate. ) ;
256             en_st = "Ca" ;
257             exit  = min( doDM, &censdate ) ;
258             if exit eq doDM then ex_st = "Ca-DM" ; else
259                                  ex_st = en_st ;
260             if entry lt exit then output ;
261             * from DM to end ;
262             entry = max( doDM, &truncdate. ) ;
263             en_st = ex_st ;
264             exit  = min( doDD, &censdate ) ;
265             if exit eq doDD then ex_st = "Dead" ; else
266                                  ex_st = en_st ;
267             if entry lt exit then output ;
268             end ;
269        run ;

NOTE: There were 992753 observations read from the data set WORK.DMCR.
NOTE: The data set WORK.TOLEX has 1031778 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           1.26 seconds
      cpu time            0.60 seconds


270
271        libname allPT xport '../data/allPTr.xpt' ;
NOTE: Libref ALLPT was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\DemoDMCA\data\allPTr.xpt
272        proc copy in = work
273                 out = allPT ;
274          select toLex ;
275        run;

NOTE: Copying WORK.TOLEX to ALLPT.TOLEX (memtype=DATA).
NOTE: There were 1031778 observations read from the data set WORK.TOLEX.
NOTE: The data set ALLPT.TOLEX has 1031778 observations and 11 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           26.50 seconds
      cpu time            0.96 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           1:08.61
      cpu time            10.96 seconds
```

The dataset is generated in Lexis-ready-format, so that it can be put into a `Lexis` object after 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
  0.323   0.005   0.328
```

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 )
```

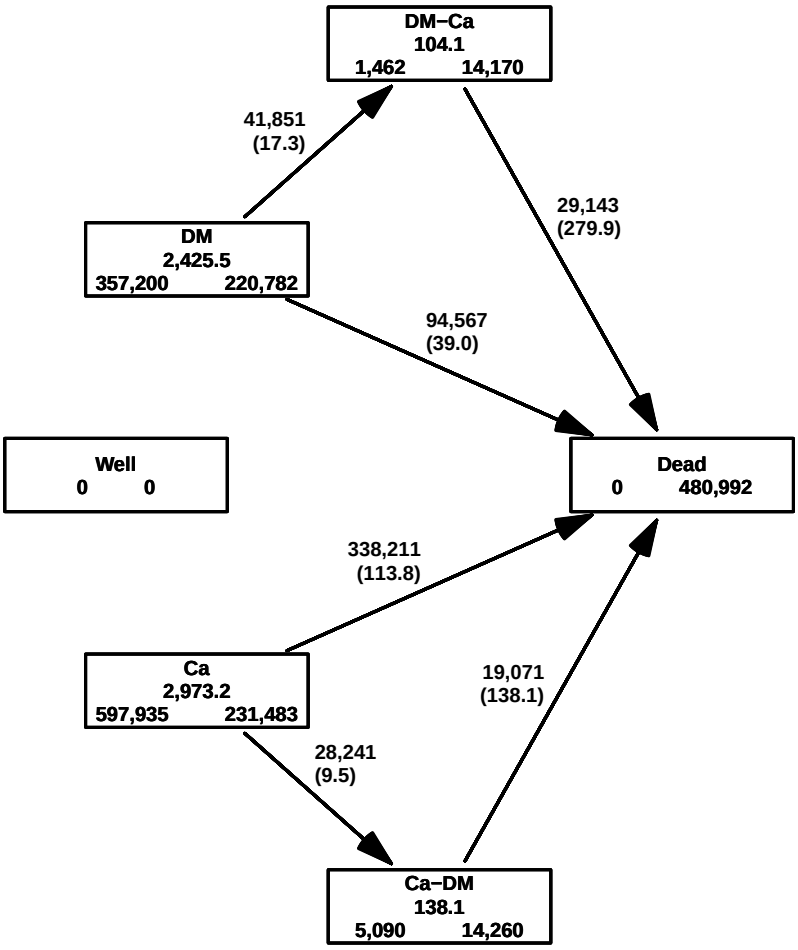

Figure 6.1: *The follow-up of the patients alone. The central number in each box is the amount of follow-up time (in 1000 PY) and the two number at the bottom are the number of persons that enter resp. exit the study in the state. Those entering also include persons that are prevalent cases as of 1.1.1995.*

6.3 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.

6.3.1 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 2015-06-03 11:42:25
Merged in chunk 2 now 20707 rows, at 2015-06-03 11:42:52
Merged in chunk 3 now 21475 rows, at 2015-06-03 11:43:18
Merged in chunk 4 now 22101 rows, at 2015-06-03 11:43:44
Merged in chunk 5 now 22480 rows, at 2015-06-03 11:44:11
Merged in chunk 6 now 22903 rows, at 2015-06-03 11:44:38
Merged in chunk 7 now 23227 rows, at 2015-06-03 11:45:06
Merged in chunk 8 now 23446 rows, at 2015-06-03 11:45:34
Merged in chunk 9 now 23633 rows, at 2015-06-03 11:46:01
Merged in chunk 10 now 23782 rows, at 2015-06-03 11:46:27
Merged in chunk 11 now 23994 rows, at 2015-06-03 11:46:54
Merged in chunk 12 now 24225 rows, at 2015-06-03 11:47:21
Merged in chunk 13 now 24358 rows, at 2015-06-03 11:47:48
Merged in chunk 14 now 24643 rows, at 2015-06-03 11:48:15
Merged in chunk 15 now 24792 rows, at 2015-06-03 11:48:42
Merged in chunk 16 now 24845 rows, at 2015-06-03 11:49:08
Merged in chunk 17 now 24948 rows, at 2015-06-03 11:49:34
Merged in chunk 18 now 25110 rows, at 2015-06-03 11:49:59
Merged in chunk 19 now 25169 rows, at 2015-06-03 11:50:26
Merged in chunk 20 now 25238 rows, at 2015-06-03 11:50:53
   user  system elapsed
533.570   2.477 535.797
```

```
> 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" )
```

6.3.2 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 ...
```

6.3.3 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 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
```

```
> library( Epi )
> library( splines )
> options( width=90,
+ #         prompt=" ", continue=" ",
+          SweaveHooks=list( fig=function()
+          par(mar=c(3,3,1,1),mgp=c(3,1,0)/1.6,bty="n",las=1) ) )
```

```
> print( sessionInfo(), l=F )
R version 3.2.0 (2015-04-16)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.2 LTS

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

other attached packages:
[1] Epi_1.1.68

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7    MASS_7.3-39     parallel_3.2.0  survival_2.38-1 etm_0.6-2
[6] grid_3.2.0      lattice_0.20-29
```

Chapter 7

Modelling of rates

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

```
> clear()
> load( file="./data/cols.Rda" )
> load( file="./data/dcd-r.Rda" )
> dcd <- subset( dcd, P<2012 & !is.na(Y) )
> 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=clx["Well"], border=clx["Well"], 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=clx["Well"], border=clx["Well"], 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=clx["Well"], border=clx["Well"], 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=clx["DM"], border=clx["DM"], 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=clx["DM"], border=clx["DM"], 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=clx["DM-Ca"], border=clx["DM-Ca"], 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=clx["Ca"], border=clx["Ca"], 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=clx["Ca"], border=clx["Ca"], 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=clx["Ca-DM"], border=clx["Ca-DM"], main="", yaxt="n",
+                           ylab="", xlab="Dead | Ca-DM" ) )
```

7.1 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,]
```

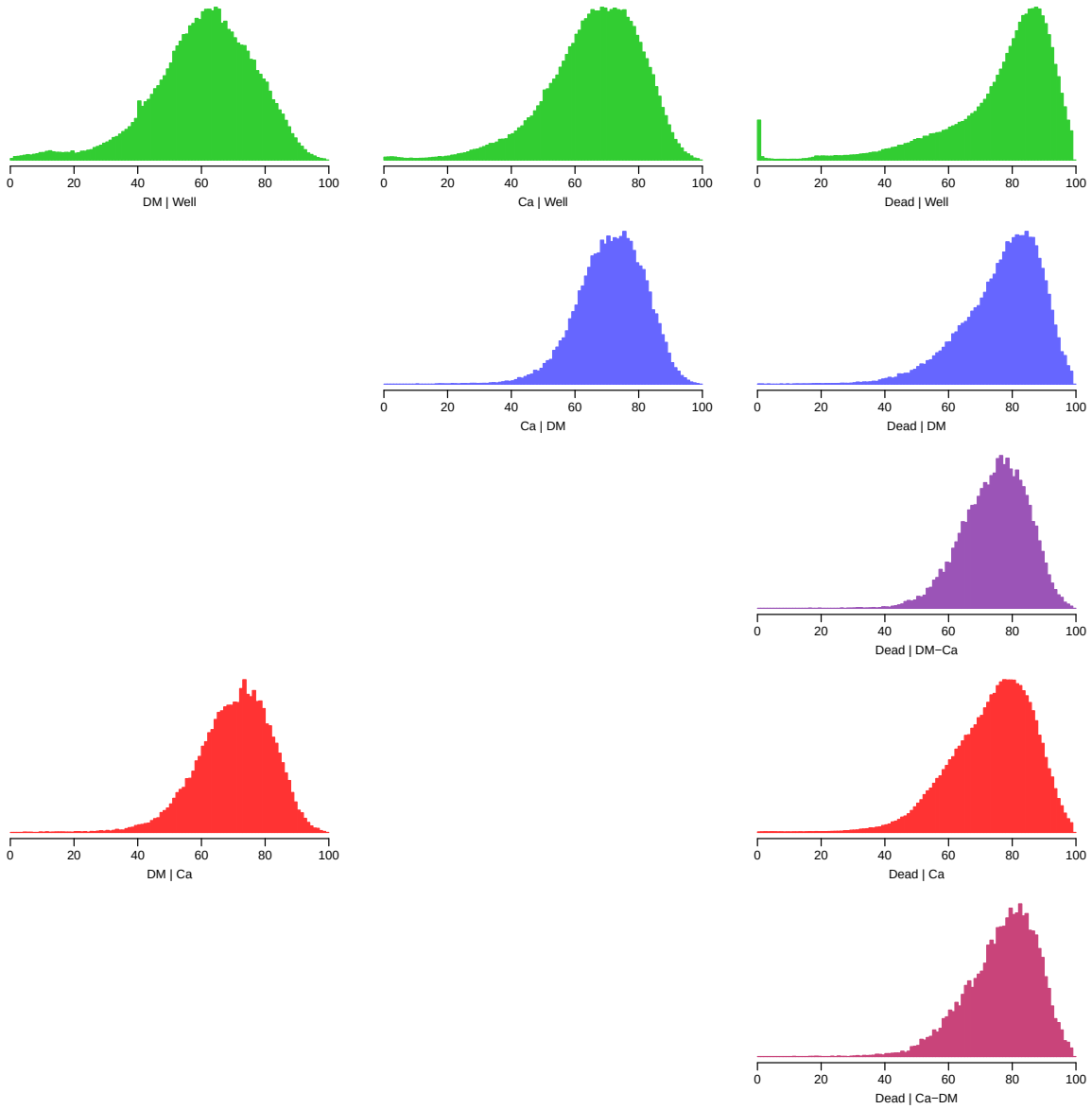

Figure 7.1: *Histograms of the age at event for the 9 possible transitions. Clearly, nothing much is happening in the younger ages, so we shall have age-knots a little closer in the older ages.*

```
+ acp$Coh <- acp$Coh[acp$Coh[,"Coh"]<1990,]
+ Kn <- apc$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(1,0,0.5,1), oma=c(3,4,2,3),
+      mgp=c(3,1,0)/1.6, las=1, bty="n" )
> inc.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.dm2ca$apc, col=clx["DM"]    , lend=1, lwd=4 )
> lines( M.w2ca$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( M.w2dm$apc , col=clx["Well"] , lend=1, lwd=4, lty="21" )
```

```
> lines( M.ca2dm$apc, col=clx["Ca"]   , lend=1, lwd=4, lty="21" )
> text(rep(101,2), c((rev(M.ca2dm$apc$Age[,2])[1]+
+                     rev( M.w2dm$apc$Age[,2])[1])/2,
+                    (rev(M.dm2ca$apc$Age[,2])[1]+
+                     rev( M.w2ca$apc$Age[,2])[1])/2), c("DM inc.","Ca inc."), adj=0 )
> text( c(145,208), c(0.25,0.25), c("Cohort","Period") )
> mort.fr(sides=1:2,col.grid=gray(0.9))
> lines( M.w2dd$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( M.dm2dd$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( M.ca2dd$apc, col=clx["Ca"]   , lend=1, lwd=4 )
> lines( M.cd2dd$apc, col=clx["Ca-DM"], lend=1, lwd=4 )
> lines( M.dc2dd$apc, col=clx["DM-Ca"], lend=1, lwd=4 )
> text( c(145,208), c(2.5,2.5), c("Cohort","Period") )
> mtext( "Age", at=55, side=1, line=2 )
> mtext( "Calendar time", at=165, side=1, line=2 )
> inc.fr(sides=1,col.grid=gray(0.9))
> lines( F.dm2ca$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( F.w2ca$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( F.w2dm$apc , col=clx["Well"] , lend=1, lwd=4, lty="21" )
> lines( F.ca2dm$apc, col=clx["Ca"]   , lend=1, lwd=4, lty="21" )
> text(rep(101,2), c((rev(F.ca2dm$apc$Age[,2])[1]+
+                     rev( F.w2dm$apc$Age[,2])[1])/2,
+                    (rev(F.dm2ca$apc$Age[,2])[1]+
+                     rev( F.w2ca$apc$Age[,2])[1])/2), c("DM inc.","Ca inc."), adj=0 )
> text( c(145,208), c(0.25,0.25), c("Cohort","Period") )
> mort.fr(sides=1,col.grid=gray(0.9))
> lines( F.w2dd$apc , col=clx["Well"] , lend=1, lwd=4 )
> lines( F.dm2dd$apc, col=clx["DM"]   , lend=1, lwd=4 )
> lines( F.ca2dd$apc, col=clx["Ca"]   , lend=1, lwd=4 )
> lines( F.cd2dd$apc, col=clx["Ca-DM"], lend=1, lwd=4 )
> lines( F.dc2dd$apc, col=clx["DM-Ca"], lend=1, lwd=4 )
> text( c(145,208), c(2.5,2.5), c("Cohort","Period") )
> 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 )
> mtext( "Age", at=55, side=1, line=2 )
> mtext( "Calendar time", at=165, side=1, line=2 )


> 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 7.2 and 7.3 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" )
```

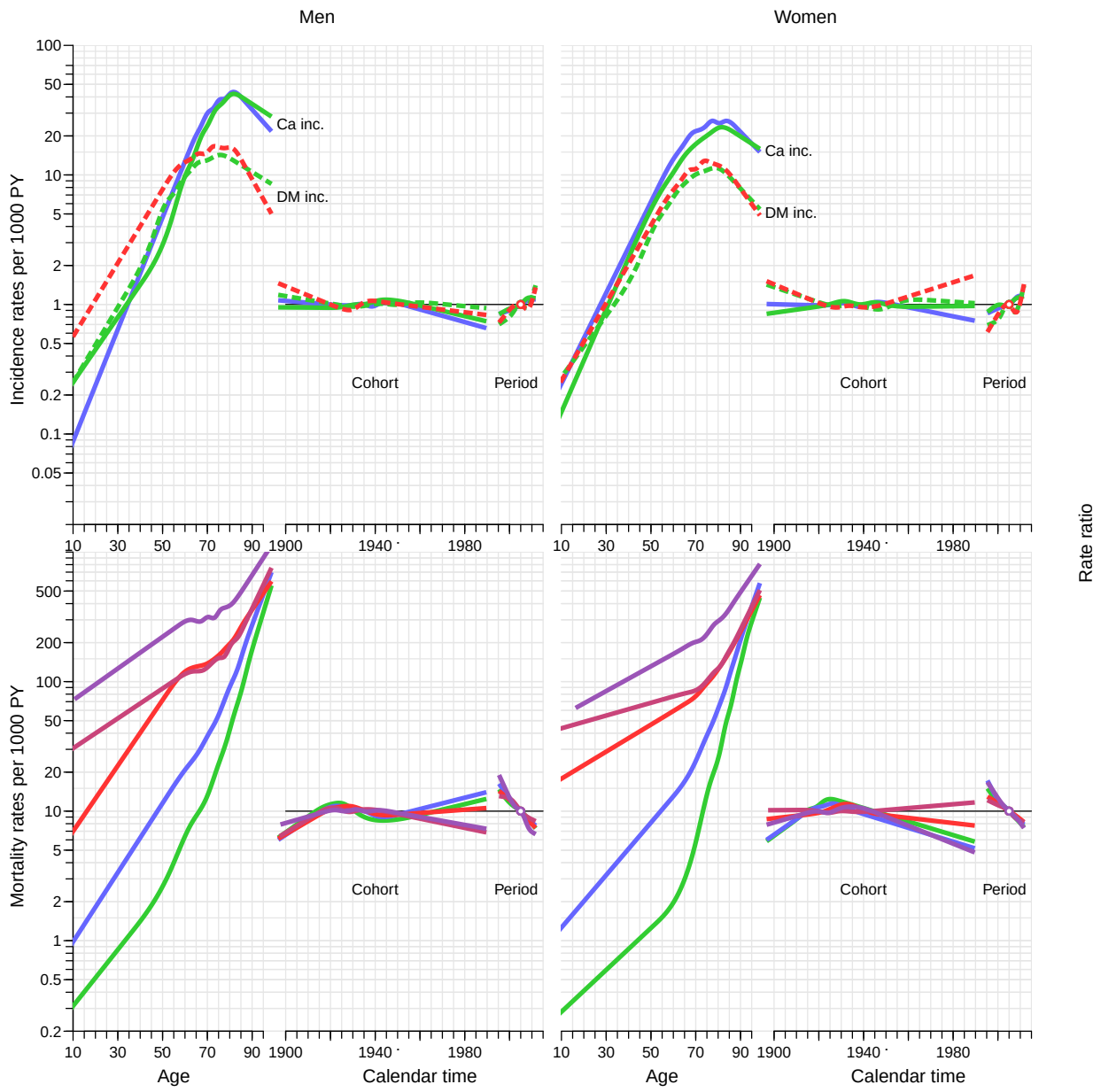

Figure 7.2: *Parameters from the fitted APC-models for rates using the period effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes 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.

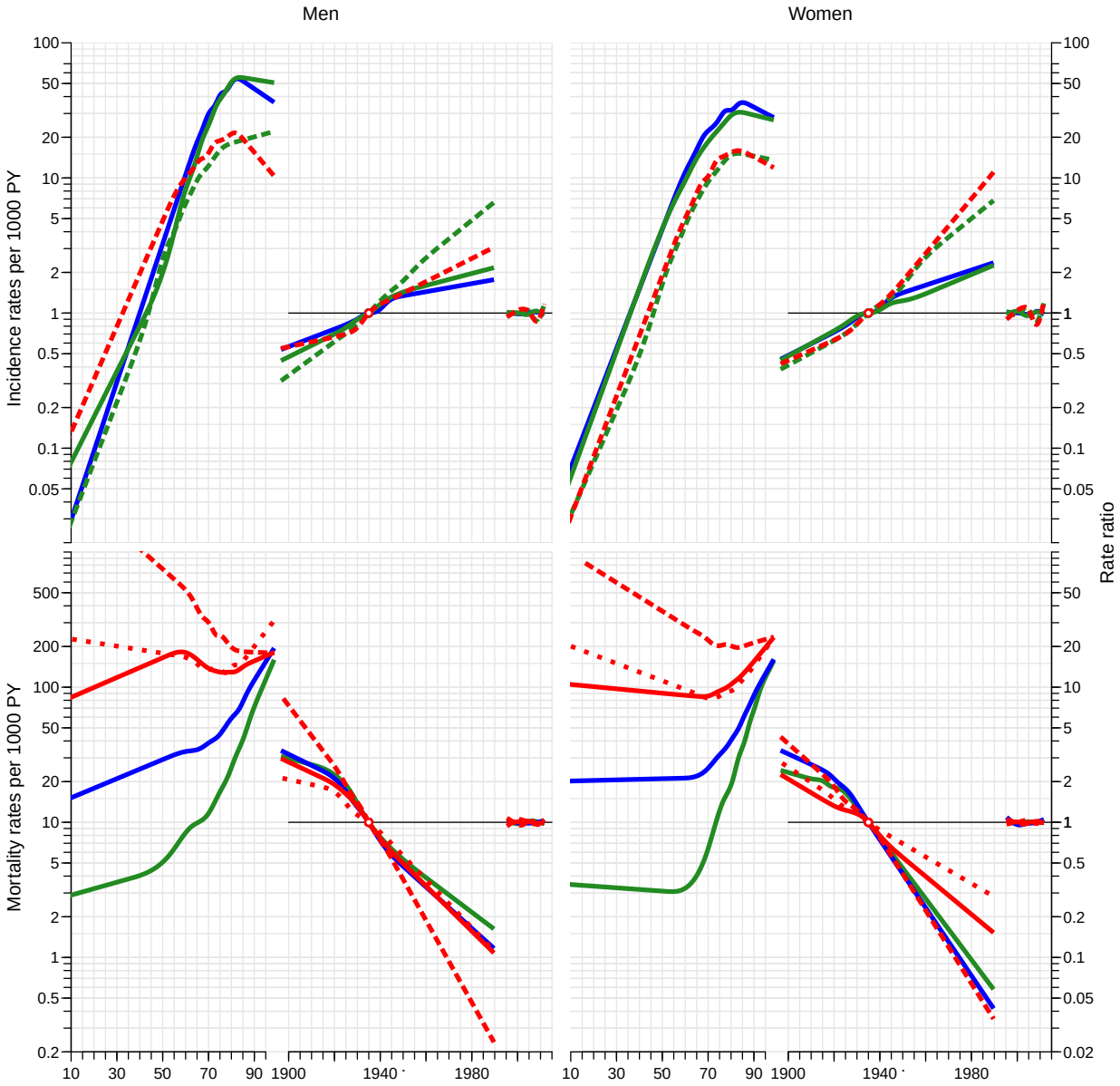

Figure 7.3: *Parameters from the fitted APC-models for rates using the cohort effect as the primary secular trend:*
Top panels: green lines: rates from the "Well" state; blue lines: rates from the "DM" state; red lines: rates form the "Ca" state; full lines: cancer incidence rates; broken lines: diabetes 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.

7.1.1 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.

```
> library( Epi )
> library( splines )
> clear()
> options( width=130,
+ #         prompt=" ", continue=" ",
+          SweaveHooks=list( fig=function()
```

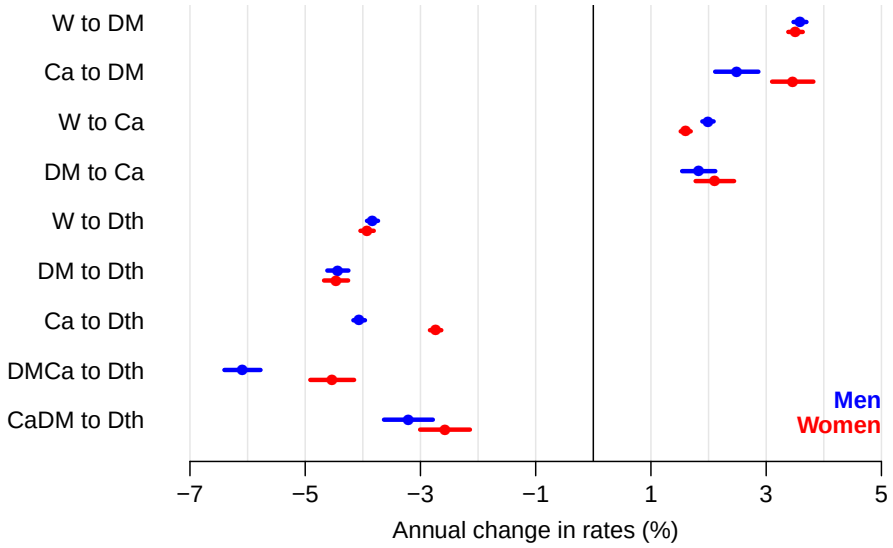

Figure 7.4: *Annual changes in the 9 incidence and mortality rates considered. Blue: men, red: women.*

```
+          par(mar=c(3,3,1,1),mgp=c(3,1,0)/1.6,bty="n",las=1) ) )
> load( file="./data/APC-r.Rda" )
> load( file="./data/dcd-r.Rda" )
> load( file="./data/cols.Rda" )
```

Chapter 8

State probabilities (modified def.)

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 of 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.

8.1 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

8.2 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$model, newdata=nd )[,1]
+ }
+ )
   user  system elapsed
  2.529   0.028   2.557
> for( bb in dimnames(CR)[[4]] )
+ {
+ nd <- data.frame( A=a.pt, P=as.numeric(bb)+a.pt, Y=int )
+
+ CR["Well" ,"DM"   ,,bb,"M"] <- ci.pred( M.w2dm$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="./data/rates-r.Rda" )
> load(         file="./data/rates-r.Rda" )
```

8.2.1 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 estimated age-specific 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 8.2 and 8.4 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.

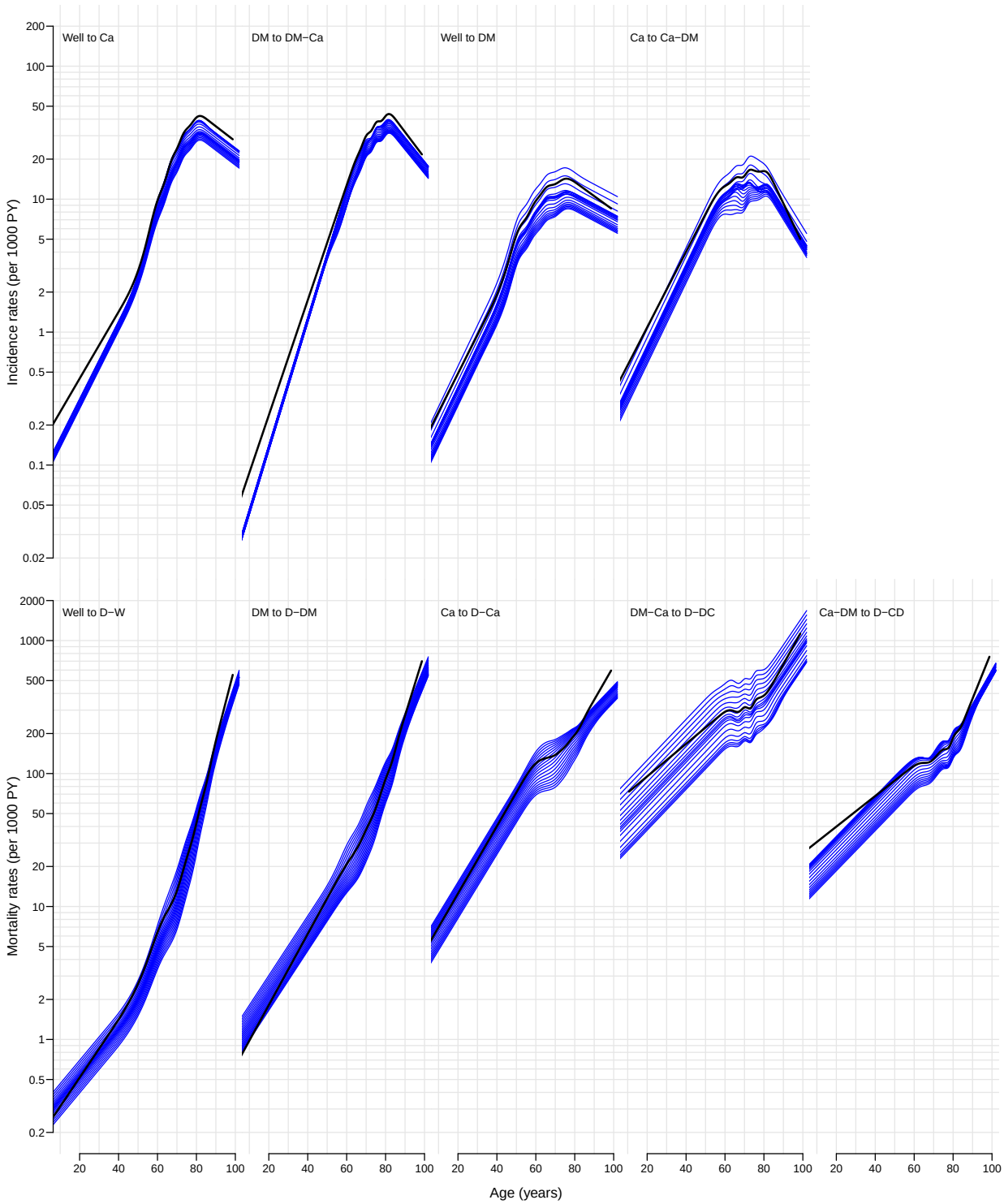

Figure 8.1: *Cross-sectional rates 1995–2012 for men, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

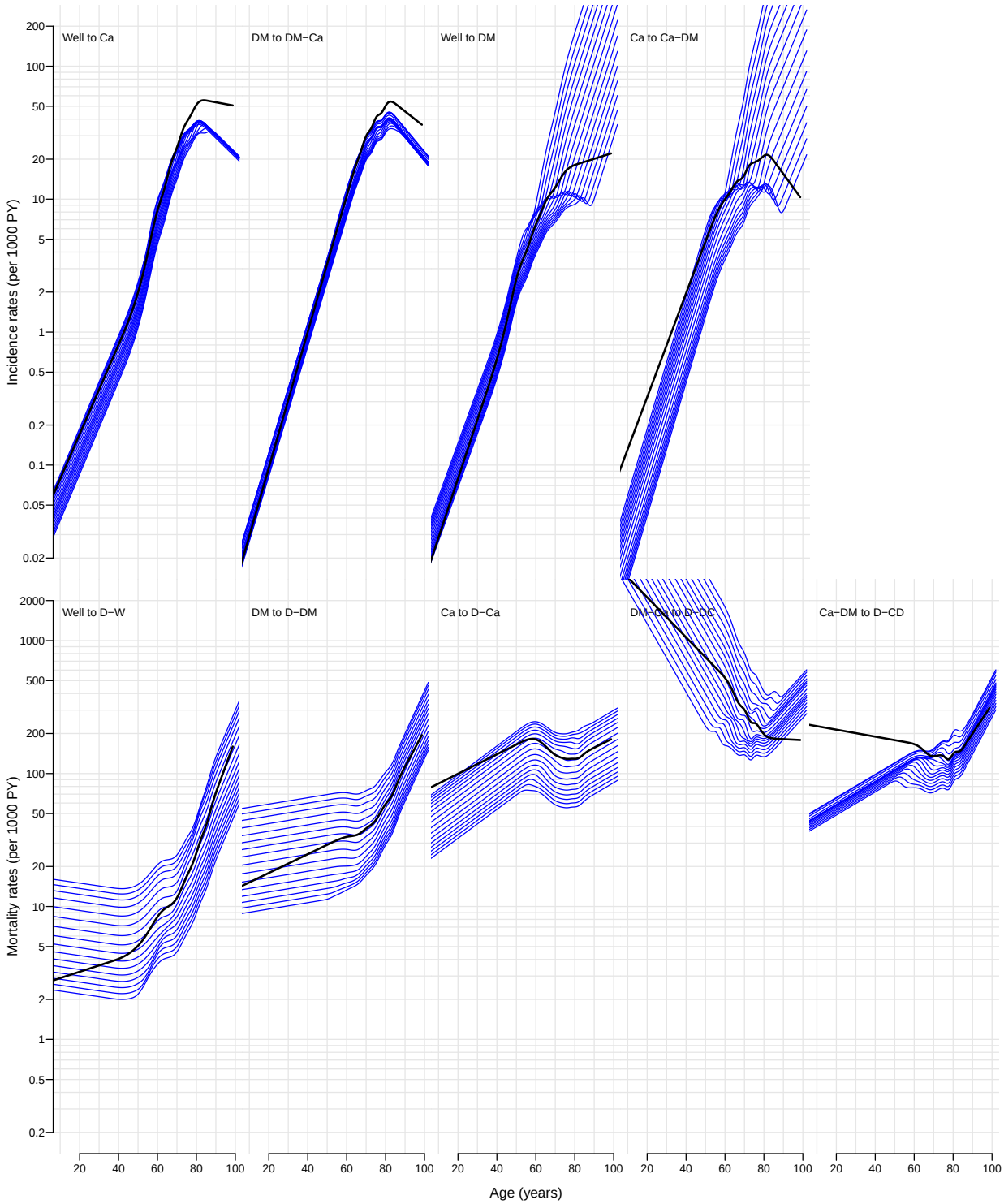

Figure 8.2: *Longitudinal rates 1930–1970 for men, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

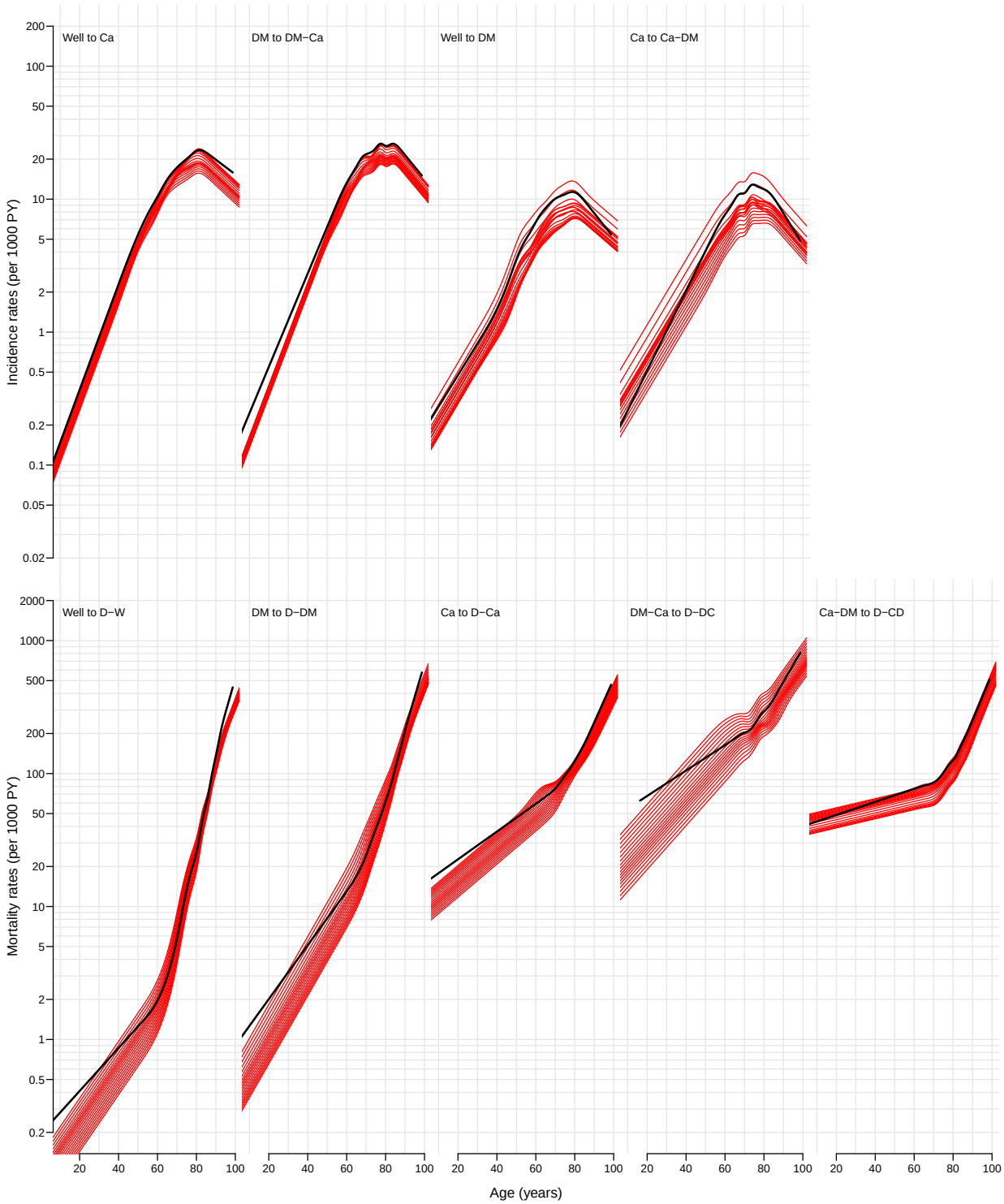

Figure 8.3: *Cross-sectional rates 1995–2012 for women, 1995–2012. Black curve is the age-effect at 2005, for flat cohort curve.*

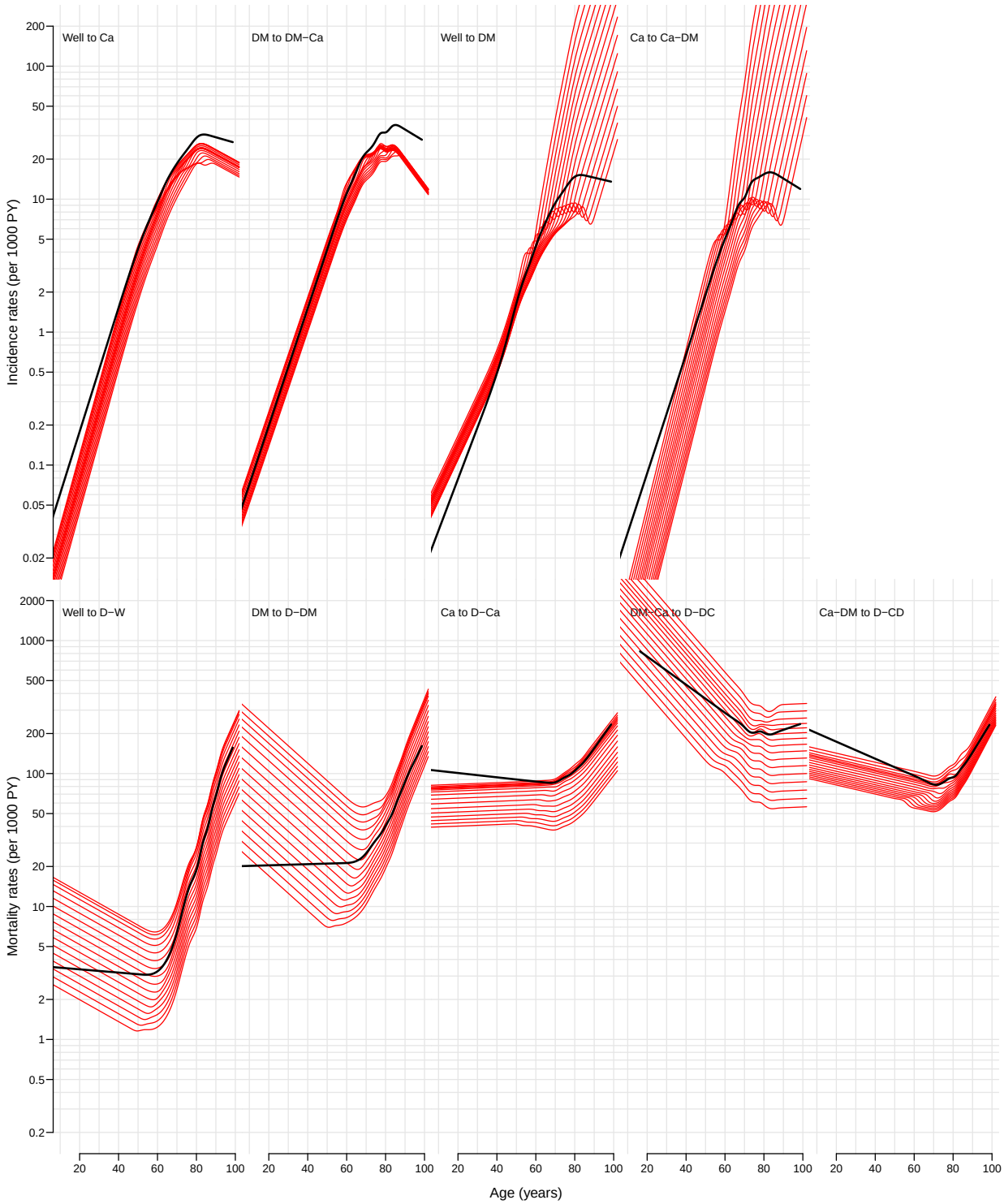

Figure 8.4: *Longitudinal rates 1930–1970 for women, cohorts 1930, 1932, . . . 1970. Black curve is the age-effect for 1935 cohort, for flat period curve.*

8.3 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.

8.4 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"
> system.time(
+ 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]
+    } )
   user  system elapsed
  0.382   0.000   0.382
```

```
> 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
```

8.4.1 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:

```
> 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 )
```

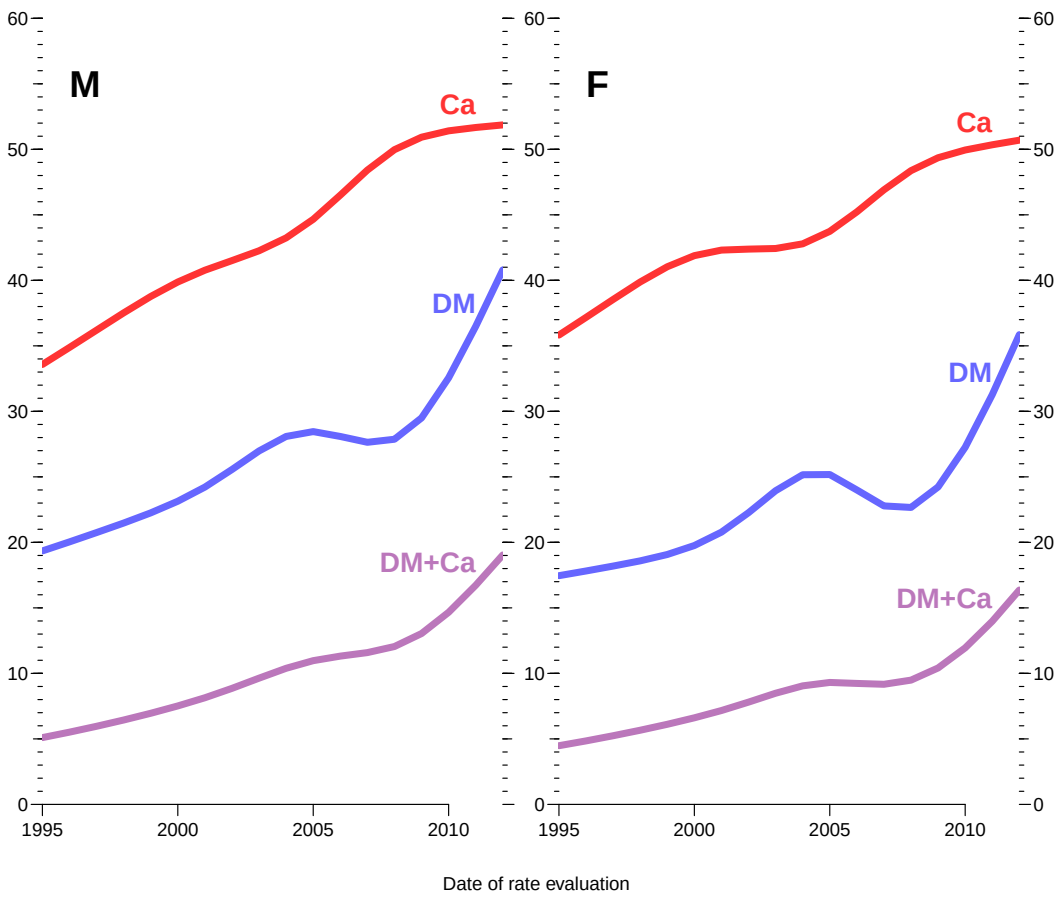

Figure 8.5: *Lifetime risk of DM (blue), cancer(red) and both diseases (purple) by year of rate calculation.*

For the corresponding cumulative plots we also define an array of cumulative lifetime probabilities over the states in the order: DM / DM+Ca / Ca / Well. 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
     state        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 cooresponding 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 )
```

We do the same thing, cunulating in a different order: Ca / DM+Ca / DM / Neither:

```
> LcP <- apply( PV[c("D-DM","D-DC","D-CD","D-Ca","D-W")[c(4:1,5)],nA,,], 2:3, cumsum )
> LcP <- LcP[c(1,1:5),,]
> LcP[1,,] <- 0
```

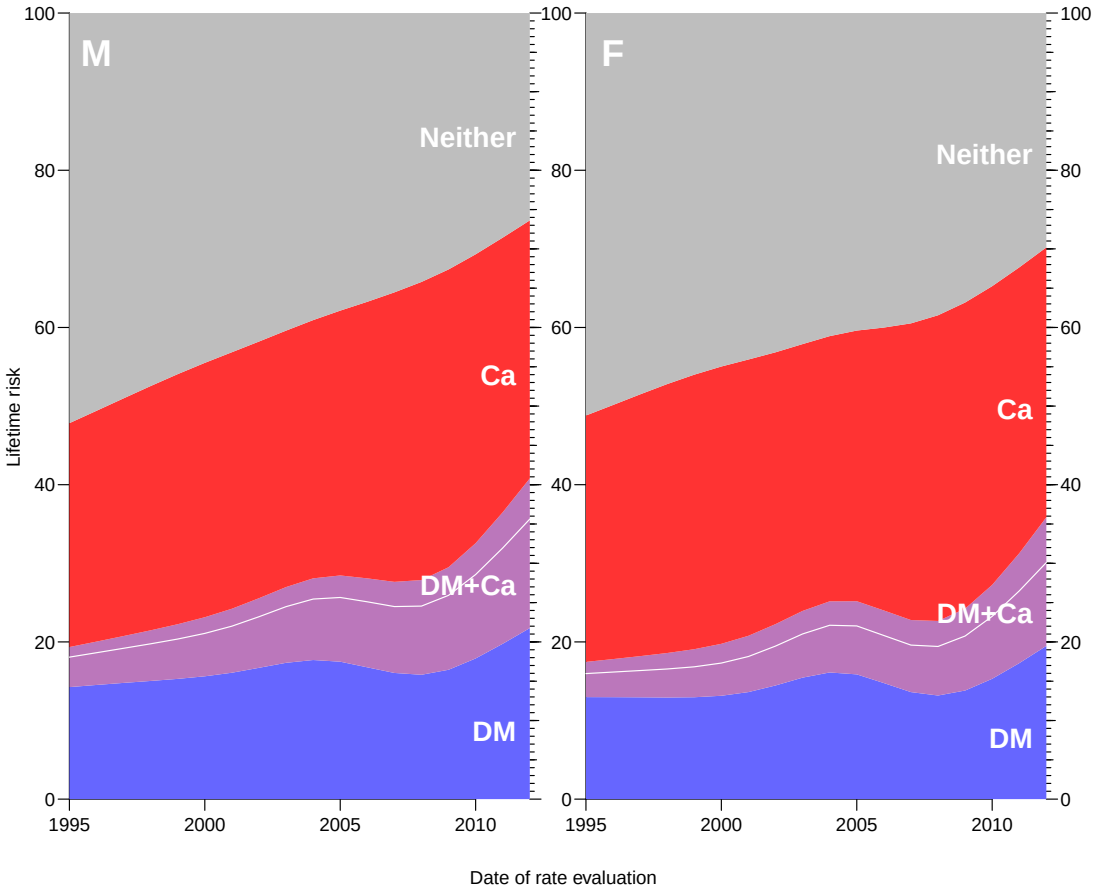

Figure 8.6: *Lifetime risk of DM and cancer by year of rate calculation.*

```
> LcP[6,,] <- 1 # Compensate the rounding errors
> dimnames( LcP )[[1]][1] <- ""
> ftable( round( LcP*100, 1 ), col.vars=c(3,1) )
     sex        M                                  F
     state         D-Ca  D-CD  D-DC  D-DM   D-W         D-Ca  D-CD  D-DC  D-DM   D-W
per
1995          0.0  28.5  29.8  33.6  47.8 100.0    0.0  31.3  32.8  35.8  48.8 100.0
1996          0.0  29.4  30.8  34.9  49.4 100.0    0.0  32.3  34.0  37.2  50.1 100.0
1997          0.0  30.2  31.8  36.2  51.0 100.0    0.0  33.3  35.1  38.5  51.5 100.0
1998          0.0  31.1  32.8  37.5  52.5 100.0    0.0  34.2  36.2  39.9  52.8 100.0
1999          0.0  31.8  33.7  38.8  54.1 100.0    0.0  34.9  37.2  41.0  54.0 100.0
2000          0.0  32.4  34.4  39.9  55.5 100.0    0.0  35.3  37.7  41.9  55.0 100.0
2001          0.0  32.6  34.8  40.8  56.9 100.0    0.0  35.1  37.8  42.3  55.9 100.0
2002          0.0  32.7  35.0  41.5  58.2 100.0    0.0  34.6  37.4  42.4  56.8 100.0
2003          0.0  32.6  35.1  42.3  59.6 100.0    0.0  33.9  36.9  42.4  57.9 100.0
2004          0.0  32.9  35.5  43.2  60.9 100.0    0.0  33.7  36.8  42.8  58.9 100.0
2005          0.0  33.7  36.5  44.7  62.1 100.0    0.0  34.4  37.6  43.7  59.6 100.0
2006          0.0  35.2  38.2  46.5  63.3 100.0    0.0  36.0  39.2  45.2  60.0 100.0
2007          0.0  36.8  40.0  48.4  64.5 100.0    0.0  37.7  40.9  46.9  60.5 100.0
2008          0.0  37.9  41.2  50.0  65.8 100.0    0.0  38.9  42.1  48.4  61.6 100.0
2009          0.0  37.9  41.5  50.9  67.4 100.0    0.0  38.9  42.4  49.4  63.2 100.0
2010          0.0  36.7  40.7  51.4  69.3 100.0    0.0  38.0  42.0  49.9  65.3 100.0
2011          0.0  34.9  39.5  51.7  71.4 100.0    0.0  36.4  41.2  50.3  67.6 100.0
2012          0.0  32.8  38.0  51.9  73.6 100.0    0.0  34.3  40.1  50.7  70.2 100.0

> 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")[c(3:1,4)],
+       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 )
```

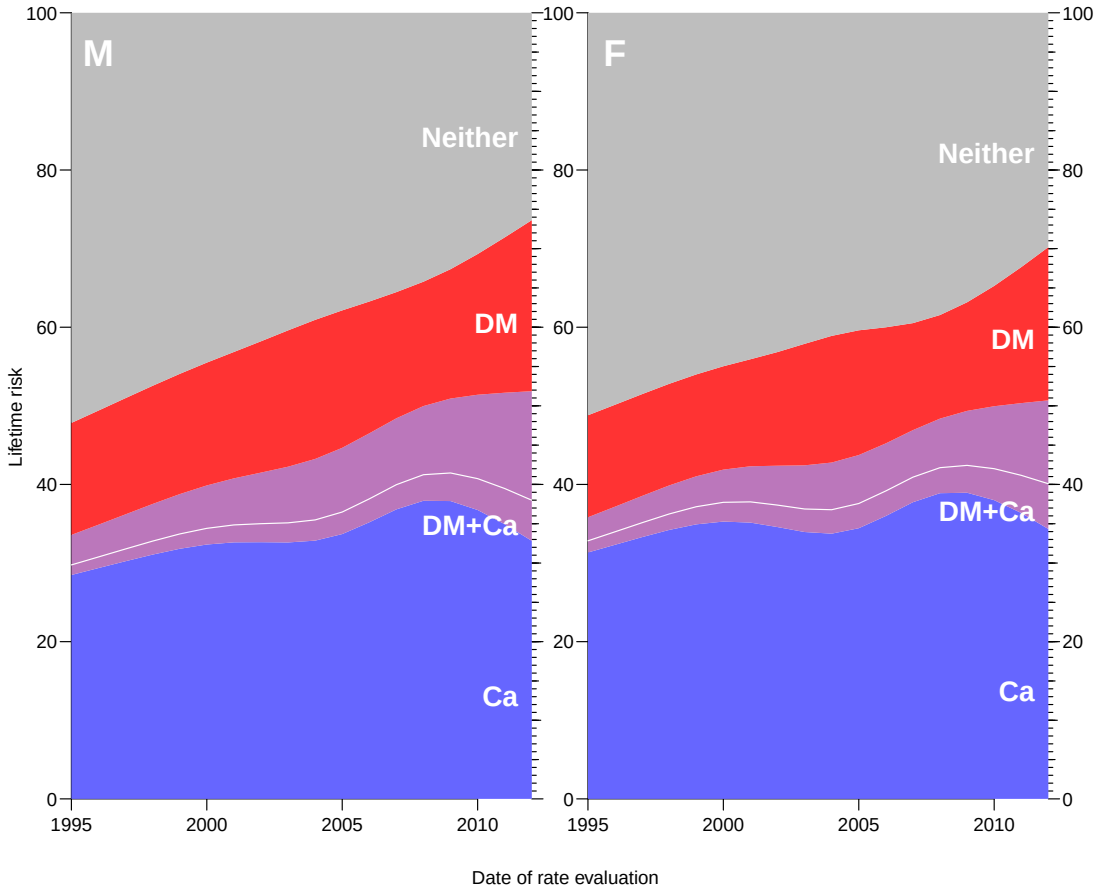

Figure 8.7: *Lifetime risk of DM and cancer by year of rate calculation.*

From the figures 8.5 and 8.6 we see that there is a dramatic 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.

8.5 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
  ..$ state: 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 )
                  state
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


> crapl <- function( sc, aa, sx="M" ) # sc is the year of rate evaluation,
+                                     # aa the age
+    {
+ an <- aa*12
+ plot( NA, xlim=c(50,100), ylim=c(0,100),
+           xlab="Age", ylab="Probability (%)", xaxs="i", yaxs="i" )
+ csq <- clx[c("DM","DM-Ca","Ca-DM","Ca","Well","Dead","Ca","Ca-DM","DM-Ca","DM")]
+ if( aa>95 ) csq[1:5] <- "transparent"
+ for( i in 1:10)
+ rect( aa-1, cPV[i  ,an,sc,sx]*100,
+       aa+1, cPV[i+1,an,sc,sx]*100,
+       col = csq[i], border="transparent")
+ segments( aa-1, cPV[6,an,sc,sx]*100,
```

```
+           aa+1, cPV[6,an,sc,sx]*100, lwd=2 )
+ pm <- ( aa<80 ) - ( aa>=80 )
+ text( rep(aa+pm*4,10), seq(5,95,,10),
+       c("DM","DM-Ca","Ca-DM","Ca","Well","Dead(W)","D(Ca)","D(Ca-DM)","D(DM-Ca)","D(DM)"),
+       col=csq, cex=1.1, font=2, adj=(1-pm)/2 )
+ segments( rep(aa+pm*1.0,10), (cPV[1:10,an,sc,sx]+cPV[1:10+1,an,sc,sx])/2*100,
+           rep(aa+pm*3.8), seq(5,95,,10), col=csq, lwd=2 )
+ }
> for( a in seq(55,100,5) )
+    {
+ pdf( paste("./graph/demo-r-crh-",a,".pdf",sep=""),
+      height=5, width=6 )
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
+ crapl( sc="2010", aa=a )
+ dev.off()
+    }
```

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

```
> aa <- as.numeric( dimnames(CR)[["age"]] )
> nul <- aa * 0
> crpl <- function( sc ) # sc is the year of rate evaluation
+    {
+ 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 = clx["DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[2,,sc,sx],
+                        rev(cPV[3,,sc,sx]))*100,
+          col = clx["DM-Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[3,,sc,sx],
+                        rev(cPV[4,,sc,sx]))*100,
+          col = clx["Ca-DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[4,,sc,sx],
+                        rev(cPV[5,,sc,sx]))*100,
+          col = clx["Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[5,,sc,sx],
+                        rev(cPV[6,,sc,sx]))*100,
+          col = clx["Well"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[6,,sc,sx],
+                        rev(cPV[7,,sc,sx]))*100,
+          col = clx["Dead"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[7,,sc,sx],
+                        rev(cPV[8,,sc,sx]))*100,
+          col = clx["Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[8,,sc,sx],
+                        rev(cPV[9,,sc,sx]))*100,
+          col = clx["Ca-DM"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[9,,sc,sx],
+                        rev(cPV[10,,sc,sx]))*100,
+          col = clx["DM-Ca"], border="transparent")
+ polygon( c(aa,rev(aa)), c(cPV[10,,sc,sx],
+                        rev(cPV[11,,sc,sx]))*100,
+          col = clx["DM"], border="transparent")
+ matlines( aa, 100*cPV[6,,sc,sx],
+           lty=1, col="black", lwd=3, 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( "demo-r-film.pdf", width=11, height=8 )
> for( sc in dimnames(cPV)[[3]] ) crpl( sc )
> dev.off()

pdf
  2
```

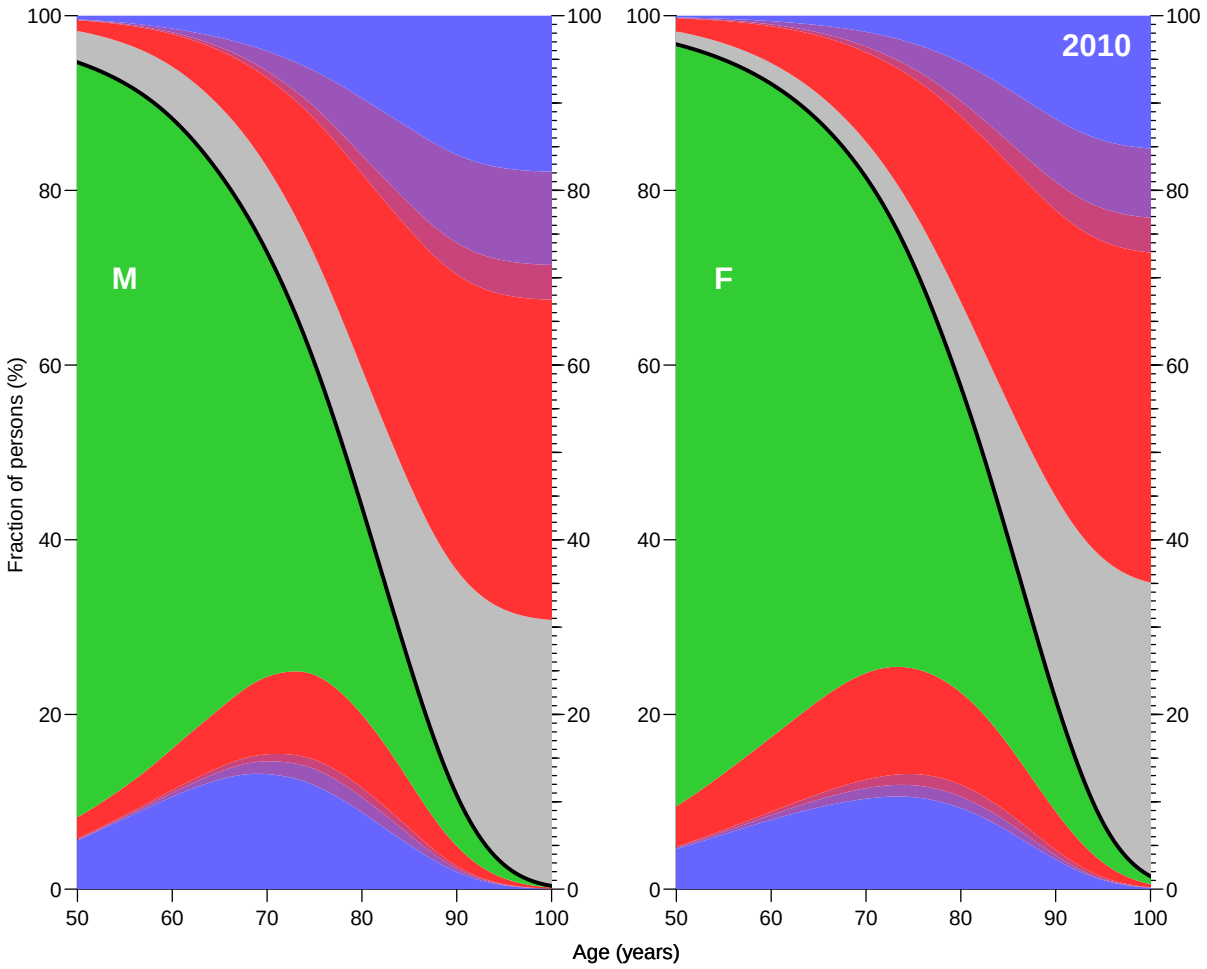

Figure 8.8: *Occupation probabilities for the various states in figure 1 at various ages, assuming all start in "Well" at age 0. Based on cross-sectional rates from 2010.*
The thick black line is the overall survival curve, with "Dead" states are above and "Alive" below the line. The blue states are 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.

8.5.1 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 these 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 )
+     }
```

8.5.2 Conditional state probabilities

We can of course also make the same exercise *conditional* on 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] <-
```

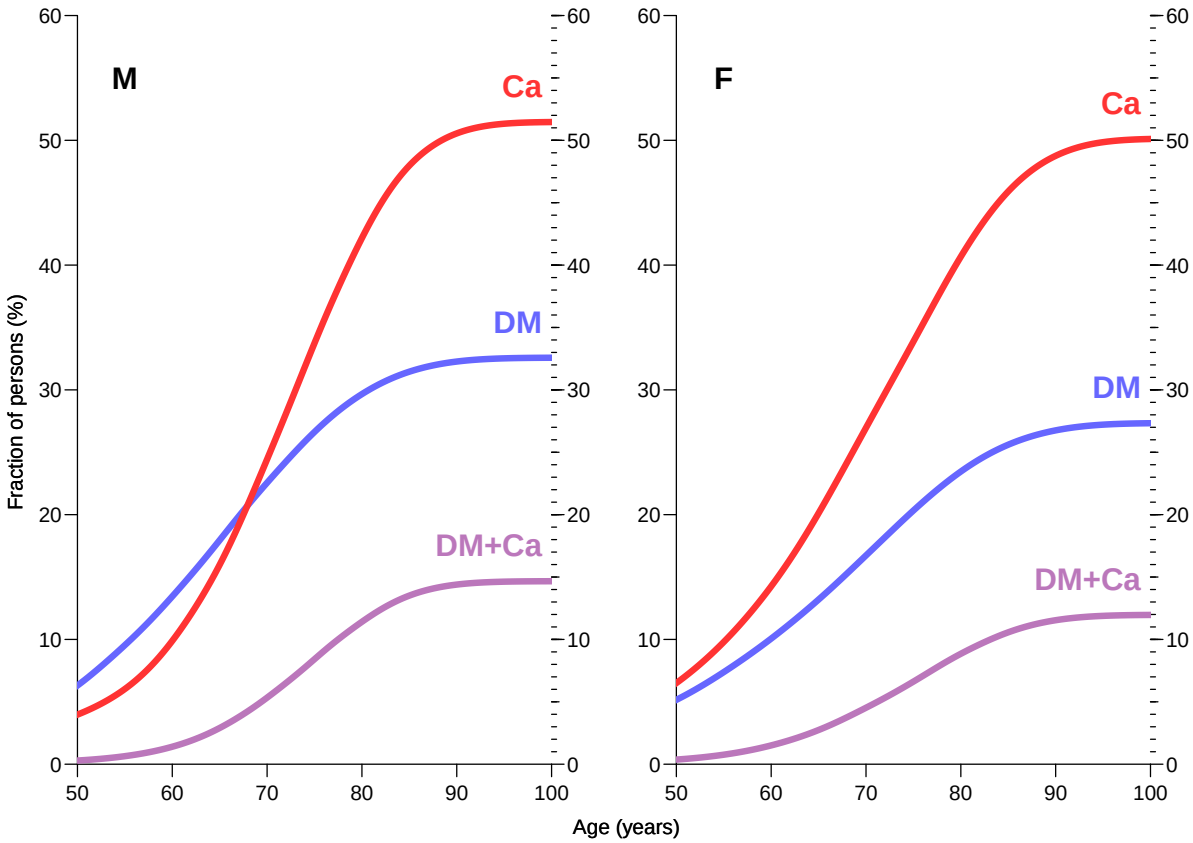

Figure 8.9: *Cumulative risk of acquiring DM, cancer or both before a given age, using rates as of 2010.*

```
+     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]
+        }
+     }
```

8.5.3 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 make the same plot as above, calssified by sex, conditioning age (50, 60, 70) and state conditioned on (DM/no DM), in total 12 combinations:

```
> 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 )
```

8.5.4 Lifetime risk

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

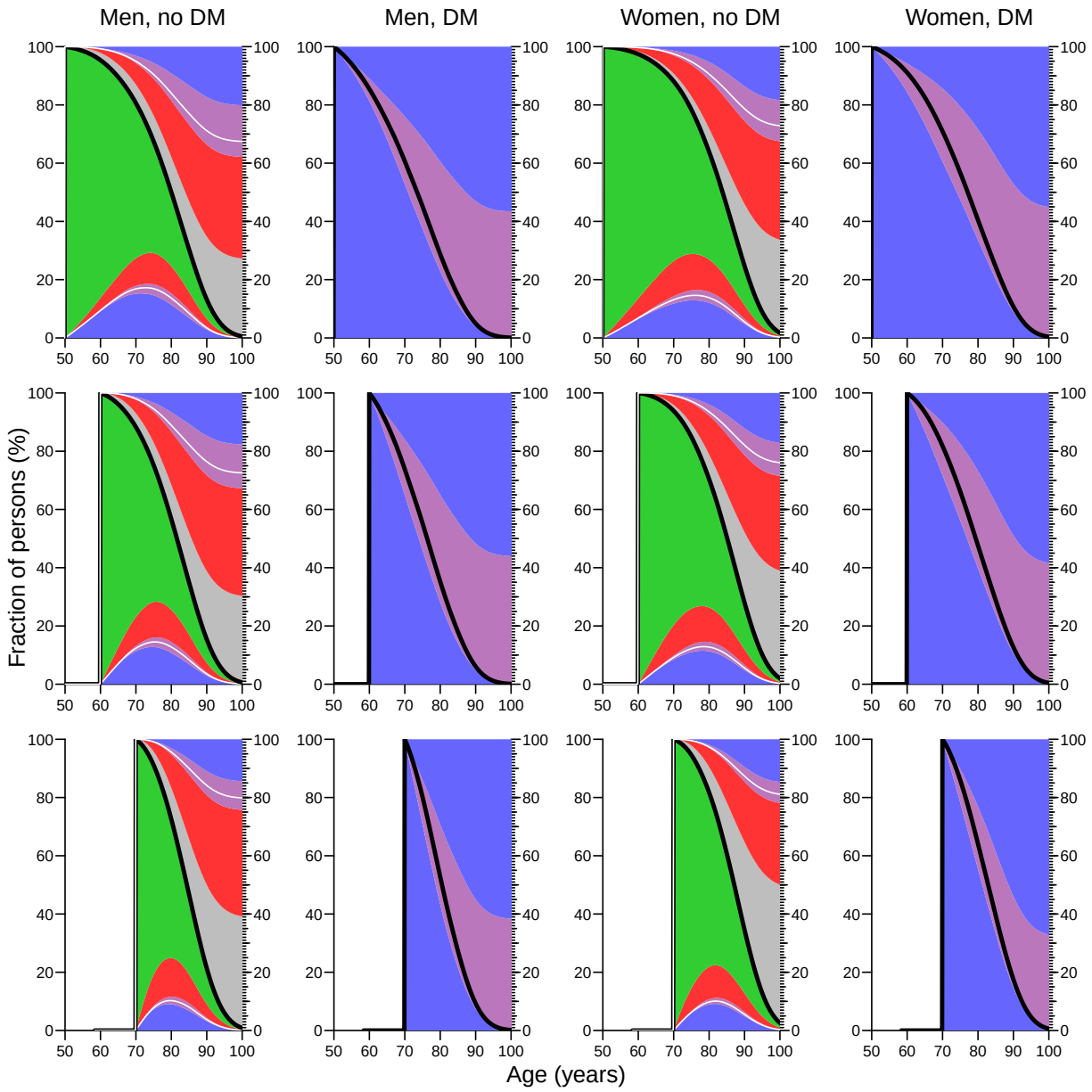

Figure 8.10: *Plots of state occupancies conditional on being either well or diabetic at different ages. Based on cross-sectional rates as of 2012-01-01.*

```
> 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"
> 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"
  ..$ : NULL
> 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
```

8.5.5 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] "0.041666666666667" "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] "0.041666666666667" "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( PY )[-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" "200

$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" "200

$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 well as the percentage of total life spent with diabetes, as a function of the date at which we evaluated rates:

```
> par( mfrow=c(1,2), mar=c(1,1,1,1), mgp=c(3,1,0)/1.6, oma=c(2,2,0,1) )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,],
+                                                100*PY["anyDM",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,2),each=2), lwd=4, col=c("blue","red"),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "Diabetes", adj=0 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyCa",,],
+                                                100*PY["anyCa",,]/PY["All",,] ),
+          type="l", lty=rep(c(1,2),each=2), lwd=4, col=c("blue","red"),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "Cancer", adj=0 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> mtext("Date of rate evaluation", side=1, line=1, cex=1.0, outer=TRUE )
> mtext("Years / % of life spent with disease", side=2, line=1, cex=1.0,
+       outer=TRUE, las=0 )
```

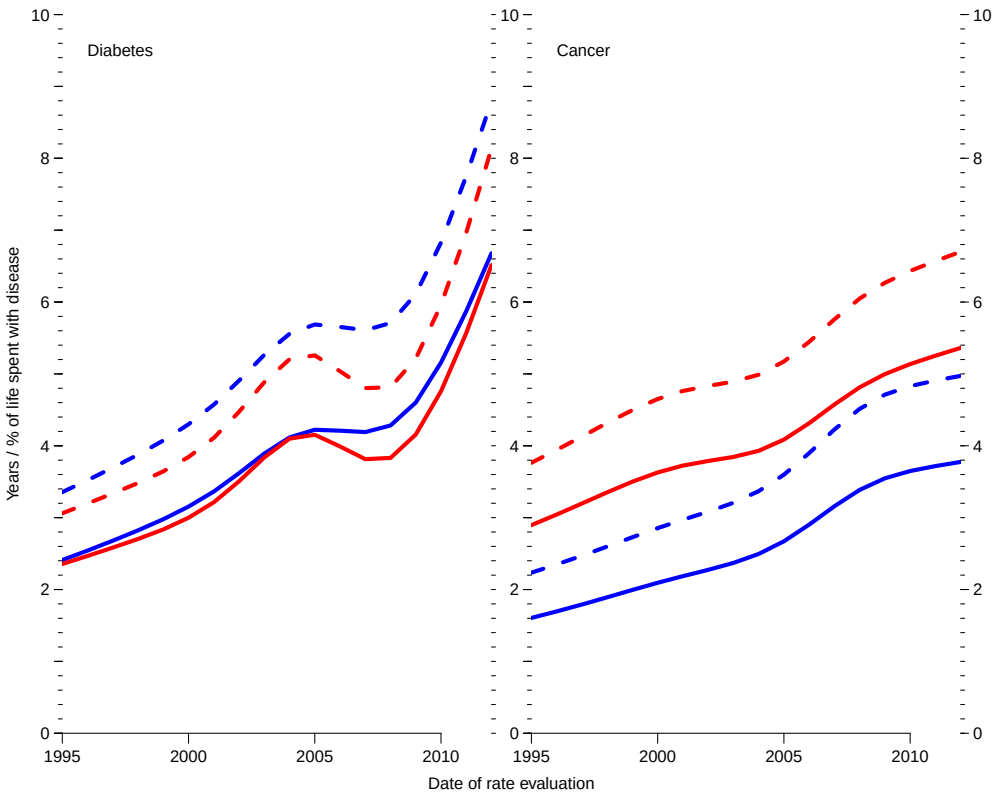

Figure 8.11: *Years (full lines) and percent of life (broken lines) spent with disease (daiebets or cancer); red: women, blue: men.*

```
> par( mfrow=c(1,2), mar=c(1,1,1,1), mgp=c(3,1,0)/1.6, oma=c(2,2,0,1) )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,"M"], PY["anyDM",,"M"],
+                                                100*PY["anyDM",,"M"]/PY["All",,"M"],
+                                                    PY["anyCa",,"M"], PY["anyCa",,"M"],
+                                                100*PY["anyCa",,"M"]/PY["All",,"M"] ),
+          type="l", lty=c("F1","11F1","11"), lwd=6, col=rep(clr[c(2,3)],each=3),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "M", adj=0, font=2, cex=2 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> matplot( as.numeric(dimnames(PY)[[2]]), cbind( PY["anyDM",,"F"], PY["anyDM",,"F"],
+                                                100*PY["anyDM",,"F"]/PY["All",,"F"],
+                                                    PY["anyCa",,"F"], PY["anyCa",,"F"],
+                                                100*PY["anyCa",,"F"]/PY["All",,"F"] ),
+          type="l", lty=c("F1","11F1","11"), lwd=6, col=rep(clr[c(2,3)],each=3),
+          xaxs="i", xlab="",
+          yaxs="i", ylab="", yaxt="n", ylim=c(0,10) )
> text( 1996, 9.5, "F", adj=0, font=2, cex=2 )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.2, at=0:50/5, labels=FALSE )
> axis( side=2, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:10, labels=FALSE )
> axis( side=4, lwd=0, lwd.ticks=1, tcl=-0.4, at=0:5*2 )
> mtext("Date of rate evaluation", side=1, line=1, cex=1.0, outer=TRUE )
> mtext("Years / % of life spent with disease", side=2, line=1, cex=1.0,
+       outer=TRUE, las=0 )
```

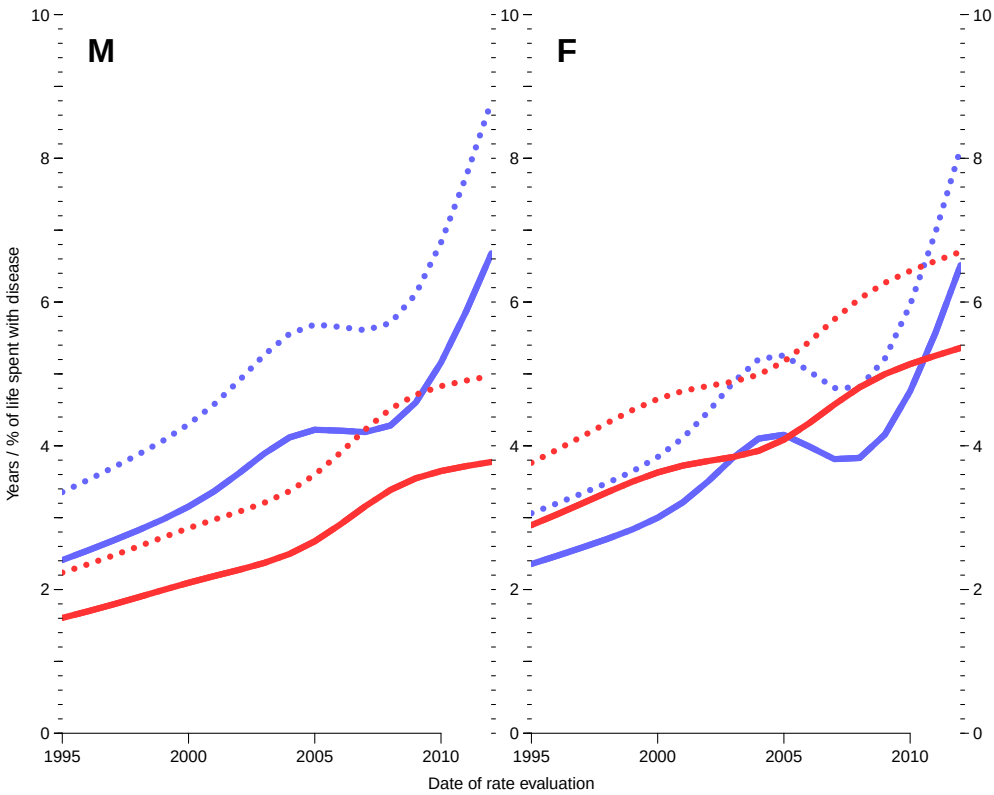

Figure 8.12: *Years (full lines) and percent of life (broken lines) spent with disease; diabetes (blue) or cancer (red).*

8.5.5.1 Diseased ages of life

A little more tricky is measures of the ages in which the time with diabetes/cancer is spent. The probabilities in `PV` gives the *distributions* of persons in states in each age. Since this refers to the distribution of *all* initial persons in the population, normalization of the age-specific occupancy probabilities of, say, the state "DM" to an age-distribution will represent the distribution of time alive spent in the state. This of course is meaningless for the corresponding death states.

Thus we devise a function that does this for a 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`, `Age Diseased`) 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" "200

$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" "200

$sex
[1] "M" "F"
> AD[,"Well" ,,] <- apply( PV["Well",,,],
+                          2:3, aqnt )
> AD[,"DM"   ,,] <- apply( apply( PV[c("DM","DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> AD[,"Ca"   ,,] <- apply( apply( PV[c("Ca","DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> AD[,"DM+Ca",,] <- apply( apply( PV[c("DM-Ca","Ca-DM"),,,],
+                                  2:4, sum ),
+                          2:3, aqnt )
> str( AD )
 num [1:5, 1:4, 1:18, 1:2] 6.81 17.09 34.6 53.35 67.27 ...
 - attr(*, "dimnames")=List of 4
  ..$ 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 then show the distribution of the ages with DM, and augment the plot with an indication of the expected *length* of time spent diseased, arbitrarily allocated around the median age spent with disease:

```
> par( mfrow=c(1,2), mar=c(2,2,1,1) )
> 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="",
+          xlab="Date of rates used", ylim=c(0,100), yaxs="i")
> 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), yaxs="i")
> 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" )
> mtext( "Age with diabetes (10,25,50,75,90 percentiles)", side=2,
+        outer=TRUE, line=0 )
```

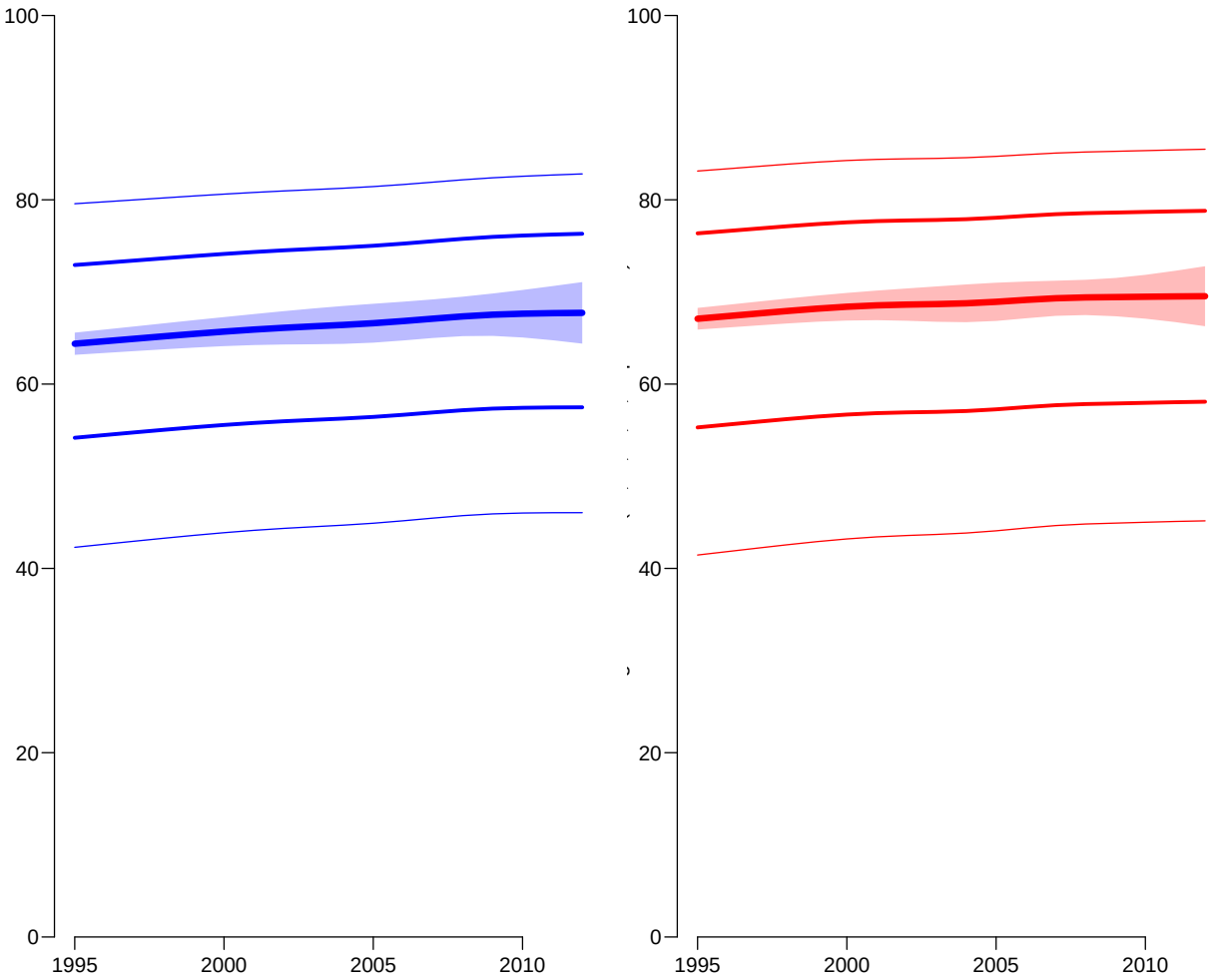

Figure 8.13: *Percentiles of ages in which persons suffer from diabetes for men (blue) and women (red). The* height *of shaded area indicate the expected* length *of time spent with diabetes.*

The comparison in figure 8.13 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.